Bayesian modelling of organ deformations in radiotherapy

Øyvind Lunde Rørtveit

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2023

UNIVERSITY OF BERGEN

Bayesian modelling of organ deformations in radiotherapy

Øyvind Lunde Rørtveit

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 13.06.2023

Year: 2023
Title: Bayesian modelling of organ deformations in radiotherapy

Name: Øyvind Lunde Rørtveit
Print: Skipnes Kommunikasjon / University of Bergen

We do these things not because they are easy, but because we thought they were going to be easy

-J.F. Kennedy (paraphrased)

Scientific environment

This PhD study is part of the project "Establishing a framework for interdisciplinary clinical particle therapy research at Haukeland University Hospital" funded by Trond Mohn Foundation (formerly Bergen Research Foundation).

The research has been performed at Haukeland University Hospital (HUH) with the research group at the section for medical physics under the department of oncology and medical physics. The ph.d. project has been approved by and associated with the institute of physics and technology (IFT) at the University of Bergen (UiB). The work has been supervised by Sara Pilskog and Liv Bolstad Hysing at HUH and IFT, and by Andreas Størksen Stordal at NORCE Research and the mathematical institute of UiB.

Acknowledgements

I am grateful for having been given the opportunity to engage in this PhD-project by Helse Bergen, IFT at UiB and the Trond Mohn Foundation. The subjectss of radiation therapy and medical physics were new to me, and it has been engaging to get to learn so much new from an interesting and important field. The working environment at the section of medical physics has been inspiring, with exceedingly friendly and interesting colleagues.

I would like to extend my deepest gratitude to my advisors. Sara, you are always immensely helpful, ready to assist despite being constantly busy, and massively knowledgeable, always ready with a reference for any topic. Despite the nuisance of having every written sentence corrected, I must admit the writing does improve afterwards. Liv, you are also extremely knowledgeable, consistently helpful, and the reason why practicalities have never been an issue during this project. You are also the reason why the medical physics research group is equal parts professional and friendly. Thanks Andreas, for being one of my best friends in addition to my advisor. Your guidance is the reason why I dare publish these methods, the mathematical derivations of which I would not trust without your affirmation.

Most of all, I wish to thank my family, Mildrid, Vebjørn and Hallvard, most of whom were created during the work on this project. You are the motivation for all my work. Mildrid, you are the reason I get up in the morning. Vebjørn, you are the reason I *wake* up in the morning (though often it is more night than morning). Hallvard, your constant good mood is an inspiration. Without the two of you, this project would have been completed a long time ago. I also need to thank Vilnius, who's support, insight and hunger I can always count on.

Special thanks to Mamma, Pappa, Oddrun, Jori and Olaf who helped me get to work by hanging out with Vebjørn in the daytime. And of course thanks to Bodil, Sigrun, Agnes, Ingeborg, Vilde, Magnus, Askild and Vidar for being the best family.

Abstract in English

Modern radiotherapy tends to be highly conformal, meaning that a high and uniform dose is delivered to the target volume and as little dose as possible to the surrounding normal tissue. The total radiation dose is delivered across several smaller daily fractions, typically spanning several weeks. During and between these fractions, internal organs are constantly in motion due to factors such as breathing, changes to bladder filling state, intestinal movement and external influences. Nevertheless, the position of the target and relevant organs at risk (OARs) are determined based on a static 3D scan acquired before start of treatment. A common safeguard which is used to take such motion into account is the addition of margins around the target. These margins reduce the chance of missing parts of the target, yet increases dose to the healthy tissue surrounding the target. The margin size is based on statistics from previous patients. However, for the most part, the statistical methods used are very simple, and typically based on an assumption of rigid patient motion. Similarly, motion of the OARs is commonly neglected. For estimation of dose to the OARs, it is common to assume that the organ shape at the static scan is representative for its shape during treatment.

The work in this thesis concerns the use of techniques from Bayesian statistics for modelling inter-fraction organ motion and deformation. The goal is to estimate accurately the statistical distribution of shapes for one or more organs for a given patient. The distribution provides knowledge of how the patient's organs might move and deform during the radiotherapy course. This information is useful for the evaluation of radiotherapy plans, prediction of adverse effects, so-called motion-robust radiotherapy planning, the generation of margins and more. The methods presented in this thesis have been evaluated for predicting deformations of the rectum of prostate cancer patients. For these patients, the rectum is a crucial OAR that is affected by both early and late side effects including leakage, bleeding and pain.

Compared to existing methods, the Bayesian approach developed and implemented in this thesis offers two advantages: first, combining population statistics and individual data leads to more accurate estimates of the patient-specific distribution. Secondly, the new methods estimate the distribution of the so-called systematic error in addition to variations from fraction to fraction.

The systematic error is the difference between the estimated shape/position of an organ at the planning stage and its average shape/position during therapy, and was the subject of paper I. Here, we were able to reduce the systematic error of the rectum in 33 out of 37 prostate cancer patients using a straightforward method to combine the shape of the rectum at the planning CT with the population mean shape. We also evaluated the impact of this improvement on the estimation of dose to the rectum. We found no significant improvement on the estimation of two presumably relevant dose parameters (equivalent uniform dose and D5%). However, we did find significant reduction in the bias of the estimated dose-volume histogram in the range from 52.5 Gy to 65 Gy.

Paper II contains the central work of this project. It presents two organ deformation models based on Bayesian methods. The input data to these algorithms are organ shapes derived from 3D scans. The methods can take a varying number of such inputs from a given patient, and will produce more accurate results the more inputs they are given. They provide an estimate of the mean shape of the organ, as well as the uncertainty of this mean, in addition to the distribution of the variation of shapes from fraction to fraction. The methods were evaluated in the task of estimating coverage probabilities, i.e. the probability that the organ will cover a certain point in the patient coordinate system, for the rectum of prostate cancer patients. For this evaluation, tens of thousands of organ shapes needed to be converted to so-called binary masks, which are 3D arrays of points in the patient coordinate system where the value of each point is 1 if the point is inside the organ and 0 if it is outside. This was enabled by the highly efficient point-in-polyhedron software presented in paper III, which was developed for this project.

The models were given varying number of scans, from 1 to 10, as input, and compared to two existing (non-Bayesian) models. The estimates of the coverage probability produced by the new models were significantly more similar to the ground truth than those produced by the existing models, at least up to three input scans. The main differences between the two new algorithms are their of conceptual complexity and accuracy, and the choice of method in a given application will therefore come down to a trade-off between these qualities.

An application for the models derived in paper II, concerning patients receiving re-irradiation for recurrent prostate cancer, is presented in paper IV. We introduce a way of estimating the expectation and uncertainty of the accumulated dose to the rectum from the two treatment courses. The method is based on “representative shapes” of the rectum, that is, shapes that are probable and also particularly favourable or unfavourable

in terms of dose. The advantage is that these shapes can be used as a visual aid for the oncologist or dose planner, and that the method can be implemented using existing features of treatment planning systems.

Overall, this thesis provides novel solutions to the central challenge of organ motion mitigation in RT. The presented models are the first to simultaneously exploit population and patient specific organ motion and addressing both systematic and random errors.

Samandrag på norsk

Moderne strålebehandling mot kreft er skreddarsydd for å gje ein høg stråledose tilpassa svulsten (målvolumet), mens så lite dose som mogleg vert gitt til det friske vevet omkring. Den totale dosen vert levert over nokre veker i daglege “fraksjonar”, noko som reduserer biverknader. Under og mellom desse fraksjonane rører dei indre organa på seg heile tida på grunn av pust, fylling av blæra, tarmar si rørsle og ekstern påverknad. Likevel vert posisjonen til målvolumet og relevante risikoorgan bestemt på grunnlag av eit statisk 3D-skann som er tatt før behandlinga startar. Den vanlege måten å sikre seg mot konsekvensar av denne rørsla er å legge til marginar rundt svulsten. Slik sikrar ein å treffe målvolumet, men til gjengjeld får det friske vevet meir dose. Marginane sin storleik er fastsett ved hjelp av statistikk over tidlegare behandla pasienter. Dei statistiske metodane som vert brukte er ofte enkle, og tek berre omsyn til rigid rørsle, altså at heile kroppen rører seg i eitt. Dessutan vert det ikkje teke omsyn til rørsla til risikoorgan. For å berekne dose til risikoorgana er det vanleg å anta at forma til organa i planleggingsskannet er representative for forma deira under behandling.

Arbeidet i denne avhandlinga handlar om å bruka teknikkar frå Bayesiansk statistikk for å modellere korleis organ rører og deformerer seg mellom fraksjonane. Målet er å estimere nøyaktig den statistiske fordelinga av rørsle for eit eller fleire organ til ein pasient. Fordelinga gjev innsikt i korleis organa forandre seg medan behandlinga går for seg. Denne innsikta er nyttig for evaluering av stråleterapiplanar, statistisk prediksjon av biverknader, såkalla robust planlegging og å berekna størrelsen på marginar. Metodane som vert presentert er evaluerte for endetarmen (rektum) sine rørsler hjå prostatakreftpasientar. For desse pasientane er rektum eit viktig risikoorgan, som kan bli ramma både av akutte og seine biverknader, som lekkasje, bløding og smerter.

Samanlikna med eksisterande metodar har den Bayesianske tilnærminga to fordelar: For det første gir kombinasjonen av populasjonsstatistikk og individuelle data meir nøyaktige anslag av den pasientspesifikke fordelinga. For det andre estimerer dei nye metodane den såkalla systematiske feilen i tillegg til variasjonar frå fraksjon til fraksjon.

Den systematiske feilen er forskjellen mellom den estimerte forma på organet under

planlegging, og gjennomsnittsforma til organet under bestråling. Denne typen feil var tema for artikkel I. Her fekk vi til å redusere den systematiske feilen til rektum hjå 33 av 37 prostatakreftpasientar ved å bruke ein metode som kombinerer forma på rektum under planlegginga og gjennomsnittsforma i populasjonen. Vi vurderte og om denne forbetringa hadde påverknad på estimering av summert dose til rektum. Metoden gav ikkje signifikant forbetring for to antatt relevante parametrar (ekvivalent uniform dose og D5%), men gav signifikant reduksjon av bias på det estimerte dose-volum-histogrammet i intervallet 52.5 Gy til 65 Gy.

Hovudarbeidet i dette prosjektet er publisert i artikkel II. Der presenterer vi to modellar for organrørsle basert på Bayesianske metodar. Inndata til desse metodane er organformer som er henta frå 3D-skanningar. Metodane kan ta ulikt tal slike former, og produserer meir nøyaktige resultat jo fleire former dei får. Dei gjev anslag av gjennomsnittsforma og kor stor uvissa om denne forma er, i tillegg til anslag av fordelinga av variasjon av former frå fraksjon til fraksjon. Vi evaluerte metodane etter kor godt dei kunne berekne "dekningssannsyn", altså sannsynet for at organet skal dekke eit gitt punkt i pasientkoordinatsystemet til ei gitt tid. For denne berekninga måtte titusenvis av organformer gjerast om til såkalla binærmasker, som er 3D-matriser av punkter i pasient-koordinatsystemet der verdien til eit punkt er 1 dersom punktet er inne i organet, og 0 elles. Denne berekninga var mogleg på grunn av programvare som blei implementert for dette prosjektet, og som er presentert i artikkel III. Også her var det prostatakreftpasientar sitt rektum som vart brukt til evaluering. Berekningane til dei nye metodane var likare det sanne dekningssannsynet enn tilsvarande berekningar frå tidlegare metodar, i signifikant grad, i alle fall opp til tre input. Forskjellen mellom dei to nye algoritmane er i hovudsak kompleksiteten og nøyaktigheita, og valet mellom algoritmane i ein gitt bruk vil vere ei avveging mellom desse faktorane.

Vi viste ein måte modellane kan verte brukte i artikkel IV, som handlar om pasientar som får re-bestråling for tilbakefall av prostatakreft. Her brukte vi modellane til å berekne forventa akkumulert dose til rektum frå dei to behandlingane, og også uvissa rundt den forventa dosen. Metoden er basert på "representative former" av rektum, altså former som rektum kan ta som er sannsynlege, men lite fordelaktige. Desse formene kan brukast som visuell hjelp for onkologar og doseplanleggjarar, og metoden kan implementerast ved hjelp av eksisterande funksjonar i programvaren for behandlingsplanlegging.

Overordna gir denne avhandlinga nye løysingar for den sentrale utfordringa med å redusere konsekvensar av organrørsle i stråleterapi. Dei presenterte modellane er dei første som utnyttar statistikk for populasjonen og data frå den enkelte pasienten samstundes, og som tar omsyn til både systematiske og tilfeldige feil.

List of publications

Papers included in this thesis:

- **Paper I.** Rørtveit ØL, Hysing LB, Stordal AS, Pilskog S, *Reducing systematic errors due to deformation of organs at risk in radiotherapy*, Med Phys. 2021;48:6578–6587
- **Paper II.** Rørtveit ØL, Hysing LB, Stordal AS, Pilskog S, *An organ deformation model using Bayesian inference to combine population and patient-specific data*, Phys. Med. Biol. 2023;68 055009.
- **Paper III.** Rørtveit ØL, *InsidePolyhedron - Fast point-in-polyhedron test on a grid*, Submitted to the Journal of Open Research Software
- **Paper IV.** Rørtveit ØL, Hysing LB, Stordal AS, Ekanger C, Pilskog S, *Calculating cumulative dose and uncertainty for OARs of re-irradiation patients using a Bayesian motion model*, Manuscript in preparation

Conference contributions

- **Rørtveit ØL**, Hysing LB, Stordal AS, Pilskog S, *Prediction of inter-fractional rectum shape changes in prostate cancer radiotherapy*, NACP 2021, Online conference, May 2021 (Oral presentation)
- **Rørtveit ØL**, Hysing LB, Stordal AS, Pilskog S, *A statistical method for reducing systematic errors in the presence of organ at risk deformations,* ESTRO 2021, Madrid, August 2021 (Oral presentation)
- **Rørtveit ØL**, Hysing LB, Stordal AS, Pilskog S, *A deformation model for the rectum using Bayesian inference,* BiGART 2021, Aarhus/Oslo/Online, October 2021 (Poster and oral presentation)
- **Rørtveit ØL**, Hysing LB, Stordal AS, Pilskog S, *A deformation model for the rectum which combines population and patient-specific motion patterns,* ESTRO 2022, Copenhagen, May 2022 (Oral presentation)
- **Rørtveit ØL**, Hysing LB, Stordal AS, Pilskog S *Re-irradiation in recurring prostate cancer: Estimating motion-inclusive DVHs for the rectum,* NACP 2023, Reykjavik, March 2023 (To be presented as a poster)
- Thu SB, **Rørtveit ØL**, Pettersen HES, Ekanger C, Pilskog S *NTCP analysis from DVHs following treatment of locally recurring prostate cancer*, NACP 2023, Reykjavik, March 2023 (To be presented as a poster)
- Pilskog S, Hysing LB, Stordal AS, **Rørtveit ØL** *Combining patient and population rectal shapes for re-irradiation of recurring prostate cancer,* ESTRO 2023, Vienna, May 2023 (To be presented as a poster discussion)

List of abbreviations

3D-CRT	3D Conformal Radiation Therapy. 1, 2
AI	Artificial Intelligence. 56
ART	Adaptive Radiotherapy. 5, 58
BLD	Bi-directional Local Distance. 38
CBCT	Cone Beam Computed Tomography. 2, 5, 8
CPM	Coverage Probability Matrix. 39, 45, 52
CT	Computed Tomography. 1, 2, 6, 7, 21–25, 39, 52, 54, 55, 57
CTV	Clinical Target Volume. 3, 4, 7, 52
DIR	Deformable Image Registration. 24
DVF	Displacement Vector Field. 26
DVH	Dose-Volume Histogram. 6, 51–53, 61
EBRT	External Beam Radiation Therapy. 1, 53
EQD$_2$	Equivalent dose in 2 Gy fractions. 39
EUD	Equivalent Uniform Dose. 41
FEM	Finite Element Modelling. 6
Gy	Gray. 21, 22, 41
HD	Hausdorff Distance. 38, 41
IGRT	Image Guided Radiation Therapy. 51, 54, 58
IMRT	Intensity Modulated Radiation Therapy. 1, 21
LOOCV	Leave-One-Out Cross Validation. 40, 41, 51
MCMC	Markov-Chain Monte-Carlo. 14, 15
NIG	Normal-Inverse-Gamma. 33, 35, 36
NIW	Normal-Inverse-Wishart. 36, 50
NTCP	Normal Tissue Complication Probability. 2, 8
OAR	Organ At Risk. vii, 2–4, 9, 51
PC	Principal Component. 29, 30, 32, 34–37, 54
PCA	Principal Component Analysis. 5–7, 29–33, 35, 36, 54, 56

pCT	Planning CT. 2, 3, 6, 21, 22, 25, 51, 53, 54, 58, 59
pdf	Probability Density Function. 12, 14–16
PDM	Point Distribution Model. 5
PRV	Planning Risk Volume. 4
PTV	Planning Target Volume. 3, 4, 52
RT	Radiotherapy. 1, 2, 4, 8, 21, 22, 51, 52, 56, 57, 59, 61, 62
SVD	Singular Value Decomposition. 30
TPS	Treatment Planning System. 2, 5, 51, 53, 61, 62
VMAT	Volumetrically Modulated Arc Therapy. 1, 4

Contents

Chapter 1

Introduction

Radiotherapy (RT) is currently one of the four main modalities of cancer treatment, along with surgery, chemotherapy and immunotherapy. In Norway, 38% of cancer patient were treated with RT in 2012 (Borras et al. 2015), while studies indicate that 50% of patients would benefit from RT (Barton et al. 2014; Borras et al. 2015). RT as a treatment for cancer dates back to the end of the 19th century, with the discovery of x-rays in 1895 by Wilhelm Röntgen, and the discovery of natural radioactivity and its sources by Henri Becquerel, Marie Curie and Pierre Curie in 1896-1898. Experiments with the equivalent of today's external beam radiation therapy EBRT for treatment of breast cancer using Röntgen tubes were started as early as 1896. In the early days of EBRT, the radiation was given with very low dose rates due to limitations of the available technology. As a result, the treatment was necessarily "fractionated" as it had to be given over several days or weeks to achieve anything near sufficient radiation dose. The discovery of the advantages of fractionation is usually credited to French biologist Claude Regaud in the 1910s (Foray 2012). Today, fractionation is the norm for most curative RT. Typically, 25-40 fractions are given in total, with five fractions per week over the course of 5-8 weeks, although the current trend is towards shorter therapy courses. The main advantage of fractionation is that differences in radiosensitivity between tumors and normal tissue lead to better recovery of normal tissue under smaller daily fractions of radiation.

A crucial step in the advent of precision EBRT was the invention 3D x-ray imaging through computed tomography (CT) in the 1970s. Together with the multi-leaf collimator and treatment-planning software, this later enabled 3D conformal RT (3D-CRT), where each radiation field is sculpted to the target. Further advances enabled *dose modulation*, which, coupled with with inverse plan optimization, made possible the highly conformal modern techniques of intensity-modulated RT (IMRT) and volumetrically modulated arc therapy (VMAT). Figure 1.1 illustrates how VMAT tailors dose to the

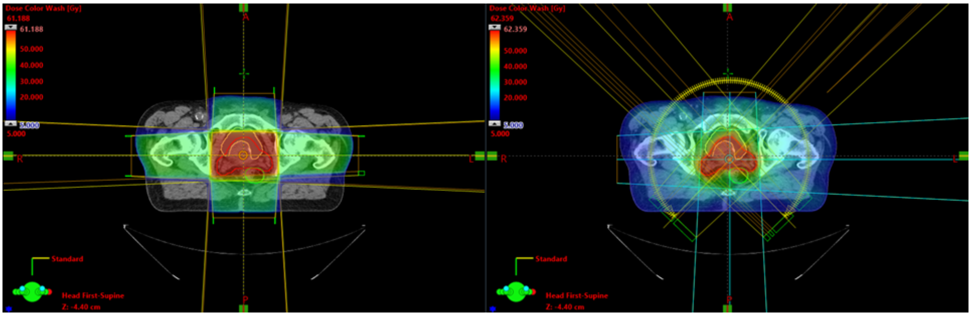

Figure 1.1: Example of dose distributions for 3D-CRT (left) and VMAT (right) plans for the same patient.

target (right) compared to 3D-CRT (left).

In a typical RT workflow, after diagonsis and prescripition, an initial CT scan of the patient, called the planning CT (pCT), is acquired. On the pCT, oncologists or other experts delineate the prescription doses and any organs at risk (OARs). A planning expert then uses a treatment planning system (TPS) to create an optimized RT plan, which contains machine instructions for the delivery of RT, according to prescribed dose to the target and constraints on the dose to the OARs. The quality and soundness of the plan is then verified in a rigorous quality assurance process. Finally, treatment is delivered according to the quality-assured plan over the course of the following weeks.

While higher precision in RT is generally seen as a great advantage, the combination of conformity and fractionation can cause problems. High precision requires that patients are set up identically for treatment delivery each day as they were during the pCT scan. Any discrepancy, or setup error, will cause a geometrical miss of the delivered dose. Image guided RT, where an in-treatment-room image (e.g. a cone beam CT (CBCT)) is taken before delivery, and the patient position is shifted to match the pCT to the CBCT, has decreased setup errors considerably in the last decades. However, setup errors are only one part of the daily changes between treatment fractions. Breathing motion, filling and emptying of the stomach, bladder and intestines and so forth cause displacements and deformations of internal organs. These motions should be taken into account when administering RT.

With more patients being cured from their disease, there is an increased focus on limiting the risk of permanent damage of normal tissue. To ensure that an RT-plan is within acceptable risk, the normal tissue complication probability (NTCP) is sometimes calculated from models based largely on the dose to OARs as seen in the pCT. Due to the use of the pCT to calculate dose, these models are also affected by geometrical uncertainties (Ong et al. 2022; Scaife et al. 2015).

Geometrical errors are commonly divided into systematic and random errors. Systematic errors are errors that are constant over all treatment fractions, while random errors are caused by random shifts in geometry around the mean position or shape from fraction to fraction. Systematic errors are introduced during the planning phase, and the typical example is the setup error in the pCT, i.e. the shift in the geometric origin in the pCT compared to the average position of the origin during treatment. Deformations of the patient geometry also cause systematic errors, due to the difference between the shape in the pCT and the average shape during treatment. A relevant example is described in Fontenla et al. (2001b):

> *Consider the following two cases: (1) the rectum happened to be full at the time of the planning scan, thereby displacing the prostate anteriorly, and (2) the rectum happened to be empty at the time of the planning scan, thereby allowing the prostate to move posteriorly. In the first case, the motions of the prostate over the course of therapy, relative to its location on the planning scan, will tend to be systematically posterior over the entire treatment course. In the second case, they will be systematically anterior over the entire treatment course.*

Systematic errors may lead to consistent overdosage to OARs or underdosage to targets at all fractions, while random errors causes a blurring of the dose (van Herk et al. 2000). The models proposed in this thesis are the first of their kind to actually reduce the systematic errors by estimating the mean organ shape from a combination of the pCT and population data. In this case, the remaining systematic errors derive from the difference between the true mean shape and the estimated mean shape, rather than the shape at the pCT.

The approach taken to safeguard against geometrical uncertainties in the clinic is usually to introduce margins around the clinical target volume (CTV). The CTV extended with margins is referred to as the planning target volume (PTV). The margins are commonly based on the well known margin-recipes by Stroom et al. (1999) and van Herk et al. (2000), which require estimates of the population standard deviations of both random and systematic errors for the specific target. These margin recipes are based on exact solutions to certain statistical uncertainty problems that are really only valid under special assumptions such as a rigid and spherical CTV. However, margins can also be generated by statistical models such as those presented in this thesis. The method of Stroom et al. (1999) is especially well suited for this, since it relies on constant *coverage probability*, which is easy to compute with these models.

There are several weaknesses of handling geometrical uncertainties in the PTV concept

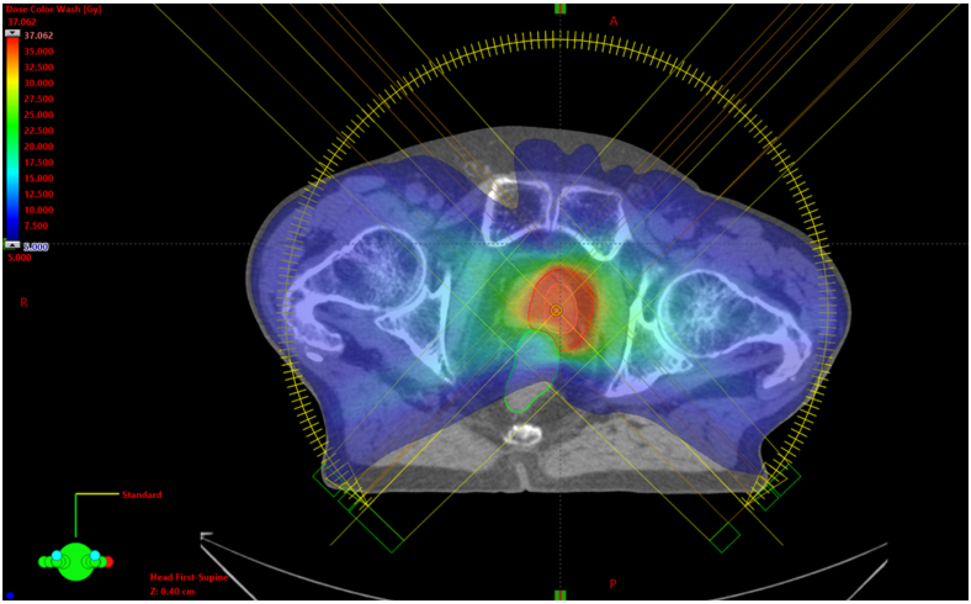

Figure 1.2: Example of a VMAT RT-plan for re-irradiation of recurrent prostate cancer. The PTV has been cropped at the rectum, according to the local clinical procedure for these patients.

(Unkelbach et al. 2018): An underlying assumption is that the dose is conformal within the PTV, so that if the CTV moves only within the PTV, the CTV is given the same dose in every fraction. Heterogeneity of the dose is therefore not accounted for. Further, the same dose is prescribed to the whole PTV even if the probability of the tumor moving to specific locations within the PTV may be low. The concept of margins around OARs, named *planning risk volumes* (PRVs) was proposed in ICRU report 62 (1999). However, the problem of overlap between the PTV and OARs is aggravated by PRVs, which have not gained mainstream acceptance in the clinic (Stroom and Heijmen 2006). Instead, the dominating paradigm for OARs is to assume that the shape of the organ at the pCT is representative, so that dose assessment for that shape is still valid during treatment delivery. Figure 1.2 shows the CTV, PTV and rectum for a patient receiving re-irradiation for recurring prostate cancer. The figure illustrates how the PTV, and therfore the dose, has been cropped at the rectum for fear of inducing adverse effects.

A practical alternative to the PTV concept is *robust radiotherapy planning*, where geometric uncertainties of both targets and OARs are included in the objective function used to optimize the plan. Most early robust RT planning techniques assumed rigid shifts of either the patient (i.e. setup errors) or organs (Li and Xing 2000; Lind et al. 1993; Lof et al. 1995; Unkelbach and Oelfke 2004; Unkelbach and Oelfke 2005). Birkner et al. (2003) introduced deformations into such models, decomposing the organs into a set of deforming tetrahedron-shaped volume elements, but without the ability to actually measure the required probability distribution of the deformation. Baum et al. (2006) presented an efficient approach to robust optimization based on coverage probabilities, which in practice can be derived from both rigid and deformable models. Treatment course simulations, where a simulated anatomy is sampled for each fraction, and the optimization takes all simulated fractions into account (Fredriksson et al. 2021; Xu et al. 2014) is also well suited for statistical deformation models, and so is the method of Sobotta et al. (2010), which is based on an objective function which is evaluated for multiple realizations of the patient anatomy in each iteration. Today, most commercial TPSs offer some kind of robust optimization (Biston et al. 2020).

Geometrical uncertainties in a fractionated treatment can also be reduced if the radiotherapy plan is adjusted to fit the daily anatomical variations, which is the subject of adaptive radiotherapy (ART) (Yan et al. 1999). Examples of ART strategies include re-planning when existing plans fail to meet given criteria, and plan-of-the-day strategies, where, after a preliminary scan (typically CBCT) before each fraction, a plan can be selected from a library of pre-existing plans prepared before start of treatment . Organ deformation models can play a role in adaptive workflows, e.g. through generating likely organ shapes for plan library creation (Nie et al. 2012; Rigaud et al. 2019) or modelling intra-fraction deformations (Fransson et al. 2021; Li et al. 2011b; Liu et al. 2021). For the rectum, intra-fraction motion can be nearly as large as inter-fraction motion if the time of delivery is on the order of 10-20 minutes (Kleijnen et al. 2015). Furthermore, robust planning and evaluation can be valuable in reducing the effort invested in re-planning, as each additional plan comes with a cost (McComas et al. 2022).

1.1 Statistical motion models

Statistical models of organ motion and deformation can be used to improve the precision of margins, for robust optimization and robust evaluation, to increase understanding and visualization of organ motion and to generate plan libraries for adaptive radiotherapy. There has been extensive research into organ deformation models in the recent decades. An early precursor to the algorithms used in this thesis are the 2D point distribution models (PDMs) presented by Cootes et al. (1992, 1994). The statistics of shape coordinate vectors were calculated using principal component analysis (PCA), similarly to the present use. However, all the points needed to be manually drawn and labeled for correspondence in a set of training images. The PDMs were used together with statistics of grayscale levels around each point to locate similar shapes in new images. As such, the algorithm performs 2D auto-segmentation based on local deformable image registration.

Many deformation models in practical use rely heavily on deformable registration of images or contours. Yan et al. (1999) was the first to use deformable registration in organ motion modelling, although that term was not used at the time. They created a mesh representation of the organ volume (not just the surface), and used a physical model of the elastic properties of organ tissue together with finite element modelling (FEM) to deform the mesh of the planning CT into the organ shapes at later CTs. Accumulated dose was calculated by bootstrapping from these shapes together with random sampling from an estimated distribution of setup errors, and summing dose to each node in the mesh over all samples. An early fully deformable model for the rectum in particular was devised by Hoogeman et al. (2002). The rectum is modelled as a deformable tube and unwrapped, slice by slice, to a sheet. Thus one rectum-sheet can be mapped to another, and statistics can be calculated. The method was used to quantify patient-specific variations in rectum shape and dose to the rectal wall.

A patient-specific precursor to the methods in this thesis was presented in Söhn et al. (2005), who used a point distribution model on top of deformable contour registration based on the FEM algorithm of Liang and Yan (2003). They used PCA to visualize the deformation modes of individual patient's rectum, bladder and prostate simultaneously. The method, with variations in registration methods (both image based and contour based) has reached wide adoption in the literature[1]. The downside of the method is that many individual scans are needed to model the patient-specific deformation patterns.

Probably the earliest population-based organ deformation model was presented in Mageras et al. (1996). They used a simplistic algorithm to find point-to-point correspondence in each 2D-slice of organ shapes. Instead of assuming any distribution, the shape changes of a set of reference patients were applied directly to the pCT shape of a new patient, and dose-volume histograms (DVHs) were calculated for each deformation. The end goal was generating uncertainty bars on the new patient's DVHs. Another population based model was derived by Budiarto et al. (2011), based largely on techniques from the intra-patient model by Söhn et al. (2005). This model requires a set of training data from previous patients, and inter-patient contour registration is used in addition to intra-patient registration. The underlying assumption is that a new patient has the same deformation modes as those observed in the training data, on top of a different mean organ shape. They applied the model to the prostate and seminal vesicles, and showed that about 75% of the variance of out-of-sample patients could be explained by the first 15 eigenmodes from the training data. The method can be used when only a

---

[1] E.g. Fransson et al. (2021), Hysing et al. (2011), Söhn et al. (2012), Thor et al. (2013b), Thörnqvist et al. (2013a,b), and Xu et al. (2014) (pelvic organs), Badawi et al. (2010), Dhou et al. (2020), and Li et al. (2011b) (lung), Hysing et al. (2011) (small bowel), Nie et al. (2012) and Tsiamas et al. (2018) (head and neck).

planning CT from the new patient is available. With minor variations, this method has also been widely adopted[2]. A potential improvement to the population model of Budiarto et al. (2011) was presented in Rios et al. (2017). They used a mixed-effect model where the variance of the PCA-modes are calculated based on a combination of the training data and the current patient's data. As such, the goal is similar to that of this thesis. However, there are some major weaknesses of their proposed method: No variance information can be computed from a single scan, and this means that the method is no better than the original population model when only the planning CT is available. Since only the intra-patient deformations are modeled, the method cannot compensate for systematic errors. Further, as in the model by Budiarto et al. (2011), the deformation modes are based on the population only, and not updated for the individual patient, so only deformation patterns present in the training data can be modeled.

The combination of population and individual data was proposed as early as 2001, in Fontenla et al. (2001b), who developed a non-parametric model of the parameters of deformation functions for CT images from a population of patients. The model could combine population and individual variations by including deformations from the current patient in the non-parametric distribution. This has not been shown in practical use though; the model was applied to the prostate, bladder and rectum in Fontenla et al. (2001a), but only for patient-specific data, and only using affine transformations as opposed to fully deformable registration.

In later years, computational methods in medical image processing have largely focused on deep learning. The deep learning studies tend to focus on the application rather than the underlying model, however, in several applications (e.g. autosegmentation (Minaee et al. 2022), deformable image registration (Xiao et al. 2021)), there must necessarily be an understanding of the statistics of organ deformation within the neural network. But the network itself is considered a black box, and we can only examine its input and output. A possible route to modelling deformations using deep learning is the generation of multiple likely organ shapes or CT images from one or more inputs, as proposed by Pastor-Serrano et al. (2022, preprint). A major advantage of their method is that it generates full CT images, which is difficult with the traditional PCA population model (a lone example is Szeto et al. (2017)).

---

[2] E.g. Tilly et al. (2017) (cervix), Argota-Perez et al. (2022) and Tsiamas et al. (2018) (head and neck), Owens et al. (2022) (colon and rectum), Rios et al. (2017) (bladder), Bondar et al. (2014) (rectal CTV) Magallon-Baro et al. (2019) (stomach, duodenum, bowel) Furmanová et al. (2021) (pelvic organs) and Szeto et al. (2017) (thorax).

1.2 Bayesian methods in motion management

Combination of population and individual statistics through Bayesian methods have previously been applied to the problem of setup errors, i.e. rigid shifts of the whole body, in radiotherapy. A Bayesian model for setup errors was first introduced by Lam et al. (2005), further explored by Herschtal et al. (2012, 2015) and van Herk et al. (2009) and recently revisited in Fornacon-Wood et al. (2022a) and Sevillano et al. (2019). The model treats shifts in the x, y and z-direction as independent and normal distributed, and applies a normal-inverse-gamma prior to the mean and variance of the error. If the errors in each direction can be measured during therapy (e.g. through fiducial markers seen on daily CBCTs), the expected mean and variance of the error can be updated, and margins adapted accordingly. The model presented in section 4.6.2 is partially inspired by the work of Lam et al. (2005).

1.3 Rectal toxicity and re-irradiation

The models in this thesis have been evaluated for the rectum of prostate cancer patients. Under prostate cancer RT, rectal toxicity is a major concern. Acute toxicity includes diarrhea, increased frequency of stools, incontinence, rectal bleeding, abdominal cramps, urgency, mucus loss and pain (De Meerleer et al. 2004). The most common late toxicities are rectal bleeding and mucous discharge/leakage (Zelefsky et al. 2008). Giordano et al. (2006) found that 51.3% of patients had some form of gastrointestinal complications five years after RT, versus 31.9 percent of patient who did not receive RT. The the risk of late complications increases exponentially with the volume of the rectum irradiated with high dose (Huang et al. 2002). There may also be fractionation effects (Widmark et al. 2019) and differences between IGRT regimes (McNair et al. 2008; Rudat et al. 2016; Silverman et al. 2016; Singh et al. 2013).

In the calculation of NTCP for the rectum of prostate cancer patients, accounting for motion may provide better predictions (Shelley et al. 2017). Accumulating dose over multiple CT or CBCT scans can provide good results (Ong et al. 2022; Scaife et al. 2015; Shelley et al. 2017), but depend on accurate deformable registration in addition to the drawback of requiring several scans. A model based purely on rigid shifts of the organ shape at the pCT was proposed by Thor et al. (2013a), but did not improve prediction.

Prostate cancer patients with recurrent primary tumor sometimes receive a second round of RT. These patients are at particularly high risk for toxicity (Munoz et al. 2021). The tissue repair that occurs between the primary treatment and re-irradiation will increase radiation tolerance to some extent, but exactly how much is uncertain (Abusaris et al. 2012). There is also significant difficulty in estimating the accumulated dose of the two treatments locally in the OARs, and inter-fraction organ deformation increases this complexity further. Organ deformation models therefore have a potential to improve decision making in this context.

Chapter 2

Background: Bayesian inference

Inference means deriving conclusions from reasoning and evidence, and Bayesian inference is the process of using data and models to derive conclusions based on Bayes' rule. In this process, Bayes' rule is used to update belief in a hypothesis when new data becomes available. The antithesis of Bayesian inference is *frequentist inference*, which deals with p-values and confidence intervals. For the interested reader, discussions of the differences are found in e. g. Efron (2005) and Fornacon-Wood et al. (2022b).

Bayes' rule (or Bayes' theorem) is given by

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{Pr(A)}, \tag{2.1}$$

where A and B are two events, $Pr(\cdot)$ represents a probability of an event, and $Pr(A|B)$ refers to the probability of A *conditioned on* B, i.e. the probability of event A given that B has occurred. This formula, named after the English Reverend Thomas Bayes (1701-1761), can be used to find conditional probabilities given some prior information. In (2.1), $Pr(B)$ is the prior probability of the event B, i.e. the probability of B without information about A.

Bayes' theorem can be used for discrete numerical random variables, for example, A could be the event that $X = 1$ while B is the event that $Y = 2$. However, for continuous random variables, the probability of a specific value is always zero. Luckily, Bayes theorem also works for probability *densities*. The probability density of X at a value x is defined as

$$f_X(x) = \lim_{\Delta x \to 0} \frac{Pr(X \geq x \quad \text{and} \quad X < x + \Delta x)}{\Delta x}. \tag{2.2}$$

In words, the probability density at x is the probability that the value of X falls in an interval around x, divided by the size of that interval, as the size of the interval goes to zero. The probability density-version of Bayes theorem is

$$f_{X|Y}(x|Y=y) = \frac{f_{Y|X}(y|X=x) f_X(x)}{f_Y(y)}, \tag{2.3}$$

The shorthand $f(x|y)$ is commonly used for $f_{X|Y}(x|Y=y)$. In Bayesian inference, the y in equation (2.3) is observed data. The denominator $f(y)$ is constant as a function of x, and the equation can often be solved without knowledge of $f(y)$ since the left-hand side of (2.3) must adhere to the nature of probability density functions (pdfs), that all probabilities must integrate to 1, i.e. $\int_{-\infty}^{+\infty} f(x|y)dx = 1$. Taking the integral with respect to x of both sides of (2.3), we then find $f(y) = \int_{-\infty}^{+\infty} f(y|x)f(x)dx$.

Example

A classical example application of Bayes rule is the probability that a person carries a disease given that some test for the disease comes out positive. The accuracy of a test is often specified through its sensitivity (the probability of correctly identifying a positive individual) and its specificity (the probability of correctly identifying a negative individual). Let A be the event that the test comes out positive, and B be the event that a the test subject actually carries the disease. Then $Pr(A|B)$ is the the sensitivity of the test, $Pr(B)$ is the prevalence of the disease in the population, and $Pr(A)$ is the overall probability of a positive test. The specificity is the probability of a negative test given that the individual does not carry the disease, i. e. $Pr(\neg A|\neg B)$. The overall probability can be calculated using the specificity, since

$$\begin{aligned} Pr(A) &= Pr(B)Pr(A|B) + Pr(\neg B)Pr(A|\neg B) \\ &= Pr(B)Pr(A|B) + (1 - Pr(B))(1 - Pr(\neg A|\neg B)). \end{aligned}$$

Given a test with a perfect sensitivity of 100% and a specificity of 99%, for a disease with a prevalence of 1 in 1000 people, we can insert these numbers into Bayes theorem:

$$Pr(B|A) = \frac{(1/1000) \cdot 100\%}{100\% \cdot (1/1000) + (1 - 1/1000) \cdot (1 - 99\%)}. \tag{2.4}$$

The (perhaps surprising) result is that an arbitrary person who gets a positive test result actually has only 9.1% probability of actually having the disease.

In equation (2.3), $f(x)$ is called the *prior* distribution, since the pdf of X represents the prior knowledge of X, i.e. what we know of X without any information about Y. Similarly, $f(x|y)$ is called the *posterior* distribution, since it represents our knowledge of X when the data y, which is a realization of Y, is given. The factor $f(y|x)$ in the numerator looks like a pdf, but in reality, when it is used to calculate the posterior, y is constant and we consider $f(y|x)$ a function of x. Since $f(y|x)$ as a function of x is not a probability density, yet is strongly linked to probability, it has been given the name *likelihood*. To emphasize the fact that the likelihood is a function of x, it is sometimes written as $\mathcal{L}(x; y)$ or variants thereof. Using this terminology, we say that the posterior is proportional to the prior times the likelihood,

$$f(x|y) \propto f(x)\mathcal{L}(x; y), \tag{2.5}$$

and, again, proportionality is sufficient to determine a distribution.

To take the simplest possible example, say a scalar Y is normally distributed with unknown mean m and known variance σ^2:

$$Y|m \sim \mathcal{N}(m, \sigma^2). \tag{2.6}$$

In the Bayesian context, we treat the unknown m as a realization of a random variable M. Let the prior distribution of M also be normal

$$M \sim \mathcal{N}(\mu_0, \sigma_0^2). \tag{2.7}$$

The parameters of the prior, in this case μ_0 and σ_0^2, are called *hyperparameters*. An expression for the posterior distribution of M given a realization y of Y is found by inserting (2.6) and (2.7) into (2.5):

$$f(m|y) \propto \mathcal{N}(m; \mu_0, \sigma_0^2) \cdot \mathcal{N}(y; m, \sigma^2) \tag{2.8}$$

The calculations are somewhat intricate, but it turns out that the posterior is itself normal:

$$f(m|y) = \mathcal{N}(m; \mu_0', \sigma_0'^2), \tag{2.9}$$

with

$$\sigma_0'^2 = (\sigma^{-2} + \sigma_0^{-2})^{-1} \tag{2.10}$$

$$\mu_0' = \sigma_0'^2 (\frac{\mu_0}{\sigma_0^2} + \frac{y}{\sigma^2}). \tag{2.11}$$

Instead of calculating an estimate of M, as we might have done in a frequentist context, we have found its distribution. The expected value is μ_0', but it is associated with a variance $\sigma_0'^2$ which we interpret as *uncertainty*. This uncertainty can be used to calculate the the *credible interval*, the Bayesian parallel to the frequentist confidence interval.

In this example, the prior for M was normal, and its posterior is also normal. When the posterior is in the same distribution family as the prior, it is said that the prior is *conjugate* to the likelihood. Conjugate priors have the advantage that their parameters can be calculated directly from the data and the hyperparameters, as in (2.10)-(2.11).

Mathematical tractability is therefore the reason conjugate priors are usually chosen when no strong considerations suggest other priors. With non-conjugate priors, on the other hand, it is usually difficult or impossible to find the posterior analytically. Instead, one can generate samples from the posterior using Markov Chain-Monte Carlo (MCMC) techniques. For example, a Metropolis sampler can generate samples for any distribution as long as pdf-ratios (such as $f(\mu_1)/f(\mu_2)$) can be calculated. Since constants cancel in this ratio, sampling can be performed for a posterior distribution through Bayes rule when the prior and likelihood is given.

The posterior is the distribution of the mean M, which is a parameter of the likelihood. However, we are often more interested in the distribution of Y given a realization y. This is called the *posterior predictive* distribution. An analytical expression for the posterior predictive cannot always be found, however, samples of the posterior predictive can be drawn by first drawing a parameter value from the posterior distribution and then drawing a new value from the conditional distribution (2.6) using this parameter value. In our simple example, the posterior predictive can be found analytically, and is given by

$$Y|y \sim \mathcal{N}(\mu_0', \sigma^2 + \sigma_0'^2). \tag{2.12}$$

The posterior predictive has a higher variance than the sampling distribution (2.6) because it takes into account the uncertainty of M. When the posterior predictive is unknown, or overly complex, it is nevertheless not uncommon to simply plug in point-estimates of the unknown parameter(s) into the sampling distribution (2.6), thus ignoring the increased variance due to model uncertainty. This is part of the empirical Bayes framework.

The example with an unknown mean and known variance for a univariate Gaussian likelihood is one of the the simplest possible cases. In this thesis, distributions of high-dimensional vectors with unknown vector and matrix-valued parameters are used. The concepts are nevertheless the same as those shown here.

2.1 Shrinkage estimation

Bayesian models tend to lead to so-called *shrinkage estimates* of parameters. Consider the previous example: if we let the prior mean μ_0 be zero, then equation (2.11), the estimate of the posterior mean, becomes

$$\mu_0' = \frac{y}{1 + \sigma^2 \sigma_0^{-2}}. \tag{2.13}$$

Since $\sigma^2\sigma_0^{-2}$ is positive, the absolute value of μ_0' is guaranteed to be smaller than that of y. Since y itself is the frequentist estimate of the mean, this leads to the term *shrinkage estimation.* If the prior mean is not zero, the shrinkage estimate is not guaranteed to be smaller, but rather to be closer to the prior mean. This challenges our intuitive understanding of the term shrinkage, but we still say the estimate has been shrunk towards the prior mean. Paper 1 is based on simple shrinkage towards the prior mean in high dimensions.

2.2 Mean-field variational inference

In high dimensions, and especially when random matrices are involved, MCMC-sampling is very computationally intensive. An alternative set of techniques that can sometimes be used when there is no analytical expression for the posterior are *variational Bayes* methods. These are techniques used to find a closed form function that approximates the pdf of the true posterior. The name comes from the application of calculus of variations in the derivations of the approximations. In variational inference, we seek an approximate posterior of the latent variable x given data y, i. e. $q(x) \approx f(x|y)$. Given a class Q of candidates for $q(x)$, the approximation is cast as an optimization problem: Minimize the Kullback-Leibler divergence from $q(x)$ to $f(x|y)$,

$$q^*(x) = \arg\min_{q \in Q} \mathrm{KL}(q(x)||f(x|y)). \tag{2.14}$$

The Kullback-Leibler divergence is an information-theoretic asymetric measure of difference between two pdfs, say f and g, defined as

$$\mathrm{KL}(f||g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx, \tag{2.15}$$

which is zero when the pdfs are identical.

The most common form of variational Bayes is the *mean field approximation*, where the elements of the vector x is split into different subsets, and the approximation is such that each subset is independent from the others. Thus, if $x = \{x_1, x_2, \dots x_n\}$, where x_i are subsets subset of the variables of the full set x, then

$$f(x|y) \approx q(x) = \prod_{i=1}^{n} q_i(x_i) \tag{2.16}$$

Here, the class Q of allowable densities are those that satisfy the equality. A *coordinate ascent* algorithm exists for updating each $q_i(x_i)$ based on the latest updates of the other functions $q_j(x_j)$ for $j \neq i$ (Blei et al. 2017, p. 10). In the general case, this algorithm is purely theoretical, as it updates continuous functions as opposed to their parameters — not something that can be done on a computer. However, in the special case where the distributions of x_i are *conditionally conjugate* with respect to the likelihood, and are in the exponential family of distributions, the algorithm becomes computationally feasible. Recall that a prior distribution is said to be conjugate to a specific likelihood if the posterior is in the same distribution family as the prior. A distribution is said to be conditionally conjugate if the posterior distribution of one partition conditioned on other partitions is in the same family $\mathcal{F}_i$ as the prior for that partition (Gelman et al. 1995), in other words if:

$$\begin{aligned} f(x_i|x_{j\neq i}) &\in \mathcal{F}_i \text{ for all } i && (2.17) \\ \rightarrow f(x_i|y, x_{j\neq i}) &\in \mathcal{F}_i \text{ for all } i. && (2.18) \end{aligned}$$

This concept becomes clearer with an example: Let the likelihood be gaussian $\mathcal{N}(y; \mu, \sigma^2)$ with unknown mean and variance with prior pdf $\mathcal{N}(\mu; \mu_0, \sigma_0^2)\Gamma^{-1}(\sigma^2; \alpha, \beta)$, i.e. the mean is normal distributed, the variance is inverse-gamma distributed, and they are independent. This is not a conjugate prior. However, when conditioning the posteriors of each of the parameters on the other parameter, they appear conjugate:

$$\begin{aligned} \mu|y, \sigma^2 &\sim \mathcal{N}(\mu_0', \sigma_0^{2\prime}) && (2.19) \\ \sigma^2|y, \mu &\sim \Gamma^{-1}(\alpha', \beta'). && (2.20) \end{aligned}$$

Expressions can be found for the parameters $\mu_0', \sigma_0^{2\prime}, \alpha'$ and β', but these will depend on the given μ or σ^2. The mean field approximation replaces these conditional distributions with unconditional ones. The procedure works as follows: First, find q_u and q_σ by taking the expectation of the logarithm of the pdf of (2.19) with respect to σ^2 and of (2.20) with respect to μ:

$$\begin{aligned} \ln q_\mu(\mu) &= E_{\sigma^2}[\ln \mathcal{N}(\mu; \mu_0', \sigma_0'^2)] + \text{const} && (2.21) \\ \ln q_\sigma(\sigma^2) &= E_\mu[\ln \Gamma^{-1}(\sigma^2; \alpha', \beta')] + \text{const} && (2.22) \end{aligned}$$

This results in expressions for the posterior hyperparameters of μ which depends on moments of σ^2 and vice versa. The coordinate ascent algorithm then proceeds as follows: Calculate the parameters of the pdf of μ based on the current guess of the relevant moments of σ^2. Calculate the relevant moments of μ based on these parameters. Calculate the parameters of the pdf of σ^2 based on the moments of μ. Calculate the relevant moments of σ^2 from the parameters. Repeat until convergence. More detailed explanations with examples can be found in e.g. Bishop (2006) and Blei et al. (2017).

Mean field variational inference was used in paper 2, and the derivation of the algorithm for the specific model is found in paper 2, Appendix B.

Chapter 3

Aims of the thesis

There is room for improvement in organ deformation modelling, as current models either require many individual scans as input, or lack the ability to tailor the model to the individual deformation patterns. The overall aim of this project was therefore to improve upon existing statistical models for organ deformation by introducing Bayesian concepts, and to utilize both population and patient-specific data in these models.

Specifically, the aims of the papers were:

Paper I: To reduce systematic errors by estimating the patient-specific mean shape of recti of prostate cancer patients using a single measurement, i.e. the pCT, combined with population data.

Paper II: To introduce new specific methods for statistical organ deformation modelling using Bayesian techniques to combine population and patient-specific data.

Paper III: Enabling the evaluation of deformation models using coverage probability matrices by introducing a software package for fast solving of the point-in-polyhedron problem.

Paper IV: To estimate accumulated dose with uncertainties for the rectum of patients receiving re-irradiation for recurrent prostate cancer, using models from paper 2.

Chapter 4

Materials and methods

4.1 Image data

All methods presented in the papers were evaluated for recti of prostate cancer patients. For implementation and evaluation of the methods in paper I and II, we used a data set consisting of CT scans from prostate cancer patients enrolled in a clinical study at Haukeland University Hospital in 2007[1](Ekanger et al. 2020; Hysing et al. 2018). The patients were treated with moderately hypofractionated IMRT with three dose levels delivered as a simultaneous integrated boost. The treatment consisted of 25 fractions of 2.7 Gy each. Twice a week, a CT scan (resolution 0.7mm × 0.7mm × 2mm) was acquired to monitor organ changes. The patients did not follow any particular preparation protocol before the pCTs, repeat CTs or treatment fractions. In each CT, the RT-targets comprising the prostate, seminal vesicles and lymph nodes, were contoured by a senior oncologist. The organs at risk, specifically the rectum, bladder, large and small bowel, were contoured by expert physicists. For fear of adverse effects due to the moderate hypofractionation, the PTV was cropped at the rectum. The study showed that there was in practice little rectal toxicity, and this practice of cropping has since been abandoned in our clinic.

The repeat CT-regime was adopted for 41 of the 97 patients enrolled in the study. Of these, 37 patients were included in the final data set. In total, 373 scans (including pCTs) were used in the evaluation.

In paper III, the same data set was used for training the deformation model. The model was then applied to a data set from a different study, consisting of pCTs and RT-plans

---

[1] Approved by the regional ethics committee in western Norway, REK no. 2006/15727. For the current project, all patient data had been anonymized.

for 39 patients who received re-irradiation for recurring prostate cancer 5-10 years after primary RT[2]. A pCT for the primary RT could be retrieved for 28 of the patients, for which the treatment was heterogeneous. The re-irradiation consisted of 5 fractions of 7 Gy to the recurrent tumour in the prostate.

4.2 Data representation

The data used for the statistical models in this thesis are the hand-segmented contours of structures seen in the CT-scans. During contouring, the expert draws the structure slice by slice in 2D. The structure is stored together with the CT-image in the DICOM file format (NEMA 2022). Each slice of the structure is stored as one or more 2D-polygons (even though the 3D-structure is continuous, a in a single 2D-slice it can appear as several unconnected structures). A visual example of this kind of representation is shown in figure 4.1 A.

The polygonal representation is not ideal for deformation model, because the three dimensions are not represented equivalently. While a vertex of one of the polygons can be moved around freely in the x-y plane without breaking the structure, it cannot be moved in the z-direction. We therefore wish to convert the structure to what is usually referred to as a "mesh" representation. In reality, the mesh is a polyhedron; a 3D-surface defined by a set of contiguous polygonal (in practice triangular) faces. The computer representation of the polyhedron typically comprises two arrays: one array of vertex coordinates, and one of triangular faces (in terms of pointers to the three connected vertices for each face). In the deformation models in this thesis, only the vertices are used. As such, the organ representations within the deformation models can be thought of as "point clouds". However, the faces are necessary for volume calculation, and in general for processing the results and converting back to DICOM format.

There are different ways of converting from the format using multiple 2D polygons to a polyhedron, and the relative straightforward procedure we used is described in the following section.

4.2.1 Conversion from slices to mesh

To convert a 2D-polygonal structure to a 3D-mesh, the structure was first converted to a binary mask, i.e. a 3D array of voxels where each voxel has a value of 0 if it

---

[2] Approved by the regional ethics committee in western Norway, REK no. 2012/1868.

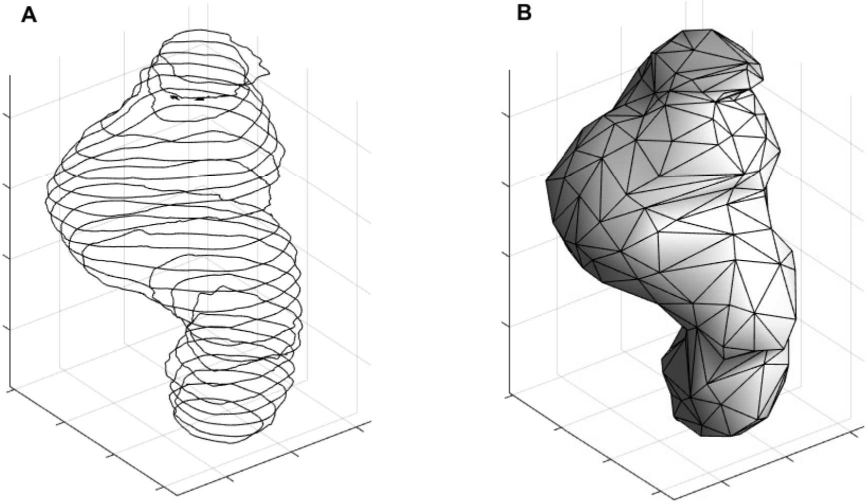

Figure 4.1: Data representations for an example rectal shape. A: 2D polygon representation (DICOM). B: Polyhedral (mesh) representation. A coarse mesh is used for illustration purposes.

is outside the structure and 1 if it is inside the structure. This was done in 2D slice by slice by finding pixels inside or outside the polygon structure, using a ray tracing algorithm. The slices were then stacked to form a 3D structure. The resolution in the x-y plane is arbitrary, and is chosen when running the ray-tracing algorithm, while the resolution in the z-direction is determined by the input structure (in practice by the axial CT-resolution).

Now the binary mask is interpreted as a discrete-valued function $s = f(x, y, z)$, where the value s is either zero or one. We are looking for a contour, also known as an isosurface, of this function. An isosurface is the region where the value of the function crosses a given threshold — a well known example in 2D are height contours of a map. To find a smooth surface, the discrete valued function f was converted to a continuous-valued function by applying a 3D low-pass filter. Finally, the *marching cubes* (Lorensen and Cline 1987), a well known and widely available algorithm, was used to find the isosurface of the function as a polyhedron.

The properties of the filter and the marching cubes threshold were chosen so that the result was a reasonably smooth structure, and such that the original polygon vertices were as close as possible to the polyhedron surface. In practice, high precision of this transformation is not of great importance to the end result, due to the two-iteration deformable registration described in section 4.3.1

4.2.2 Conversion from mesh to slices

For exporting structures to a treatment planning system, they must first be converted back to DICOM format. For each slice, we find the polygon around the polyhedron at a constant z-level. Every vertex in the polygon is a location where the z-plane crosses an edge of one of the faces of the polyhedron. This is conceptually relatively straightforward, yet programmatically a bit more involved. Matlab™-functions for this procedure have been published to the mathworks file exchange (Rørtveit 2022).

4.3 Contour registration

4.3.1 Rigid registration

Rigid registration is the process of aligning the coordinate systems between images. For each patient, all CT images were rigidly aligned to gold markers implanted in the patients' prostates. Between patients, a "center of gravity" approach was used: The planning CTs were aligned such that the average of all vertex coordinates matched. Only the lower (caudal) half of the recti was used when calculating the center of gravity. This was done because we observed that the cranial region, where the variation and uncertainty is greatest, negatively affected the matching.

No rotation was performed in the rigid registration.

4.3.2 Deformable registration

Deformable registration is a key element in all the methods in this thesis. Deformable registration is the process of deforming a shape so that it matches a different shape. In this thesis, deformable *contour* registration is used, as opposed to deformable *image* registration (DIR). In DIR, the image intensities are used as a guide to deform one image to resemble another, whereas contour registration is purely shape-based, and the shape of a point cloud is deformed to match the shape of another point cloud. The registration was performed with the *Matterhorn* software, developed at the Erasmus MC cancer institute in Rotterdam. The deform registration in Matterhorn uses a Thin-Plate-Spline Robust Point Matching algorithm (Chui and Rangarajan 2003) based on the work by Vásquez Osorio et al. (2009).

The algorithm finds a transform function (a thin-plate-spline transform) which "warps" the original shape into a shape similar to the target shape. The thin-plate-spline transform warps not just the surface shape, but all of space, as shown in Figure 4.2. Therefore, it can be used to interpolate and extrapolate the motion of points interior and exterior of the organ as well.

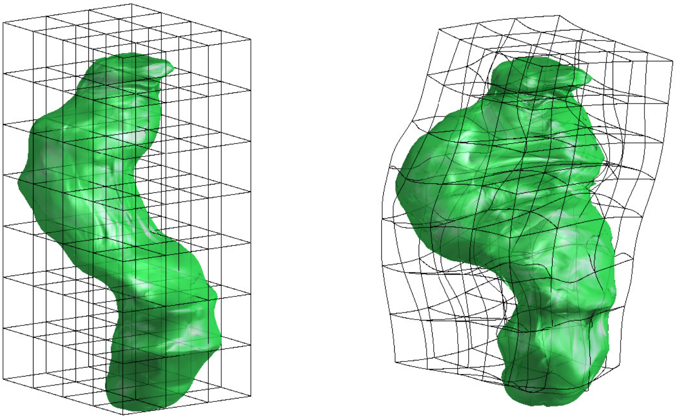

Figure 4.2: Thin-plate-spline-transformation of the rectum from the planning CT to the first treatment fraction for an example patient. The grid illustrates how the thin plate spline deforms not only the surface points, but all of space, and thus can be used for interpolation of interior points.

By replacing the target shape by the warped original shape, point-to-point correspondence between the two shapes is found, as illustrated in Figure 4.3.

The process of deformably registering every shape in the data set is shown in Figure 4.4. For each patient, deformable contour registration was performed between the pCT and all repeat CTs (intra-patient registration). Furthermore, inter-patient registration was performed by registering a reference patient's pCT to all the other pCTs in the data set. The reference pCT was warped to recreate the other pCTs, and the results were further warped using the intra-patient transform functions to recreate the repeat CTs. The precision of the inter-patient registration is lower than that of the intra-patient registration, because there is greater variation in shapes between patients than within the data for one patient. Therefore, we performed a second iteration of inter-patient registration based on a new reference shape built by averaging all shapes in the data set, as seen in Figure 4.4. The warping was then repeated from the new reference shape to recreate all shapes in the data set. This second iteration improved the precision of the inter-patient registration so that it was on par with the intra-patient registration.

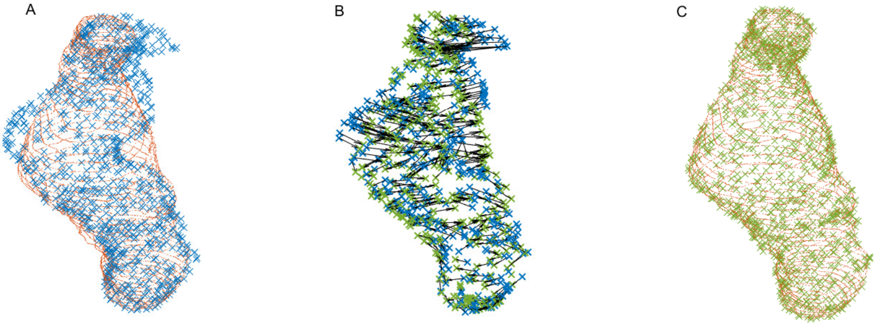

Figure 4.3: Illustration of how deformable registration finds point-to-point correspondence. A: The original shape as blue crosses, the target shape as red dots. B: Displacement vector field (DVF) found by the deformable registration algorithm as black arrows, with original shape as blue crosses, and warped original shape as green crosses. Point-to-point correspondence is found because the target shape is replaced by the warped original shape. C: "Warped" original shape as green crosses, target shape as red dots.

4.4 Statistical distribution of shapes

To create a statistical model for shapes represented by P surface points, we gather the x, y and z coordinates of the points into *shape vectors*:

$$s = [x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_P, y_P, z_P]^T. \tag{4.1}$$

With this representation, we can use standard multivariate statistical distributions. An assumption for all models discussed in this thesis is that, for a specific patient, the shape coordinates follow a multivariate Gaussian distribution:

$$s \sim \mathcal{N}(\mu, R). \tag{4.2}$$

The mean shape vector μ represents the patient's mean organ shape, and the covariance matrix R describes the variance of the coordinates as well as the covariance between each pair of coordinates.

This data representation and data distribution underlies the patient-specific model by Söhn et al. (2005), the population model by Budiarto et al. (2011) and the Bayesian models in this thesis and the associated papers. The essential difference between the models is how μ and R are estimated.

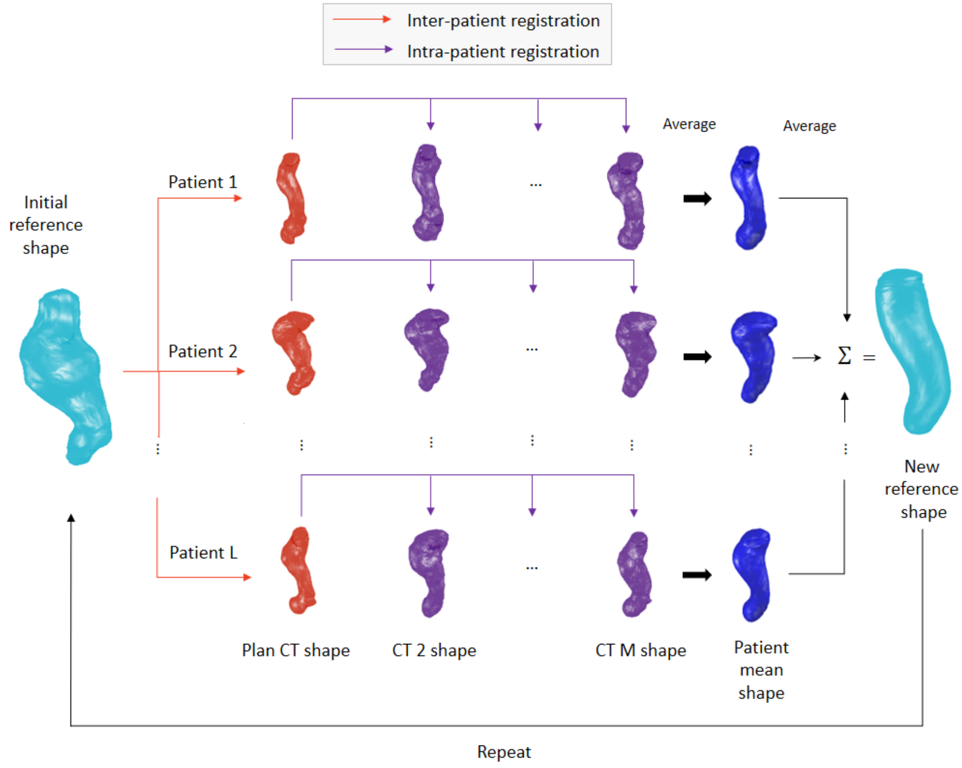

Figure 4.4: Illustration of the inter- and intra-patient deformable registration process.

4.4.1 Patient-specific covariance matrix

The (unbiased) sample covariance matrix for a specific patient is given by

$$R_{\text{ps}} = \frac{1}{J-1}\sum_{j=1}^{J}(s_j - \bar{s})(s_j^T - \bar{s}^T), \tag{4.3}$$

where s_j, $j = 1 \ldots J$ is the sample of shapes for that patient, and $\bar{s}$ is the sample average, i.e. $\bar{s} = \frac{1}{J}\sum_{j=1}^{J} s_j$. Equation (4.3) can be written in matrix form as

$$R_{\text{ps}} = \frac{1}{J-1} SS^T, \tag{4.4}$$

where S is a matrix whose columns are the vectors $s_j - \bar{s}$, i.e.

$$S = [s_1 - \bar{s} \quad s_2 - \bar{s} \quad \ldots \quad s_J - \bar{s}]. \tag{4.5}$$

In high-dimensional applications, the sample covariance matrix is in practice never constructed. Instead, the covariance matrix is represented implicitly by the sample matrix S, which contains the same information.

The model by Söhn et al. (2005), which we refer to as the patient-specific model, uses this covariance matrix for R and the individual sample mean for μ in equation (4.2).

4.4.2 Population covariance matrix

While the sample covariance matrix discussed in section 4.4.1 is well known and common across a wide range of applications, the population covariance matrix discussed here is specific for the problem at hand. Under the assumption that the shapes of patients' organs differ in the mean, but share the same pattern of variation around this mean, we can estimate the common population covariance matrix as the average of all patient-specific covariance matrices in a data set:

$$R_{\text{pop}} = \frac{1}{M} \sum_{i=1}^{M} R_{\text{ps},i}, \tag{4.6}$$

where $R_{\text{ps},i}$ is the sample covariance matrix of the ith patient.

Equation (4.6) can also be written in terms of sample-matrices as

$$R_{\text{pop}} = \frac{1}{M} S_{\text{pop}} S_{\text{pop}}^T, \tag{4.7}$$

where S_{pop} is a matrix which contains all patient-specific sample matrices S_i:

$$S_{\text{pop}} = [\frac{1}{\sqrt{J_1 - 1}} S_1 \quad \frac{1}{\sqrt{J_2 - 1}} S_2 \quad \dots \quad \frac{1}{\sqrt{J_M - 1}} S_M], \tag{4.8}$$

where J_i is the number of samples for patient i.

Despite R_{pop} being a population covariance matrix, it is still an *intra-patient* covariance matrix, since it describes the variation from fraction to fraction for one patient rather than the variation between patients.

The model by Budiarto et al. (2011), which we refer to as the population model, uses this covariance matrix for R and the individual sample mean for μ in equation (4.2).

4.5 Principal Component Analysis

PCA is a popular technique in problems involving high-dimensional vectors. PCA finds a representation of the sample covariance matrix comprising a set of orthogonal vectors, called principal components (PCs), along with the variance of the data set in the direction of each of these vectors. The first PC is the direction in which the data varies most. The second PC is the direction in which the data varies most — in the subspace orthogonal to the first PC. The third PC is the direction in which the data varies most in the subspace orthogonal to both the first and the second PC, and so on. The purpose of PCA varies between applications, but in organ deformation models the the purpose is usually one or more of the following:

Reduce overfitting The uncertainty of the sample covariance matrix depends on the number of samples and the number of dimensions. In high-dimensional problems, the number of samples is almost always much smaller than the number of dimensions, and the uncertainty is therefore very high. In PCA, the PCs with variance less than some tolerance is typically discarded. In essence, we are only modelling the subspace which we are most certain about, and simply saying we know nothing of how the data behaves in the remaining space - only that there is less variance there than in the principal subspace. This reduces overfitting of the covariance matrix, because the discarded space presumably corresponds to features that are specific to the small available data set.

Enable/simplify calculations Under PCA, data vectors can be approximated by a low-dimensional vector of *PCA scores*, which are the coefficients of the vector expressed in the basis of the PCs. This representation is easy to find due to the orthonormality of the PCs. Furthermore, all coefficients are uncorrelated to each other, which often simplify calculations by allowing the use of a few scalars rather than vectors.

Visualization The covariance matrix may contain millions of values. To gain an understanding of the information it contains, it is necessary to visualize it in some way. By reducing the covariance matrix to a set of “modes” (PCs), PCA makes it possible to visualize typical variations of the distribution. Figure 4.5, taken from paper II, shows how this works for a population model of the rectum.

The PCs are the eigenvectors of the sample covariance matrix, and can, in principle, be computed by numerical eigenvalue decomposition. Often, the covariance matrix is too large to be constructed, and the PCs are computed in some other way. For example, the

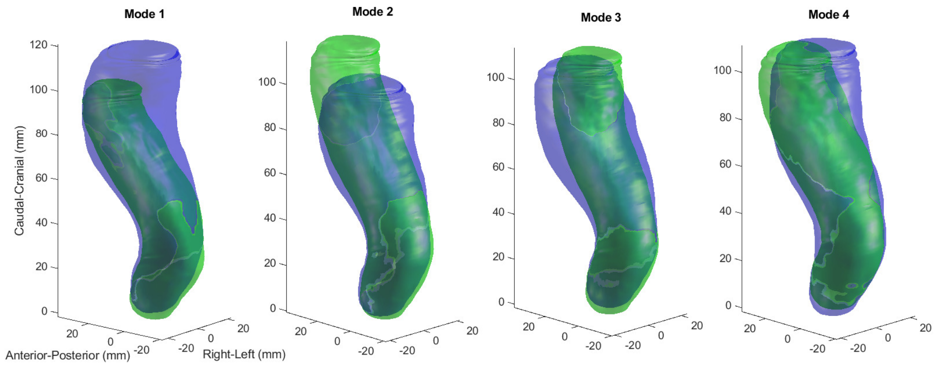

Figure 4.5: Visualization of the first four PCA-modes for a population model for the rectum. The blue and green shapes represent plus and minus two standard deviations of each deformation mode relative to the population average rectum shape.

reduced singular value decomposition (SVD) of the sample matrix S gives the PCs and their standard deviations.

Defining a matrix W whose columns are the PCs, and the diagonal matrix Σ whose diagonal elements are the corresponding eigenvalues, R is related to W and Σ through

$$R = W\Sigma W^T. \tag{4.9}$$

There are $J-1$ non-zero eigenvalues if the covariance matrix is based on J shapes. Thus, the dimensions of W are $3P \times J-1$, where $3P$ is the number of elements in the sample vectors (representing P points with three coordinates each), while Σ is $J-1 \times J-1$. However, it is common to discard some of the PCs corresponding to the smallest eigenvalues. In this case, (4.9) becomes an approximation instead of an equality.

The PCA-scores of a vector s are the coefficients of the vector expressed in the basis of W. Since W is orthonormal, we can write the score vector y as

$$y = W^T(s-\bar{s}). \tag{4.10}$$

Each element of the vector y represents a PCA-score for the corresponding PC. The vector y has mean zero and covariance matrix Σ. Since the elements of y are uncorrelated, Σ is diagonal. The vector y is commonly used as a low-dimensional approximate representation of s. The input s cannot be recreated exactly, but can be approximated:

$$s = \bar{s} + Wy + \epsilon \approx \bar{s} + Wy. \tag{4.11}$$

Here, ϵ is a vector of errors - the part of s that cannot be represented by the PCs which are the columns of W. PCA has the property that it minimizes the variance of the error vector ϵ.

4.6 Bayesian deformation models

In the following, a short presentation of three Bayesian deformation models is given. The first method, presented in section 4.6.2, was the first to be developed during the PhD-project, and has not been published. Nevertheless, it serves as an important piece of the chronicle of how the methods in both paper I and paper II came to be. The latter two, in sections 4.6.4 and 4.6.5, are the models introduced in paper II. The mean shape estimation in paper I is also briefly discussed.

As in section 4.4, the distribution of an individual patient's organ shape is modelled by the random vector s which is distributed as

$$s \sim \mathcal{N}(\mu, R). \tag{4.12}$$

In a frequentist interpretation, μ and R are constant, but unknown. They can be estimated from data, albeit with some uncertainty. In the Bayesian interpretation, there are no "true" parameters. Rather, the parameters are stochastic and governed by a probability distribution. The distribution reflects the best of our knowledge about the parameters, and is updated when we get new knowledge through new data.

4.6.1 R and μ as random and systematic error

The mean vector μ represents the average shape of the patient. Any deviation between the estimated and the true mean represents a systematic error. The covariance matrix R, on the other hand, describes how the organ shapes may change from fraction to fraction in terms of the variance of each point coordinate and the covariance between the coordinates. The variation over the mean from fraction to fraction is the random error. In this context this is a bit of a misnomer, as the systematic error is also, statistically speaking, random. It would not be quite correct to say that μ and R represent the systematic and random errors, respectively, but it is certainly true that μ is related to the systematic error while R is related to the random error.

To build a Bayesian model, either μ, R or both can be considered stochastic. Any of these choices would create a novel method. Letting R be stochastic, an update equation can be found that updates the distribution of R when given new data, thus combining population and patient-specific data to model the random error. Letting μ be stochastic, an update equation for the distribution of μ when given new data is found, which both models and reduces the systematic error; something that has not been done in previous models. In the following, we have chosen to consider both μ and R stochastic.

4.6.2 A Bayesian deformation model based on PCA

The general idea of the following model is that the mean and variance of the PCA-score of each mode follows a separate distribution. This way we avoid the complexity of matrix distributions.

The population model of Budiarto et al. (2011) assumes that all patients share the same covariance matrix. This covariance matrix is represented by PCs and their variances. We loosen this assumption: suppose that the PCs are constant, but their variances vary between patients. The independence between PCs makes this model particularly desirable, because independent scalar distributions can be used for each PC.

This method relies on the assumption that the PCs of the intra-patient variation, i.e. the variation from fraction to fraction for one patient, are the same as the PCs of the inter-patient variation, i.e. the variation of mean organ shapes between patients. The variance of the PCA-scores can, however, be different between the inter-patient and intra-patient level. Furthermore, we assume that the intra-patient PCs are the same for each patient, but that their means and variances differ from patient to patient.

Given a set of such PCs ψ_k, $k = 1 \dots K$, the PCA-scores c_k for a shape vector s are given by

$$c_k = \psi_k^T (s - \mu_0), \tag{4.13}$$

where μ_0 is the population mean shape. The shape vector s can be approximately reconstructed from the scores c_k and the components ψ_k,

$$s \approx \mu_0 + \sum_{k=1}^{K} c_k \psi_k \tag{4.14}$$

Note that the computation of PCA-scores in (4.13) differs from that of the population model by Budiarto et al. (2011), in that it is the population mean μ_0 rather than the patient mean $\bar{s}$ that is subtracted from s. This means that the scores are non-zero even when only a single observation is given. Furthermore, although the expected values of the scores are zero at a population level, the expected value for a specific patient is non-zero. We model the scores c_k for a given patient as normal distributed according to

$$c_k|\mu_k, \lambda_k \sim N(\mu_k, \lambda_k) \tag{4.15}$$

The scores $c_1 \dots c_K$ are then independent by PCA theory. At the population level, the mean of c_k is 0 by the definition of PCA, while the variance is λ_k, the kth largest eigenvalue of the population covariance matrix. However, we want to find the patient-specific mean and variance of c_k for each k. We do this in a Bayesian fashion by proposing that the mean μ_k and variance λ_k of c_k are random, and distributed according to a normal-inverse-gamma (NIG) prior distribution:

$$\mu_k, \lambda_k \sim N\Gamma^{-1}(m_k, \tau_k, \alpha_k, \beta_k). \tag{4.16}$$

In a NIG distribution, λ_k is drawn from an inverse gamma distribution,

$$\lambda_k \sim \Gamma^{-1}(\alpha_k, \beta_k), \tag{4.17}$$

where α_k and β_k are hyperparameters called the scale and shape parameters. The distribution of μ_k depends on λ_k, and is Gaussian given λ_k:

$$\mu_k|\lambda_k \sim \mathcal{N}(m_k, \frac{1}{\tau_k}\lambda_k). \tag{4.18}$$

The NIG distribution is a conjugate prior to the normal likelihood. Therefore, the posterior distribution given a sample $\{c_{k,j}\}_{j=1}^{J}$ of J realizations of c_k is

$$\mu_k, \lambda_k|c_{k,1} \dots c_{k,J} \sim N\Gamma^{-1}(m'_k, \tau'_k, \alpha'_k, \beta'_k). \tag{4.19}$$

The posterior hyperparameters are given by

$$m'_k = \frac{J\bar{c}_k}{\tau_k + J} \tag{4.20}$$

$$\tau'_k = \tau_k + J \tag{4.21}$$

$$\alpha'_k = \alpha_k + \frac{J}{2} \tag{4.22}$$

$$\beta'_k = \beta_k + \frac{1}{2}\sum_{j=1}^{J}(c_{k,j} - \bar{c}_k)^2 + \frac{J\tau_k\bar{c}_k^2}{2(\tau_k + J)}, \tag{4.23}$$

where $\bar{c}_k$ is the average of $c_{k,1} \dots c_{k,J}$. Now, to draw samples from the posterior predictive distribution of c_k, one would first draw μ_k and λ_k from the posterior distribution (4.19), and then use these to draw c_k from the likelihood (4.15). Alternatively, one may use point estimates $\hat{\mu}_k$ and $\hat{\lambda}_k$ and insert directly into the likelihood (4.15). Choosing the expected value of the posterior as the point estimate, we get

$$\hat{\mu}_k = m'_k \tag{4.24}$$

$$\hat{\lambda}_k = \frac{\beta'_k}{\alpha'_k - 1}. \tag{4.25}$$

Each mode k has different prior and posterior parameters. To draw a new shape s^*, one would draw new scores c_k^* for each k from the respective distributions. Then s^* can be synthesized as in (4.14):

$$s^* = \mu_0 + \sum_{k=1}^{K} c_k^* \psi_k. \tag{4.26}$$

However, the limited number K of PCs cannot represent the patient specific mean shape exactly, therefore the expected value of s^* is not equal to the patient specific mean shape μ. This problem does not appear in the traditional population model because the sample average $\bar{s}$ is used in place of μ_0. To correct for this bias, we can add the approximation error of the sample mean to the new sample s^*:

$$s^* = \mu_0 + \sum_{k=1}^{K} c_k^* \psi_k + \bar{\epsilon}, \tag{4.27}$$

where $\bar{\epsilon}$ is the approximation error of the sample mean,

$$\bar{\epsilon} = \bar{s} - \mu_0 - \sum_{k=1}^{K} \bar{c}_k \psi_k. \tag{4.28}$$

However, note that c_k^* is drawn from a distribution whose mean is not $\bar{c}_k$, but rather $\frac{J}{\tau_k + J}\bar{c}_k$, i.e. a *shrunk* estimate of the mean of the kth mode. It is natural to treat the approximation error in the same fashion. Therefore, we introduce a parameter τ_ϵ, and finally let the new samples be drawn as

$$s^* = \mu_0 + \sum_{k=1}^{K} c_k^* \psi_k + \frac{J}{\tau_\epsilon + J}\bar{\epsilon}. \tag{4.29}$$

The hyperparameters τ_k, α_k and β_k for all k as well as τ_ϵ can be found by maximum likelihood estimation or the method of moments based on the set of training data from which the PCs ψ_k are derived. By construction, m_k is zero for all k.

4.6.3 Shrinkage towards the population mean

When developing the method in section 4.6.2, we found that the shrunk error term in equation (4.29) was responsible for a portion of the improvement over the population model. This term is non-randomly added to every drawn shape, and therefore represents a systematic shift. The improvement thus represents an improvement in *systematic* error. In fact, this improvement can be achieved without using the PCA model at all: Suppose there are no PCs, i.e. $K = 0$. Then (4.28) becomes

$$\bar{\epsilon} = \bar{s} - \mu_0, \tag{4.30}$$

and (4.29) becomes

$$s^* = \mu_0 + \frac{J}{\tau_\epsilon + J}\bar{\epsilon}. \tag{4.31}$$

This s^* is non-random given the sample $s_1, \dots, s_J$, and can be considered an estimate $\hat{\mu}$ of the patient mean shape. Inserting (4.30) into (4.31), we get

$$\begin{aligned} \hat{\mu} &= \mu_0 + \frac{J}{\tau_\epsilon + J}(\bar{s} - \mu_0) \\ &= (1-a)\mu_0 + a\bar{s}, \end{aligned} \tag{4.32}$$

where $a = \frac{J}{\tau_\epsilon + J}$. In other words, the sample mean shape $\bar{s}$ is shrunk towards the population mean shape μ_0 by a factor $1 - a$. This discovery led to the publishing of paper I, which uses exactly this shrinkage estimate of the patient mean shape to to reduce systematic errors.

4.6.4 A model based on a full matrix distribution

The method in section 4.6.2 can be considered an enhancement of the population model, since it adds patient specific means and variances of the PCA-scores. However, it still has a weakness compared to the patient-specific model in that the PCs are never updated when new patient-specific data becomes available. Thus, while the patient-specific model can recreate patient-specific motion pattern given sufficient data, the NIG-model can never do so.

In order to get a step further, the entire covariance matrix needs to be updated when new data becomes available. To do this in a Bayesian fashion, the covariance matrix must be governed by probability distribution of matrices.

The conjugate prior of a multivariate Gaussian likelihood with unknown mean and covariance matrix is the normal-inverse-Wishart (NIW) distribution. In the scalar case (i.e. the vector we are modelling is of dimension 1), the NIW is equal to the NIG distribution. In paper II, we developed a model using the NIW as a prior and a specific point estimate for the mean and convariance matrix, which resulted in the following update equations:

$$\hat{\mu} = \frac{1}{\kappa + J}(\kappa\mu_0 + J\bar{s}) \tag{4.33}$$

$$\hat{R} = \frac{1}{\nu + J}\left(\nu\hat{R}_{\text{pop}} + \sum_{j=1}^{J}(s_j - \bar{s})(s_j - \bar{s})^T + \frac{\kappa J}{\kappa + J}(\bar{s} - \mu_0)(\bar{s} - \mu_0)^T\right), \tag{4.34}$$

where κ and ν are new hyperparameters. Interestingly, the update equation for the mean estimate (4.33) is identical to that of paper I.

The direct calculation of $\hat{R}$ is not practical because of the size of the matrix. An efficient way of implicitly representing $\hat{R}$, which also involves PCA in the representation of $\hat{R}_{\text{pop}}$ is shown in paper II.

4.6.5 A model with separate inter and intra-patient distributions

The NIW model is an attempt at getting “the best of both worlds” from the population and patient-specific models. Even so, we have seen that the NIG-model can outperform the NIW-model given a single input scan. Looking more closely at the estimated mean $\hat{\mu}$, we find a possible contributing factor to this phenomenon: The mean estimate for the NIW model (4.33) is equivalent to that of the NIG model using zero PCs (4.32). The estimated mean is a convex combination of the patient-specific sample mean and the population mean. In the NIG-model, on the other hand, there are separate weights τ_k for each PC. This is due to a more complex relationship between the inter-patient and the intra-patient covariance matrices in the NIG-model compared to the NIW-model. In the NIW prior, the inter-patient covariance matrix is $\frac{1}{\kappa}R$, where R is the intra-patient covariance matrix. In the NIG model, the intra-patient covariance matrix is

$$R = \sum_{k=1}^{K} \lambda_k \psi_k \psi_k^T, \tag{4.35}$$

whereas the inter-patient covariance matrix is

$$\Lambda = \sum_{k=1}^{K} \frac{\lambda_k}{\tau_k} \psi_k \psi_k^T, \tag{4.36}$$

where ψ_k are the combined inter/intra-patient PCs. The scalar parameters τ_k provide extra degrees of freedom in the relationship between the inter and intra-patient covariance matrices, which again leads to more degrees of freedom in the estimate of the patient mean.

This suggests that we might be able to enhance the model further by making no assumptions about the relationship between the inter-and intra-patient covariance matrices. Consider a scenario where the intra-patient covariance matrix R is known, and the patient mean μ is distributed as

$$\mu \sim \mathcal{N}(\mu_0, \Lambda) \tag{4.37}$$

for some given μ_0 and Λ. Then the posterior distribution of μ given data $s_1, \ldots, s_J$ is also Gaussian (Gelman et al. 1995):

$$\mu | s_1, \ldots, s_J \sim \mathcal{N}(\mu_0', \Lambda'), \tag{4.38}$$

with

$$\Lambda' = (\Lambda^{-1} + J R^{-1})^{-1} \tag{4.39}$$

$$\mu' = \Lambda' (\Lambda^{-1} \mu_0 + J R^{-1} \bar{s}). \tag{4.40}$$

Equation (4.40) can be used in place of the simple shrinkage estimate (4.32) from paper I, though it requires estimation of R and Λ.

This result is valid when R is deterministic. In order to model patient-specific motion patterns in a Bayesian framework, R must be considered stochastic. To untie R from Λ, let the prior of μ be given by (4.37) and let the prior of R be inverse-Wishart:

$$R \sim \mathcal{IW}(\Psi, \nu), \tag{4.41}$$

and let R and μ be independent. The resulting joint prior for R and μ is not conjugate to the Gaussian likelihood (4.12), and therefore the posterior is not easily calculated. In paper II, we used the mean field variational Bayes approximation to find an approximate posterior. In this approximation, the posteriors of μ and R are also independent, and also Gaussian and inverse Wishart distributed, respectively. An iteration for calculating the parameters of the posterior was found, and we also demonstrated a way to estimate the prior parameters Ψ and Λ, and found an efficient algorithm for performing the iteration using these estimates.

4.7 Methods used for evaluation

4.7.1 Surface and volume similarity

An important part of the evaluation of organ deformation models is comparison of organ shapes. It is desirable to have a single number to represent the similarity or dissimilarity between two shapes. Two fundamentally different concepts of similarity between object shapes are surface similarity, i.e. how far the surfaces are from each other (on average, maximum or similar), and volumetric similarity, which compares how well the inside volumes of the shapes match.

A common measure for surface similarity is the Hausdorff distance (HD). It has the advantage of being defined for any two sets of points A and B, regardless of whether there is correspondence between the points or even if they contain different number of points. The HD is determined by the one point in any of the two sets that is furthest from any point in the other set, and as such it is a "worst-case" measure. Defining the minimum distance from an arbitrary point x to any point in A as

$$d_A(x) = \min_{a \in A} \|x - a\|, \tag{4.42}$$

the HD between A and B is given by

$$HD_{A,B} = \max(\max_{a \in A} d_B(a), \max_{b \in B} d_A(b)). \tag{4.43}$$

A generalization of the Hausdorff distance to local distances, i.e. one distance per point in A instead of a single distance for all points, is the bidirectional local distance (BLD) (Kim et al. 2012). The BLD for a point a in A is the maximum of: the shortest distance from a to any point in B, and the distance from a to any point b in B for which the closest point in A is a. Mathematically, the BLD from a set A to a set B at a point a in A, denoted $BLD_{A,B}(a)$, is given by

$$a_{\min}(b) = \arg\min_{a \in A} \|b - a\| \tag{4.44}$$

$$B_{\min}(a) = \text{All } b \text{ such that } a_{\min}(b) = a \tag{4.45}$$

$$BLD_{A,B}(a) = \max(d_B(a), \max_{b \in B_{\min}(a)} d_A(b)). \tag{4.46}$$

The maximum BLD is the HD.

For volumetric similarity, the most common measure is the Dice-coefficient, which is

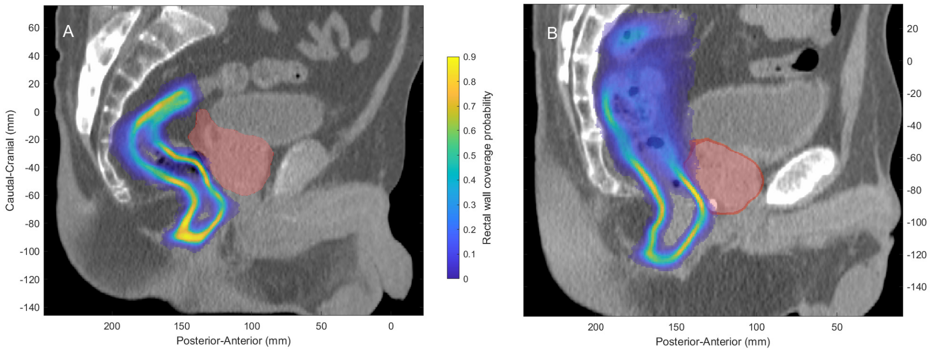

Figure 4.6: Coverage probability matrices for the rectal wall on a sagittal slice of the CT scan for two example patients — a "small mover" (A) and a "large mover" (B). The red area is the high dose volume to the prostate, that receives more than 67 Gy EQD$_2$

defined as

$$\text{Dice}(C, D) = 2\frac{|C \cap D|}{|C| + |D|} \tag{4.47}$$

Here $|C|$ and $|D|$ represent the volume of C and D, and $|C \cap D|$ represents the volume that is common between C and D. The Dice coefficient is 1 if the shapes are identical, and 0 if there is no overlap between the shapes. To compute the Dice coefficient in practice, the 3D space is divided into small voxels, and a "point containment" algorithm determines which voxels are inside and which are outside each shape. The total volume of a shape is the number of voxels inside the shape, and the shared volume is the number of voxels inside both shapes.

4.7.2 Computation of coverage probability matrices

A coverage probability matrix (CPM) for an organ is a 3D matrix of voxels, where each voxel has a value indicating the probability that the organ will cover that voxel at any given time, as illustrated in Figure 4.6 (taken from paper II). CPMs are useful in motion-robust radiotherapy planning and evaluation, among other applications. The simplest way to estimate CPMs from a deformation model is by Monte-Carlo sampling: Randomly sample many shapes based on the model, and, for each shape, find which voxels are inside and which are outside of the shape. The coverage probability at each voxel is determined by the fraction of random shapes that covered that specific voxel.

To find the points inside and outside the shape, a point-in-polyhedron algorithm is used. A fast algorithm for the point-in-polyhedron problem was implemented specifically for this use, and is presented in paper IV.

4.7.3 Cross validation

For evaluation of the methods in paper I and paper II, we used leave-one-out cross validation (LOOCV). Cross validation is used to avoid the inflated results that occur if the evaluation data is included in the training database. In LOOCV, the model is retrained for each patient in the data set, using all the other patients' data as training data. While LOOCV is more costly in terms of complexity than other methods, specifically k-fold cross validation, the advantage is that we use as much training data as possible for each patient, and that we get an isolated result for each patient and can therefore report the results as though the training database was the evaluation database.

Chapter 5

Summary of results

5.1 Paper I

In paper I, we introduced a method using shrinkage estimation (see section 2.1) to estimate the patient-specific mean organ shape by combining the shape seen in the pCT and the population mean shape. We evaluated the method using LOOCV on the recti of prostate cancer patients. The “true” mean shape for each patient was calculated by averaging the remaining 8-10 rectum shapes from the repeat CTs.

The estimated mean rectal shapes from the shrinkage method were significantly more similar than the pCT-shapes to the “true” mean shapes, both in Dice-score and HD, as seen in Figure 5.1 A and C. The Dice-score improved for 33/37 patients, while the HD improved for 28/37 patients. Additionally, the improvement was greater for patients who had large discrepancy between the pCT and the true mean shape, as seen in figure 5.1 B and D.

The improvement in similarity did not lead to a significant improvement in D5% (the minimum dose delivered to the 5% of the organ receiving highest dose) or equivalent uniform dose (EUD). However, in the DVH, we saw a significant reduction in bias in the high-dose region between 52.5 Gy and 65 Gy.

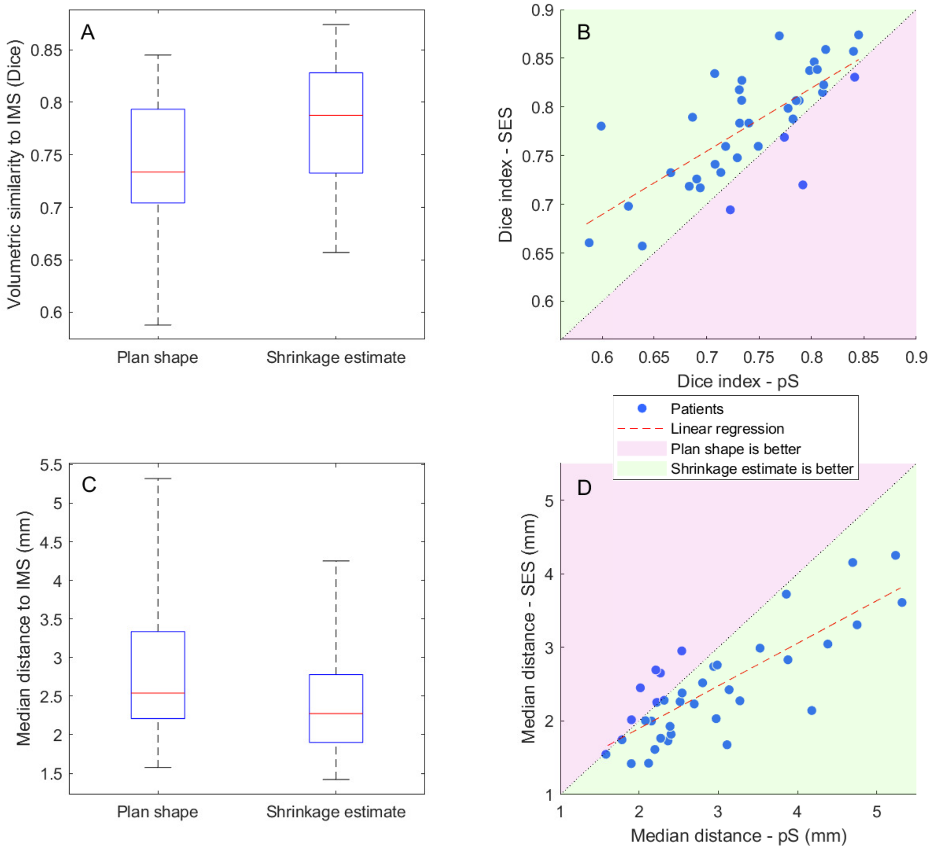

Figure 5.1: Similarity to the individual patient mean rectum shape (individual mean shape, IMS) for the "shrinkage estimated shapes" (SES) and the planning CT shapes (pS) in each patient. A and C: Box plots of the similarity to the patient mean shape of the plan shape and the shrinkage estimate as median (red line) with 25th and 75th percentiles and maximum and minimum values (whiskers). B and D: scatter plots of the similarity to the patient's mean shape for the plan shape versus the shrinkage estimate. In A and B (Dice similarity), higher is better, while in C and D (distance), lower is better. Figure taken from paper I (Rørtveit et al. 2021).

5.2 Paper II

In paper II, we presented two new organ deformation models based on Bayesian methods, the NIW model (section 4.6.4) and the variational Bayes model (section 4.6.5), and evaluated these for the rectum of prostate cancer patients. In our evaluation, the variational Bayes model produced more accurate CPMs than the NIW-model. Nevertheless, we felt that it was important to present the NIW-model since it is considerably less complex, conceptually and implementation-wise. We also compared the two new methods to the patient-specific model by Söhn et al. (2005) and the population model by Budiarto et al. (2011).

To evaluate the two new and two old methods, we estimated the CPM for each patient using a varying number of scans as input. We then correlated the estimated CPMs with reference CPMs which we considered the ground truth, and which were calculated from the remaining scans not used as input. Since there is a limited number of scans remaining to calculate the ground truth, we evaluated the method for the original data only for up to 3 input scans. To evaluate the convergence of the methods up to 10 input scans, we used a PCA-based bootstrapping technique to generate a new data set of shapes.

In the evaluation with real data, both the NIW-model and the variational Bayes-model produced CPMs with significantly higher correlation to the ground truth than the population model when using 1, 2 and 3 input scans. The variational Bayes model further produced significantly better results than the NIW model for 1, 2 and 3 input scans (Figure 5.2). Also when evaluating convergence, the two Bayesian methods outperformed both the existing methods for up to 8 input scans, after which they perform similar to the patient-specific model (Figure 5.3). The population and patient specific model produced approximately equal correlation to the ground truth at four scans. The population model was superior to the patient-specific model with fewer than four scans and vice versa.

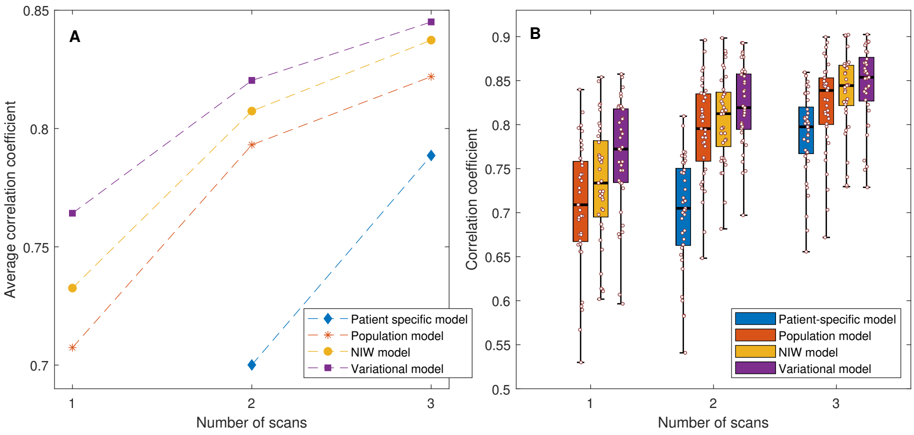

Figure 5.2: Correlation between the estimated CPMs and the references for the different methods using 1-3 input scans. A: Average correlation. B: Box plots showing median, 25th and 75th percentile and minimum and maximum values (whiskers). All individual values are also shown as circles over the box plot. Figure taken from Paper II (Rørtveit et al. 2023).

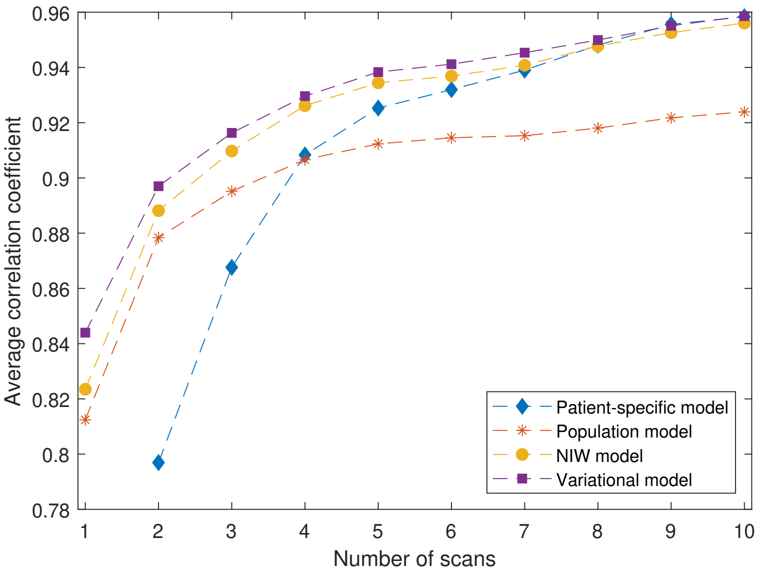

Figure 5.3: Average correlation between the estimated CPMs and the reference CPMs for the different methods using 1-10 scans, based on bootstrapped data. Figure taken from paper II (Rørtveit et al. 2023).

5.3 Paper III

To calculate the CPMs used in the evaluation of the methods of paper II, it was necessary to determine which voxels were outside and which were inside each of the thousands of shapes randomly generated for the Monte Carlo calculations. In addition, the voxel grid was quite fine, at $1 \times 1 \times 1$ mm, meaning each shape contained millions of points.

Determining whether a point is inside or outside a polyhedron is called the “point-in-polyhedron” problem, and there are many known algorithms for solving it. However, none of the available open-source software libraries could provide the performance that was needed for the evaluation in paper II. Therefore, a new software library was created, that uses optimizations that are only possible when all the points to be tested are aligned on a grid. The implemented algorithm contains several new ideas, among them a fast and simple inexact way of dealing with edge-cases, which are a major concern in most such algorithms. To maximize performance, the algorithm was implemented in C++, and has a Matlab™ interface.

The software was benchmarked against an open source implementation of the algorithm in Liu et al. (2010), on a Windows PC with an Intel(R) Core(TM) i5-9500 CPU @ 3.00GHz and 16 GB of RAM. Identifying 213 million points in a polyhedron with 73000 faces took 0.24 seconds with InsidePolyhedron, and 48.7 seconds with the reference software - i.e. a factor 200 difference. It should, however, be noted that the reference software is more general, in that it can solve the point-in-polyhedron-problem for points that are not aligned on a grid.

5.4 Paper IV

As an application of the organ deformation models presented in paper II, paper IV presents a of a method to accumulate dose and estimate dose uncertainty for the rectum of prostate cancer patients receiving re-irradiation. The method will automatically adjust the uncertainty based on the model’s assessment of the two patient-specific scans combined with population statistics. The estimates of the average dose and its uncertainty are based on dose to three rectum shapes generated by the model that represent the expected dose and plus/minus one standard deviation. The anterior region of the rectum near the prostate moved on average 2.6mm in the direction of the dose gradient for the +1 standard deviation shapes compared to the expected shape. The patients with highest and lowest uncertainty moved 3.4mm and 1.9mm in the dose gradient direction, respectively, showing that the model implicitly differentiates between small and

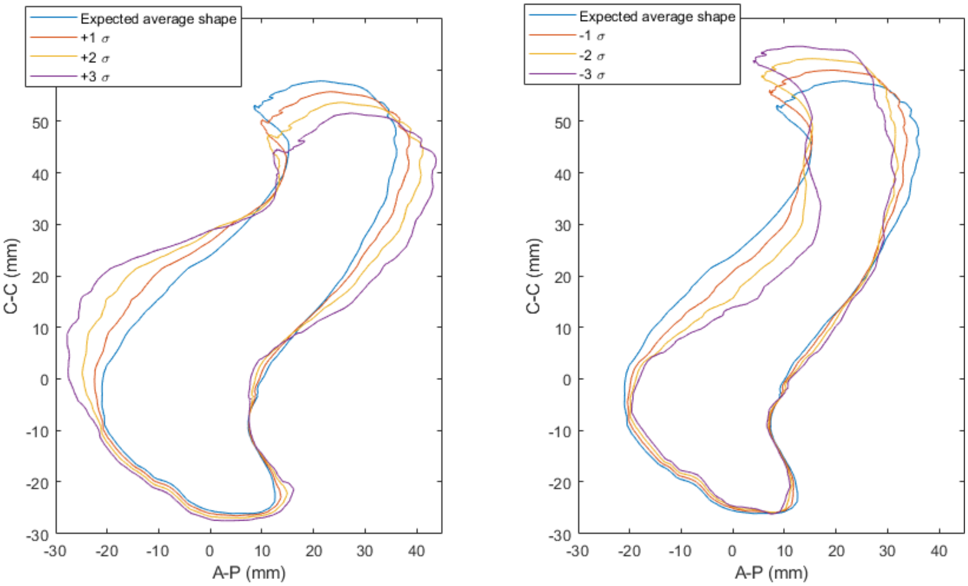

Figure 5.4: Saggital view of the estimated mean shape and plausible mean shapes representing plus and minus 1, 2, and 3 standard deviations (of dose) for an example patient.

large movers. A saggital view of representative shapes for an example patient at plus and minus 1, 2, and 3 standard deviations is shown in Figure 5.4.

For one of three example patients, the accumulated planning CT DVH underestimates the dose compared to the expected mean shape calculated by the model, as seen in figure 5.5, center. For this patient, D5% (5% of the volume gets at least this dose) is lower than the 97.5th percentile. For the two other patients, the accumulated DVHs are relatively similar to the expected DVH. The second of these patients had rectal toxicity grade 2, while the others had no rectal toxicity.

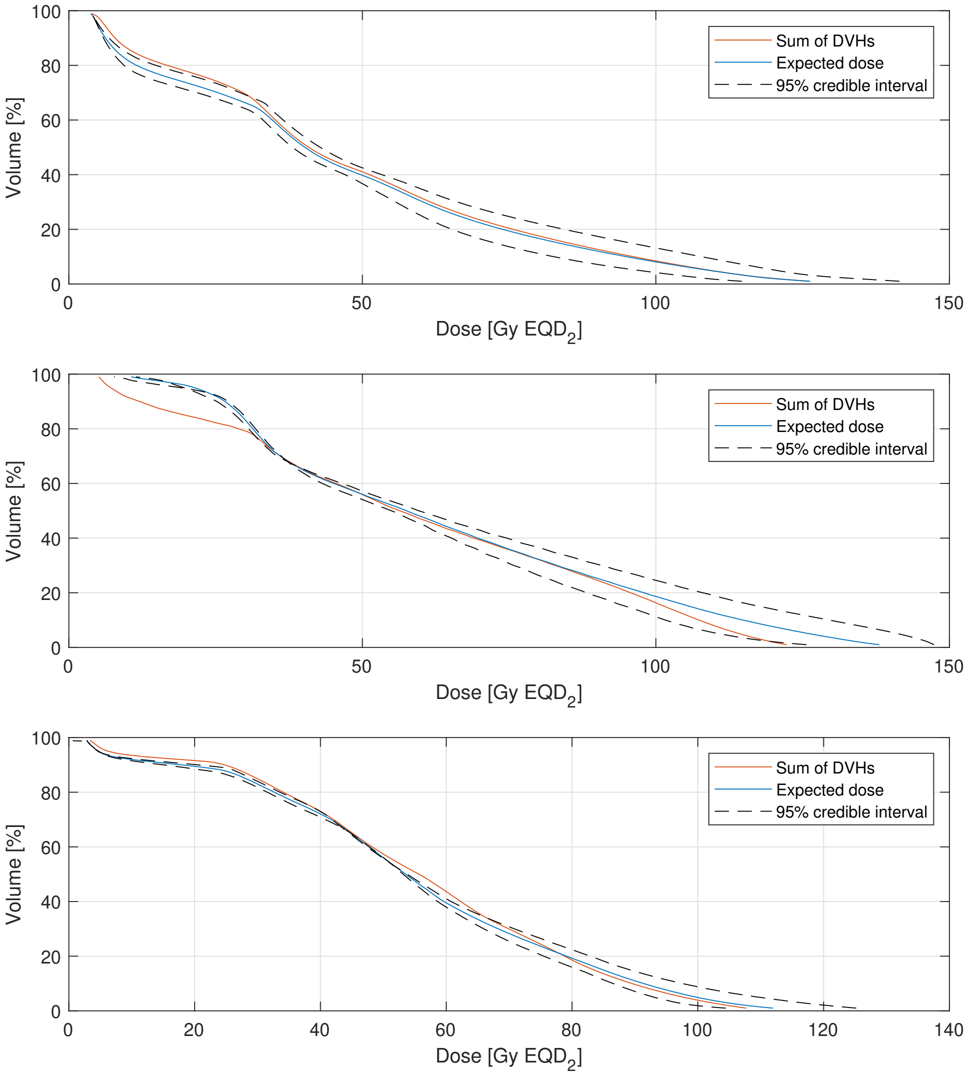

Figure 5.5: DVHs to the rectum for three example patients, showing the sum of the planning CT DVHs for both therapy courses, the expected accumulated dose and the 95% credible interval, which is the Bayesian counterpart to the confidence interval.

Chapter 6

Discussion

Current organ motion and deformation models suffer from some limitations that have been the focus of this project. An inherent issue with the patient-specific model by Söhn et al. (2005) is that it requires many scans to predict a patient's deformation patterns. It is rarely realistic to have 3-5 scans available before start of treatment, which is the main reason for the introduction of the population model by Budiarto et al. (2011). In this model, the deformation patterns of the "population" is measured from a set of training data, and the average pattern is assumed to be valid for each new patient. This solves the issue of needing many scans. On the other hand, even if given multiple scans from a new patient, this method cannot tailor the distribution of the random error to the specific patient. In this project, for the first time, both these methods were implemented and evaluated for a range of inputs. They were thereby compared to the newly developed methods from this project. The comparison and improvements are found in paper II.

The methods developed in this project solve the mentioned issues of the patient-specific model and the population model simultaneously. Using Bayesian methodology, the estimates from the population and the individual patient are merged, with the weight given to the individual estimates increasing in proportion to the number of individual measurements. With few individual scans, the population estimate dominates, with many, the individual estimate dominates. The gradual shift from population to individual weight is regulated by the parameter ν. With J measurements, the weight given to the individual estimate is $J/(\nu + J)$.

Additionally, these are the first methods to estimate the distribution of the patient's mean organ shape. Since the expected mean shape in these models is calculated from both population and patient-specific data, systematic errors are reduced as compared to previous models where only the patient-specific mean shape has been used.

As mentioned in the introduction, systematic errors are a major concern, as they will cause parts of a structure to consistently receive too high or too low dose throughout all fractions, as opposed to random errors that tend to be “blurred out” by fractionation (Amer et al. 2001; van Herk 2004). Consider the well-known and much used margin recipe by van Herk et al. (2000), which under typical considerations says the margin should be $0.7\sigma + 2.5\Sigma$, where σ is the standard deviation of the random error and Σ is the standard deviation of the systematic error. Typically, the values of σ and Σ are almost identical, yet more than 3 times more weight is given to the systematic error than to the random error to ensure proper target coverage.

Reduction of systematic errors using population statistics has previously been proposed by Hoogeman et al. (2005) who modeled the rectum as an unfolded tube using the method developed in their previous publication (2004). They had some success in predicting the rotation of the prostate and displacement of the upper anterior rectal wall based on a linear regression from the volume of the rectum. Their results motivated our work in paper I.

The improved estimation of the patient’s mean organ shape also leads to improved estimates of the random error. A single observation usually does not provide any information on the variation from fraction to fraction, i.e. the random error. This is the reason why the mixed-effect model of Rios et al. (2017) did not show improvement over the population model given a single input. However, in the Bayesian models with unknown mean shape μ, the difference between the expected mean shape and the observation is a source of information on the patient-specific deformation patterns. Thus, the Bayesian models can provide better estimates of both the systematic and the random errors even using only a single scan.

The systematic error is defined as $\hat{\mu} - \mu$, where $\hat{\mu}$ is the estimate of the mean and μ is the true mean. In the conventional methods (both the population and patient specific model), the average of the J observed shapes of a patients is used to estimate the true mean. Under the assumption that conditions are identical during scanning and therapy, the distribution of the systematic error is $\mathcal{N}(0, \frac{1}{J}R)$. Here, R is the covariance matrix of the random error, which is the same as the covariance matrix of the observations. In the Bayesian models, μ is considered random, therefore the distribution of $\hat{\mu} - \mu$ depends on the prior distribution of μ. In the NIW-model, since $\hat{\mu}$ is the mean of the posterior, the systematic error is $\mu' - \mu_0'$, which is distributed as $\mathcal{N}(0, \frac{1}{J+\kappa}R)$. Thus, the variance of the systematic error is reduced by a factor of $J/(J+\kappa)$ as compared to the conventional methods. For the variational Bayes model, the systematic error is distributed approximately as $\mathcal{N}(0, \Lambda^*)$, and it is not equally easy to see the reduction in variance because of the complex relationship between Λ^* and Ψ, but here as well a reduction compared to $\frac{1}{J}R$ is guaranteed. This of course only applies if the model and model parameters are a good fit for the data. The fit of the parameters is ensured by the training.

While the methods presented in paper I and II are general in the sense that they can be applied to any organ or RT-target, they have been evaluated for the recti of prostate cancer patients. The rectum is a highly flexible organ which undergoes major deformations, in particular in response to filling and emptying of feces and gas, but also due to the motion of other organs (Hoogeman et al. 2004; Roch et al. 2021; Scaife et al. 2015). Additionally, the rectum is usually the dose-limiting organ for prostate cancer RT due to its proximity to the prostate and the incidence of rectal toxicity (Schultheiss et al. 1997; Sripadam et al. 2009). Under IGRT with matching to the bony anatomy, rectal distension has been found to be the main cause of prostate motion (Pinkawa et al. 2006).

In all the methods in paper I and II were cross-validated using LOOCV. This means that we get a result for every patient in the data set without bias, because the training is performed separately for each patient using the remaining patient' data as training data. For a complex model, both the implementation effort and running time of LOOCV can be significant. On the other hand, LOOCV provides more data for both training and evaluation than the simpler approach of splitting the data set in two parts, one for training and one for evaluation.

In paper IV, a method for using the models from paper II to estimate doses in terms of DVHs with uncertainty bands based on "representative shapes" was developed. This method was specialized for the rectum of prostate cancer patients, but also has the potential to be generalized. One advantage of using representative shapes is that they make it easy for the oncologist or dose planner to see possible organ configurations that may be problematic. Another advantage is that the TPS's existing features for calculating dose to a structure can be used to calculate dose uncertainty. The method was applied to estimate accumulated dose and uncertainty for patients receiving re-irradiation for prostate cancer. These patients are at special risk, because the OARs have already been exposed to radiation, and the amount of repair since the previous irradiation is unknown. On the other hand, there will usually be two pCTs available for the patient, which makes it possible to model motion and deformation with improved accuracy.

6.1 Deformation model applications

Organ deformation models can be used to generate statistically meaningful margins. The margin recipe of Stroom et al. (1999) places the margin at a contour where the coverage probability is constant, such that the target rarely move outside of the PTV. The coverage probability can be calculated by organ deformation models, as done for prostate OARs in Hysing et al. (2011), and for rectal cancer targets in Bondar et al. (2014). Ramlov et al. (2017) expanded this technique to simulate the CPM-based robust optimization of Baum et al. (2006) by introducing multiple margins with decreasing coverage probability, and with reduced dose objectives for margins at lower CPM boundaries.

Another important application of deformation models is the characterization and visualization of organ motion and deformation. Lung motion and deformation has been studied by Badawi et al. (2010), Dhou et al. (2020), and Li et al. (2011a), Tsiamas et al. (2018) studied head and neck-related structures while Rios et al. (2016) studied bladder motion and deformation. The rectum, prostate and bladder was studied simultaneously to characterize correlations between their motions in Söhn et al. (2005). Motion and deformation characterization can have direct use in the clinic, e.g. Thörnqvist et al. (2013a) characterized the motion of prostate CTVs to evaluate the required magnitude of isotropic margins.

Many motion-robust plan optimization algorithms rely on distributions of the patient motion (e.g. Baum et al. 2006; Birkner et al. 2003; Unkelbach and Oelfke 2004; Unkelbach and Oelfke 2005). Such distributions can be found by organ deformation models, and directly substituted into the robust optimization. Other algorithms rely on multiple realizations of the patient geometry (Fredriksson et al. 2021; Sobotta et al. 2010; Xu et al. 2014), which can also be produced by deformation models.

When robust optimization is not available, an alternative is to evaluate the robustness of plans generated with traditional methods (Hysing et al. 2018; Söhn et al. 2012). Through Monte-Carlo simulations, one can estimate both the expected accumulated dose and its variance locally in the organ, which again can be used e.g. to estimate DVHs with uncertainty bars.

Other applications include generating plan-of-the day libraries for adaptive radiotherapy (Rigaud et al. 2019), correlating breathing motion to abdominal changes to correct for these (Liu et al. 2021; Nakao et al. 2021), and auto-segmentation (Heimann and Meinzer 2009). Recently, Owens et al. (2022) used a pure inter-patient model to reconstruct colorectal dose in childhood cancer survivors who had received RT with no CT simulation.

Despite the widespread use of statistical organ motion models in research, there is little evidence of their use in the clinic. An exception is the HYPERION dose planning software, developed at the university of Tübingen, which has implemented the coverage probability based robust optimization by Baum et al. (2006) as well as the Monte-Carlo based robust optimization method by Sobotta et al. (2010). Ramlov et al. (2017) introduced a simple way to build plans in Varian Eclipse™ that were near identical to those produced by HYPERION. This method was used in a clinical study on EBRT boost of lymph nodes in patients receiving brachytehrapy for locally advanced cervical cancer (Lindegaard et al. 2017), with promising results.

An obvious hurdle in the use in clinics is the lack of implementation of such models in commercial TPSs. Furthermore, quality assured training data needs to be available for population models, and many scans are needed for patient-specific models. There may also be a certain reluctance to using machine learning techniques in the clinic (Poon and Sung 2021), though this reluctance is gradually dwindling with the introduction of time-saving tools such as auto-segmentation and rapid automatic planning.

Bayesian models offer additional advantages and possible applications because they quantify the model uncertainty. This is exemplified by paper IV, where the uncertainty of the mean is used to produce uncertainty bands for the DVHs of patients receiving re-irradiation. While re-irradiation is becoming more common, prescriptions and procedures tend to vary from clinic to clinic (Andratschke et al. 2022). Internationally accepted guidelines are therefore needed, but such guidelines require modelling of previous radiation damage and repair (Paradis et al. 2019). Guidelines must build on reports of re-irradiated cohorts, and these reports should contain information on the cumulative dose (Andratschke et al. 2022). The algorithms presented in this thesis allow simple yet relatively accurate (as compared to a deformable image registration approach) calculation of accumulated dose by replacing organs in both pCTs by the estimated mean shape of the organ. Furthermore, dose uncertainty can be calculated, accounting for systematic error and the model uncertainty using simple means (by representative shapes) or more complex means (Monte-Carlo simulations).

6.2 Model assumptions and potential improvements

The presented deformation models explicitly do not include rigid shifts of the patient geometry. Rather, they model deformations relative to a fixed patient coordinate system. Where the origin is depends on the rigid registration of images performed before deformable registration. It is important that the model is based on data which uses the same origin as used in treatment. For pelvic treatments, fiducial markers or bony anatomy is typically used as the fixed point for rigid registration. In the daily fractions, IGRT-routines are used, where the patient is moved by a robotic treatment table to match the position of the daily portal image to the pCT. However, some amount of setup error will remain. This setup error is not included in the models, as it is not seen in the matched images. If measured, the variance of the remaining setup error can be included in the models by adding a constant to the covariance matrix.

The methods in papers I, II and IV all additionally require rigid registration between patients. This is a more challenging concept because one cannot always rely on landmarks such as fiducial markers or bony anatomy. In paper I, we used the “center of gravity” of the lower (most caudal) half of the rectum at the pCT as a fixed point between patients. This was chosen because this part of the rectum is more stable than the cranial part, because it includes the part of the rectum closest to the prostate, and because it generated satisfactory results in the estimation of the patient’s mean rectum shape. A more complex method involving optimization was tested, but yielded no better results. This center-of-gravity registration method was therefore carried over to papers II and IV.

The PCA/covariance matrix-based models are linear; they do not model e.g. rotation. To see this, consider that each PC corresponds to a linear shift in the coordinates, and a different PCA-score means a different length of the shift. A small rotation can nevertheless be approximated by a small linear shift. Nonlinearity can be taken into account by using a nonlinear parametrization instead of PCA, or by modelling the parameters of a nonlinear function (e.g. the thin-plate-spline) instead of the coordinates of surface points.

The model assumes that all scans are identically distributed. This means that it does not account for systematic changes that may occur during radiotherapy, such as tumor shrinkage, weight loss/gain, or diarrhea. In our evaluation data, the rectal volume did not change significantly over time (Hysing et al. 2018). Nevertheless, modelling systematic changes through PCA models has been shown possible by Chetvertkov et al. (2016), and may be an interesting future direction.

While this thesis has focused on deformation of organ contours, and therefore also deformable *contour* registration, the methods can equally well be applied to the deformation of voxels of the organ of interest, and use deformable *image* registration (Dhou et al. 2020; Fransson et al. 2021; Li et al. 2011b; Szeto et al. 2017; Tilly et al. 2017; Tsiamas et al. 2018). While one might imagine that this improves the accuracy of deformations in the interior of the organ, this is not necessarily so due to lack of contrast on the CT scan of an organ interior. For the rectum, registering the interior will cause unwanted warping because of the changing location of gas, so the organ interior is usually replaced by a fixed intensity level in the CT data before deformable image registration.

The sources of error for the deformation models include errors in their input data and in the assumptions and approximations. It is well known that organ/target delineation varies both between and within observers. While no reports on the inter/intra-observer variations of rectum delineations was found in a pubmed search, White et al. (2021) reported average inter-observer variations of delineations of the mesorectum at 1.7 mm (mean distance to agreement) and 11.1 mm (max distance to agreement, or Hausdorff distance) on CT scans. There are also errors in the deformable registration. We measured 0.22/0.23 mm average forward accuracy of the inter/intra patient registrations used in the papers. This may appear negligible compared to the inter-observer variation, however, for deformable registrations, a good distance-to-agreement does not necessarily mean a correct correspondence between shapes. Unless there are visible "landmarks" in the scans (Vásquez Osorio et al. 2012; Wolthaus et al. 2008), there is no definite way of quantifying the correctness of the correspondence. Nevertheless, a popular measure is the *inverse consistency*, which is found by performing the deformation in the opposite direction and checking that the result is consistent with the forward deformation. The average backward consistency of the registrations used in the papers was 0.62 mm and 0.47 mm for the intra and inter-patient registrations, respectively. See Vásquez Osorio (2012) for the definition of the forward accuracy and inverse consistency. The delineation and registration errors are propagated to the distributions estimated by the deformation models. It is important that the training data matches the input data, so contours delineated by multiple observers should be present in the training data.

In the models in paper II, the patient-specific distributions are assumed to be Gaussian. The same is true for the majority of applications of deformation models (see e.g. Söhn et al. 2012). The Gaussian assumption makes life much easier for the researcher, but there are alternatives, such as using a non-parametric distribution (Fontenla et al. 2001b). While there is no strong evidence suggesting that the distributions are in fact Gaussian, there is neither evidence of the opposite. The uncertainty of such assumptions underline the importance of evaluating the models for real data.

For the Bayesian models, the performance will depend to some amount on how well the data matches the prior. The evaluation in paper II compared the model with two different priors to the population model, and showed that in both cases the chosen priors outperformed a pure population model.

6.3 Artificial intelligence

The methods in this thesis are so-called unsupervised machine learning methods; a subcategory of artificial intelligence (AI) where the algorithm learns patterns from data. Outside the professional sphere, however, "AI" has been practically synonymous with "deep learning artificial neural networks" for the last decade. At the start of this project, most well-known deep learning applications were in the category of *supervised* machine learning. In supervised learning, the outcome of each measurement is also part of the training data, and the goal is to predict the outcome (either as a category, in classification, or as a number or vector, in regression) for new samples. These algorithms are typically trained on massive data sets of thousands or even millions of examples, and are excellent at categorizing which bird is in an image or whether the image contains a traffic light.

Since then, the diversity of applications of deep neural networks have exploded, most recently with the public release of fascinating image-from-prompt generators such as Dall-E, Stable Diffusion and Midjourney, and not least the surprisingly knowledgeable and eloquent advanced language model ChatGPT.

While it is not immediately clear how deep learning can be applied to organ deformation modelling, this will almost certainly be a research topic in the future. Only recently, a preprint detailing a deep learning structure for random generation of patient-specific deformations has been released (Pastor-Serrano et al. 2022). This method uses both supervised and unsupervised learning, and has some properties in common with the PCA-based model in section 4.6.2, though it uses a U-net instead of PCA for parametrization.

A classical criticism of deep learning methods is that they act as a "black box" — they take an input and give an output, but we have very little understanding of what happens in between. The Bayesian models in this thesis as such have an advantage in using more traditional statistics. However, an advantage of the black box is that it comes wrapped, that is, the methods are available in tried and tested software packages that the data scientist for the most part does not need to meddle with.

6.4 Hypofractionation

As the accuracy of RT increases, the trend in the clinic is toward shorter RT courses, i.e. hypofractionation (Kavanagh et al. 2011). Cher et al. (2022) documented that the average number of fractions for all cancers decreased from 17.5 to 13.6 between 2010 and 2020 at one center[1] which treated 22865 patients. This trend comes naturally with higher precision in RT: fractionation is designed to allow normal tissue to heal, and as normal tissue receives less dose, less healing is needed. But fewer fractions coupled with organ motion leads to a new issue: With few fractions, the statistically expected dose cannot be assumed to be equal to the delivered dose. The standard deviation of the delivered dose is proportional to $1/\sqrt{n}$, where n is the number of fractions. In shorter RT courses, one must therefore take the variance of the delivered dose into account. Many robust optimization algorithms already do this, e.g. Sobotta et al. (2010) and Unkelbach and Oelfke (2004). In the evaluation of the methods in this thesis, only the mean dose and the mean coverage probability matrix has been used. A possible direction for future work is therefore including the variance for evaluation of motion models in highly hypofractionated regimes.

6.5 Particle therapy

Organ motion models also have potential applications in particle therapy, where an additional challenge compared to photon RT is *range uncertainty* due to the fact that the position of the Bragg peak where the particles leave most of the dose is highly dependent on the properties (particularly the density) of the tissue through which the beam travels. Therefore, taking organ motion into account through robustness and/or adaptation is even more crucial in particle therapy. Several statistical models have been proposed, but most do not directly model organ deformations (Holloway et al. 2017; Park et al. 2013; Unkelbach et al. 2009), despite organ deformations together with systematic anatomical changes being the greatest sources of range uncertainty according to Lomax (2020). Since a deformation model used for range uncertainty must take the tissue density into account, the most promising approaches are those that generate full synthetic CT scans (Pastor-Serrano et al. 2022; Szeto et al. 2017). However, it is not clear that they conserve the important *water equivalent path length* between training data and output. The problem of statistically accounting for organ deformation is largely an unsolved problem in particle therapy, and is a natural direction for future research.

6.6 On patient preparation

The patients in our validation data did not follow any particular preparation protocol. In general, various preparation techniques and instructions are used to reduce the anatom-

---

[1] Department of Radiation Oncology at the Universiy of Michigan Medical School

ical variation for a patient, such as routines for eating/drinking and toilet visits (Graf et al. 2012; Stasi et al. 2006), special diets (McNair et al. 2011), laxatives (Nijkamp et al. 2008), insertion of rectal balloons (Wachter et al. 2002), emptying rectal gas (Ogino et al. 2008) or daily enemas (Fiorino et al. 2008; Yahya et al. 2013). Some of these techniques cause additional discomfort for the patient, and have become less common or even superfluous with the advancement of IGRT and ART. Studies have shown that a highly distended rectum at the pCT is associated with decreased local control (de Crevoisier et al. 2005; Heemsbergen et al. 2007). Silverman et al. (2016) found that modern IGRT eliminated this effect, while Engels et al. (2009) came to the opposite conclusion. One might think that the resulting recommendation should be to keep conditions as similar as possible between the pCT and the therapy fractions. As argued by Roach (2005), some patients have a consistently distended rectum. However, the recommendations of de Crevoisier et al. (2005) was to eliminate large rectums at the pCT stage using enemas or other means. This has led to a practice, in some centers, of rectal emptying for the pCT only, while no intervention is performed during therapy. This causes a (presumable large) bias of the pCT relative to the treatment. While it may be possible to compensate for this bias in a deformation model, from a statisticians point of view, the pCT and the treatment should always be performed under as similar conditions as possible.

Chapter 7

Conclusions

The main contribution of this thesis is the introduction of new methods to estimate the random and systematic errors due to organ deformation in RT. Through the evaluation of the methods, we have shown that it is possible to efficiently predict patient specific motion patterns for the rectum using Bayesian techniques to combine population and patient-specific data.

For the rectum of prostate cancer patients, it is possible to achieve a significant reduction in systematic error due to organ deformations (in both volumetric and distance measures) using a combination of the rectal shape seen in the pCT and the population mean shape. Whether the reduced systematic error resulted in significant differences in estimated dose depended on the dose level that was examined. The clinical importance of the method will therefore depend on the prescribed treatment and normal tissue dose constraints.

In this doctoral work, two new models based on Bayesian methods were developed. Both these models were shown to more accurately predict organ motion patterns than models using only population data or only patient-specific data for the rectal wall of prostate cancer patients. Due to the novel way of estimating the patient mean shape, the models performed better than the existing population model even when using a single input, i.e. the pCT. The models are further able to improve in accuracy given more inputs, thus getting the “best of both worlds” from pure population or pure patient-specific models.

We were further able to use one of the models as an application to assess dose in patients undergoing re-irradiation. We have presented a method of estimating the accumulated dose and its uncertainty based on “representative shapes” of possible rectum deformations that can be useful for visual aid.

Chapter 8

Future perspectives

Current trends in RT-research are largely focused on new technologies: Adaptive RT using MR-linacs and the Varian Ethos™ system, proton and heavy ion therapy, and (though not yet clinically) FLASH (ultra high dose-rate) RT. These are fantastic, and of course also expensive, technologies that require significant extra work and training for clinicians. The fact is, even in clinics that offer such great treatment options, the vast majority of patients are treated with photons on traditional linacs, and this will likely continue to be true for the foreseeable future (Baumann et al. 2020). One of the great advantages of modelling is that it involves only software, and can therefore offer improvements in treatment for all suitable patients practically free of cost. That is not to say that modelling cannot also provide advantages when used with new technology — as long as there are geometric uncertainties in RT, there will be a need for statistical models. One may expect that advanced motion and deformation modelling will be part of the standard RT-planning in the future, perhaps performed implicitly by the TPS, and surely marketed as "AI".

A natural path of further research is the exploration of applications of the models, including the possibility of implementing models in TPSs through external scripts. This can for example be done by creating the representative shapes presented in paper IV to produce expected DVHs with error bars. While paper IV focused on re-irradiation patients, the methods are also applicable for general patients. The advantage of this technique is that the TPSs own computation of DVHs can be used. The representative shape can even be used in the TPSs optimization algorithm.

The patients in our re-irradiation study had considerable urinary toxicity. We are currently investigating the relationship between accumulated dose to the bladder and urinary toxicity for these patients, using dose accumulation based on techniques from this thesis.

Other possible future studies include applying the models to different organs, or to multiple organs as well as RT-targets to model correlation between these. The models' potential for calculating accumulated dose and their potential use in robust optimizations should also be evaluated.

One of the obstacles for wider usage of statistical deformation models is the need to perform contour-based deformable registration, which is typically not implemented in commercial TPSs. But what if the model itself could be used for deformable registration? It is possible to create an objective function for deformable registration that includes both a metric for the difference between the target shape and the proposed deformed shape, as well as the *probability* of the proposed deformation. To achieve this, a probabilistic model is required. This might seem circular, as deformable registration is necessary to create the model in the first place. However, this is only in the training phase; for a new patient, the model is already in place and can therefore be used for deformable registration as well as prediction of individual motion patterns. Furthermore, it might be possible to train the model using an iterative approach, where a guess of the model is used for deformable registration to start with, and both the training and registration is repeated in an iterative fashion, gradually improving both.

The current models are purely data-driven, in other words, they do not include any modelling of the physical properties of tissue (though the deformable registration usually does to some extent). This has certain advantages in terms of simplicity and reduced number of assumptions. However, this also means that the models are unable to anticipate motion that is not seen in the training data. A simple way to include assumptions for possible motion patterns may be to induce constraints on the covariance matrix estimate. The precision matrix, which is the inverse of the covariance matrix, has the desirable property that the entry for any two coordinates that are *conditionally uncorrelated* is zero. Two coordinates being conditionally uncorrelated means that their correlation can be deduced from the correlation between these coordinates and other coordinates. We can assume that points that are not near each other are conditionally independent, to build a highly sparse precision matrix. This kind of precision matrix will be able to anticipate deformations not seen in the training data, which can be important in e.g. the previously mentioned application of deformable registration.

Bibliography

Abusaris, Huda, M. Hoogeman, and Joost J. Nuyttens (Dec. 2012). “Re-Irradiation: Outcome, Cumulative Dose and Toxicity in Patients Retreated with Stereotactic Radiotherapy in the Abdominal or Pelvic Region”. In: *Technology in Cancer Research & Treatment* 11.6, pp. 591–597. DOI: `10.7785/tcrt.2012.500261`.

Amer, Ali M., Ranald I. Mackay, Stephen A. Roberts, Jolyon H. Hendry, and Peter C. Williams (Nov. 1, 2001). “The required number of treatment imaging days for an effective off-line correction of systematic errors in conformal radiotherapy of prostate cancer — a radiobiological analysis”. In: *Radiotherapy and Oncology* 61.2, pp. 143–150. DOI: `10.1016/S0167-8140(01)00440-6`.

Andratschke, Nicolaus, Jonas Willmann, Ane L. Appelt, Najlaa Alyamani, Panagiotis Balermpas, Brigitta G. Baumert, Coen Hurkmans, Morten Høyer, Johannes A. Langendijk, Orit Kaidar-Person, Yvette van der Linden, Icro Meattini, Maximilian Niyazi, Nick Reynaert, Dirk De Ruysscher, Stephanie Tanadini-Lang, Peter Hoskin, Philip Poortmans, and Carsten Nieder (Oct. 1, 2022). “European Society for Radiotherapy and Oncology and European Organisation for Research and Treatment of Cancer consensus on re-irradiation: definition, reporting, and clinical decision making”. In: *The Lancet Oncology* 23.10, e469–e478. DOI: `10.1016/S1470-2045(22)00447-8`.

Argota-Perez, Raul, Jennifer Robbins, Andrew Green, Marcel van Herk, Stine Korreman, and Eliana Vásquez-Osorio (Apr. 1, 2022). “Evaluating principal component analysis models for representing anatomical changes in head and neck radiotherapy”. In: *Physics and Imaging in Radiation Oncology* 22, pp. 13–19. DOI: `10.1016/j.phro.2022.04.002`.

Badawi, Ahmed M., Elisabeth Weiss, William C. Sleeman IV, Chenyu Yan, and Geoffrey D. Hugo (2010). “Optimizing principal component models for representing interfraction variation in lung cancer radiotherapy”. In: *Medical Physics* 37.9, pp. 5080–5091. DOI: `10.1118/1.3481506`.

Barton, Michael B., Susannah Jacob, Jesmin Shafiq, Karen Wong, Stephen R. Thompson, Timothy P. Hanna, and Geoff P. Delaney (July 1, 2014). “Estimating the demand for radiotherapy from the evidence: A review of changes from 2003 to 2012”. In:

Radiotherapy and Oncology 112.1, pp. 140–144. DOI: `10.1016/j.radonc.2014.03.024`.

Baum, Christoph, Markus Alber, Mattias Birkner, and Fridtjof Nüsslin (Jan. 1, 2006). "Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities". In: *Radiotherapy and Oncology* 78.1, pp. 27–35. DOI: `10.1016/j.radonc.2005.09.005`.

Baumann, Michael, Nadja Ebert, Ina Kurth, Carol Bacchus, and Jens Overgaard (2020). "What will radiation oncology look like in 2050? A look at a changing professional landscape in Europe and beyond". In: *Molecular Oncology* 14.7, pp. 1577–1585. DOI: `10.1002/1878-0261.12731`.

Birkner, M., D. Yan, M. Alber, J. Liang, and F. Nüsslin (Oct. 2003). "Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation". In: *Medical Physics* 30.10, pp. 2822–2831. DOI: `10.1118/1.1610751`.

Bishop, Christopher M. (2006). *Pattern recognition and machine learning.* Information science and statistics. New York: Springer. ISBN: 978-0-387-31073-2.

Biston, M. -C., S. Chiavassa, V. Grégoire, J. Thariat, and T. Lacornerie (Oct. 1, 2020). "Time of PTV is ending, robust optimization comes next". In: *Cancer/Radiothérapie.* 31e Congrès national de la Société française de radiothérapie oncologique 24.6, pp. 676–686. DOI: `10.1016/j.canrad.2020.06.016`.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (Apr. 3, 2017). "Variational Inference: A Review for Statisticians". In: *Journal of the American Statistical Association* 112.518, pp. 859–877. DOI: `10.1080/01621459.2017.1285773`. arXiv: `1601.00670[cs,stat]`.

Bondar, Luiza, Martijn Intven, J. P. Maarten Burbach, Eka Budiarto, Jean-Paul Kleijnen, Marielle Philippens, Bram van Asselen, Enrica Seravalli, Onne Reerink, and Bas Raaymakers (Nov. 1, 2014). "Statistical Modeling of CTV Motion and Deformation for IMRT of Early-Stage Rectal Cancer". In: *International Journal of Radiation Oncology*Biology*Physics* 90.3, pp. 664–672. DOI: `10.1016/j.ijrobp.2014.06.040`.

Borras, Josep M., Yolande Lievens, Peter Dunscombe, Mary Coffey, Julian Malicki, Julieta Corral, Chiara Gasparotto, Noemie Defourny, Michael Barton, Rob Verhoeven, Liesbeth van Eycken, Maja Primic-Zakelj, Maciej Trojanowski, Primoz Strojan, and Cai Grau (July 1, 2015). "The optimal utilization proportion of external beam radiotherapy in European countries: An ESTRO-HERO analysis". In: *Radiotherapy and Oncology* 116.1, pp. 38–44. DOI: `10.1016/j.radonc.2015.04.018`.

Budiarto, E., M. Keijzer, P. R. Storchi, M. S. Hoogeman, L. Bondar, T. F. Mutanga, H. C. J. de Boer, and A. W. Heemink (Feb. 21, 2011). "A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases". In: *Physics in Medicine and Biology* 56.4, pp. 1045–1061. DOI: `10.1088/0031-9155/56/4/011`.

Cher, Benjamin A. Y., Michael Dykstra, Chang Wang, Matthew Schipper, James A. Hayman, Charles S. Mayo, and Reshma Jagsi (Nov. 1, 2022). "Trends in Radiation Oncology Treatment Fractionation at a Single Academic Center, 2010 to 2020". In: *Advances in Radiation Oncology* 7.6, p. 101032. DOI: `10.1016/j.adro.2022.101032`.

Chetvertkov, Mikhail A., Farzan Siddiqui, Jinkoo Kim, Indrin Chetty, Akila Kumarasiri, Chang Liu, and J. James Gordon (2016). "Use of regularized principal component analysis to model anatomical changes during head and neck radiation therapy for treatment adaptation and response assessment". In: *Medical Physics* 43.10, pp. 5307–5319. DOI: `https://doi.org/10.1118/1.4961746`.

Chui, Haili and Anand Rangarajan (Feb. 1, 2003). "A new point matching algorithm for non-rigid registration". In: *Computer Vision and Image Understanding*. Nonrigid Image Registration 89.2, pp. 114–141. DOI: `10.1016/S1077-3142(03)00009-2`.

Cootes, T. F., C. J. Taylor, D. H. Cooper, and J. Graham (1992). "Training Models of Shape from Sets of Examples". In: *BMVC92*. Ed. by David Hogg and Roger Boyle. London: Springer, pp. 9–18. ISBN: 978-1-4471-3201-1. DOI: `10.1007/978-1-4471-3201-1_2`.

Cootes, TF, A Hill, CJ Taylor, and J Haslam (July 1, 1994). "Use of active shape models for locating structures in medical images". In: *Image and Vision Computing*. Information processing in medical imaging 12.6, pp. 355–365. DOI: `10.1016/0262-8856(94)90060-4`.

De Meerleer, Gert, Luc Vakaet, Sabine Meersschout, Geert Villeirs, Antony Verbaeys, Wim Oosterlinck, and Wilfried De Neve (Nov. 1, 2004). "Intensity-modulated radiotherapy as primary treatment for prostate cancer: Acute toxicity in 114 patients". In: *International Journal of Radiation Oncology*Biology*Physics* 60.3, pp. 777–787. DOI: `10.1016/j.ijrobp.2004.04.017`.

De Crevoisier, Renaud, Susan L. Tucker, Lei Dong, Radhe Mohan, Rex Cheung, James D. Cox, and Deborah A. Kuban (July 15, 2005). "Increased risk of biochemical and local failure in patients with distended rectum on the planning CT for prostate cancer radiotherapy". In: *International Journal of Radiation Oncology*Biology*Physics* 62.4, pp. 965–973. DOI: `10.1016/j.ijrobp.2004.11.032`.

Dhou, Salam, John Lewis, Weixing Cai, Dan Ionascu, and Christopher Williams (Apr. 9, 2020). "Quantifying day-to-day variations in 4DCBCT-based PCA motion models". In: *Biomedical Physics & Engineering Express* 6.3, p. 035020. DOI: `10.1088/2057-1976/ab817e`.

Efron, Bradley (Mar. 1, 2005). "Bayesians, Frequentists, and Scientists". In: *Journal of the American Statistical Association* 100.469, pp. 1–5. DOI: `10.1198/016214505000000033`.

Ekanger, Christian, Svein Inge Helle, Daniel Heinrich, Dag Clement Johannessen, Åsa Karlsdóttir, Yngve Nygård, Ole Johan Halvorsen, Lars Reisæter, Rune Kvåle, Liv

Bolstad Hysing, and Olav Dahl (June 2020). “Ten-Year Results From a Phase II Study on Image Guided, Intensity Modulated Radiation Therapy With Simultaneous Integrated Boost in High-Risk Prostate Cancer”. In: *Advances in Radiation Oncology* 5.3, pp. 396–403. DOI: `10.1016/j.adro.2019.11.007`.

Engels, Benedikt, Guy Soete, D. Verellen, and Guy Storme (June 1, 2009). “Conformal arc radiotherapy for prostate cancer: increased biochemical failure in patients with distended rectum on the planning computed tomogram despite image guidance by implanted markers”. In: *International Journal of Radiation Oncology, Biology, Physics* 74.2, pp. 388–391. DOI: `10.1016/j.ijrobp.2008.08.007`.

Fiorino, Claudio, Nadia Di Muzio, Sara Broggi, Cesare Cozzarini, Eleonora Maggiulli, Filippo Alongi, Riccardo Valdagni, Ferruccio Fazio, and Riccardo Calandrino (June 1, 2008). “Evidence of limited motion of the prostate by carefully emptying the rectum as assessed by daily MVCT image guidance with helical tomotherapy”. In: *International Journal of Radiation Oncology, Biology, Physics* 71.2, pp. 611–617. DOI: `10.1016/j.ijrobp.2008.01.048`.

Fontenla, E., C. A. Pelizzari, J. C. Roeske, and G. T. Y. Chen (Aug. 2001a). “Numerical analysis of a model of organ motion using serial imaging measurements from prostate radiotherapy”. In: *Physics in Medicine & Biology* 46.9, p. 2337. DOI: `10.1088/0031-9155/46/9/305`.

– (Aug. 2001b). “Using serial imaging data to model variabilities in organ position and shape during radiotherapy”. In: *Physics in Medicine and Biology* 46.9, pp. 2317–2336. DOI: `10.1088/0031-9155/46/9/304`.

Foray, N. (Aug. 2012). “[Claudius Regaud (1870-1940): A pioneer of radiobiology and radiotherapy]”. In: *Cancer Radiotherapie: Journal De La Societe Francaise De Radiotherapie Oncologique* 16.4, pp. 315–321. DOI: `10.1016/j.canrad.2012.05.006`.

Fornacon-Wood, Isabella, Hitesh Mistry, Corinne Johnson-Hart, Corinne Faivre-Finn, James P. B. O’Connor, and Gareth J. Price (Nov. 1, 2022a). “Bayesian methods provide a practical real-world evidence framework for evaluating the impact of changes in radiotherapy”. In: *Radiotherapy and Oncology* 176, pp. 53–58. DOI: `10.1016/j.radonc.2022.09.009`.

– (Apr. 1, 2022b). “Understanding the Differences Between Bayesian and Frequentist Statistics”. In: *International Journal of Radiation Oncology, Biology, Physics* 112.5, pp. 1076–1082. DOI: `10.1016/j.ijrobp.2021.12.011`.

Fransson, Samuel, David Tilly, Anders Ahnesjö, Tufve Nyholm, and Robin Strand (Oct. 1, 2021). “Intrafractional motion models based on principal components in Magnetic Resonance guided prostate radiotherapy”. In: *Physics and Imaging in Radiation Oncology* 20, pp. 17–22. DOI: `10.1016/j.phro.2021.09.004`.

Fredriksson, Albin, Erik Engwall, and Björn Andersson (Feb. 21, 2021). “Robust radiation therapy optimization using simulated treatment courses for handling deformable organ motion". In: *Physics in Medicine & Biology* 66.4, p. 045010. DOI: `10.1088/1361-6560/abd591`.

Furmanová, Katarína, Ludvig P. Muren, Oscar Casares-Magaz, Vitali Moiseenko, John P. Einck, Sara Pilskog, and Renata G. Raidou (June 1, 2021). "PREVIS: Predictive visual analytics of anatomical variability for radiotherapy decision support". In: *Computers & Graphics* 97, pp. 126–138. DOI: `10.1016/j.cag.2021.04.010`.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Giordano, Sharon H., Andrew Lee, Yong-Fang Kuo, Jean Freeman, and James S. Goodwin (2006). "Late gastrointestinal toxicity after radiation for prostate cancer". In: *Cancer* 107.2, pp. 423–432. DOI: `10.1002/cncr.21999`.

Graf, Reinhold, Dirk Boehmer, Jacek Nadobny, Volker Budach, and Peter Wust (Aug. 1, 2012). "Appropriate patient instructions can reduce prostate motion". In: *Radiation Oncology (London, England)* 7, p. 125. DOI: `10.1186/1748-717X-7-125`.

Heemsbergen, Wilma D., Mischa S. Hoogeman, Marnix G. Witte, Stéphanie T. H. Peeters, Luca Incrocci, and Joos V. Lebesque (Apr. 1, 2007). "Increased Risk of Biochemical and Clinical Failure for Prostate Patients with a Large Rectum at Radiotherapy Planning: Results from the Dutch Trial of 68 GY Versus 78 Gy". In: *International Journal of Radiation Oncology*Biology*Physics* 67.5, pp. 1418–1424. DOI: `10.1016/j.ijrobp.2006.11.014`.

Heimann, Tobias and Hans-Peter Meinzer (Aug. 1, 2009). "Statistical shape models for 3D medical image segmentation: A review". In: *Medical Image Analysis* 13.4, pp. 543–563. DOI: `10.1016/j.media.2009.05.004`.

Herschtal, A., F. Foroudi, P. B. Greer, T. N. Eade, B. R. Hindson, and T. Kron (Apr. 17, 2012). "Finding the optimal statistical model to describe target motion during radiotherapy delivery—a Bayesian approach". In: *Physics in Medicine & Biology* 57.9, pp. 2743–2755. DOI: `10.1088/0031-9155/57/9/2743`.

Herschtal, A., L. Te Marvelde, K. Mengersen, Z. Hosseinifard, F. Foroudi, T. Devereux, D. Pham, D. Ball, P. B. Greer, P. Pichler, T. Eade, A. Kneebone, L. Bell, H. Caine, B. Hindson, and T. Kron (Mar. 7, 2015). "Calculating radiotherapy margins based on Bayesian modelling of patient specific random errors". In: *Physics in Medicine and Biology* 60.5, pp. 1793–1805. DOI: `10.1088/0031-9155/60/5/1793`.

Holloway, S. M., M. D. Holloway, and S. J. Thomas (Dec. 2017). "A method for acquiring random range uncertainty probability distributions in proton therapy". In: *Physics in Medicine & Biology* 63.1, 01NT02. DOI: `10.1088/1361-6560/aa9502`.

Hoogeman, Mischa S, Marcel van Herk, Di Yan, Liesbeth J Boersma, Peter C. M Koper, and Joos V Lebesque (Oct. 1, 2002). "A model to simulate day-to-day variations in rectum shape". In: *International Journal of Radiation Oncology*Biology*Physics* 54.2, pp. 615–625. DOI: `10.1016/S0360-3016(02)02946-2`.

Hoogeman, Mischa S., Marcel van Herk, Josien de Bois, and Joos V. Lebesque (Feb. 1, 2005). “Strategies to reduce the systematic error due to tumor and rectum motion in radiotherapy of prostate cancer”. In: *Radiotherapy and Oncology* 74.2, pp. 177–185. DOI: `10.1016/j.radonc.2004.12.010`.

Hoogeman, Mischa S., Marcel van Herk, Josien de Bois, Pietje Muller-Timmermans, Peter C. M. Koper, and Joos V. Lebesque (Jan. 1, 2004). “Quantification of local rectal wall displacements by virtual rectum unfolding”. In: *Radiotherapy and Oncology* 70.1, pp. 21–30. DOI: `10.1016/j.radonc.2003.11.015`.

Huang, Eugene H, Alan Pollack, Larry Levy, George Starkschall, Lei Dong, Isaac Rosen, and Deborah A Kuban (Dec. 1, 2002). “Late rectal toxicity: dose-volume effects of conformal radiotherapy for prostate cancer”. In: *International Journal of Radiation Oncology*Biology*Physics* 54.5, pp. 1314–1321. DOI: `10.1016/S0360-3016(02)03742-2`.

Hysing, Liv B., Christian Ekanger, Ándras Zolnay, Svein Inge Helle, Mana Rasi, Ben J. M. Heijmen, Marcin Sikora, Matthias Söhn, Ludvig Paul Muren, and Sara Thörnqvist (Aug. 2018). “Statistical motion modelling for robust evaluation of clinically delivered accumulated dose distributions after curative radiotherapy of locally advanced prostate cancer”. In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 128.2, pp. 327–335. DOI: `10.1016/j.radonc.2018.06.004`.

Hysing, Liv Bolstad, Matthias Söhn, Ludvig P. Muren, and Markus Alber (Sept. 2011). “A coverage probability based method to estimate patient-specific small bowel planning volumes for use in radiotherapy”. In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 100.3, pp. 407–411. DOI: `10.1016/j.radonc.2011.08.037`.

ICRU (1999). *ICRU Report 62: Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50).* International Commission on Radiation Units and Measurements.

Kavanagh, Brian D., Moyed Miften, and Rachel A. Rabinovitch (June 2011). “Advances in Treatment Techniques: Stereotactic Body Radiation Therapy and the Spread of Hypofractionation”. In: *The Cancer Journal* 17.3, pp. 177–181. DOI: `10.1097/PPO.0b013e31821f7dbd`.

Kim, Hak Soo, Samuel B. Park, Simon S. Lo, James I. Monroe, and Jason W. Sohn (Nov. 2012). “Bidirectional local distance measure for comparing segmentations”. In: *Medical Physics* 39.11, pp. 6779–6790. DOI: `10.1118/1.4754802`.

Kleijnen, Jean-Paul J. E., Bram van Asselen, Johannes P. M. Burbach, Martijn Intven, Marielle E. P. Philippens, Onne Reerink, Jan J. W. Lagendijk, and Bas W. Raaymakers (Nov. 25, 2015). “Evolution of motion uncertainty in rectal cancer: implications for adaptive radiotherapy”. In: *Physics in Medicine & Biology* 61.1, p. 1. DOI: `10.1088/0031-9155/61/1/1`.

Lam, Kwok L., Randall K. Ten Haken, Dale Litzenberg, James M. Balter, and Stephen M. Pollock (Aug. 21, 2005). “An application of Bayesian statistical methods to adaptive radiotherapy”. In: *Physics in Medicine and Biology* 50.16, pp. 3849–3858. DOI: `10.1088/0031-9155/50/16/013`.

Li, Jonathan G. and Lei Xing (2000). “Inverse planning incorporating organ motion”. In: *Medical Physics* 27.7, pp. 1573–1578. DOI: `10.1118/1.599023`.

Li, Ruijiang, John H. Lewis, Xun Jia, Tianyu Zhao, Weifeng Liu, Sara Wuenschel, James Lamb, Deshan Yang, Daniel A. Low, and Steve B. Jiang (Aug. 2011a). “On a PCA-based lung motion model”. In: *Physics in Medicine & Biology* 56.18, p. 6009. DOI: `10.1088/0031-9155/56/18/015`.

Li, X. Allen, Feng Liu, An Tai, Ergun Ahunbay, Guangpei Chen, Tracy Kelly, Colleen Lawton, and Beth Erickson (Sept. 2011b). “Development of an online adaptive solution to account for inter- and intra-fractional variations”. In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 100.3, pp. 370–374. DOI: `10.1016/j.radonc.2011.08.027`.

Liang, J. and D. Yan (2003). “Reducing uncertainties in volumetric image based deformable organ registration”. In: *Medical Physics* 30.8, pp. 2116–2122. DOI: `10.1118/1.1587631`.

Lind, Bengt K., Patric Kållman, BjÖRn Sundelin, and Anders Brahme (Jan. 1, 1993). “Optimal Radiation Beam Profiles Considering Uncertainties in Beam Patient Alignment”. In: *Acta Oncologica* 32.3, pp. 331–342. DOI: `10.3109/02841869309093605`.

Lindegaard, Jacob Chr., Marianne Assenholt, Anne Ramlov, Lars Ulrik Fokdal, Markus Alber, and Kari Tanderup (Nov. 2, 2017). “Early clinical outcome of coverage probability based treatment planning for simultaneous integrated boost of nodes in locally advanced cervical cancer”. In: *Acta Oncologica* 56.11, pp. 1479–1486. DOI: `10.1080/0284186X.2017.1349335`.

Liu, Jianfei, Y. Q. Chen, José M. Maisog, and George Luta (Dec. 1, 2010). “A new point containment test algorithm based on preprocessing and determining triangles”. In: *Computer-Aided Design* 42.12, pp. 1143–1150. DOI: `10.1016/j.cad.2010.08.002`.

Liu, Lianli, Adam Johansson, Yue Cao, Rojano Kashani, Theodore S. Lawrence, and James M. Balter (Apr. 2021). “Modeling intra-fractional abdominal configuration changes using breathing motion-corrected radial MRI”. In: *Physics in Medicine & Biology* 66.8, p. 085002. DOI: `10.1088/1361-6560/abef42`.

Lof, J., B. K. Lind, and A. Brahme (Dec. 1995). “Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy”. In: *Inverse Problems* 11.6, p. 1189. DOI: `10.1088/0266-5611/11/6/005`.

Lomax, Antony John (Mar. 2020). “Myths and realities of range uncertainty”. In: *The British Journal of Radiology* 93.1107, p. 20190582. DOI: `10.1259/bjr.20190582`.

Lorensen, William E. and Harvey E. Cline (Aug. 1, 1987). “Marching cubes: A high resolution 3D surface construction algorithm”. In: *ACM SIGGRAPH Computer Graphics* 21.4, pp. 163–169. DOI: `10.1145/37402.37422`.

Magallon-Baro, Alba, Mauro Loi, Maaike T. W. Milder, Patrick V. Granton, Andras G. Zolnay, Joost J. Nuyttens, and Mischa S. Hoogeman (May 1, 2019). “Modeling daily changes in organ-at-risk anatomy in a cohort of pancreatic cancer patients”. In: *Radiotherapy and Oncology* 134, pp. 127–134. DOI: `10.1016/j.radonc.2019.01.030`.

Mageras, Gig S., Gerald J. Kutcher, Steven A. Leibel, Michael J. Zelefsky, Edward Melian, Radhe Mohan, and Zvi Fuks (May 1, 1996). “A method of incorporating organ motion uncertainties into three-dimensional conformal treatment plans”. In: *International Journal of Radiation Oncology*Biology*Physics* 35.2, pp. 333–342. DOI: `10.1016/0360-3016(96)00008-9`.

McComas, Kyra N., Adam Yock, Kaleb Darrow, and Eric T. Shinohara (Oct. 26, 2022). “Online Adaptive Radiation Therapy and Opportunity Cost”. In: *Advances in Radiation Oncology*, p. 101034. DOI: `10.1016/j.adro.2022.101034`.

McNair, Helen A., Vibeke N. Hansen, Christopher C. Parker, Phil M. Evans, Andrew Norman, Elizabeth Miles, Emma J. Harris, Louise Del-Acroix, Elizabeth Smith, Richard Keane, Vincent S. Khoo, Alan C. Thompson, and David P. Dearnaley (May 1, 2008). “A Comparison of the Use of Bony Anatomy and Internal Markers for Offline Verification and an Evaluation of the Potential Benefit of Online and Offline Verification Protocols for Prostate Radiotherapy”. In: *International Journal of Radiation Oncology*Biology*Physics* 71.1, pp. 41–50. DOI: `10.1016/j.ijrobp.2007.09.002`.

McNair, Helen A., Linda Wedlake, Gerard P. McVey, Karen Thomas, Jervoise Andreyev, and David P. Dearnaley (Dec. 2011). “Can diet combined with treatment scheduling achieve consistency of rectal filling in patients receiving radiotherapy to the prostate?” In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 101.3, pp. 471–478. DOI: `10.1016/j.radonc.2011.08.003`.

Minaee, Shervin, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos (July 2022). “Image Segmentation Using Deep Learning: A Survey”. In: *IEEE transactions on pattern analysis and machine intelligence* 44.7, pp. 3523–3542. DOI: `10.1109/TPAMI.2021.3059968`.

Munoz, Fernando, Francesco Fiorica, Luciana Caravatta, Consuelo Rosa, Letizia Ferella, Luca Boldrini, Bruno Fionda, Anna Rita Alitto, Alessia Nardangeli, Francesco Dionisi, Stefano Arcangeli, Alessandro Di Marzo, Antonio Pontoriero, Vittorio Donato, and Mariangela Massaccesi (Apr. 1, 2021). “Outcomes and toxicities of re-irradiation for prostate cancer: A systematic review on behalf of the Re-Irradiation Working Group of the Italian Association of Radiotherapy and Clinical Oncology (AIRO)”. In: *Cancer Treatment Reviews* 95, p. 102176. DOI: `10.1016/j.ctrv.2021.102176`.

Nakao, Megumi, Mitsuhiro Nakamura, Takashi Mizowaki, and Tetsuya Matsuda (Jan. 1, 2021). "Statistical deformation reconstruction using multi-organ shape features for pancreatic cancer localization". In: *Medical Image Analysis* 67, p. 101829. DOI: `10.1016/j.media.2020.101829`.

NEMA (2022). *NEMA PS3 / ISO 12052: Digital Imaging and Communications in Medicine (DICOM) Standard*. Rosslyn, VA, USA: National Electrical Manufacturers Association. URL: `http://medical.nema.org/`.

Nie, Xiaobo, Jian Liang, and Di Yan (2012). "Organ sample generator for expected treatment dose construction and adaptive inverse planning optimization". In: *Medical Physics* 39.12, pp. 7329–7337. DOI: `https://doi.org/10.1118/1.4765457`.

Nijkamp, Jasper, Floris J. Pos, Tonnis T. Nuver, Rianne de Jong, Peter Remeijer, Jan-Jakob Sonke, and Joos V. Lebesque (Jan. 1, 2008). "Adaptive Radiotherapy for Prostate Cancer Using Kilovoltage Cone-Beam Computed Tomography: First Clinical Results". In: *International Journal of Radiation Oncology*Biology*Physics* 70.1, pp. 75–82. DOI: `10.1016/j.ijrobp.2007.05.046`.

Ogino, Ichiro, Hiroji Uemura, Tomio Inoue, Yoshinobu Kubota, Kazuto Nomura, and Naoyuki Okamoto (Oct. 1, 2008). "Reduction of prostate motion by removal of gas in rectum during radiotherapy". In: *International Journal of Radiation Oncology, Biology, Physics* 72.2, pp. 456–466. DOI: `10.1016/j.ijrobp.2008.01.004`.

Ong, Ashley Li Kuan, Kellie Knight, Vanessa Panettieri, Mathew Dimmock, Jeffrey Kit Loong Tuan, Hong Qi Tan, and Caroline Wright (2022). "Predictive modelling for late rectal and urinary toxicities after prostate radiotherapy using planned and delivered dose". In: *Frontiers in Oncology* 12, p. 1084311. DOI: `10.3389/fonc.2022.1084311`.

Owens, Constance A., Bastien Rigaud, Ethan B. Ludmir, Aashish C. Gupta, Suman Shrestha, Arnold C. Paulino, Susan A. Smith, Christine B. Peterson, Stephen F. Kry, Choonsik Lee, Tara O. Henderson, Gregory T. Armstrong, Kristy K. Brock, and Rebecca M. Howell (Sept. 5, 2022). "Development and validation of a population-based anatomical colorectal model for radiation dosimetry in late effects studies of survivors of childhood cancer". In: *Radiotherapy and Oncology*. DOI: `10.1016/j.radonc.2022.08.027`.

Paradis, Kelly C., Charles Mayo, Dawn Owen, Daniel E. Spratt, Jason Hearn, Benjamin Rosen, Rojano Kashani, Jean Moran, Daniel S. Tatro, Whitney Beeler, Karen Vineberg, Dylan C. Smith, and Martha M. Matuszak (Oct. 1, 2019). "The Special Medical Physics Consult Process for Reirradiation Patients". In: *Advances in Radiation Oncology* 4.4, pp. 559–565. DOI: `10.1016/j.adro.2019.05.007`.

Park, Peter C., Joey P. Cheung, X. Ronald Zhu, Andrew K. Lee, Narayan Sahoo, Susan L. Tucker, Wei Liu, Heng Li, Radhe Mohan, Laurence E. Court, and Lei Dong (Aug. 1, 2013). "Statistical Assessment of Proton Treatment Plans Under Setup and Range

Uncertainties”. In: *International Journal of Radiation Oncology*Biology*Physics* 86.5, pp. 1007–1013. DOI: `10.1016/j.ijrobp.2013.04.009`.

Pastor-Serrano, Oscar, Steven Habraken, Mischa Hoogeman, Danny Lathouwers, Dennis Schaart, Yusuke Nomura, Lei Xing, and Zoltán Perkó (Sept. 20, 2022). *A probabilistic deep learning model of inter-fraction anatomical variations in radiotherapy.* DOI: `10.48550/arXiv.2209.09583`. arXiv: `2209.09583[physics]`.

Pinkawa, Michael, Jaroslav Siluschek, Bernd Gagel, Cengiz Demirel, Branka Asadpour, Richard Holy, and Michael J. Eble (Dec. 1, 2006). “Influence of the initial rectal distension on posterior margins in primary and postoperative radiotherapy for prostate cancer”. In: *Radiotherapy and Oncology* 81.3, pp. 284–290. DOI: `10.1016/j.radonc.2006.10.028`.

Poon, Aaron I. F. and Joseph J. Y. Sung (Mar. 2021). “Opening the black box of AI-Medicine”. In: *Journal of Gastroenterology and Hepatology* 36.3, pp. 581–584. DOI: `10.1111/jgh.15384`.

Ramlov, Anne, Marianne S. Assenholt, Maria F. Jensen, Caroline Grønborg, Remi Nout, Markus Alber, Lars Fokdal, Kari Tanderup, and Jacob Chr. Lindegaard (Apr. 1, 2017). “Clinical implementation of coverage probability planning for nodal boosting in locally advanced cervical cancer”. In: *Radiotherapy and Oncology* 123.1, pp. 158–163. DOI: `10.1016/j.radonc.2017.01.015`.

Rigaud, B., A. Simon, M. Gobeli, J. Leseur, L. Duvergé, D. Williaume, J. Castelli, C. Lafond, O. Acosta, P. Haigron, and R. De Crevoisier (Feb. 2019). “Statistical Shape Model to Generate a Planning Library for Cervical Adaptive Radiotherapy”. In: *IEEE Transactions on Medical Imaging* 38.2, pp. 406–416. DOI: `10.1109/TMI.2018.2865547`.

Rios, R., J. D. Ospina, C. Lafond, O. Acosta, J. Espinosa, and R. de Crevoisier (Nov. 1, 2016). “Characterization of Bladder Motion and Deformation in Prostate Cancer Radiotherapy”. In: *IRBM* 37.5, pp. 276–283. DOI: `10.1016/j.irbm.2016.09.001`.

Rios, Richard, Renaud De Crevoisier, Juan D. Ospina, Frederic Commandeur, Caroline Lafond, Antoine Simon, Pascal Haigron, Jairo Espinosa, and Oscar Acosta (May 1, 2017). “Population model of bladder motion and deformation based on dominant eigenmodes and mixed-effects models in prostate cancer radiotherapy”. In: *Medical Image Analysis* 38, pp. 133–149. DOI: `10.1016/j.media.2017.03.001`.

Roach, Mack (July 15, 2005). “Commentary on increased risk of biochemical and local failure in patients with a distended rectum on the planning CT for prostate cancer radiotherapy”. In: *International Journal of Radiation Oncology, Biology, Physics* 62.4, pp. 949–950. DOI: `10.1016/j.ijrobp.2005.03.067`.

Roch, M., A. Zapatero, P. Castro, D. Hernández, M. Chevalier, and F. García-Vicente (Nov. 2021). “Dosimetric impact of rectum and bladder anatomy and intrafractional prostate motion on hypofractionated prostate radiation therapy”. In: *Clinical & Trans-*

lational Oncology: Official Publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico 23.11, pp. 2293–2301. DOI: `10.1007/s12094-021-02628-3`.

Rørtveit, Øyvind Lunde (2022). *PolyhedronToPolygons (Software).* Version 1.0.0. URL: `https://se.mathworks.com/matlabcentral/fileexchange/119313-polyhedrontopolygons`.

Rørtveit, Øyvind Lunde, Liv Bolstad Hysing, Andreas Størksen Stordal, and Sara Pilskog (2021). "Reducing systematic errors due to deformation of organs at risk in radiotherapy". In: *Medical Physics* 48.11, pp. 6578–6587. DOI: `10.1002/mp.15262`.

– (Feb. 2023). "An organ deformation model using Bayesian inference to combine population and patient-specific data". In: *Physics in Medicine & Biology* 68.5, p. 055009. DOI: `10.1088/1361-6560/acb8fc`.

Rudat, Volker, A. Nour, M. Hammoud, A. Alaradi, and A. Mohammed (Feb. 2016). "Image-guided intensity-modulated radiotherapy of prostate cancer: Analysis of interfractional errors and acute toxicity". In: *Strahlentherapie Und Onkologie: Organ Der Deutschen Rontgengesellschaft ... [et Al]* 192.2, pp. 109–117. DOI: `10.1007/s00066-015-0919-y`.

Scaife, Jessica E., Simon J. Thomas, Karl Harrison, Marina Romanchikova, Michael P. F. Sutcliffe, Julia R. Forman, Amy M. Bates, Raj Jena, M. Andrew Parker, and Neil G. Burnet (Oct. 2015). "Accumulated dose to the rectum, measured using dose-volume histograms and dose-surface maps, is different from planned dose in all patients treated with radiotherapy for prostate cancer". In: *The British Journal of Radiology* 88.1054, p. 20150243. DOI: `10.1259/bjr.20150243`.

Schultheiss, T. E., W. R. Lee, M. A. Hunt, A. L. Hanlon, R. S. Peter, and G. E. Hanks (Jan. 1, 1997). "Late GI and GU complications in the treatment of prostate cancer". In: *International Journal of Radiation Oncology, Biology, Physics* 37.1, pp. 3–11. DOI: `10.1016/s0360-3016(96)00468-3`.

Sevillano, David, Ana B. Capuz, Alberto Gómez, Rafael Colmenares, Rafael Morís, Juan D. García, Maddalen Alonso, Miguel Cámara, Ana M. Martínez, María J. Béjar, Daniel Prieto, Sonsoles Sancho, Margarita Chevalier, and Feliciano García-Vicente (2019). "On the use of Bayesian statistics in the application of adaptive setup protocols in radiotherapy". In: *Medical Physics* 46.10, pp. 4622–4630. DOI: `10.1002/mp.13745`.

Shelley, L. E. A., J. E. Scaife, M. Romanchikova, K. Harrison, J. R. Forman, A. M. Bates, D. J. Noble, R. Jena, M. A. Parker, M. P. F. Sutcliffe, S. J. Thomas, and N. G. Burnet (June 2017). "Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy". In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 123.3, pp. 466–471. DOI: `10.1016/j.radonc.2017.04.008`.

Silverman, Rafael, Kerstie Johnson, Christina Perry, and Santhanam Sundar (2016). "Degree of Rectal Distension Seen on Prostate Radiotherapy Planning CT Scan Is

Not a Negative Prognostic Factor in the Modern Era of Image-Guided Radiotherapy". In: *Oncology* 90.1, pp. 51–56. DOI: `10.1159/000441225`.

Singh, Jasmeet, Peter B. Greer, Martin A. White, Joel Parker, Jackie Patterson, Colin I. Tang, Anne Capp, Christopher Wratten, and James W. Denham (Mar. 15, 2013). "Treatment-Related Morbidity in Prostate Cancer: A Comparison of 3-Dimensional Conformal Radiation Therapy With and Without Image Guidance Using Implanted Fiducial Markers". In: *International Journal of Radiation Oncology*Biology*Physics* 85.4, pp. 1018–1023. DOI: `10.1016/j.ijrobp.2012.07.2376`.

Sobotta, B., M. Söhn, and M. Alber (Aug. 2010). "Robust optimization based upon statistical theory". In: *Medical Physics* 37.8, pp. 4019–4028. DOI: `10.1118/1.3457333`.

Söhn, M., M. Birkner, D. Yan, and M. Alber (Dec. 21, 2005). "Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation". In: *Physics in Medicine and Biology* 50.24, pp. 5893–5908. DOI: `10.1088/0031-9155/50/24/009`.

Söhn, M., B. Sobotta, and M. Alber (June 21, 2012). "Dosimetric treatment course simulation based on a statistical model of deformable organ motion". In: *Physics in Medicine and Biology* 57.12, pp. 3693–3709. DOI: `10.1088/0031-9155/57/12/3693`.

Sripadam, Raj, Julie Stratford, Ann M. Henry, Andrew Jackson, Chris J. Moore, and Pat Price (Mar. 2009). "Rectal motion can reduce CTV coverage and increase rectal dose during prostate radiotherapy: A daily cone-beam CT study". In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 90.3, pp. 312–317. DOI: `10.1016/j.radonc.2008.07.031`.

Stasi, Michele, Fernando Munoz, Claudio Fiorino, Massimo Pasquino, Barbara Baiotto, Piergiorgio Marini, Giuseppe Malinverni, Riccardo Valdagni, and Pietro Gabriele (Sept. 1, 2006). "Emptying the rectum before treatment delivery limits the variations of rectal dose–volume parameters during 3DCRT of prostate cancer". In: *Radiotherapy and Oncology* 80.3, pp. 363–370. DOI: `10.1016/j.radonc.2006.08.007`.

Stroom, J. C., H. C. de Boer, H. Huizenga, and A. G. Visser (Mar. 1, 1999). "Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability". In: *International Journal of Radiation Oncology, Biology, Physics* 43.4, pp. 905–919. DOI: `10.1016/s0360-3016(98)00468-4`.

Stroom, Joep C. and Ben J. M. Heijmen (Sept. 1, 2006). "Limitations of the planning organ at risk volume (PRV) concept". In: *International Journal of Radiation Oncology, Biology, Physics* 66.1, pp. 279–286. DOI: `10.1016/j.ijrobp.2006.05.009`.

Szeto, Yenny Z., Marnix G. Witte, Marcel van Herk, and Jan-Jakob Sonke (Apr. 1, 2017). "A population based statistical model for daily geometric variations in the thorax". In: *Radiotherapy and Oncology* 123.1, pp. 99–105. DOI: `10.1016/j.radonc.2017.02.012`.

Thor, Maria, Aditya Apte, Joseph O. Deasy, Àsa Karlsdóttir, Vitali Moiseenko, Mitchell Liu, and Ludvig Paul Muren (Dec. 1, 2013a). "Dose/volume–response relations for rectal morbidity using planned and simulated motion-inclusive dose distributions". In: *Radiotherapy and Oncology* 109.3, pp. 388–393. DOI: `10.1016/j.radonc.2013.10.021`.

Thor, Maria, Lise Bentzen, Liv B. Hysing, Christian Ekanger, Svein-Inge Helle, Ásа Karlsdóttir, and Ludvig Paul Muren (May 2013b). "Prediction of rectum and bladder morbidity following radiotherapy of prostate cancer based on motion-inclusive dose distributions". In: *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 107.2, pp. 147–152. DOI: `10.1016/j.radonc.2013.03.029`.

Thörnqvist, Sara, Liv B. Hysing, Andras G. Zolnay, Matthias Söhn, Mischa S. Hoogeman, Ludvig P. Muren, Lise Bentzen, and Ben J. M. Heijmen (Dec. 1, 2013a). "Treatment simulations with a statistical deformable motion model to evaluate margins for multiple targets in radiotherapy for high-risk prostate cancer". In: *Radiotherapy and Oncology* 109.3, pp. 344–349. DOI: `10.1016/j.radonc.2013.09.012`.

Thörnqvist, Sara, Liv B. Hysing, Andras G. Zolnay, Matthias Söhn, Mischa S. Hoogeman, Ludvig P. Muren, and Ben J. M. Heijmen (Oct. 2013b). "Adaptive radiotherapy in locally advanced prostate cancer using a statistical deformable motion model". In: *Acta Oncologica (Stockholm, Sweden)* 52.7, pp. 1423–1429. DOI: `10.3109/0284186X.2013.818249`.

Tilly, David, Agustinus J. A. J. van de Schoot, Erik Grusell, Arjan Bel, and Anders Ahnesjö (Apr. 2017). "Dose coverage calculation using a statistical shape model—applied to cervical cancer radiotherapy". In: *Physics in Medicine and Biology* 62.10, pp. 4140–4159. DOI: `10.1088/1361-6560/aa64ef`.

Tsiamas, Panagiotis, Hassan Bagher-Ebadian, Farzan Siddiqui, Chang Liu, Christian A. Hvid, Joshua P. Kim, Stephen L. Brown, Benjamin Movsas, and Indrin J. Chetty (Dec. 2018). "Principal component analysis modeling of Head-and-Neck anatomy using daily Cone Beam-CT images". In: *Medical Physics* 45.12, pp. 5366–5375. DOI: `10.1002/mp.13233`.

Unkelbach, J. and U. Oelfke (Aug. 20, 2004). "Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions". In: *Physics in Medicine & Biology* 49.17, p. 4005. DOI: `10.1088/0031-9155/49/17/013`.

Unkelbach, Jan, Markus Alber, Mark Bangert, Rasmus Bokrantz, Timothy C. Y. Chan, Joseph O. Deasy, Albin Fredriksson, Bram L. Gorissen, Marcel van Herk, Wei Liu, Houra Mahmoudzadeh, Omid Nohadani, Jeffrey V. Siebers, Marnix Witte, and Huijun Xu (Nov. 2018). "Robust radiotherapy planning". In: *Physics in Medicine & Biology* 63.22, 22TR02. DOI: `10.1088/1361-6560/aae659`.

Unkelbach, Jan, Thomas Bortfeld, Benjamin C. Martin, and Martin Soukup (2009). "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning". In: *Medical Physics* 36.1, pp. 149–163. DOI: `10.1118/1.3021139`.

Unkelbach, Jan and Uwe Oelfke (2005). "Incorporating organ movements in IMRT treatment planning for prostate cancer: Minimizing uncertainties in the inverse planning process". In: *Medical Physics* 32.8, pp. 2471–2483. DOI: `https://doi.org/10.1118/1.1929167`.

Van Herk, Marcel (Jan. 1, 2004). "Errors and margins in radiotherapy". In: *Seminars in Radiation Oncology.* High-Precision Radiation Therapy of Moving Targets 14.1, pp. 52–64. DOI: `10.1053/j.semradonc.2003.10.003`.

Van Herk, Marcel, Peter Remeijer, Coen Rasch, and Joos V. Lebesque (July 1, 2000). "The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy". In: *International Journal of Radiation Oncology*Biology*Physics* 47.4, pp. 1121–1135. DOI: `10.1016/S0360-3016(00)00518-6`.

Van Herk, Marcel, M. Witte, and P. Remeijer (2009). "Performance of patient specific margins derived using a Bayesian statistical method". In: *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany.* Ed. by Olaf Dössel and Wolfgang C. Schlegel. IFMBE Proceedings. Berlin, Heidelberg: Springer, pp. 769–771. ISBN: 978-3-642-03474-9. DOI: `10.1007/978-3-642-03474-9_217`.

Vásquez Osorio, Eliana M. (Apr. 5, 2012). "Improving Feature-based Non-rigid Registration for Applications in Radiotherapy". PhD thesis. Erasmus University Rotterdam. URL: `https://repub.eur.nl/pub/32083/` (visited on 10/22/2022).

Vásquez Osorio, Eliana M., Mischa S. Hoogeman, Luiza Bondar, Peter C. Levendag, and Ben J. M. Heijmen (2009). "A novel flexible framework with automatic feature correspondence optimization for nonrigid registration in radiotherapy". In: *Medical Physics* 36.7, pp. 2848–2859. DOI: `https://doi.org/10.1118/1.3134242`.

Vásquez Osorio, Eliana M., Mischa S. Hoogeman, Alejandra Mendez Romero, Piotr Wielopolski, Andras Zolnay, and Ben J. M. Heijmen (May 2012). "Accurate CTaMR vessel-guided nonrigid registration of largely deformed livers". In: *Medical Physics* 39.5, pp. 2463–2477. DOI: `10.1118/1.3701779`.

Wachter, Stefan, Natascha Gerstner, Daniela Dorner, Gregor Goldner, Adriana Colotto, Andre Wambersie, and Richard Pötter (Jan. 1, 2002). "The influence of a rectal balloon tube as internal immobilization device on variations of volumes and dose-volume histograms during treatment course of conformal radiotherapy for prostate cancer". In: *International Journal of Radiation Oncology*Biology*Physics* 52.1, pp. 91–100. DOI: `10.1016/S0360-3016(01)01821-1`.

White, Ingrid, Arabella Hunt, Thomas Bird, Sarah Settatree, Heba Soliman, Dualta Mcquaid, David Dearnaley, Susan Lalondrelle, and Shree Bhide (Dec. 2021). "Interobserver variability in target volume delineation for CT/MRI simulation and MRI-guided adaptive radiotherapy in rectal cancer". In: *The British Journal of Radiology* 94.1128, p. 20210350. DOI: `10.1259/bjr.20210350`.

Widmark, Anders, Adalsteinn Gunnlaugsson, Lars Beckman, Camilla Thellenberg-Karlsson, Morten Hoyer, Magnus Lagerlund, Jon Kindblom, Claes Ginman, Bengt Johansson, Kirsten Björnlinger, Mihajl Seke, Måns Agrup, Per Fransson, Björn Tavelin, David Norman, Björn Zackrisson, Harald Anderson, Elisabeth Kjellén, Lars Franzén, and Per Nilsson (Aug. 3, 2019). "Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomised, non-inferiority, phase 3 trial". In: *The Lancet* 394.10196, pp. 385–395. DOI: `10.1016/S0140-6736(19)31131-6`.

Wolthaus, J. W. H., J.-J. Sonke, M. van Herk, and E. M. F. Damen (2008). "Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registrationa)". In: *Medical Physics* 35.9, pp. 3998–4011. DOI: `10.1118/1.2966347`.

Xiao, Haonan, Xinzhi Teng, Chenyang Liu, Tian Li, Ge Ren, Ruijie Yang, Dinggang Shen, and Jing Cai (Dec. 2021). "A review of deep learning-based three-dimensional medical image registration methods". In: *Quantitative Imaging in Medicine and Surgery* 11.12, pp. 4895–4916. DOI: `10.21037/qims-21-175`.

Xu, Huijun, Douglas J. Vile, Manju Sharma, J. James Gordon, and Jeffrey V. Siebers (2014). "Coverage-based treatment planning to accommodate deformable organ variations in prostate cancer treatment". In: *Medical Physics* 41.10, p. 101705. DOI: `https://doi.org/10.1118/1.4894701`.

Yahya, S, A Zarkar, E Southgate, P Nightingale, and G Webster (Nov. 2013). "Which bowel preparation is best? Comparison of a high-fibre diet leaflet, daily microenema and no preparation in prostate cancer patients treated with radical radiotherapy to assess the effect on planned target volume shifts due to rectal distension". In: *The British Journal of Radiology* 86.1031, p. 20130457. DOI: `10.1259/bjr.20130457`.

Yan, Di, D. A. Jaffray, and J. W. Wong (June 1, 1999). "A model to accumulate fractionated dose in a deforming organ". In: *International Journal of Radiation Oncology*Biology*Physics* 44.3, pp. 665–675. DOI: `10.1016/S0360-3016(99)00007-3`.

Zelefsky, Michael J., Emily J. Levin, Margie Hunt, Yoshiya Yamada, Alison M. Shippy, Andrew Jackson, and Howard I. Amols (Mar. 15, 2008). "Incidence of Late Rectal and Urinary Toxicities After Three-Dimensional Conformal Radiotherapy and Intensity-Modulated Radiotherapy for Localized Prostate Cancer". In: *International Journal of Radiation Oncology*Biology*Physics* 70.4, pp. 1124–1129. DOI: `10.1016/j.ijrobp.2007.11.044`.

Publications

Paper I

Reducing systematic errors due to deformation of organs at risk in radiotherapy

Øyvind Lunde Rørtveit, Liv Bolstad Hysing, Andreas Størksen Stordal, Sara Pilskog

Medical Physics **48** 6578–6587 (2021).
DOI: 10.1002/mp.15262
Open access

Received: 17 June 2021 | Revised: 24 August 2021 | Accepted: 16 September 2021

DOI: 10.1002/mp.15262

RESEARCH ARTICLE

MEDICAL PHYSICS

Reducing systematic errors due to deformation of organs at risk in radiotherapy

Øyvind Lunde Rørtveit[1,2] | **Liv Bolstad Hysing**[1,2] | **Andreas Størksen Stordal**[3,4] | **Sara Pilskog**[1,2]

[1] Department of Oncology and Medical Physics, Haukeland University Hospital, Bergen, Norway

[2] Department of physics and technology, University of Bergen, Bergen, Norway

[3] NORCE Research, Bergen, Norway

[4] Department of Mathematics, University of Bergen, Bergen, Norway

Correspondence
Øyvind Lunde Rørtveit, Department of Oncology and Medical Physics, Haukeland University Hospital, P.O.B. 1400, 5021 Bergen, Norway.
Email: oyvind.lunde.rortveit@helse-bergen.no

Funding information
Trond Mohn stiftelse, Grant/Award Number: BFS2017TMT07

Abstract

Purpose: In radiotherapy (RT), the planning CT (pCT) is commonly used to plan the full RT-course. Due to organ deformation and motion, the organ shapes seen at the pCT will not be identical to their shapes during RT. Any difference between the pCT organ shape and the organ's mean shape during RT will cause systematic errors. We propose to use statistical shrinkage estimation to reduce this error using only the pCT and the population mean shape computed from training data.

Methods: The method was evaluated for the rectum in a cohort of 37 prostate cancer patients that had a pCT and 7–10 treatment CTs with rectum delineations. Deformable registration was performed both within-patient and between patients, resulting in point-to-point correspondence between all rectum shapes, which enabled us to compute a population mean rectum. Shrinkage estimates were found by combining the pCTs linearly with the population mean.

The method was trained and evaluated using leave-one-out cross validation. The shrinkage estimates and the patient mean shapes were compared geometrically using the Dice similarity index (DSI), Hausdorff distance (HD), and bidirectional local distance. Clinical dose/volume histograms, equivalent uniform dose (EUD) and minimum dose to the hottest 5% volume (D5%) were compared for the shrinkage estimate and the pCT.

Results: The method resulted in moderate but statistically significant increase in similarity to the patient mean shape over the pCT. On average, the HD was reduced from 15.6 to 13.4 mm, while the DSI was increased from 0.74 to 0.78. Significant reduction in the bias of volume estimates was found in the DVH-range of 52.5–65 Gy, where the bias was reduced from −1.3 to −0.2 percentage points, but no significant improvement was found in EUD or D5%,

Conclusions: The results suggest that shrinkage estimation can reduce systematic errors due to organ deformations in RT. The method has potential to increase the accuracy in RT of deformable organs and can improve motion modeling.

KEYWORDS
organ deformation, organ motion, rectum, systematic errors, treatment uncertainties

1 INTRODUCTION

The planning CT (pCT) is used as a representation of the patient anatomy during treatment. Changes to the patient anatomy and its organ shapes during treatment cause differences between the planned and the delivered doses.[1] Discrepancies between the pCT anatomy and the average anatomy during treatment are called systematic errors, since they affect every fraction. While this term is often used to describe the discrepancy in setup position, systematic errors also occur when an organ of interest has a different shape in the pCT as compared to its average shape during treatment. This study focuses on this latter systematic shape error. Potential systematic errors in organs-at-risk (OARs) can be accounted for through robust optimization,[2] or with margins (planning OAR volumes, PRVs).[3] Unlike these methods, the presented method aims to predict and correct for the systematic error. Correcting for systematic setup error (i.e., without deformation) by using several scans, typically taken during the first few fractions of RT, has been investigated.[4] A similar method can be used to handle deformation, using deformable registration from multiple scans to find an average shape,[5] but this would require adaptive re-planning unless all scans are taken before the first treatment. The presented method, on the other hand, requires only a single scan. Image-guidance has been successful in reducing the systematic errors for the target volumes, but for shape changes and for many OARs, mitigation strategies based on rigid re-alignment are insufficient, calling for more resource demanding adaptive RT.[6,7] Even adaptive RT with replanning at every fraction is not a perfect remedy, since intra-fraction motion can be considerable.[8,9]

For the rectum, one of the dose-limiting OARs in both prostate and cervical cancer RT, the shape of the organ seen in the pCT directly influences the dose that can be safely administered to the tumor. Dose objectives used for planning come from dose-response models, and the majority of these models are based on the dose to the pCT-shape of the rectum.[10–12] Stronger response prediction has been achieved by accounting for rectal motion,[13–15] but estimates of the average delivered dose to the rectum are resource expensive, demanding both frequent images and complex software, limiting its use.

In this study, we aim to derive a model that enables estimation of a patient's mean rectal shape from the pCT scan only. To solve the obstacle of requiring multiple imaging input to assess the average shape, we apply the statistical method of shrinkage estimation combined with information from a deformable population model of the rectum.

2 MATERIALS AND METHODS

In statistics, shrinkage estimation is a well-known technique used to reduce variance in estimates,[16] not to be confused with physical/anatomical shrinkage. The organ shape used in planning can be considered an estimate of the mean organ shape. Reducing the variance of this estimate means reducing the average difference between the planning shape and the average shape during treatment and therefore corresponds to reducing systematic shape errors. One important aspect of the proposed method is the calculation and use of a mean organ shape across the population – the population mean shape (PMS). Intuitively, the shrinkage method can be understood through the principle of "regression toward the mean": if a single sample taken from an individual is extreme, the next sample taken from the same individual is likely to be closer to the population mean. Our hypothesis is therefore that the true mean shape of an organ can be estimated with lower variance by combining information from the pCT shape (pS) and the PMS.

For computation of the PMS, we rely on deformable registration of organ contours. We used a variant of the thin plate spline - robust point matching (TPS-RPM) algorithm.[17,18] Each shape is represented by a set of points on the organ surface (a mesh representation). The algorithm finds a transform function (a TPS function in this case) which can be used to transform a reference structure into a shape as similar as possible to a target structure. Transforming the reference structure is commonly referred to as «warping». By replacing the target structure by the warped reference structure, we end up with a one-to-one correspondence of points in the two shapes.

To find point-to-point correspondence between the surface points on all rectum shapes in the dataset, we used the method first presented in Budiarto et al.[19] The method is illustrated in Figure 1. First, we performed registration from a reference patient's pCT rectum structure to each of the other patients' pCT rectum structures (inter-patient registration). Since our data set contains multiple scans for each patient, we then performed registration from the rectum structure in each pCT to each of the same patient's other rectum shapes (intra-patient registration). Since there is now a point-to-point correspondence between all shapes, we were able to average the coordinates of the points across all CTs to find a population mean rectum shape.

After the (preliminary) PMS had been built, the procedure was repeated, with the preliminary PMS taking the place of the global reference shape. This was done to increase the accuracy of the intra-patient registrations, and similar to the procedure used in Budiarto et al.[19] In practice, the deformable registration was split in two steps – finding the transform function and applying it (warping). In the intra-patient step, the transform function was found between the original pS (not the warped reference shape) and the treatment CT shapes. To find the final treatment CT shapes, the global reference was warped twice, first from its original shape into the pCT shape of the specific patient,

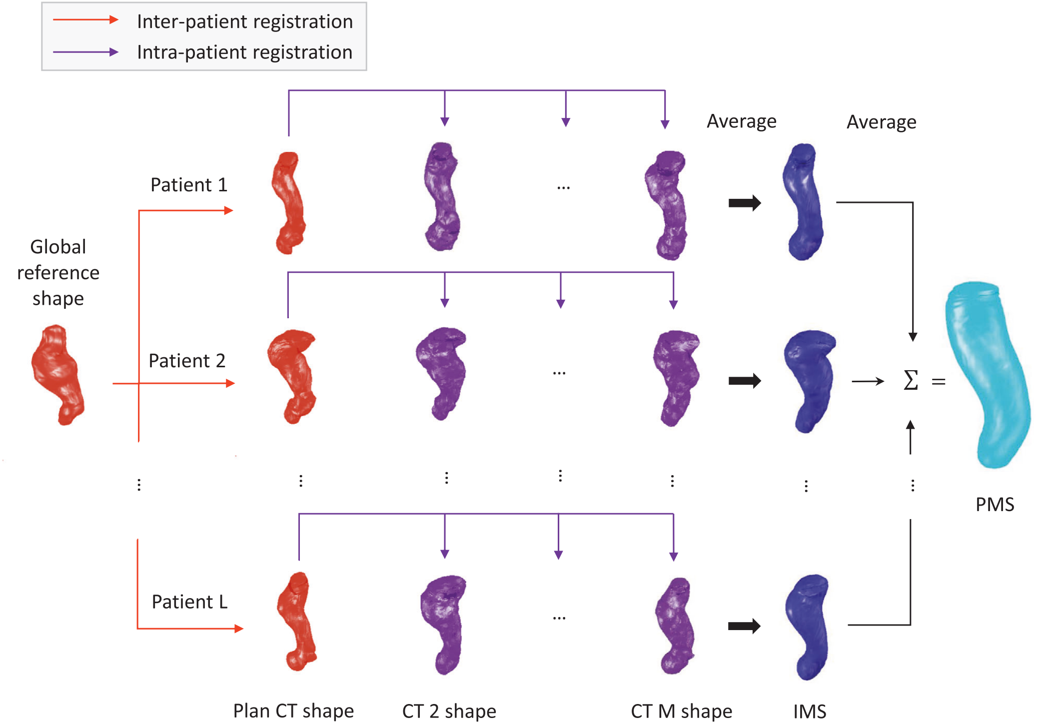

FIGURE 1 Procedure for using inter- and intra-patient registration to compute the individual mean shapes (IMS) and the population mean shape (PMS)

and then from the pCT shape to the treatment CT shape.

For a new patient, deformable registration needs only be performed between the patient's pS and the PMS. Thus, the procedure in Figure 1 is only needed in order to build the PMS and to evaluate the method.

With point-to-point correspondence, we can compare and combine rectal shapes. For each shape, the coordinates of the points were gathered into a vector:

$$[x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N]^T.$$

Let the PMS and the pS be described by such vector representations. The shrinkage estimated shape (SES) of the patient mean shape can then be computed as:

$$\text{SES} = pS + \lambda(\text{PMS} - pS).$$

Here, λ is the shrinkage factor—a value between 0 and 1 that determines weighing between the pS and the PMS. A small λ means little shrinkage, that is, that the SES is close to the pS, while a large λ means that the SES is closer to the PMS.

Matching the coordinate systems of the PMS and the pS was secured by a rigid shift of the PMS before combining with the pS. The shift was performed such that the center of gravity, that is, the coordinate average, matched between the PMS and the PS. We saw that the cranial region, where the variation and uncertainty is greatest, negatively affected the matching; therefore, we left out the 50% most cranial points when calculating the center of gravity.

2.1 | Computing the shrinkage factor

The value of the shrinkage factor λ that minimizes the mean squared error (MSE) over the training set can be found analytically. The treatment CT rectum shapes (but not the pCT shapes) was used to estimate each patient's individual mean shape (IMS), which was used as the target in the optimization of the shrinkage factors. Given the PMS vector, the pS_i vector, and the IMS_i vector for patient i, the MSE is given by

$$\text{MSE}\ (\lambda) = \frac{1}{LN} \sum_{i=1}^{L} \|pS_i + \lambda(PMS - pS_i) - IMS_i\|^2,$$

with a minimum at

$$\lambda_{\mathrm{MMSE}} = \frac{\sum_{i=1}^{L} (IMS_i - pS_i)^T (PMS - pS_i)}{\sum_{i=1}^{L} \|PMS - pS_i\|^2}.$$

2.2 | Validation

The method was evaluated using the rectum shapes of 37 prostate cancer patients, who each had a pCT and seven to 10 treatment CTs acquired during their course of RT – a total of 373 CT scans. The CT resolution was 0.7 mm in-plane and 2 mm in the z-direction. The rectum shapes for all CTs were contoured by expert physicists and quality assured by two independent physicists.

The SESs used for validation were produced using leave-one-out cross validation (CV), where for each patient, the data were separated into a test set containing this patient only, and a training set comprising the remaining 36 patients. The training set was used to compute the PMS and the shrinkage factor for the test patient's SES. The SESs computed this way were used in all validation.

2.2.1 | Volumetric similarity

To reduce the systematic error, the SES must resemble the IMS more than the pS does. We used the Dice similarity index (DSI) to assess the similarity between shapes. The DSI between two shapes X and Y is defined as $\frac{2|X \cap Y|}{|X|+|Y|}$, where $|\cdot|$ indicates volume, and $\cap$ is the intersection operator. The DSI ranges from 0 to 1, with a higher value indicating more overlapping shapes. For each patient, both the pS and the SES were compared to the IMS to evaluate improvement in the novel method.

2.2.2 | Surface similarity

The DSI is related to the proportion of the volume that is shared between two structures, but the actual numbers can be hard to interpret. A more tangible measure might be to compare distances between surfaces, in our case between the pS and the IMS and between the SES and the IMS, with results in mm. We used the bidirectional local distance (BLD) as a distance metric. The BLD, introduced by Kim et al.,[20] is an extension to the Hausdorff distance (HD) to include local, pointwise distances. If the one-directional local distance from a point a in mesh A to a mesh B is defined as

$$\mathrm{OLD}\ (a, B) = \min_{b \in B} \|a - b\|,$$

then the bi-directional local distance from a to B is

$$\mathrm{BLD}\ (a,\ B) = \max\left(OLD\,(a,\ B)\,, \max S_{a,\ B}\right).$$

Where $S_{a,B}$ is the set of all local distances $OLD(b \in B, A)$ where the endpoint in A is a.

We present results for the median, mean, and maximum BLD, where the latter is the same as the HD.

We also used the BLD to study the spatial distribution of the systematic error. We averaged the distance from the pS to the IMS for each point on the organ surface across the population. Changes to the spatial distribution from the shrinkage estimate were analyzed by comparing the pointwise distances between the pS and the IMS to the pointwise distances between the SES and the IMS.

2.2.3 | Evaluation of dosimetric impact

Dosimetric evaluation was based on a retrospective analysis of clinical IMRT plans for locally advanced prostate cancer, including treatment to the prostate, seminal vesicles, and the pelvic lymph nodes. Hypofractionated RT was prescribed in 25 fractions simultaneously delivering fraction doses of 2.7 Gy to the prostate clinical target volume, 2.4 Gy to prostate and seminal vesicles and 2.0 Gy to a larger target also including the pelvic lymph nodes, see Hysing[1] for details. For all patients, the structures pS, IMS, and SES were imported into Varian Eclipse for dosimetric analysis. We do not have available the true accumulated dose over all fractions. As a substitute, we used the dose-volume histogram (DVH) based on the IMS as a representation of the ground truth. This removes systematic errors, but does not take into account the random variations that occur from fraction to fraction.

For dosimetric comparison, the population average DVH and its 95% confidence interval was computed separately for the pS, SES, and IMS. In addition, two parameters were extracted from the DVHs: The equivalent uniform dose (EUD) with a volume factor of 12 and the minimum dose to the hottest 5% volume (D5%). These parameters have been shown to correlate with late rectal toxicity by Söhn et al.[32] and Thor et al,[33] respectively. The average differences between these parameters from the PS to the IMS and from the SES to the IMS were computed.

2.2.4 | Statistical tests

All tests and calculations were performed using Matlab v R2020b with the statistics and machine learning toolbox. All significance levels were set at $\alpha = 0.05$.

The 95% CI of all the geometric similarity and dosimetric metrics was calculated by the bootstrap method

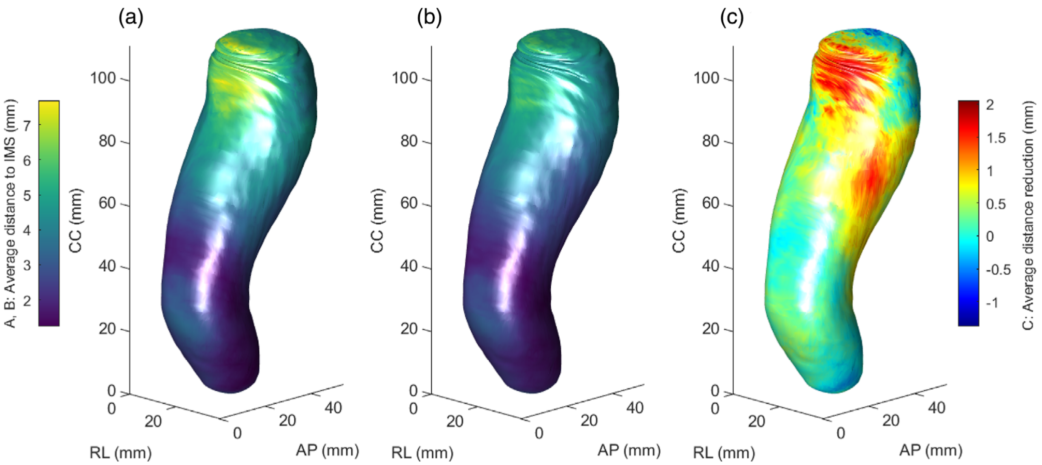

FIGURE 2 Population average systematic error over the rectum (represented by the population mean shape [PMS] rectum), for the pS versus the shrinkage estimated shape (SES). (a) Local distance (bidirectional local distance [BLD]) between the pS and the individual mean shape (IMS). (b) Local distance (BLD) between the SES and the IMS. (c) Difference between (a) and (b), that is, improvement when using SES

TABLE 1 Results for the geometric comparison metrics: Dice similarity index (DSI), median and mean bidirectional local distance (BLD), and Hausdorff distance (HD)

	pS to IMS		**SES to IMS**		**Improvement**			
	μ	σ	μ	σ	μ	*CI*	*p-value*	*% +*
DSI	0.74	0.07	0.78	0.06	0.04 (15%)	0.03–0.06	5.8 e-6	89%
Median BLD (mm)	2.9	1.0	2.4	0.7	0.5 (17%)	0.3–0.7	2.0 e-5	84%
Mean BLD (mm)	3.6	1.1	3.1	1.0	0.5 (13%)	0.3–0.7	2.7 e-5	84%
HD (mm)	15.6	5.8	13.4	4.9	2.2 (14%)	1.2–3.2	1.8 e-4	76%

Note: For DSI, the average improvement in percent was calculated by dividing the absolute improvement (0.04) by the maximum achievable improvement (1–0.74 = 0.26).

Abbreviations: CI, confidence interval; IMS, individual mean shape; SES, shrinkage estimated shape.

The column symbols are μ: mean; σ: standard deviation; % +: percentage of of patients with improvement.

with one million samples. The Wilcoxon signed rank test was used to test for difference in the median values in DSI, since the DSI is bounded between 0 and 1 and therefore not normally distributed. For the other values, the paired *t*-test was used, after testing for evidence against normality with the Anderson-Darling test.

3 | RESULTS

The PMS shape in Figure 2 shows the spatial distribution of systematic errors being heterogeneously, distributed over the rectum. The population average pointwise distance from the pS to the IMS was largest in the cranial end, with values in the range of 5–7 mm (Figure 2a). The average distance decreased steadily toward the caudal part of the rectum to below 2 mm.

The minimum MSE shrinkage factor for the whole 37-patient dataset was 0.37. Under CV, where one patient was held out of the training data for each validation, the shrinkage factors ranged from 0.34 to 0.38. Using these factors to estimate the SES for each patient decreased the average distance between the SES and the IMS across the population for most parts of the rectum (Figure 2c). The population average improvement was greatest, an error reduction of 2 mm, at the cranial-anterior part of the rectum.

The results for the geometric similarity metrics are shown in Table 1. All metrics showed moderate (13%–17%) but statistically significant improvement. The individual results for the DSI and the median BLD are shown in Figure 3. Although the average improvement is moderate, a very high percentage of patients did show improvement – 33/37 had improvement in DSI, 31/37 in median and mean BLD, and 28/37 in HD.

The rectum shapes (pS, SES, and IMS) of three example patients are shown in Figure 4. These patients represent the 10th, 50th, and 90th percentiles in terms of DSI improvement and illustrate the SES model for different

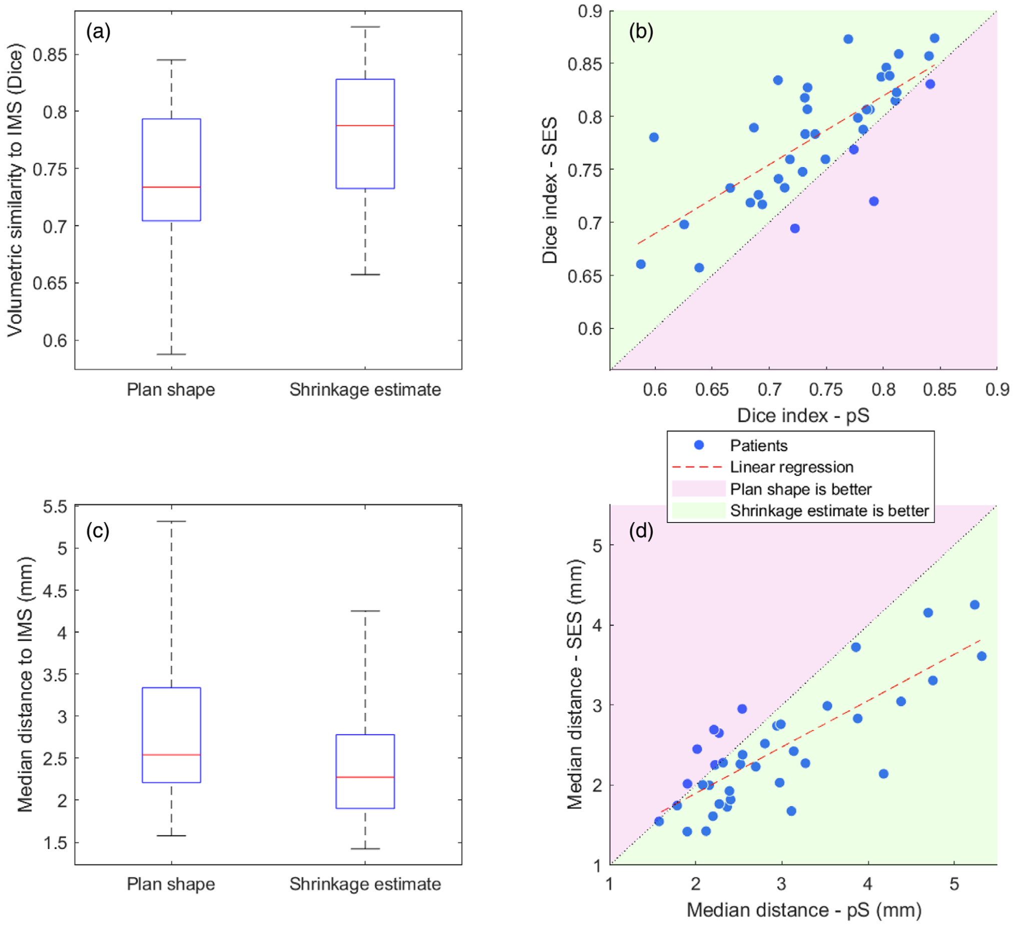

FIGURE 3 Similarity to the individual patient mean for the shrinkage estimated recti and the planning CT recti in each patient. (a and c) Box plots of the similarity to the patient mean shape of the plan shape and the shrinkage estimate as median (red line) with 25th and 75th percentiles and maximum and minimum values (whiskers). (b and c) scatter plots of the similarity to the patient's mean shape for the plan shape versus the shrinkage estimate. In (a) and (b) (Dice similarity), higher is better, while in (c) and (d) (distance), lower is better

geometries. In the 90th percentile patient, the pS shows more bending in the lower (anorectal) flexure than the average, but the shrinkage estimate has overcompensated this feature. In the 50th percentile patient where the mean rectum volume was reduced compared to the pS, the shrinkage estimate reduced the volume further, and thereby achieved better conformity to the IMS. In the 10th percentile patient, the mean rectum was larger, and the shrinkage method compensated for this.

Population average DVHs derived from the pS, SES, and IMS structures are shown in Figure 5. In the low dose region, the pS and SES show similar bias. In the higher dose region, the SES is closer to the IMS. The differences were significant ($p < 0.05$) for doses in the range of 52.5–65.0 Gy. In this range, the bias was, on average, reduced from −1.3 percentage points to −0.2 percentage points.

The results for the dose metrics EUD and D5% are shown in Table 2. For both metrics, the SES gave a better dose estimate on average, but the improvement was not significant. This can be anticipated by investigating Figure 6, which shows the structures with dose for an example patient. The higher doses are restricted to the anterior-middle part of the rectum. The high dose (D5%) region is where the similarity between the SES and PS to the IMS is the greatest.

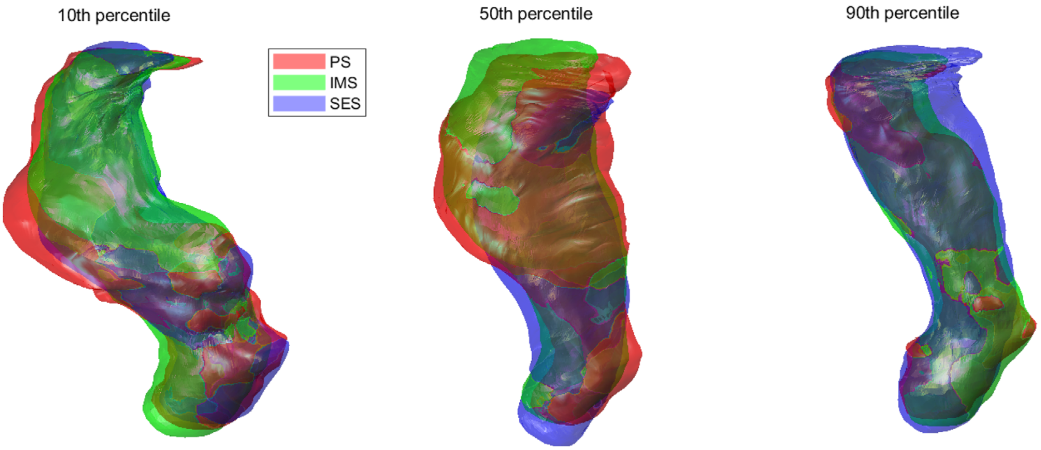

FIGURE 4 Example recti in three different patients with the planning structure (red), the shrinkage estimated rectum (green), and the individual mean shape (blue). The patients were chosen based on their percentiles in terms of improvement in Dice index to the mean rectum, ranging from poorer (left) to better (right)

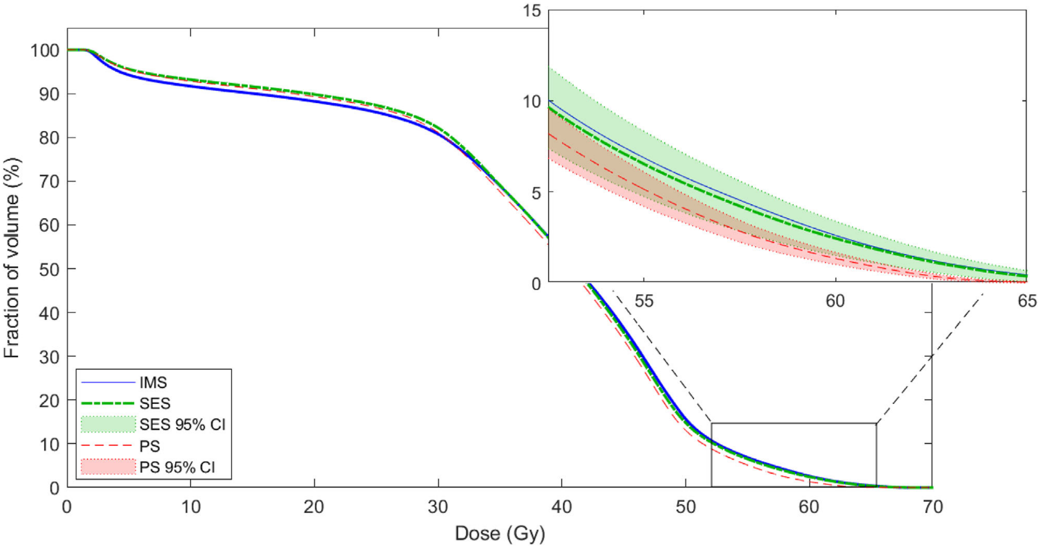

FIGURE 5 Population average DVHs of the three rectum structures individual mean shape (IMS) (blue), planning shape (pS) (red) and shrinkage-estimated shape (SES) (green). The magnification shows the region where the SES showed significantly less bias than the IMS, that is, the dose range from 52.5 to 65.0 Gy

TABLE 2 Comparison of dosimetric parameters between the pS and the SES. The column symbols are μ: mean; σ: standard deviation

	IMS		pS error		SES error		Improvement		
	μ	σ	μ	σ	μ	σ	μ	*CI*	*p-value*
D5% (Gy)	55.8	3.6	−1.4	3.3	−0.6	3.4	0.8	0.0–1.7	0.070
EUD (Gy)	49.0	2.5	−1.1	2.1	−0.6	2.4	0.5	0.0–1.2	0.11

Abbreviations: CI, confidence interval; EUD, equivalent uniform dose; IMS, individual mean shape; SES, shrinkage estimated shape.

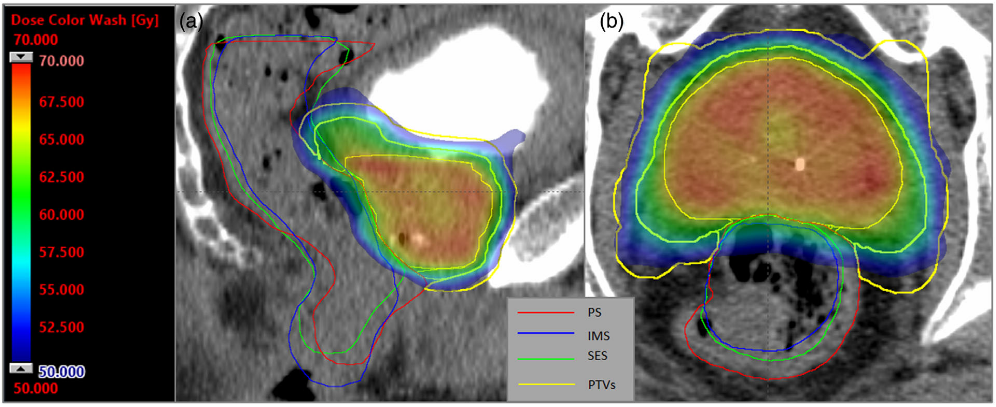

FIGURE 6 2D view of rectum and PTV structures with dose color wash for an example patient. The PTVs, from outermost to innermost, are PTV50, PTV60, and PTV67.5. (a) sagittal view, (b) transversal view. The dose threshold is 50 Gy, identical to the prescription to the PTV50 and to the lowest dose level used to assess rectal exposure in our planning procedure

4 | DISCUSSION

The proposed method showed significantly improved geometric similarity to the patient mean rectum. The average improvement is about 15% in terms of both distance and volume. However, this is just the average; several patients had much greater improvement (Figure 3). A positive trait is that the patients that have the greatest systematic error in the pCT are those that typically see the greatest improvement. Also, the method is low-risk, as few patients see worse systematic errors.

Whether this improvement is important will depend on the application, the dose distribution, and other factors. In our dataset, the improvements in geometric similarity did not translate into significant improvement of estimated accumulated dose to the rectum. This may be due to the procedures of image-guidance that we used. All treatment CTs were rigidly aligned to gold markers in the prostate, following our clinical procedures of RT for these patients. This means that all movements are relative to the prostate, and there are no fixed points in the rectum. The volume of the rectum near the prostate is therefore rather stable, which contributes to the relatively low improvement in the anterior region where the highest doses are located (Figure 2c). For the example patient seen in Figure 6, the SES is closer to the IMS almost everywhere except in the high-dose region, where the SES, the pS, and the IMS are all very similar. Because the patients were part of a trial in hypofractionation, extra care was taken to avoid rectal toxicity by cropping the planning target volume (PTV) where it would otherwise overlap with the rectum, also seen in Figure 6. Clinical results have proven this fear unfounded, and the practice has since been changed into cropping the rectum in such cases.[1]

Systematic shape errors due to rectal deformation have previously been studied by Hoogeman et al.[21] and Haekal et al.[22] There, the treatment CTs were not registered to gold markers, and, contrary to our results, both studies found the largest systematic errors in the anterior region. As such, the current results are a testament to the success of fiducial markers. If the higher doses were delivered to volumes where the systematic shape errors are greater, we would likely have seen higher impact from the method also in the dosimetric analysis. We can therefore expect to see different results if the method is applied to different patient groups, for example, bladder or cervical cancer, or different OARs.

The amount of dose degradation in EUD and D5% that is acceptable is not known for the rectum. This is likely dependent on several factors additional to organ motion, for example, dose distribution and delivery modality, nonuniform radio-sensitivity of the rectum.[13,23,24] In our clinical procedure, the patients in the present study were planned for a maximum of 60 Gy (about 62GyEDQ2) delivered to 10 cc of the rectal volume and with restrictions to the volume receiving 50 Gy.[1] This resulted in very little rectal toxicity, indirectly indicating that the obtained changes to the EUD from pCT in these patients were acceptable.[25]

Since dose accumulation across fractions is difficult, we used the dose to the IMS as a substitute for the true dose when comparing dose-volume parameters. Under the assumption that the rectum is in a dose region with a constant gradient, the dose to the IMS will be the same as the expected accumulated dose to the deforming organ. This is because each voxel moves both into higher and lower dose regions, which, in the mean, cancel each other out.[a] Although this assumption is false,

---

[a] Formally, if x is the (random) position of a voxel, with distribution $p(x)$, and $D(x) = Gx$ is the dose at position x, with constant gradient G, then $E[D(x)] = \int_{-\infty}^{\infty} D(x)p(x)dx = G\int_{-\infty}^{\infty} xp(x)dx = GE[x] = D(E[x])$.

the smoothness of the gradient outside the PTV justifies the use of the dose to the IMS as an approximation of the true accumulated dose.

We have chosen to present both volumetric (DSI) and distance-based (DS and BLD) similarity metrics for the evaluation of our proposed method. Distance-based metrics are easier to interpret and are more relevant for radiotherapy (RT), for example, to assess the influence of deformation in relation to a dose gradient. The DSI, on the other hand, includes the whole structure as opposed to a single point and is a common metric to evaluate the performance of deform registration.

The deformable registration will introduce some geometric distortion. The distortion is greater for inter-patient registration than the intra-patient, since there is no “true” transformation between patients. However, the improvement in DSI is itself a proof-of-concept of the inter-patient registration. Our registrations were optimized to yield high forward accuracy and backward consistency. The average forward accuracy (see Osorio et al.[18] for definition) was measured at 0.23 mm for the intra-patient registrations and 0.22 mm for the second iteration of inter-patient registrations (0.56 mm in the first iteration). In addition, the obtained vector fields were checked visually. Our results thus reflect a rather meticulous verification of the registrations, which may be more difficult to achieve in commercial registration software.

The choice to use the 50% most caudal points in the rigid alignment between the pS and the PMS was relatively arbitrary; we saw the need to exclude the cranial part of the rectum and at the same time include the points near the prostate, but apart from that, the exact percentage of points chosen did not make a great difference. The method can potentially be improved by optimizing the rigid shift w.r.t geometry or dose using training data.

Hoogeman et al.[26] have previously investigated the possibility of predicting systematic changes to the rectum and prostate based on the pCT rectal volume alone. For the rectum, they achieved best results in the upper anterior part, with a 30% improvement in one coordinate (AP). Their results are, however, not directly comparable to ours, as their images were aligned to the bony anatomy instead of gold markers.

In this paper, we have looked at the possibility of measuring dose on the SES to estimate the accumulated dose to the rectum. In clinical applications, discrepancies between the SES and the pS — either geometrically overall, locally at key regions or in calculated dose — could be used for screening of patients for further imaging, or for additional verification based on in-treatment CBCTs, which may again indicate adaptive RT.

Other applications include motion as well as normal-tissue complication probability modeling. Many deformable organ motion models do not account for systematic error or require multiple CTs to do so.[19,27–31] Such models can be improved or simplified by replacing the pS or the multiple-CT average by the SES. For voxel-based normal tissue complication probability (NTCP) models, the method can be used to reduce the bias in the pCT.

It is also possible to use the method in plan optimization: A robust optimization algorithm may for example consider the pS and the SES as separate scenarios. One complicating factor is the need to take into account the correlation between target motion and the motion of OARs; for example, the SES rectum may overlap with the target. However, the method does not need to be restricted to a single organ. A potential solution is therefore to include multiple organs (in this case, at least the rectum and the prostate) in the shrinkage method. Still, inclusion of the target is not straightforward, as tumor shrinkage must be taken into account. The accuracy of the predicted anatomy should be verified through daily CBCT scans.

5 | CONCLUSIONS

The shrinkage method was successfully applied to improve estimates of the individual mean rectum shapes of prostate cancer patients when only the planning CT is available. The method reduced the systematic shape changes, especially for patients where the similarity between the planning CT and the individual mean shape was poor. The method could be applied to more complex motion modeling, plan optimization as well as toxicity assessments.

ACKNOWLEDGMENTS

We gratefully thank Prof. Heijmen and Prof. Hoogeman and PhD Zolnay at Erasmus MC Cancer Institute, Rotterdam for valuable discussions and advice on motion modeling as well as sharing the deformable registration software. We also thank Maria Thor and Lise Fjellsbø for rectum contouring. Thanks to Helge Egil Seime Pettersen for invaluable help in importing structures to Eclipse. The work was funded by Trond Mohn Foundation (grant number: BFS2017TMT07).

CONFLICT OF INTEREST

The authors have no relevant conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data representing the population mean rectum shape are available in the supplementary material of this article. The raw data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

REFERENCES

1. Hysing LB, Ekanger C, Zolnay Á, et al. Statistical motion modelling for robust evaluation of clinically delivered accumulated dose distributions after curative radiotherapy of locally advanced prostate cancer. *Radiother Oncol*. 2018;128(2):327-335.

2. Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning. *Phys Med Biol*. 2018;63(22):22TR02.
3. Muren LP, Ekerold R, Kvinnsland Y, Karlsdottir À, Dahl O. On the use of margins for geometrical uncertainties around the rectum in radiotherapy planning. *Radiother Oncol*. 2004;70(1):11-19.
4. de Boer HC, Heijmen BJ. A protocol for the reduction of systematic patient setup errors with minimal portal imaging workload. *Int J Radiat Oncol Biol Phys*. 2001;50(5):1350-1365.
5. Söhn M, Birkner M, Yan D, Alber M. Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. *Phys Med Biol*. 2005;50(24):5893-5908.
6. Scaife JE, Thomas SJ, Harrison K, et al. Accumulated dose to the rectum, measured using dose-volume histograms and dose-surface maps, is different from planned dose in all patients treated with radiotherapy for prostate cancer. *Br J Radiol*. 2015;88(1054):20150243.
7. Thörnqvist S, Petersen JBB, Høyer M, Bentzen LN, Muren LP. Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration. *Acta Oncol*. 2010;49(7):1023-1032.
8. Kleijnen J-PJE, van Asselen B, Burbach JPM, et al. Evolution of motion uncertainty in rectal cancer: implications for adaptive radiotherapy. *Phys Med Biol*. 2015;61(1):1.
9. Liu L, Johansson A, Cao Y, Kashani R, Lawrence TS, Balter JM. Modeling intra-fractional abdominal configuration changes using breathing motion-corrected radial MRI. *Phys Med Biol*. 2021;66(8):085002.
10. Jackson A, Marks LB, Bentzen SM, et al. The lessons of QUANTEC: recommendations for reporting and gathering data on dose-volume dependencies of treatment outcome. *Int J Radiat Oncol Biol Phys*. 2010;76:S155-S160.
11. Michalski JM, Gay H, Jackson A, Tucker SL, Deasy JO. Radiation dose-volume effects in radiation-induced rectal injury. *Int J Radiat Oncol Biol Phys*. 2010;76:S123-S129.
12. Söhn M, Alber M, Yan D. Principal component analysis-based pattern analysis of dose-volume histograms and influence on rectal toxicity. *Int J Radiat Oncol Biol Phys*. 2007;69(1):230-239.
13. Casares-Magaz O, Bülow S, Pettersson NJ, et al. High accumulated doses to the inferior rectum are associated with late gastro-intestinal toxicity in a case-control study of prostate cancer patients treated with radiotherapy. *Acta Oncol*. 2019;58(10):1543-1546.
14. Shelley LEA, Scaife JE, Romanchikova M, et al. Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy. *Radiother Oncol*. 2017;123(3):466-471.
15. Thor M, Bentzen L, Hysing LB, et al. Prediction of rectum and bladder morbidity following radiotherapy of prostate cancer based on motion-inclusive dose distributions. *Radiother Oncol*. 2013;107(2):147-152.
16. Fourdrinier D, Strawderman WE, Wells MT. Decision theory preliminaries. In: Fourdrinier D, Strawderman WE, Wells MT, eds. *Shrinkage Estimation*. Springer International Publishing; 2018:1-28.
17. Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. *Comput Vision Image Understanding*. 2003;89(2):114-141.
18. Osorio EMV, Hoogeman MS, Bondar L, Levendag PC, Heijmen BJM. A novel flexible framework with automatic feature correspondence optimization for nonrigid registration in radiotherapy. *Medical Physics*. 2009;36(7):2848-2859.
19. Budiarto E, Keijzer M, Storchi PR, et al. A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases. *Phys Med Biol*. 2011;56(4):1045-1061.
20. Kim HS, Park SB, Lo SS, Monroe JI, Sohn JW. Bidirectional local distance measure for comparing segmentations. *Med Phys*. 2012;39(11):6779-6790.
21. Hoogeman MS, van Herk M, de Bois J, Muller-Timmermans P, Koper PCM, Lebesque JV. Quantification of local rectal wall displacements by virtual rectum unfolding. *Radiother Oncol*. 2004;70(1):21-30.
22. Haekal M, Arimura H, Hirose T, et al. Computational analysis of interfractional anisotropic shape variations of the rectum in prostate cancer radiation therapy. *Physica Medica*. 2018;46:168-179.
23. Troeller A, Yan D, Marina O, et al. Comparison and limitations of DVH-based NTCP models derived from 3D-CRT and IMRT data for prediction of gastrointestinal toxicities in prostate cancer patients by using propensity score matched pair analysis. *Int J Radiat Oncol Biol Phys*. 2015;91(2):435-443.
24. Pedersen J, Flampouri S, Bryant C, et al. Cross-modality applicability of rectal normal tissue complication probability models from photon- to proton-based radiotherapy. *Radiother Oncol*. 2020;142:253-260.
25. Ekanger C, Helle SI, Heinrich D, et al. Ten-year results from a phase II study on image guided, intensity modulated radiation therapy with simultaneous integrated boost in high-risk prostate cancer. *Adv Radiat Oncol*. 2020;5(3):396-403.
26. Hoogeman MS, van Herk M, de Bois J, Lebesque JV. Strategies to reduce the systematic error due to tumor and rectum motion in radiotherapy of prostate cancer. *Radiother Oncol*. 2005;74(2):177-185.
27. Bondar L, Intven M, Burbach JPM, et al. Statistical modeling of CTV motion and deformation for IMRT of early-stage rectal cancer. *Int J Radiation Oncol Biol Phys*. 2014;90(3):664-672.
28. Fontenla E, Pelizzari CA, Roeske JC, Chen GTY. Using serial imaging data to model variabilities in organ position and shape during radiotherapy. *Phys Med Biol*. 2001;46(9):2317-2336.
29. Hoogeman MS, van Herk M, Yan D, Boersma LJ, Koper PCM, Lebesque JV. A model to simulate day-to-day variations in rectum shape. *Int J Radiat Oncol Biol Phys*. 2002;54(2):615-625.
30. Magallon-Baro A, Loi M, Milder MTW, et al. Modeling daily changes in organ-at-risk anatomy in a cohort of pancreatic cancer patients. *Radiother Oncol*. 2019;134:127-134.
31. Rios R, De Crevoisier R, Ospina JD, et al. Population model of bladder motion and deformation based on dominant eigenmodes and mixed-effects models in prostate cancer radiotherapy. *Medical Image Analysis*. 2017;38:133-149.
32. Söhn M, Yan D, Liang J, Meldolesi E, Vargas C, Alber M. Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose–based and dose–volume–based normal tissue complication probability models. *International Journal of Radiation Oncology*Biology*Physics*. 2007;67:(4):1066–1073. http://doi.org/10.1016/j.ijrobp.2006.10.014
33. Thor M, Deasy JO, Paulus R, et al. Tolerance doses for late adverse events after hypofractionated radiotherapy for prostate cancer on trial NRG Oncology/RTOG 0415. *Radiotherapy and Oncology*. 2019;135:19–24. http://doi.org/10.1016/j.radonc.2019.02.014

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Rørtveit ØL, Hysing LB, Stordal AS, Pilskog S. Reducing systematic errors due to deformation of organs at risk in radiotherapy. *Med Phys.* 2021;48:6578–6587. https://doi.org/10.1002/mp.15262

Paper II

An organ deformation model using Bayesian inference to combine population and patient-specific data

Øyvind Lunde Rørtveit, Liv Bolstad Hysing, Andreas Størksen Stordal, Sara Pilskog

Physics in Medicine and Biology, **68** 055009 (2023)
DOI: 10.1088/1361-6560/acb8fc
Open access

Physics in Medicine & Biology

An organ deformation model using Bayesian inference to combine population and patient-specific data

To cite this article: Øyvind Lunde Rørtveit *et al* 2023 *Phys. Med. Biol.* **68** 055009

IOP Publishing Phys. Med. Biol. 68 (2023) 055009 https://doi.org/10.1088/1361-6560/acb8fc

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

CrossMark

OPEN ACCESS

RECEIVED
25 October 2022

REVISED
20 December 2022

ACCEPTED FOR PUBLICATION
3 February 2023

PUBLISHED
21 February 2023

PAPER

An organ deformation model using Bayesian inference to combine population and patient-specific data

Øyvind Lunde Rørtveit[1,2,*], **Liv Bolstad Hysing**[1,2], **Andreas Størksen Stordal**[3,4] **and Sara Pilskog**[1,2]

[1] Department of Oncology and Medical Physics, Haukeland University Hospital, Bergen, Norway
[2] Department of Technology and Physics, University of Bergen, Norway
[3] NORCE Norwegian Research Centre, Bergen, Norway
[4] Department of Mathematics, University of Bergen, Norway
[*] Author to whom any correspondence should be addressed.

E-mail: oyvind.rortveit@uib.no

Keywords: organ motion, Bayesian modelling, deformable registration, radiotherapy, personalized therapy

Supplementary material for this article is available online

Abstract

Objective. Organ deformation models have the potential to improve delivery and reduce toxicity of radiotherapy, but existing data-driven motion models are based on either patient-specific or population data. We propose to combine population and patient-specific data using a Bayesian framework. Our goal is to accurately predict individual motion patterns while using fewer scans than previous models. *Approach.* We have derived and evaluated two Bayesian deformation models. The models were applied retrospectively to the rectal wall from a cohort of prostate cancer patients. These patients had repeat CT scans evenly acquired throughout radiotherapy. Each model was used to create coverage probability matrices (CPMs). The spatial correlations between these estimated CPMs and the ground truth, derived from independent scans of the same patient, were calculated. *Main results.* Spatial correlation with ground truth were significantly higher for the Bayesian deformation models than both patient-specific and population-derived models with 1, 2 or 3 patient-specific scans as input. Statistical motion simulations indicate that this result will also hold for more than 3 scans. *Significance.* The improvement over previous models means that fewer scans per patient are needed to achieve accurate deformation predictions. The models have applications in robust radiotherapy planning and evaluation, among others.

1. Introduction

In radiotherapy (RT), the dose is carefully shaped to the patient anatomy as seen in the CT acquired before start of treatment (plan CT), to achieve a good compromise between disease control and risk of inducing complications. Since the variability of the organ positions and deformations is unknown before start of treatment, different measures have been adopted to safeguard against motion uncertainties through planning margins (Stroom *et al* 1999, van Herk *et al* 2000), robust optimization (Unkelbach *et al* 2018) and/or treatment plan adaptation (Yan *et al* 1997).

A statistical model for the deformation of organs of individual patients using principal component analysis (PCA) of the organ's surface shape vectors was first proposed by Söhn *et al* (2005). The main drawback of the patient-specific model is that the number of data samples (in the form of organ contours derived from 3D images) per patient is often low, which limits the robustness of the motion estimates (Thörnqvist *et al* 2013b). Budiarto *et al* (2011) proposed a population based statistical model, under the assumption that, although the size, shape and position of organs differ greatly between patients, the patterns of deformation are generally the same. The advantage is that an estimate of a patient's deformation patterns exists even when only a single observation is available. When applied to prostate target deformation, they showed that about 50% of the variation could be explained by 15 population deformation modes (i.e. principal components). Subsequent uses

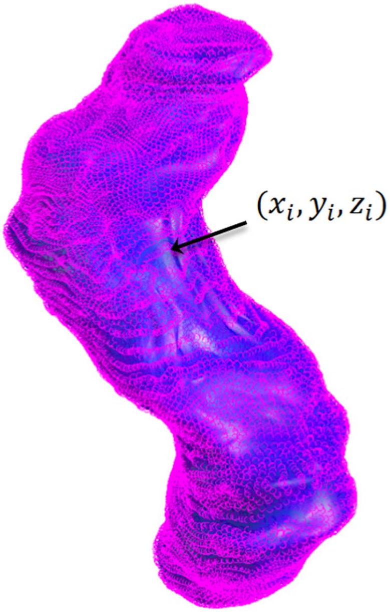

Figure 1. A rectum shape represented by a set of organ surface points.

of the population model include Bondar *et al* (2014), who used it to create margins for rectal cancer patients, Rios *et al* (2017), who modeled bladder deformation for prostate cancer RT, Szeto *et al* (2017) who modeled daily variations in the thorax, and Magallon-Baro *et al* (2019), who modeled deformation in the stomach, duodenum and bowel for pancreatic cancer RT. A weakness of the population model is its inability to model patient-specific deformation patterns, even when multiple scans are available for the patient in question. The aim of the current work is to combine the strengths of the population and patient-specific models by introducing Bayesian models that take in to account both the population deformation patterns (in terms of a *prior* distribution) and patient-specific measurements, forming an individualized *posterior* distribution. Bayesian models have previously been applied to the problem rigid shifts of the patient, termed setup errors (Lam *et al* 2005, Herschtal *et al* 2012).

In this paper, we introduce two Bayesian models, which differ in their choice of priors. The choice of model to use will be a trade-off between accuracy and simplicity. We derive necessary algorithms to efficiently calculate the approximate posterior distributions in high dimensions. We apply the introduced models to a realistic example with complex motion, in terms of the rectal wall of prostate cancer patients. We use the models to estimate coverage probability matrices (CPMs), i.e. 3D-arrays of voxels where the value in each voxel is the probability that the voxel will be covered by the rectal wall at any given time. We compare the accuracy of CPMs estimated using the two Bayesian methods, the patient-specific model by Söhn *et al* (2005) and the population model by Budiarto *et al* (2011). In addition to the presentation of new models, this is to our knowledge the first comparison between these two previous models, as well as the first time such an organ deformation model has been applied to the rectum.

2. Methods

In the class of deformation models that we study, an organ shape is represented by a set of points on the organ surface, as illustrated in figure 1. These representations are derived from organ contours segmented from 3D images. The x, y and z coordinates of all P points are gathered into a *shape vector* s:

$$s = [x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_P, y_P, z_P]^T. \tag{1}$$

With this representation, we can use standard multivariate statistical distributions.

To compare organs across scans, we need corresponding points between all shapes in the data set. This correspondence is found using deformable and rigid contour registration both within and between patients. Details are beyond the scope of the current work, but can be found in Rørtveit *et al* (2021).

Due to the random character of the organ shape, a set of shape vectors $s_1, \ldots, s_J$ derived from J scans of a patient is described as J realizations of the random variable s. For all the following methods, the shape coordinates for a specific patient are assumed to follow a multivariate Gaussian distribution:

$$s \sim \mathcal{N}(\mu, R). \tag{2}$$

The mean shape vector μ represents the patient's mean organ shape, and the covariance matrix R describes the variance of the coordinates as well as the covariance between each pair of coordinates. When μ and R are given, we can use the distribution to draw new random organ shapes for the patient. The difference between the previous patient-specific and population models and the Bayesian models introduced in section 2.3 is how μ and R are estimated. In the Bayesian methods, μ and R are considered random samples from specific prior distributions, whose parameters are calculated from the training data. *Point estimates* of μ and R are derived from the posterior distributions. Due to the high dimensions of the shape vectors, all covariance matrices are parametrized using principal component analysis (PCA), see e.g. Fujikoshi *et al* (2010, chapter 10). Under PCA, a covariance matrix is represented by a few eigenvectors and corresponding eigenvalues. These are usually found through singular value decomposition (SVD) of a *data matrix* D, whose columns are normalized mean-subtracted samples, such that $R = DD^T$.

In the following sections, we show how μ and R are estimated in the previous and the new models.

2.1. Patient-specific model

In the patient-specific model introduced by Söhn *et al* (2005), only data from the patient under consideration is used. The mean shape μ is thus set to the average of the J available shapes $s_1, s_2, \ldots, s_J$ for that patient;

$$\mu = \bar{s} = \frac{1}{J}\sum_{j=1}^{J} s_j, \tag{3}$$

while R is set to the patient-specific sample covariance matrix $\hat{R}_{\mathrm{ps}}$:

$$\hat{R}_{\mathrm{ps}} = \frac{1}{J-1}\sum_{j=1}^{J}(s_j - \hat{\mu})(s_j - \hat{\mu})^T. \tag{4}$$

2.2. Population model

The population model introduced by Budiarto *et al* (2011) rests on the assumption that the covariance matrix is the same for all patients, and only the mean differs. The mean is calculated as the mean shape vector for the individual patient as in (3). The covariance matrix is the average of the sample covariance matrices $\hat{R}_i$ for each patient i in the training set. Given M patients, where patient i has J_i shapes denoted $s_{i,1} \ldots s_{i,J_i}$, the estimated population covariance matrix is

$$\hat{R}_{\mathrm{pop}} = \frac{1}{M}\sum_{i=1}^{M}\hat{R}_i = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{J_i - 1}\sum_{j=1}^{J_i}(s_{i,j} - \bar{s}_i)(s_{i,j} - \bar{s}_i)^T. \tag{5}$$

2.3. Bayesian models

In Bayesian inference, new data is combined with prior knowledge (such as population statistics) in the form of a *prior distribution*, which describes how we would expect a quantity to behave before any specific evidence is taken into account. The result of the combination of the prior and data is a *posterior distribution*.

In the following, the mean and covariance matrix for a given patient are considered random parameters that vary across the population according to a prior distribution defined by the probability density function (pdf) $f(\mu, R)$. When data for a new patient is available, we can compute the posterior pdf of μ and R given $\boldsymbol{s}$, where $\boldsymbol{s} = \{s_1, s_2, \ldots, s_J\}$, denoted $f(\mu, R|\boldsymbol{s})$, through Bayes theorem:

$$f(\mu, R|\boldsymbol{s}) = \frac{f(\boldsymbol{s}|\mu, R)f(\mu, R)}{f(\boldsymbol{s})}. \tag{6}$$

Bayes theorem gives us a distribution of the possible values of μ and R, as opposed to single values. Nevertheless, due to the complexity of the posterior distributions in our subject matter, we shall resort to looking at point estimates of μ and R, such as the expected value or mode of the posterior.

The Bayesian models we present differ in the selection of the prior distribution. We resort to priors that result in computationally feasible posterior distributions, since Markov Chain-Monte Carlo methods are computationally expensive in high dimensions. In the following sections, we present two priors which each represent a Bayesian model.

2.3.1. Normal-inverse-wishart prior

2.3.1.1. Background

We present a short background to aid the intuitive understanding of the normal-inverse-wishart (NIW) distribution. More details can be found in e.g. Bishop (2006).

A combined population and patient-specific covariance matrix $\hat{R}$ can be calculated by a simple weighted average,

$$\hat{R} = \lambda\hat{R}_{\text{pop}} + (1 - \lambda)\hat{R}_{\text{ps}}, \tag{7}$$

for some weight λ between 0 and 1. The weight should be proportional to the number J of scans used to compute the estimates. By setting $\lambda = \frac{\nu}{\nu + J}$ for some parameter ν, we obtain

$$\hat{R} = \frac{1}{\nu + J}(\nu\hat{R}_{\text{pop}} + J\hat{R}_{\text{ps}}). \tag{8}$$

We can achieve the same result by assuming an inverse Wishart (IW) prior for R and using a specific point estimate for the posterior, as shown below.

IW is a matrix distribution, and a *conjugate prior* to the multivariate Gaussian likelihood with known mean and unknown covariance matrix. This means that the posterior distribution for R is also IW, and the parameters are obtained from equations involving the prior parameters and the data. The parameters of the IW are the *scale matrix* Ψ and the *degrees of freedom* ν. Formally, if μ is given, and the prior for R is IW,

$$R \sim \mathcal{IW}(\Psi, \nu), \tag{9}$$

and the likelihood is Gaussian,

$$s|R \sim \mathcal{N}(\mu, R), \tag{10}$$

then the posterior $R|\boldsymbol{s}$, where $\boldsymbol{s} = \{s_1, s_2, \ldots, s_J\}$ is also IW,

$$R|\boldsymbol{s} \sim \mathcal{IW}(\Psi', \nu'), \tag{11}$$

with posterior parameters

$$\Psi' = \Psi + \sum_{j=1}^{J}(s_j - \mu)(s_j - \mu)^T \tag{12}$$

$$\nu' = \nu + J. \tag{13}$$

In order to obtain (8) as a point estimate for R, we define $\Psi = \nu\hat{R}_{\text{pop}}$ and set the posterior point estimate to $\hat{R} = \frac{1}{\nu'}\Psi'$. Inserting both these expressions into (12), we get

$$\hat{R} = \frac{1}{\nu + J}\left(\nu\hat{R}_{\text{pop}} + \sum_{j=1}^{J}(s_j - \mu)(s_j - \mu)^T\right). \tag{14}$$

The parameter ν determines the weight between the population covariance matrix and the sample covariance matrix of the new patient, and can be selected either by tuning or by optimization. One can think of ν as encoding the strength of our belief that $\hat{R}_{\text{pop}}$ can represent our new patient's covariance matrix.

In reality, μ is not given. One could replace μ by $\hat{\mu}$ from (3), but this will lead to bias in the covariance matrix estimate when J is small (to see this, consider equation (14) when $J = 1$ and therefore $\hat{\mu} = s_1$). Instead, we consider both μ and R random, and look for a joint prior distribution.

2.3.1.2. Normal-Inverse-Wishart distribution

The conjugate prior for the multivariate Gaussian likelihood with both unknown mean and covariance is the Normal-Inverse-Wishart (NIW) distribution. In the NIW, R is IW-distributed as in (9) , but μ and R are not independent. The conditional distribution of μ given R is Gaussian:

$$\mu|R \sim \mathcal{N}(\mu_0, \frac{1}{\kappa}R). \tag{15}$$

Here, μ_0 is the population mean, and the scalar κ represents the ratio of the variance between scans of the same patient (intra-patient) to the variance between patients (inter-patient). Thus, the NIW has the parameters μ_0, κ, Ψ and ν, and we write

$$\mu, R \sim \mathcal{NIW}(\mu_0, \kappa, \Psi, \nu). \tag{16}$$

Since this is a conjugate prior, the posterior is also NIW, and we can write

$$\mu, R|\boldsymbol{s} \sim \mathcal{NIW}(\mu_0', \kappa', \Psi', \nu') \tag{17}$$

with

$$\mu_0' = \frac{1}{\kappa + J}(\kappa\mu_0 + J\bar{s}) \tag{18}$$

$$\kappa' = \kappa + J \tag{19}$$

$$\nu' = \nu + J \tag{20}$$

$$\Psi' = \Psi + \sum_{j=1}^{J}(s_j - \bar{s})(s_j - \bar{s})^T + \frac{\kappa J}{\kappa + J}(\bar{s} - \mu_0)(\bar{s} - \mu_0)^T. \tag{21}$$

Note the similarity between (18) and (14): Both are weighted averages between population and patient-specific estimates, with the weight of the patient-specific estimate proportional to the number of patient-specific samples J. Hence, both ν and κ are parameters which determine the weight between the population and patient-specific estimates.

The final term of (21) can be considered a correction for the uncertainty of the sample mean, which makes the equation different from (12), where the mean was assumed to be known.

The maximum a-posteriori (MAP) estimate of μ is the expected value of the posterior, μ_0', so we let

$$\hat{\mu} = \frac{1}{\kappa + J}(\kappa\mu_0 + J\bar{s}). \tag{22}$$

When only a single observation for the new patient is available, i. e. $J = 1$, (22) becomes identical to the shrinkage estimation from Rørtveit *et al* (2021).

As for the IW-case, we let $\Psi = \nu\hat{R}_{\text{pop}}$ and $\hat{R} = \frac{1}{\nu'}\Psi'$. Inserting this into (21) yields

$$\hat{R} = \frac{1}{\nu + J}\left(\nu\hat{R}_{\text{pop}} + \sum_{j=1}^{J}(s_j - \bar{s})(s_j - \bar{s})^T + \frac{\kappa J}{\kappa + J}(\bar{s} - \mu_0)(\bar{s} - \mu_0)^T\right). \tag{23}$$

In practice, we never construct the full covariance matrix $\hat{R}$. Instead, it is represented by a data matrix which is augmented with extra columns, such that $D'D'^T = \hat{R}$. Given the population data matrix D, where $DD^T = \hat{R}_{\text{pop}}$, and the patient-specific data matrix S whose columns are $s_j - \bar{s}$ for $j = 1 \ldots J$, the augmented data matrix is

$$D' = \frac{1}{\sqrt{\nu + J}}\left[\sqrt{\nu}D \quad \sqrt{\frac{kJ}{k + J}}(\bar{s} - \mu_0) \quad S\right]. \tag{24}$$

2.3.2. Variational bayes model

The covariance matrix of μ describes how the individual mean varies from patient to patient, and we shall refer to it as the *inter-patient covariance matrix*. In the NIW-model, this matrix is $\frac{1}{\kappa}R$, according to (15). But the assumption that the intra-patient covariance R is proportional to the inter-patient covariance may in practice not be fulfilled. A more flexible approach is to separate the two, which motivates the following model.

Assume that the mean μ is Gaussian distributed according to

$$\mu \sim \mathcal{N}(\mu_0, \Lambda). \tag{25}$$

Here, μ_0 is the population mean, and Λ is the inter-patient covariance matrix. Assume further that R is IW distributed according to (15), and μ and R are independent (unlike in the NIW model); i.e.

$$f(\mu, R) = \mathcal{N}(\mu; \mu_0, \Lambda) \cdot \mathcal{IW}(R; \Psi, \nu). \tag{26}$$

Unfortunately, this prior is not conjugate to the Gaussian likelihood (2), and there is no simple expression for the posterior. However, both μ and R follow tractable posterior distributions *when conditioned on the other*, namely

$$\mu | R, \boldsymbol{s} = \mathcal{N}(\mu_0', \Lambda') \tag{27}$$

and

$$R | \mu, \boldsymbol{s} = \mathcal{IW}(\Psi', \nu'). \tag{28}$$

Prior distributions with this property are said to be *conditionally conjugate* to the likelihood. The conditional posterior parameters μ', Λ', Ψ' and ν' are

$$\mu_0' = (\Lambda^{-1} + JR^{-1})^{-1}(\Lambda^{-1}\mu_0 + JR^{-1}\bar{s}) \tag{29}$$

$$\Lambda' = (\Lambda^{-1} + JR^{-1})^{-1} \tag{30}$$

$$\Psi' = \Psi + \sum_{j=1}^{J}(s_j - \mu)(s_j - \mu)^T \tag{31}$$

$$\nu' = \nu + J. \tag{32}$$

The derivation of (27)–(32) is given in appendix A.

Since both μ and R are unknown, the left hand sides of (29)–(31) cannot be computed directly from the right hand sides. An alternative is to use an approximative method, known as Mean Field Variational Bayes (MFVB) (Gelman *et al* 1995). This method is applicable for conditionally conjugate priors, and is a technique used to approximate a complicated posterior distribution by a simpler distribution. The joint posterior distribution of the dependent parameters are approximated by two marginal posterior distributions by assuming independence. In our case, we are looking for densities $q_\mu()$ and $q_R()$ such that

$$q_\mu(\mu)q_R(R) \approx f(\mu, R|S). \tag{33}$$

In appendix B, we show that $q_\mu()$ is a multivariate Gaussian pdf, and $q_R()$ is an inverse Wishart pdf,

$$f(\mu, R|\mathbf{s}) \approx N(\mu; \mu_0^*, \Lambda^*) \cdot \mathcal{IW}(R; \Psi^*, \nu^*), \tag{34}$$

where the parameters are

$$\Lambda^* = (\Lambda^{-1} + J\nu^*\Psi^{*-1})^{-1} \tag{35}$$

$$\mu_0^* = \Lambda^*(\Lambda^{-1}\mu_0 + J\nu^*\Psi^{*-1}\bar{s}) \tag{36}$$

$$\Psi^* = \Psi + \sum_{j=1}^{J}(s_j - \mu_0^*)(s_j - \mu_0^*)^T + J\Lambda^* \tag{37}$$

$$\nu^* = \nu + J. \tag{38}$$

Equations (35)–(37) must be solved for Ψ^*, Λ^* and μ^*, but solving them analytically is not possible. We use instead a common iterative technique, where, starting at an initial guess for the parameters, the equations are iterated until convergence. If $\Psi^{*(0)}$ is the initial guess for Ψ^*, we get the following algorithm:

```
for $i = 1\ldots$ (until convergence) do
  $\Lambda^{*(i)} = (\Lambda^{-1} + J\nu^*\Psi^{*(i-1)-1})^{-1}$
  $\mu_0^{*(i)} = \Lambda^{*(i)}(\Lambda^{-1}\mu_0 + J\nu^*\Psi^{*(i-1)-1}\bar{s})$
  $\Psi^{*(i)} = \Psi + \sum_{j=1}^{J}(s_j - \mu_0^{*(i)})(s_j - \mu_0^{*(i)})^T + J\Lambda^{*(i)}$
end for
```

The iteration is guaranteed to converge to a local optimum, but not necessarily to the global optimum. Whether we find the global optimum or not depends on the starting point. In our case, the prior and the approximate posterior have the same parameters, so the obvious choice of starting point is the corresponding parameter of the prior, i.e $\Psi^{*(0)} = \Psi$.

Finally, we extract point estimates of μ and R. We let $\hat{\mu} = \mu_0^*$. For the point estimate of $\hat{R}$, see section 2.4.4. Although we are not directly interested in Λ^*, it is needed in order to calculate the other parameters. Λ^* represents the uncertainty about the mean μ_0^*, and as such still contains information that may be valuable depending on application. Equation (35) contains the inversion of 3 matrices, all of which are of dimension $P \times P$. This is not practical; e.g. in our validation data, P is over 50000, so such an inversion would require on the order of 10^{14} floating point operations. However, these matrices are highly redundant, as they are estimated from limited data. In practice, we have found that all three update equations (35), (36) and (37) can be computed efficiently without ever constructing any $P \times P$ matrices, and with inversion of much smaller matrices only. The details of the efficient computation are given in appendix C.

2.3.3. Workflow

When new data for a patient becomes available in the form of organ contours derived from 3D-scans, the first step is to obtain point-to-point correspondence between this patient's shapes and the shapes in the training data by deformable registration to the global reference shape. Next, the resulting shape vectors $s_1, \ldots, s_J$ are used as input to one of the algorithms in this section to produce patient-specific estimates of the posterior mean and covariance matrix. How to use these further depends on the specific application.

However, the algorithms require additional parameters, specifically the hyper-parameters μ_0, Ψ and ν as well as κ or Λ depending on the model. In this section, these parameters have been assumed given. In the next section, we show how we can obtain μ_0, Ψ and Λ from training data.

2.4. Estimating model parameters from training data

Bayesian algorithms require specification of the hyperparameters of the prior. For the present models, these are μ_0, κ, Λ, Ψ and ν, with κ specific to the NIW-model and Λ specific to the variational Bayes model.The vector and matrix valued parameters μ_0, Λ and Ψ are estimated from training data. Assume that data in the form of shape vectors $s_{i,j}$ from M patients are available, where i is the patient number and j is the scan number, and patient i has J_i scans.

2.4.1. Population mean

The prior mean μ_0 is the population mean shape, which is simply calculated as the average of all the individual mean shapes in the training data:

$$\mu_0 = \frac{1}{M}\sum_{i=1}^{M} \bar{s}_i = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{J_i}\sum_{j=1}^{J_i} s_{i,j}. \tag{39}$$

2.4.2. Population covariance matrix

The population covariance matrix R_{pop}, defined in (5), is in practice represented by its principal components and their variances. PCA of such a matrix has been dubbed 'simultaneous component analysis' (SCA) (Timmerman and Kiers 2003), since all patients are assumed to share the same principal components. The data matrix which is input to SCA contains all the columns from the patient-specific data matrices in the training data:

$$D_{\text{pop}} = \frac{1}{\sqrt{M}}[D_1 \quad D_2 \quad \dots \quad D_M], \tag{40}$$

where D_i is

$$D_i = \frac{1}{\sqrt{J_i - 1}}[s_{i,1} - \bar{s}_i \quad s_{i,2} - \bar{s}_i \quad \dots \quad s_{i,J_i} - \bar{s}]. \tag{41}$$

The covariance matrix $\hat{R}_{\text{pop}} = D_{\text{pop}}D_{\text{pop}}^T$ is used for both the classical population model and the NIW-model.

In the variational Bayes model, the scale matrix Ψ needs to be invertible. We will use a regularization approach for this model, where we add a constant δ_Ψ times the identity matrix, I, to the scaled sample covariance matrix:

$$\Psi = \nu\hat{R}_{\text{pop}} + \delta_\Psi I = \nu D_{\text{pop}}D_{\text{pop}}^T + \delta_\Psi I. \tag{42}$$

This structure, together with the similar structure of the inter-patient covariance matrix, makes it possible to compute the update equations(35)–(37) efficiently through the procedure detailed in appendix C.

2.4.3. Inter-patient covariance matrix

In the variational Bayes model, we also need to estimate the covariance matrix Λ of μ, the inter-patient covariance matrix. This matrix describes the uncertainty of μ. By definition,

$$\Lambda = E[(\mu - \mu_0)(\mu - \mu_0)^T], \tag{43}$$

where E[] is the expected value operator. We do not have direct observations of μ, but we have estimates, $\bar{s}_i$. A natural extension of the sample covariance matrix suggests an estimator of the form

$$\hat{\Lambda}_{\text{b}} = \frac{1}{M-1}\sum_{i=1}^{M}(\bar{s}_i - \hat{\mu}_0)(\bar{s}_i - \hat{\mu}_0)^T. \tag{44}$$

This estimate of Λ is biased, since the sample mean $\bar{s}_i$ is not equal to the true mean μ. We show in appendix D that the expected value of $\hat{\Lambda}_{\text{b}}$ is

$$E[\hat{\Lambda}_{\text{b}}] = \Lambda + cE[R], \tag{45}$$

where $c = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{J_i}$. The bias is therefore inversely proportional to the number of scans per patient. Since R_{pop} is an unbiased estimate of $E[R]$, we can get an unbiased estimate of Λ as

$$\hat{\Lambda} = \hat{\Lambda}_{\text{b}} - cR_{\text{pop}}. \tag{46}$$

However, since both $\hat{\Lambda}$ and $\hat{R}_{\text{pop}}$ are low rank, and they range over different subspaces, the resulting matrix is not positive semidefinite. This makes PCA a bit more complicated, but it is still possible. Details are given in appendix E. As for the intra-patient covariance matrix, the inter-patient covariance matrix must also be invertible, therefore we add a regularization factor $\delta_\Lambda I$. Additionally, since Λ expresses our uncertainly about the mean estimate, we want to have the possibility of increasing its overall size, so we introduce a constant multiplier α, which finally leads to

Table 1. Parameter values for all models. K-intra is the number of principal components used to compute the intra-patient covariance matrix, K-inter is the same for the inter-patient covariance matrix, ν and κ are scalar hyperparameters of the IW/NIW distributions, δ_Λ and δ_Ψ are regularization parameters for the matrices used in the variational Bayes iteration, and α is the weight of the inter-patient covariance matrix.

K-intra	K-inter	ν	κ	δ_Ψ	δ_Λ	α
12	20	6	0.25	240 000	80 000	4

$$\Lambda = \alpha\hat{\Lambda} + \delta_\Lambda I. \tag{47}$$

2.4.4. Probabilistic PCA

In the NIW-model, we used the point estimate $\frac{1}{\nu'}\Psi'$ for R, where Ψ' is the posterior scale matrix, and ν' is the posterior degrees of freedom. In the variational Bayes model, this is less straightforward. The posterior Ψ^* can be expressed as $D^*D^{*T} + \delta_\Psi^* I$ for some D^* and some δ_Ψ^*. The posterior δ_Ψ^* is approximately proportional to the prior δ_Ψ, and with a large δ_Ψ, the estimate $\frac{1}{\nu^*}\Psi^* = \frac{1}{\nu^*}D^*D^{*T} + \frac{1}{\nu^*}\delta_\Psi^* I$ places an unreasonable amount of variance on the shape coordinates. For this reason, we introduce a new parameter δ_R, and set the point estimate of R to

$$\hat{R} = \frac{1}{\nu^*}D^*D^{*T} + \frac{\nu}{\nu^*}\delta_R I. \tag{48}$$

For the prior distribution, the point estimate for R is found by replacing the posterior parameters values in (48) by the equivalent prior parameters. This yields

$$\hat{R}_0 = D_{\text{pop}}D_{\text{pop}}^T + \delta_R I. \tag{49}$$

When D_{pop} is found through PCA, this structure fits the description of *probabilistic PCA* (PPCA) introduced by Tipping and Bishop (1999). Their method provides a maximum likelihood estimate for δ_R given by

$$\delta_R = \frac{1}{P-K}\sum_{k=K+1}^{P}\lambda_k, \tag{50}$$

where λ_k is the kth largest eigenvalue of the population covariance matrix in (5) (i.e. the variance of the kth principal component), and K is the number of eigenpairs not discarded in PCA. In other words, δ_R is the average variance of the discarded dimensions.

3. Evaluation

3.1. Material

For evaluation, we used data from 37 patients with locally advanced prostate cancer. Each patient had 9-11 CT scans taken during treatment (typically 2 per week), including the plan CT used for RT dose planning. No laxatives were administered to the patients before or during treatment. The rectum was defined with content from the recto-sigmoid flexure to the anal verge. One single expert physicist contoured rectum on all CT scans for all patients, and all contours were reviewed and corrected by another expert physicist. This yielded a total of 373 rectum shapes, which were used in leave-one-out cross-evaluation. Details about the patients and treatment can be found in Hysing *et al* (2018). All shapes from the CT scans were converted to mesh representations with corresponding vertices, using deformable registration. Since toxicity is related to dose to the rectal wall and not its content, we evaluated the methods on the rectal wall. Since the inner wall is not seen on CT scans, we assumed 3 mm wall thickness, as in Sanguineti *et al* (2020).

3.2. Parameter values

The values of the scalar parameters were tuned manually. The values we used are shown in table 1. For the parameters K-intra and ν, which are applicable to multiple algorithms, we used the same value for all models.

3.3. Coverage probability matrices

To calculate predicted CPMs, μ_i and R_i was first estimated for each patient i using the patient-specific, population, and two Bayesian methods. For each method, 500 random rectal wall shapes per patient were then generated based on the distributions $\mathcal{N}(\mu_i,\ R_i)$. For each generated shape, we found which voxels (on a $1 \times 1 \times 1$ mm grid) were covered by the rectal wall using an in-house developed ray-tracing algorithm. The coverage probability of each voxel was defined as the fraction of generated rectal walls covering that voxel. This procedure was repeated using one, two and three input scans for each method.

We used the remaining independent $J_i - 3$ scans for each patient to compute reference CPMs. Since relatively few scans (6–8) were then available, we used the bootstrapping procedure detailed in section 3.4 with this data to generate smooth CPMs. The reference CPM for each patient was computed by drawing 500 bootstrapped rectal wall shapes, and setting the coverage probability of each voxel equal to the proportion of these shapes that covered the voxel.

The predicted CPMs and reference CPMs (the ground truth) were compared in terms of their normalized cross-correlation:

$$c = \frac{\sum_{v \in V} p_{\text{predict}}(v) p_{\text{true}}(v)}{\sqrt{\sum_{v \in V} p_{\text{predict}}^2(v) \sum_{v \in V} p_{\text{true}}^2(v)}}, \tag{51}$$

where V is the set of all voxels, and $p_{\text{predict}}(v)$ and $p_{\text{true}}(v)$ are the predicted and true coverage probabilities at voxel v, respectively.

3.4. Convergence behaviour

To analyse convergence of the four methods without re-using structures for both training and testing, we created a virtual data set for each patient in the original data set by using a PCA-based bootstrapping procedure: For each patient, we first calculated the principal components using all the patient's available shapes. We then calculated the PCA-scores for each shape: $c_{i,j,k}$, where i is the patient number, j is the scan number and k is the component number. To generate a new random scan for patient i, a new PCA-score c_k^* was drawn for each component number k, and a new shape s_i^* was synthesized according to

$$s_i^* = \bar{s}_i + \sum_{k=1}^{J_i} c_k^* w_{i,k}, \tag{52}$$

where $w_{i,k}$ is the kth principal component vector for patient i. The c_k^* values were drawn randomly from the existing values $c_{i,j,k}$ for $j = 1 \ldots J_i$, i.e. by bootstrapping. Since the principal component scores are uncorrelated, such mixing of the scores should create realistic new shapes. The bootstrapping procedure means that no specific distribution has been assumed.

For each patient, we generated 10 shapes using this procedure. These shapes were used as input to the models to estimate CPMs. The estimated CPM for each patient was compared to the reference CPM for that patient, which was generated using all individual scans.

3.5. Impact of the uncertainty parameter δ_R

For the variational Bayes model, the parameter δ_R naturally occured from the equations and the requirement that the covariance matrix must be non-singular. The PPCA method that we used to find δ_R can also be used for the other methods. We therefore tested the effect of δ_R on the the population model, the NIW model and the variational Bayes model, and compared the result to non-probabilistic PCA, i. e. $\delta_R = 0$. PPCA is not practical for the patient-specific model with as few as 3 input scans, since it requires that some principal components are not used. For the population model, δ_R was set constant, while for the NIW and variational method, it was updated according to the update equations for Ψ, which leads to

$$\delta_R(n) = \frac{n}{n + \nu} \delta_R(0), \tag{53}$$

where n is the number of scans.

The motivation for this additional evaluation was to avoid a bias in favour of the variational Bayes model.

4. Results

Visual comparison of the four first population intra-patient modes fits with anatomical expectations (figure 2). The first mode is mainly bending of the anorectal flexure; in the bent state, the rectum is less filled than in the straight state. The second mode shows stretching and compressing of the rectum in the caudal–cranial direction. The third mode shows mainly stretching of the top of the rectum in the left-right direction, while the fourth mode shows bending left-right of the top of the rectum. A general finding is that the most caudal third of the rectum, up to slightly above the anorectal flexure, moves very little. This is corroborated by figure 3, which shows coverage probabilities of the rectum wall for two example patients, a 'small mover' and a 'large mover'.

The Bayesian models take advantage of population data also when estimating the patient-specific mean rectum $\hat{\mu}$. Figure 4 shows how the mean estimates may differ with the Bayesian models for an example patient,

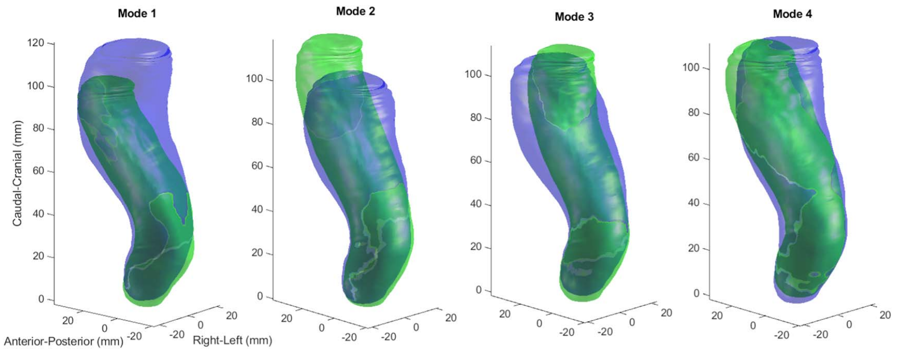

Figure 2. The four greatest population intra-patient deformation modes. The green and blue shapes represent +2 and −2 standard deviations of the deformation mode from the population mean rectum shape. A video with animation of the deformation modes is available in the supplementary material.

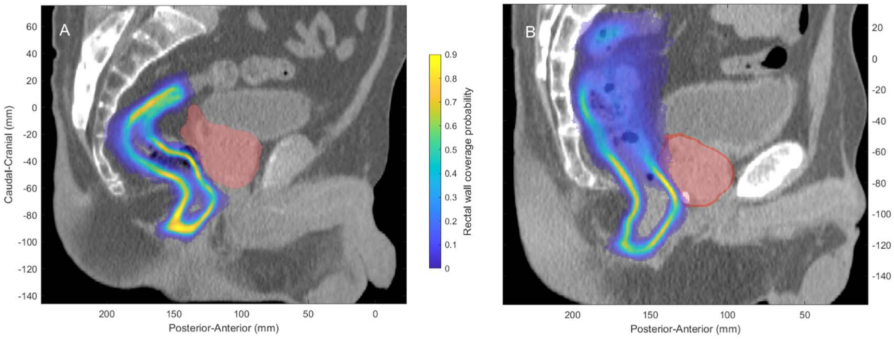

Figure 3. Coverage probability matrices for the rectal wall on a sagittal slice of the CT scan for two example patients—a 'small mover' (A) and a 'large mover' (B). The red area is the high dose volume to the prostate, that receives more than 67 Gy EQD$_2$.

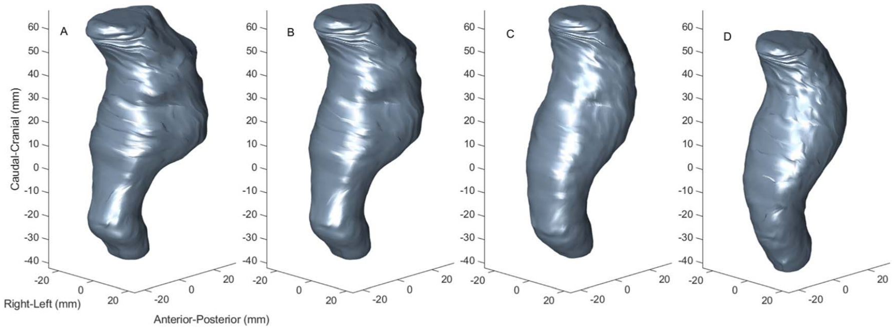

Figure 4. Estimated means and actual mean rectum shape for an example patient. A: Plan CT rectum shape, input to estimation algorithms. B: Mean rectum shape estimated by NIW model. C: Mean rectum shape estimated by variational Bayes model. D: Actual mean rectum shape over 9 scans.

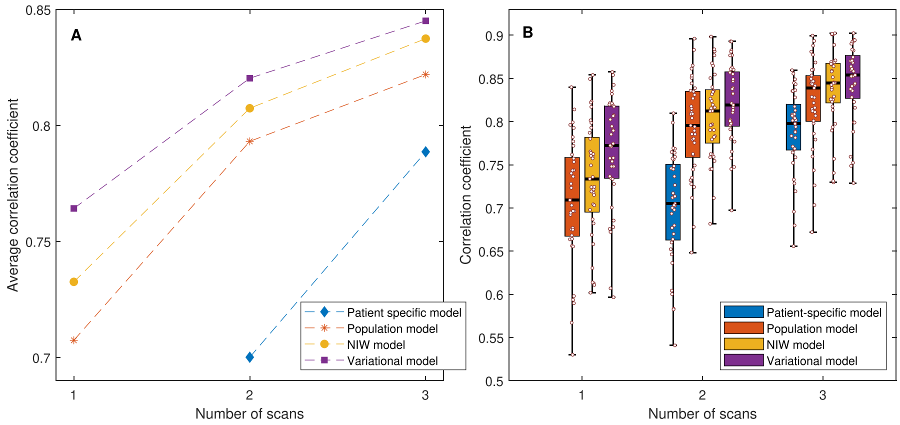

Figure 5. Correlation between the estimated CPMs and the references for the different methods using 1–3 input scans. A: Average correlation. B: Box plots showing median, 25th and 75th percentile and minimum and maximum values (whiskers). All individual values are also shown as circles over the box plot.

Table 2. Difference in CPM correlation between the population, NIW and variational models using one, two and three scans. Here, $\Delta\mu$ is the difference in average value of the CPM correlations, and %+ is the percentage of patients that saw improvement with the first method over the second.

	NIW versus pop. model			Variational versus pop. model			Variational versus NIW		
	$\Delta\mu$	p-value	%+	$\Delta\mu$	p-value	%+	$\Delta\mu$	p-value	%+
1 scan	0.026	6.2e-5	78	0.058	1.2e-8	95	0.032	2.2e-6	81
2 scans	0.014	1.8e-4	81	0.027	2.5e-6	86	0.013	1.2e-3	70
3 scans	0.015	2-2e-6	89	0.023	8.0e-7	89	0.008	0.01	62

given a single input scan. For this patient, the mean shape from variational Bayes model had the greatest similarity with the true mean shape.

The average correlation between the estimated CPMs and the references is shown in figure 5(A), while figure 5(B) shows the spread of the results among the individual patients. The two Bayesian methods outperform both the existing models, with the variational Bayes model showing superior results to the NIW-model. The results are summarized in table 2, where the patient-specific model has been left out since it performs poorly with as few as three scans. The differences between the population, NIW and variational Bayes model were consistently significant ($p < 0.05$). In comparison to the best existing model (the population model), the variation Bayes model improved correlation with the reference CPM in 35 out of 37 patients when using a single input scan (figure 6).

4.1. Convergence behaviour

The two Bayesian methods both outperform the patient-specific model with up to 6 scans, and outperforms the population model for any number of scans (figure 7). As the number of input scans increases, the patient-specific model and the two Bayesian models appear to converge toward the true CPM, while the population model improves only moderately. This is to be expected, since, in the population model, the covariance matrix representing the random error is never updated. All improvement seen in the population model is therefore from reduction of error in the mean estimate, often referred to as systematic error. The performance of the patient-specific model is comparable to that of the population model when both are given 4 scans. For more than 4 scans, the patient-specific model outperforms the population model. The variational Bayes model consistently performs slightly better than the NIW-model.

4.2. Impact of the uncertainty parameter δ_R

For all the models, PPCA through the addition of the δ_R parameter increases correlation as compared to ordinary PCA, as shown in figure 8. The difference between the models with and without the uncertainty

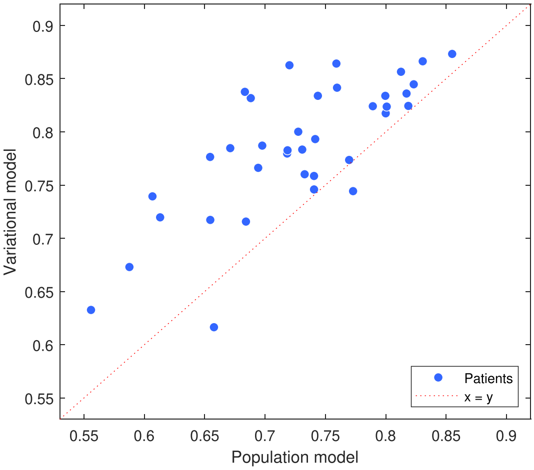

Figure 6. Correlation between the estimated CPMs and the references; comparison between the traditional population model and the proposed variational Bayes model. All points above the '$x = y$' line represent patients for which the variational Bayes method produced a better CPM estimate than the population model.

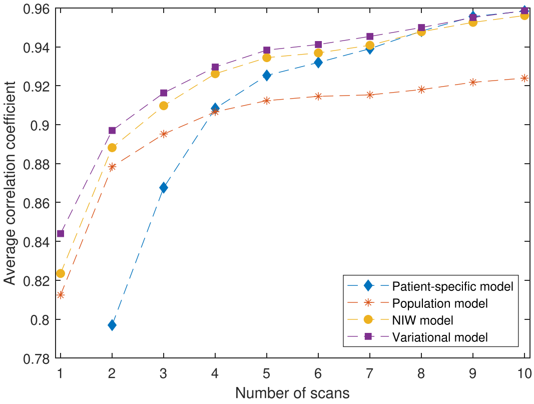

Figure 7. Average correlation between the estimated CPMs and the reference CPMs for the different methods using 1–10 scans, based on bootstrapped data.

parameter is greatest when using a single scan. Although the differences between the models decreased, both Bayesian methods with ordinary PCA still perform the same as, or better than the population model with PPCA.

5. Discussion

Both the new models outperform the existing population model significantly. Conceptually, the NIW model is only slightly more complex than the population model, so there is little rationale for rather using the population model. Additionally, figure 6 shows that the Bayesian models are robust, as evidenced by the fact that 35 out of 37 patients had improved result with the variational Bayes model over the population model (29/37 for the NIW-model without PPCA). There is therefore very little risk involved in moving to a Bayesian model.

It is to be expected that the new algorithms will perform worse for some patients due to the random nature of the data. Nevertheless, we examined the data for the two patients who performed worse with the variational Bayes than the population model using one scan to see if there were notable patterns. While no conclusion can be

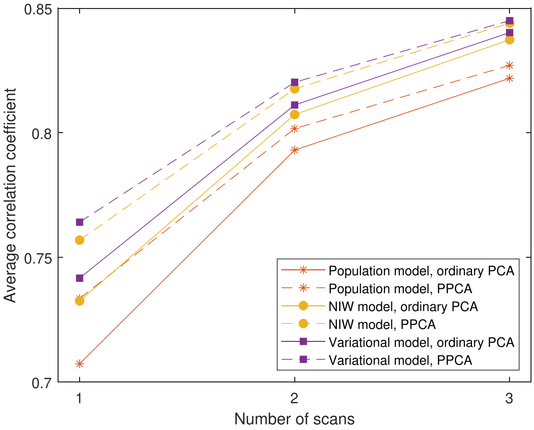

Figure 8. Comparison of ordinary and probabilistic PCA for different models. Each symbol represents the average correlation between the estimated CPMs and the references.

reached, it seems that, for these patients, the rectal shape in the pCT is coincidentally similar to the mean shape over all CTs.

The choice between the two Bayesian methods is a tradeoff between model accuracy and complexity. The main concern with the variational Bayes model is the conceptual rather than the computational complexity—it is more challenging to implement and requires more parameters than the NIW model. When using PPCA, the NIW models performance gets close to that of the variational Bayes model.

As expected, the patient-specific model cannot compete with the other models when few scans are available. This model still has an advantage in that no training data from the population is required. Additionally, deformable registration is more readily available between contours of the same patient than between contours of different patients. There are therefore applications where the patient-specific model is the only available option. However, in these cases, care should be taken that sufficient scans are available, as shown in figures 5 and 7.

The convergence analysis in figure 7 shows that we have achieved the goal of combining the advantages of both models; requiring few scans to achieve good accuracy while also improving accuracy with more scans. At around eight scans, the patient-specific model catches up with the Bayesian models. This is to be expected -at that point, the Bayesian models put very little weight on the population data since there is sufficient patient-specific data for an accurate model.

We have evaluated the model for the rectum, a highly flexible and deformable organ. The ability of the method to model other organs will depend on the amount of individual variation and the ability of the training data to replicate the variations that appear in the population. The fact that the models combine patient-specific data with the training data suggests that they should out-perform purely population based methods when there is great variability in the individual deformation. It is also possible to model multiple organs simultaneously, as done with the individual model in Söhn *et al* (2005). This may be advantageous, as correlations between the deformations of the different organs and their relative positions are taken into account.

As far as our experience goes, the variational Bayes iteration is not sensitive to the selected starting guess of the scale matrix Ψ^*, it appears to converge to the same solution regardless of starting point. The iteration takes less than a second to run for a single patient. Generating a CPM with a resoluton of 1 mm (about 3 million points) from 500 generated shapes took about 5 seconds on a standard PC. In practice, the main computational effort will be spent on deformable registration, which takes about 2 minutes for a single registration in our setup[5]

5.1. Applications

The calculation of CPMs play a key role in many applications of organ deformation models (Price and Moore 2007). The CPMs can be used for robust RT planning (Baum *et al* 2006), or to calculate margins based on the formula of Stroom *et al* (1999), as in Hysing *et al* (2011), Thörnqvist *et al* (2013a), Magallon-Baro *et al* (2019). In Ramlov *et al* (2017), Lindegaard *et al* (2017), CPMs were used clinically to reduce toxicity in nodal boosting of cervical cancer RT. Applications besides CPMs include robust evaluation through treatment course simulation (Söhn *et al* 2012, Hysing *et al* 2018), generation of plan libraries for RT personalized to motion (Rigaud *et al* 2019)

---

[5] Matterhorn software from Erasmus MC (Rotterdam), running on an Intel i7-4600U 2.1 GHz CPU.

and motion-robust optimization (Sobotta *et al* 2010, Unkelbach *et al* 2018). Recently, Owens *et al* (2022) used a pure inter-patient model to reconstruct colorectal dose in childhood cancer survivors who had received RT with no CT simulation. Thus, applications also extends to improving evaluation of complications from RT.

The Bayesian approach offers additional advantages because it quantifies the model uncertainty. Consider for example the robust evaluation in Söhn *et al* (2012), Hysing *et al* (2018): predicting dose-volume histograms (DVHs) with uncertainties (such as 5th and 95th percentiles). When using a non-Bayesian deformation model, the correctness of the predicted values rely on the correctness of the model's parameters. With a Bayesian model, the uncertainty of the parameters will translate to additional uncertainty regarding the dose-volume histogram, thus increasing the difference between the expected value and the 5/95 percentiles.

Interfractional geometrical errors in RT are often divided into systematic and random errors. The random error is the motion around the mean shape and position at each fraction, while the systematic error is the difference between the actual mean and the estimated mean, usually the shape and position at the plan CT. In terms of the deformation models, the systematic error is the difference between the estimated and the true patient mean, $\hat{\mu} - \mu$. The presented Bayesian models reduces the systematic error as compared to the previous methods by utilizing population data when estimating $\hat{\mu}$ (see figure 4). In addition, the new models provide a personalized *distribution* for the systematic error in terms of the posterior inter-patient distribution. The widely applied margin recipe by van Herk *et al* (2000) uses the formula $2.5\Sigma + 0.7\sigma$, where Σ and σ are the standard deviations of the systematic and random errors, respectively. Because the distribution of both the systematic and random errors are modeled under the Bayesian framework, it is in principle possible to use similar recipes for margins due to deformation.

5.2. Choice of evaluation metric

The cross-correlation metric puts proportionally higher weight on voxels that have a high coverage probability. Since a large portion of the organ tends to overlap in most or all shapes for one patient, all methods will tend to produce relatively high correlation values. Therefore, the differences between the methods may seem small. We still choose to use this metric because of its simplicity and ease of reproduction.

5.3. Gaussian likelihood

Both the Bayesian models and the models we compare to make the assumption that the data for a given patient is multivariate Gaussian distributed. This has been a standard assumption in applications of deformation models (e.g. Söhn *et al* 2012, Rios *et al* 2017). In the high dimensions that we operate in, it would require unrealistically many individual scans to disprove Gaussianness. Nevertheless, this assumption is a possible source of error, which showcases the need to evaluate the model against real data.

It should be possible to adapt the patient-specific and population models to use a nonparametric distribution of the PCA-scores as in Fontenla *et al* (2001), but this has not yet been demonstrated. In a Bayesian model, a non-Gaussian likelihood would make calculating the posterior mathematically intractable.

5.4. Parameter values

The values of the scalar parameters in table 1 were hand tuned with the objective to maximize the CPM correlations. Since it is not possible to evaluate the accuracy of the estimated distribution for a new patient without having many individual scans, one must in practice trust that parameter values that worked well for the training data also works well for new patients. If new data source is in some way different from the training data (e.g. a different image modality or IGRT routine, a different diagnosis or otherwise different type of patient), the parameters should at least be evaluated for this kind of data. However, in such cases Bayesian inference should perform better than a pure population approach, as it tailors the distribution to the data at hand.

The parameter κ for the NIW-model was set to 0.25. Using equation (18), we find that, given one input scan, this represents an *shrinkage factor* of 0.2; i.e the estimated mean is 'shrunk' by a factor 0.2 towards the population mean (Rørtveit *et al* 2021). The parameter ν, the number of degrees of freedom of the Wishart distribution, was set to 6 for both the NIW and the variational model. Normally, ν represents the number of samples from which Ψ was computed. However, this is under the assumption that these samples were all drawn from the same multivariate Gaussian distribution. In our case, the samples were drawn from M different Gaussian distributions with covariance matrices R_i, none of which match a future patient's covariance matrix. Therefore, we are much less certain about R, and we need to choose a value for ν that is much smaller than the total number of observations in the training data.

When tuning the values of δ_Ψ and δ_Λ, we found that these needed to be set surprisingly large to achieve satisfactory results. Possibly, some assumptions or parts of the model do not actually fit the data well, and increasing the regularization values then compensates for the poor fit. This underscores the importance of evaluating the models with realistic data, and tailoring the parameters to the case at hand.

5.5. Degenerate inverse Wishart distribution

The inverse Wishart distribution is usually defined in terms of the (forward) Wishart distribution: If a random $n \times n$ matrix G is Wishart distributed with $G \sim \mathcal{W}(\Psi, \nu)$, then its inverse G^{-1} is inverse Wishart distributed with $G^{-1} \sim \mathcal{IW}(\Psi^{-1}, \nu)$. However, when $\nu < n$, the Wishart distribution is degenerate, as any matrix G with a non-zero probability density has rank ν and is therefore singular. Then this definition of the IW distribution does not work. A singular inverse Wishart distribution is defined through the pseudo-inverse of W (Cook and Forzani 2011, Bodnar *et al* 2016). Unfortunately, this distribution is not well behaved, and does not have a finite expected value. Since we do not explicitly use the distribution, but rather a point estimate, this does not make a difference when using the models as described in this paper. However, care must be taken if using the full Bayesian model as described in section 5.6, as individual realizations of G can have very large eigenvalues.

5.6. Extensions

We have applied the models to the rectum alone, however, for use in e.g. robust optimization, it would be advantageous to model several structures simultaneously so that the correlation between structures are taken into account.

In the evaluation of the algorithms, we used point estimates for μ and R as opposed to a full distribution. We have thus ignored the uncertainty in the model itself, and therefore sinned against the Bayesian philosophy. We chose to do this for the sake of computational complexity. However, it is possible to account for the additional uncertainty: When performing Monte-Carlo sampling, one would first sample μ and R from the posterior distribution every time before sampling s from $\mathcal{N}(\mu, R)$. The resulting distribution of s is called the *posterior predictive* distribution. Particularly the sampling of R is computationally intensive. An alternative approach might therefore be to use a point estimate for R while sampling μ, as systematic errors are often of greater importance than random errors.

The presented models have been applied to deformably registered organ surfaces. A more common form of deformable registration is the deformation of 3D-images with image intensities. Since both types of registration produce deformation vector fields, it is possible, with some adaptions, to apply these models to deformed images as well.

6. Conclusions

We have implemented and evaluated two Bayesian methods for modelling organ deformation occuring during RT treatment. The NIW and the variational Bayes models both outperformed previous organ deformation models when applied to the rectal wall of prostate cancer patients.

Acknowledgments

The authors would like to thank Markus Alber for his contribution with ideas and discussions in the early phase of this project. We also thank Andras Zolnay at Erasmus MC Cancer center for valuable insight and discussion.

Funding

This work was funded by Trond Mohn Foundation [grant number BFS2017TMT07].

Ethical statement

All patients gave their consent before being enrolled in a phase II dose-escalation trial delivered with moderately hypo-fractionated pelvic IMRT at Haukeland University Hospital, Bergen, Norway. The study protocol was approved by the Regional Ethical Committee (REK 2006/15727) before enrollment starting in 2007. The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

Appendix A. Derivation of the conditional posteriors

The pdf for the multivariate Gaussian distribution for a vector x of dimension p is

$$\mathcal{N}(x;\, \mu,\, R) = \frac{1}{\sqrt{(2\pi)^p |R|}} \exp\left(-\frac{1}{2}(x - \mu)^T R^{-1}(x - \mu)\right). \tag{A.1}$$

The pdf of the inverse Wishart distribution of a $p \times p$ matrix Q is

$$\mathcal{IW}(Q;\, \Psi,\, \nu) = \frac{|\Psi|^{\nu/2}}{2^{\nu p/2}\Gamma_p(\frac{\nu}{2})}|Q|^{-(\nu+p+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\Psi Q^{-1})\right). \tag{A.2}$$

The joint pdf of μ, R and the samples $S = \{s_1, s_2, \ldots, s_n\}$, based on our prior and our likelihood is

$$f(\mu,\, R,\, S) = \mathcal{N}(\mu;\, \mu_0,\, \Lambda)\mathcal{IW}(R;\, \Psi,\, \nu) \prod_{i=1}^{n} \mathcal{N}(s_i;\, \mu,\, R). \tag{A.3}$$

Writing this out using (A.1) and (A.2), and leaving out any constant factors (factors that do not contain μ, R or S), we find

$$\begin{aligned} f(\mu,\, R,\, S) \propto \frac{|R|^{-(\nu+p+1)/2}}{|R|^{n/2}} &\exp\left(-\frac{1}{2}(\mu - \mu_0)^T \Lambda^{-1}(\mu - \mu_0)\right. \\ &\left. -\frac{1}{2}\mathrm{tr}(\Psi R^{-1}) - \frac{1}{2}\sum_{i=1}^{n}(s_i - \mu)^T R^{-1}(s_i - \mu\right). \end{aligned} \tag{A.4}$$

Using the property of the trace $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$ and the fact that a scalar is its own trace, the sum within the exponential can be written as

$$\sum_{i=1}^{n}(s_i - \mu)^T R^{-1}(s_i - \mu) = \mathrm{tr}\left([\sum_{i=1}^{n}(s_i - \mu)(s_i - \mu)^T]R^{-1}\right). \tag{A.5}$$

Furthermore, since $\mathrm{tr}(A) + \mathrm{tr}(B) = \mathrm{tr}(A + B)$, we can write

$$\begin{aligned} &\mathrm{tr}(\Psi R^{-1}) + \sum_{i=1}^{n}(s_i - \mu)^T R^{-1}(s_i - \mu) \\ &= \mathrm{tr}([\Psi + \sum_{i=1}^{n}(s_i - \mu)(s_i - \mu)^T]R^{-1}) \end{aligned} \tag{A.6}$$

To condition (A.4) on μ and S, we can leave out any factors not containing R - that is, the first term in the exponenial. Using (A.6), we find

$$\begin{aligned} &f(R|\mu,\, S) \\ &\propto |R|^{-(\nu+p+1+n)/2} \exp\left(-\frac{1}{2}\mathrm{tr}([\Psi + \sum_{i=1}^{n}(s_i - \mu)(s_i - \mu)^T]R^{-1})\right) \\ &\propto \mathcal{IW}(R;\, \Psi',\, \nu'), \end{aligned} \tag{A.7}$$

with

$$\Psi' = \Psi + \sum_{i=1}^{n}(s_i - \mu)(s_i - \mu)^T \tag{A.8}$$

and

$$\nu' = \nu + n, \tag{A.9}$$

which concludes the derivation of the conditional posterior for R.

Next, we condition (A.4) on R and S to find

$$\begin{aligned} &f(\mu|R,\, S) \\ &\propto \exp\left(-\frac{1}{2}(\mu - \mu_0)^T \Lambda^{-1}(\mu - \mu_0) - \frac{1}{2}\sum_{i=1}^{n}(s_i - \mu)^T R^{-1}(s_i - \mu)\right) \end{aligned} \tag{A.10}$$

Looking actively for a Gaussian distribution, we want to find that the terms inside the exponential are equal to

$$-\frac{1}{2}(\mu - \mu_0')^T \Lambda'^{-1}(\mu - \mu_0') + c, \tag{A.11}$$

for some Λ' and μ', with any constant term c. Grouping the terms that are quadratic in μ, we find

$$-\frac{1}{2}\mu^T\Lambda^{-1}\mu - \frac{1}{2}\sum_{i=1}^{n}\mu^T R^{-1}\mu = -\frac{1}{2}\mu^T(\Lambda^{-1} + nR^{-1})\mu, \tag{A.12}$$

therefore, if this is a Gaussian distribution, we must have

$$\Lambda'^{-1} = \Lambda^{-1} + nR^{-1} \rightarrow \Lambda' = (\Lambda^{-1} + nR^{-1})^{-1}. \tag{A.13}$$

Grouping the linear terms, we find

$$-\mu^T\Lambda^{-1}\mu_0 - \sum_{i=1}^{n}\mu^T R^{-1}s_i = -\mu^T(\Lambda^{-1}\mu_0 + nR^{-1}\bar{s}). \tag{A.14}$$

Setting this equal to the linear terms in (A.11), we have

$$\mu^T\Lambda'^{-1}\mu_0' = \mu^T(\Lambda^{-1}\mu_0 + nR^{-1}\bar{s}), \tag{A.15}$$

which is true for any μ if and only if

$$\mu_0' = \Lambda'(\Lambda^{-1}\mu_0 + nR^{-1}\bar{s}). \tag{A.16}$$

The constant terms can be ignored, as they will be absorbed by the normalization. Finally, this gives us

$$f(\mu|R, S) \propto \exp\left(-\frac{1}{2}(\mu - \mu_0')^T\Lambda'^{-1}(\mu - \mu_0')\right) \tag{A.17}$$

$$\propto \mathcal{N}(\mu; \mu_0', \Lambda'), \tag{A.18}$$

with μ_0' as in (A.16) and Λ' as in (A.13). ☐

Appendix B. Variational approximation

To find the functions q_μ and q_R, we follow the procedure presented in Gelman *et al* (1995). The minimizing functions are given by

$$\log q_\mu(\mu) = E_R[\log f(\mu|R, S)] + \text{const} \tag{B.1}$$

and

$$\log q_R(R) = E_\mu[\log f(R|\mu, S)] + \text{const}, \tag{B.2}$$

where E_R and E_μ indicate an average over R only or μ only, respectively.

Inserting (A.10) into (B.1), we get

$$\log q_\mu(\mu) = E_R\left[-\frac{1}{2}(\mu - \mu_0)^T\Lambda^{-1}(\mu - \mu_0) - \frac{1}{2}\sum_{i=1}^{n}(s_i - \mu)^T R^{-1}(s_i - \mu)\right] + \text{const} \tag{B.3}$$

$$= -\frac{1}{2}(\mu - \mu_0)^T\Lambda^{-1}(\mu - \mu_0) - \frac{1}{2}\sum_{i=1}^{n}(s_i - \mu)^T E[R^{-1}](s_i - \mu) + \text{const}. \tag{B.4}$$

Following the lines of the derivation in appendix A, we find

$$q_\mu(\mu) = \mathcal{N}(\mu; \mu_0^*, \Lambda^*), \tag{B.5}$$

with

$$\mu_0^* = (\Lambda^{-1} + nE[R^{-1}])^{-1}(\Lambda^{-1}\mu_0 + nE[R^{-1}]\bar{s}). \tag{B.6}$$

and

$$\Lambda^* = (\Lambda^{-1} + nE[R^{-1}])^{-1}. \tag{B.7}$$

Similarly, we insert (A.7) into (B.2) to find

$$\log q_R(R) = E_\mu[\log(|R|^{-(\nu+p+1+n)/2}) - \frac{1}{2}\text{tr}([\Psi + \sum_{i=1}^{n}(s_i - \mu)(s_i - \mu)^T]R^{-1})\Bigg] + \text{const} \tag{B.8}$$

$$= \log(|R|^{-(\nu+p+1+n)/2}) \\ - \frac{1}{2}\mathrm{tr}([\Psi + \sum_{i=1}^{n} E[(s_i - \mu)(s_i - \mu)^T]]R^{-1}) + \mathrm{const} \tag{B.9}$$

The term within the expectation operator is

$$\sum_{i=1}^{n} E[(s_i - \mu)(s_i - \mu)^T] \\ = \sum_{i=1}^{n} (s_i s_i^T + E[\mu\mu^T] - E[\mu]s_i^T - s_i E[\mu]^T) \tag{B.10}$$

$$= \sum_{i=1}^{n} (s_i - E[\mu])(s_i - E[\mu])^T + n(E[\mu\mu^T] - E[\mu]E[\mu]^T) \tag{B.11}$$

$$= \sum_{i=1}^{n} (s_i - E[\mu])(s_i - E[\mu])^T + n \cdot \mathrm{cov}(\mu). \tag{B.12}$$

This leads to

$$q_R(R) = \mathcal{IW}(R; \Psi^*, \nu^*), \tag{B.13}$$

with

$$\nu^* = \nu + n \tag{B.14}$$

and

$$\Psi^* = \Psi + \sum_{i=1}^{n} (s_i - E[\mu])(s_i - E[\mu])^T + n \cdot \mathrm{cov}(\mu). \tag{B.15}$$

Finally, we replace the moments in (B.6), (B.7) and (B.15) by the moments from the approximate distributions q_μ and q_R. Since R, according to (B.13), is inverse-Wishart distributed with scale matrix Ψ^* and $\nu^* = \nu + n$ degrees of freedom, its inverse R^{-1} is Wishart-distributed with scale matrix Ψ^{*-1} and $\nu + n$ degrees of freedom. Its expectation is $E[R^{-1}] = \nu^* \Psi^{*-1}$. Therefore we find

$$\mu_0^* = (\Lambda^{-1} + n(\nu + n)\Psi^{*-1})^{-1}(\Lambda^{-1}\mu_0 + n(\nu + n)\Psi^{*-1}\bar{s}) \tag{B.16}$$

and

$$\Lambda^* = (\Lambda^{-1} + n(\nu + n)\Psi^{*-1})^{-1}. \tag{B.17}$$

By (B.5), the mean and covariace of μ is μ_0^* and Λ^*, therefore (B.15) becomes

$$\Psi^* = \Psi + \sum_{i=1}^{n} (s_i - \mu_0^*)(s_i - \mu_0^*)^T + n\Lambda^*. \tag{B.18}$$

Appendix C. Efficient computation of the update iteration

The key to finding the estimated mean and covariance matrix for a patient is iteration over the update equations, repeated here for convenience:

$$\Lambda^* = (\Lambda^{-1} + n(\nu + n)\Psi^{*-1})^{-1} \tag{C.1}$$

$$\mu_0^* = \Lambda^*(\Lambda^{-1}\mu_0 + n(\nu + n)\Psi^{*-1}\bar{s}) \tag{C.2}$$

$$\Psi^* = \Psi + \sum_{j=1}^{n} (s_j - \mu_0^*)(s_j - \mu_0^*)^T + n\Lambda^* \tag{C.3}$$

$$\nu^* = \nu + n \tag{C.4}$$

Only ν^* can be calculated directly. The other parameters rely on each other, and therefore require an iteration to converge to the correct values.

Putting the iteration number i in a superscript (replacing $\cdot^*$), we can write the iteration as

$$\Lambda^{(i)} = (\Lambda^{-1} + n(\nu + n)\Psi^{(i-1)-1})^{-1} \tag{C.5}$$

$$\mu_0^{(i)} = \Lambda^{(i)}(\Lambda^{-1}\mu_0 + n(\nu + n)\Psi^{(i-1)-1}\bar{s}) \tag{C.6}$$

$$\Psi^{(i)} = \Psi + \sum_{j=1}^{n} (s_j - \mu_0^{(i)})(s_j - \mu_0^{(i)})^T + n\Lambda^{(i)} \tag{C.7}$$

We can see that we need to supply a starting guess for the first value $\Psi^{(0)}$.[5] A natural starting guess is $\Psi^{(0)} = \Psi$.

In theory, the iteration represented by equations (C.5)–(C.7) can be implemented directly in any numerically oriented programming language. However, this would require storing and inverting very large $P \times P$ matrices, which is not attainable in practice. However, due to the structure of Ψ and Λ (when estimated as in sections 2.4.2 and 2.4.3), memory and computation requirements can be drastically reduced.

Both matrices Λ and Ψ can be represented as an outer product of a data matrix with itself plus a scalar multiple of the identity matrix:

$$\Lambda = D_\Lambda D_\Lambda^T + \delta_\Lambda I \tag{C.8}$$

$$\Psi = D_\Psi D_\Psi^T + \delta_\Psi I. \tag{C.9}$$

Here, D_Ψ and D_Λ are $P \times N_\Psi$ and $P \times N_\Lambda$ matrices, with $N_\Lambda, N_\Psi \ll P$. Multiplying a vector a by such a matrix is much faster than the general $O(P^2)$ figure, since e. g.

$$\Lambda a = (D_\Lambda D_\Lambda^T + \delta_\Lambda I)a = D_\Lambda(D_\Lambda^T a) + \delta_\Lambda a, \tag{C.10}$$

which is easily computed in $O(N_\Lambda P)$ time. Furthermore, it is also fast to solve an equation such as $\Lambda x = b$.

Throughout this derivation we shall make heavy use of the following special case of the *Woodbury matrix identity*, which holds for any matrices A and B and scalar δ as long as the involved inversions are possible:

$$(\delta I + ABA^T)^{-1} = \delta^{-1}I - \delta^{-1}A(\delta B^{-1} + A^T A)^{-1}A^T. \tag{C.11}$$

This means that the inverses of Λ and Ψ can also be written in the form $DCD^T + \delta I$ for some D, C and δ.

C.1. Computing $\Lambda^{(i)}$

We shall show later that $\Psi^{(i)}$ can be written for any i as

$$\Psi^{(i)} = D^{(i)}G^{(i)}D^{(i)T} + \delta_\Psi^{(i)}I, \tag{C.12}$$

for some $\delta_\Psi^{(i)}$ and $G^{(i)}$, and where

$$D^{(i)} = [D_\Lambda \quad D_\Psi^{(i)}] \tag{C.13}$$

for some $D_\Psi^{(i)}$ of dimension $P \times (N_\Psi + n)$. Inserting (C.8) and (C.12) into (C.5), we get

$$\Lambda^{(i)} = [(D_\Lambda D_\Lambda^T + \delta_\Lambda I)^{-1} + n(\nu + n)(D^{(i-1)}G^{(i-1)}D^{(i-1)T} + \delta_\Psi^{(i-1)}I)^{-1}]^{-1}. \tag{C.14}$$

Using the matrix inversion lemma (C.11) on both the inner inverses of (C.14), we get

$$\begin{aligned}\Lambda^{(i)} = [&\delta_\Lambda^{-1}I - \delta_\Lambda^{-1}D_\Lambda(\delta_\Lambda I + D_\Lambda^T D_\Lambda)^{-1}D_\Lambda^T + n(\nu + n)\delta_\Psi^{(i-1)-1}I \\ &- n(\nu + n)\delta_\Psi^{(i-1)-1}D^{(i-1)}(\delta_\Psi^{(i-1)}G^{(i-1)-1} + D^{(i-1)T}D^{(i-1)})^{-1}D^{(i-1)T}]^{-1}\end{aligned} \tag{C.15}$$

In order to group the terms, note that

$$-\delta_\Lambda^{-1}D_\Lambda(\delta_\Lambda I + D_\Lambda^T D_\Lambda)^{-1}D_\Lambda^T = D^{(i-1)}QD^{(i-1)T}, \tag{C.16}$$

where Q is a block-diagonal matrix

$$Q = \begin{bmatrix} -\delta_\Lambda^{-1}(\delta_\Lambda I + D_\Lambda^T D_\Lambda)^{-1} & \\ & 0_{N_\Psi + n \times N_\Psi + n} \end{bmatrix}. \tag{C.17}$$

We also define

$$L^{(i)} = -\delta_\Psi^{(i)-1}(\delta_\Psi^{(i)}G^{(i)-1} + D^{(i)T}D^{(i)})^{-1} \tag{C.18}$$

and

$$F^{(i)} = Q + n(\nu + n)L^{(i-1)} \tag{C.19}$$

Now, we can write

$$\Lambda^{(i)} = (\delta_\Lambda^{-1} + n(\nu + n)\delta_\Psi^{(i-1)-1})I + D^{(i-1)}F^{(i)}D^{(i-1)T})^{-1}, \tag{C.20}$$

Applying the matrix inversion lemma again, we find

$$\Lambda^{(i)} = \delta^{(i)}I - \delta^{(i)}D^{(i-1)}H^{(i)}D^{(i-1)T}, \tag{C.21}$$

---

[5] Given that the iteration starts with the equation for $\Lambda^{(1)}$. If we had started with one of the other equations, a starting guess for at least one other parameter would need to be provided.

where

$$H^{(i)} = [D^{(i-1)T}D^{(i-1)} + \delta^{(i)-1}F^{(i)-1}]^{-1} \tag{C.22}$$

and

$$\delta^{(i)} = (\delta_\Lambda^{-1} + n(\nu + n)\delta_\Psi^{(i-1)-1})^{-1}. \tag{C.23}$$

equation (C.21) gives us an expression for $\Lambda^{(i)}$ using only lower dimensional matrices and scalars. In practice, we never construct $\Lambda^{(i)}$—it is represented implicitly by $D^{(i)}$, $H^{(i)}$ and $\delta^{(i)}$ through (C.21).

C.2. Computing $\mu_0^{(i)}$

Through the derivation of $\Lambda^{(i)}$, we have already come a long way towards computing $\mu_0^{(i)}$. We can write (C.6) as

$$\mu_0^{(i)} = \Lambda^{(i)}r^{(i)}, \tag{C.24}$$

with

$$r^{(i)} = \Lambda^{-1}\mu_0 + n(\nu + n)\Psi^{(i-1)-1}\bar{s}. \tag{C.25}$$

The first term of (C.25) is constant, and can be computed once. Using the matrix inversion lemma on (C.8), we find

$$\Lambda^{-1}\mu_0 = \delta_\Lambda^{-1}\mu_0 - \delta_\Lambda^{-1}D_\Lambda(\delta_\Lambda I + D_\Lambda^T D_\Lambda)^{-1}(D_\Lambda\mu_0) \tag{C.26}$$

The last term needs to be computed for each iteration. We find it by using the matrix inversion lemma on (C.12):

$$\begin{aligned}\Psi^{(i)-1}\bar{s} &= (\delta_\Psi^{(i)-1}I - \delta_\Psi^{(i)-1}D^{(i)}(\delta_\Psi^{(i)}G^{(i)-1} + D^{(i)T}D^{(i)})^{-1}D^{(i)T})\bar{s} &\text{(C.27)}\\ &=\delta_\Psi^{(i)-1}\bar{s} + D^{(i)}L^{(i)}(D^{(i)T}\bar{s}) &\text{(C.28)}\end{aligned}$$

Finally, inserting (C.21) into (C.24), we find

$$\mu_0^{(i)} = \delta^{(i)}r^{(i)} - \delta^{(i)}D^{(i-1)}H^{(i)}(D^{(i-1)T}r^{(i)}). \tag{C.29}$$

C.3. Computing $\boldsymbol{\Psi}^{(i)}$

The update equation for Ψ is

$$\begin{aligned}\Psi^{(i)} &= \Psi + \sum_{j=1}^{n}(s_j - \mu_0^{(i)})(s_j - \mu_0^{(i)})^T + n\Lambda^{(i)}\\ &= D_\Psi D_\Psi^T + \delta_\Psi I + \sum_{j=1}^{n}(s_j - \mu_0^{(i)})(s_j - \mu_0^{(i)})^T + n\Lambda^{(i)}. &\text{(C.30)}\end{aligned}$$

We can augment the data matrix D_Ψ by inserting new columns which are the mean-subtracted data vectors;

$$D_\Psi^{(i)} = [D_\Psi \quad s_1 - \mu_0^{(i)} \quad s_2 - \mu_0^{(i)} \quad \dots \quad s_n - \mu_0^{(i)}], \tag{C.31}$$

and we find

$$\Psi^{(i)} = D_\Psi^{(i)}D_\Psi^{(i)T} + \delta_\Psi I + n\Lambda^{(i)}. \tag{C.32}$$

Inserting (C.21), we get

$$\Psi^{(i)} = D_\Psi^{(i)}D_\Psi^{(i)T} + \delta_\Psi I + n(\delta^{(i)}I - \delta^{(i)}D^{(i-1)}H^{(i)}D^{(i-1)T}) \tag{C.33}$$

We want to group the terms of this equation, but run into a slight problem: One term contains $D_\Psi^{(i)}$, while another term contains $D^{(i-1)}$ (which contains $D_\Psi^{(i-1)}$). In practice, this can easily be resolved by replacing $D^{(i-1)}$ by $D^{(i)}$; this is in line with the algorithm philosophy of always using the most recent guess of each parameter, and also guarantees that the equations (C.1)–(C.3) hold at convergence (at convergence, we have $D^{(i)} = D^{(i-1)}$). Now, to group the terms, first note that

$$D_\Psi^{(i)}D_\Psi^{(i)T} = D^{(i)}KD^{(i)T}, \tag{C.34}$$

where

$$K = \begin{bmatrix} 0_{N_\Lambda \times N_\Lambda} & \\ & I_{N_\Psi + n} \end{bmatrix}. \tag{C.35}$$

Thus, we can write

$$\Psi^{(i)} = D^{(i)}(K - n\delta^{(i)}H^{(i)})D^{(i)T} + (\delta_\Psi + n\delta^{(i)})I. \tag{C.36}$$

We now see that we must have

$$G^{(i)} = K - n\delta^{(i)}H^{(i)} \tag{C.37}$$

and

$$\delta_{\Psi}^{(i)} = \delta_{\Psi} + n\delta^{(i)} \tag{C.38}$$

in order for $\Psi^{(i)}$ to be written as

$$\Psi^{(i)} = D^{(i)}G^{(i)}D^{(i)T} + \delta_{\Psi}^{(i)}I. \tag{C.39}$$

C.4. Initial values

Initially, we want to get $\Psi^{(0)} = \Psi$, i. e. $D^{(0)}G^{(0)}D^{(0)T} + \delta_{\Psi}^{(0)}I = D_{\Psi}D_{\Psi}^{T} + \delta_{\Psi}I$ which achieve by setting

$$\delta_{\Psi}^{(0)} = \delta_{\Psi} \tag{C.40}$$

$$D_{\Psi}^{(0)} = [D_{\Psi} \quad 0_{P\times n}] \tag{C.41}$$

$$G^{(0)} = K. \tag{C.42}$$

However, $G^{(0)}$ is not invertible, which makes it impossible to compute $L^{(0)}$ as in (C.18). Instead, $L^{(0)}$ must be initialized to

$$L^{(0)} = \begin{bmatrix} 0_{N_{\Lambda}\times N_{\Lambda}} & \\ & -\delta_{\Psi}^{-1}(\delta_{\Psi}I + D_{\Psi}^{(i)T}D_{\Psi}^{(i)})^{-1} \end{bmatrix}. \tag{C.43}$$

C.5. Algorithm summary

```
**Input:** $\mu_0, D_{\Lambda}, D_{\Psi}, \delta_{\Psi}, \delta_{\Lambda}, s_1 \ldots s_n, \nu$
**Output:** $\mu_0^{*}, D^{*}, G^{*}, \delta_{\Psi}^{*}, \delta^{*}, H^{*}$
$K = \begin{bmatrix} 0_{N_{\Lambda}\times N_{\Lambda}} & \\ & I_{N_{\Psi}+n} \end{bmatrix}.$
$Q = \begin{bmatrix} -\delta_{\Lambda}^{-1}(\delta_{\Lambda}I + D_{\Lambda}^{T}D_{\Lambda})^{-1} & \\ & 0_{N_{\Psi}+n\times N_{\Psi}+n} \end{bmatrix}$
$q \leftarrow \delta_{\Lambda}^{-1}\mu_0 - \delta_{\Lambda}^{-1}D_{\Lambda}(\delta_{\Lambda}I + D_{\Lambda}^{T}D_{\Lambda})^{-1}(D_{\Lambda}^{T}\mu_0)$ /* $q$ is $\Lambda^{-1}\mu_0$ */
$D_{\Psi}^{(0)} = [D_{\Psi} \quad 0_{P\times n}]$
$D^{(0)} \leftarrow [D_{\Lambda} \quad D_{\Psi}^{(0)}]$
$\delta_{\Psi}^{(0)} \leftarrow \delta_{\Psi}$
$\mu_0^{(0)} \leftarrow \mu_0$
$L^{(0)} \leftarrow \begin{bmatrix} 0_{N_{\Lambda}\times N_{\Lambda}} & \\ & -\delta_{\Psi}^{-1}(\delta_{\Psi}I + D_{\Psi}^{(0)T}D_{\Psi}^{(0)})^{-1} \end{bmatrix}$
$i \leftarrow 0$
**repeat**
    $i \leftarrow i + 1$
    $\delta^{(i)} \leftarrow (\delta_{\Lambda}^{-1} + n(\nu + n)\delta_{\Psi}^{(i-1)-1})^{-1}$
    $F^{(i)} \leftarrow Q + n(\nu + n)L^{(i-1)}$
    $H^{(i)} \leftarrow (D^{(i-1)T}D^{(i-1)} + \delta^{(i)-1}F^{(i)-1})^{-1}$
    $r^{(i)} \leftarrow q + n(\nu + n)(\delta_{\Psi}^{(i-1)-1}I + D^{(i-1)}L^{(i-1)}D^{(i-1)T})\bar{s}$
    $\mu_0^{(i)} \leftarrow \delta^{(i)}r^{(i)} - \delta^{(i)}D^{(i-1)}H^{(i)}(D^{(i-1)T}r^{(i)})$
    $D_{\Psi}^{(i)} \leftarrow [D_{\Psi} \quad s_1 - \mu_0^{(i)} \quad s_2 - \mu_0^{(i)} \quad \ldots \quad s_n - \mu_0^{(i)}]$
    $D^{(i)} \leftarrow [D_{\Lambda} \quad D_{\Psi}^{(i)}]$
    $\delta_{\Psi}^{(i)} \leftarrow \delta_{\Psi} + n\delta^{(i)}$
    $G^{(i)} \leftarrow K - n\delta^{(i)}H^{(i)}$
    $L^{(i)} \leftarrow -\delta_{\Psi}^{(i)-1}(\delta_{\Psi}^{(i)}G^{(i)-1} + D^{(i)T}D^{(i)})^{-1}$
**until** $\|\mu_0^{(i)} - \mu_0^{(i-1)}\| < \epsilon$
$\mu_0^{*} \leftarrow \mu^{(i)}, D^{*} \leftarrow D^{(i)}, G^{*} \leftarrow G^{(i)}, \delta_{\Psi}^{*} \leftarrow \delta_{\Psi}^{(i)}, \delta^{*} \leftarrow \delta^{(i)}, H^{*} \leftarrow H^{(i)}$
/* Implicit, not computed: $\Lambda^{*} = \delta^{*}I - \delta^{*}D^{*}H^{*}D^{*T}$ */
/*Implicit, not computed: $\Psi^{*} = D^{*}G^{*}D^{*T} + \delta_{\Psi}^{*}I$ */
```

Appendix D. Bias of the inter-patient covariance matrix estimate

We estimate the inter-patient covariance matrix as

$$\hat{\Lambda} = \frac{1}{M-1}\sum_{i=1}^{M}(\bar{s}_i - \hat{\mu}_0)(\bar{s}_i - \hat{\mu}_0)^T. \tag{D.1}$$

This is the sample covariance matrix of $\bar{s}_i$, as opposed to μ which we are interested in. But $\bar{s}_i$ are not identically distributed if J_i varies. We can show that

$$E[\hat{\Lambda}] = \frac{1}{M}\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i). \tag{D.2}$$

To avoid clutter, the proof of this result is given at the end of the appendix.

The covariance matrix of a sample mean based on n i.i.d. samples is always given by $1/n$ times the covariance matrix of one sample. In other words,

$$\mathrm{cov}(\bar{s}_i|\mu, R) = \frac{1}{J_i}R. \tag{D.3}$$

Now we can use the *law of total covariance*, which states, for two scalar random variables a and b,

$$\mathrm{cov}(a, b) = E[\mathrm{cov}(a, b|c)] + \mathrm{cov}(E[a|c], E[b|c]). \tag{D.4}$$

In our case, we get

$$\mathrm{cov}(\bar{s}_i) = E[\mathrm{cov}(\bar{s}_i|\mu, R)] + \mathrm{cov}(E[\bar{s}_i|\mu, R]) \tag{D.5}$$

$$= E\left[\frac{1}{J_i}R\right] + \mathrm{cov}(\mu) \tag{D.6}$$

$$= \frac{1}{J_i}E[R] + \Lambda \tag{D.7}$$

since, by definition, $\mathrm{cov}(\mu) = \Lambda$. Inserting (D.7) into (D.2) yields

$$E[\hat{\Lambda}] = \frac{1}{M}\sum_{i=1}^{M}\left(\Lambda + \frac{1}{J_i}E[R]\right) = \Lambda + cE[R], \tag{D.8}$$

where

$$c = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{J_i}. \tag{D.9}$$

☐

Proof of (D.2):

We start by manipulating (D.1):

$$\hat{\Lambda} = \frac{1}{M-1}\sum_{i=1}^{M}(\bar{s}_i - \hat{\mu}_0)(\bar{s}_i - \hat{\mu}_0)^T \tag{D.10}$$

$$= \frac{1}{M-1}\left(\sum_{i=1}^{M}\bar{s}_i\bar{s}_i^T + \sum_{i=1}^{M}\hat{\mu}_0\hat{\mu}_0^T - \sum_{i=1}^{M}\bar{s}_i\hat{\mu}_0^T - \sum_{i=1}^{M}\hat{\mu}_0\bar{s}_i^T\right) \tag{D.11}$$

$$= \frac{1}{M-1}\left(\sum_{i=1}^{M}\bar{s}_i\bar{s}_i^T + \sum_{i=1}^{M}\hat{\mu}_0\hat{\mu}_0^T - (\sum_{i=1}^{M}\bar{s}_i)\hat{\mu}_0^T - \hat{\mu}_0(\sum_{i=1}^{M}\bar{s}_i^T)\right) \tag{D.12}$$

$$= \frac{1}{M-1}\left(\sum_{i=1}^{M}\bar{s}_i\bar{s}_i^T + M\hat{\mu}_0\hat{\mu}_0^T - M\hat{\mu}_0\hat{\mu}_0^T - M\hat{\mu}_0\hat{\mu}_0^T\right) \tag{D.13}$$

$$= \frac{1}{M-1}\left(\sum_{i=1}^{M}\bar{s}_i\bar{s}_i^T - M\hat{\mu}_0\hat{\mu}_0^T\right), \tag{D.14}$$

where we used $\hat{\mu}_0 = \frac{1}{M}\sum_{i=1}^{M}\bar{s}_i$. Taking the expectation, and using the general formula $E[xx^T] = \mathrm{cov}(x) + E[x]E[x]^T$, we find

$$E[\hat{\Lambda}] = \frac{1}{M-1}\left(\sum_{i=1}^{M}E[\bar{s}_i\bar{s}_i^T] - ME[\hat{\mu}_0\hat{\mu}_0^T]\right) \tag{D.15}$$

$$= \frac{1}{M-1}\sum_{i=1}^{M}(\mathrm{cov}(\bar{s}_i) + E[\bar{s}_i]E[\bar{s}_i]^T)$$

$$- \frac{M}{M-1}(\mathrm{cov}(\hat{\mu}_0) + E[\hat{\mu}_0]E[\hat{\mu}_0]^T) \tag{D.16}$$

$$= \frac{1}{M-1}\left(\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i) - M\mathrm{cov}(\hat{\mu}_0)\right), \tag{D.17}$$

since $E[\bar{s}_i] = E[E[\bar{s}_i|\mu]] = E[\mu] = \mu_0 = E[\hat{\mu}_0]$. Looking at $\mathrm{cov}(\hat{\mu}_0)$, we find

$$\mathrm{cov}(\hat{\mu}_0) = \mathrm{cov}\left(\frac{1}{M}\sum_{i=1}^{M}\bar{s}_i\right) \tag{D.18}$$

$$= \frac{1}{M^2}\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i), \tag{D.19}$$

since $\bar{s}_i$ are independent (though not identically distributed). Inserting (D.19) into (D.17) yields

$$E[\hat{\Lambda}] = \frac{1}{M-1}\left(\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i) - \frac{M}{M^2}\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i)\right) \tag{D.20}$$

$$= \frac{1}{M-1}\left(1 - \frac{1}{M}\right)\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i) \tag{D.21}$$

$$= \frac{1}{M}\sum_{i=1}^{M}\mathrm{cov}(\bar{s}_i). \tag{D.22}$$

□

Appendix E. PCA for the bias-corrected inter-patient covariance matrix

The bias-corrected inter-patient covariance matrix estimate is given by

$$\tilde{\Lambda} = \hat{\Lambda} - c\hat{R}_{\mathrm{pop}}, \tag{E.1}$$

where $c = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{J_i}$. This matrix is not positive semidefinite, and cannot be expressed with a real-valued data matrix D as $\tilde{\Lambda} = DD^T$. It can, however, be expressed as

$$\tilde{\Lambda} = AB^T, \tag{E.2}$$

where $A = [D_\Lambda \quad \sqrt{c}D_{\mathrm{pop}}]$ and $B = [D_\Lambda \quad -\sqrt{c}D_{\mathrm{pop}}]^T$.

As usual $\tilde{\Lambda}$ is too big to practically perform eigenvalue decompostion on. However, there is a relation between the eigenvalue decomposition of AB^T and that of B^{TA}. The latter is a small matrix, and its eigenvalue decomposition can easily computed using any numerical software package. Given the kth eigenvalue λ_k and the kth eigenvector v_k of B^{TA}, the kth eigenvalue of $\tilde{\Lambda}$ is λ_k, and the kth eigenvector is

$$w_k = Av_k. \tag{E.3}$$

A proof of this result is given at the end of the appendix. The scale of w_k is arbitrary, so we want to normalize it as

$$w_k' = \frac{w_k}{\|w_k\|}. \tag{E.4}$$

As usual in PCA, we discard the eigenpairs corresponding to the smallest eigenvalues. In this case, since the matrix is not positive semidefinite, several of the eigenvalues will be negative. We need to discard all eigenpairs corresponding to negative eigenvalues, since we cannot have negative variance for any of the modes (which would lead to a complex data matrix). The PCA-reduced covariance matrix can now be represented by a data matrix $\tilde{D}_{PCA}$ as

$$\tilde{\Lambda}_{PCA} = \tilde{D}_{PCA}\tilde{D}_{PCA}^T, \tag{E.5}$$

with

$$\tilde{D}_{PCA} = [\sqrt{\lambda_1}w_1' \quad \sqrt{\lambda_2}w_2' \quad \ldots \quad \sqrt{\lambda_K}w_K']. \tag{E.6}$$

Proof of (E.3):

Let w_k be an eigenvector of AB^T, and λ_k be the corresponding eigenvalue, i. e.

$$AB^T w_k = \lambda_k w_k. \tag{E.7}$$

We can transform AB^T into B^{TA} by what we may call a pseudo-similarity transformation:

$$A^+(AB^T)A = (A^TA)^{-1}A^TAB^TA = B^TA, \tag{E.8}$$

where A^+ denotes the pseudo-inverse of A. Also note that AA^+ is a projection matrix onto the subspace spanned by A. Since w_k, as an eigenvector of AB^T, is already in this subspace, we have

$$AA^+w_k = w_k. \tag{E.9}$$

Using the three previous equations, we can now write

$$B^TA(A^+w_k) = A^+AB^TAA^+w_k = A^+AB^Tw_k = \lambda_k A^+w_k. \tag{E.10}$$

This shows that λ_k is an eigenvalue of B^{TA}, with corresponding eigenvector $v_k = A^+w_k$. However, we want to find w_k given v_k. Using (E.9) again, we find

$$v_k = A^+w_k \tag{E.11}$$

$$\rightarrow Av_k = AA^+w_k = w_k. \tag{E.12}$$

□

ORCID iDs

Øyvind Lunde Rørtveit https://orcid.org/0000-0001-6545-663X
Liv Bolstad Hysing https://orcid.org/0000-0002-7593-7549
Sara Pilskog https://orcid.org/0000-0002-3475-7939

References

Baum C, Alber M, Birkner M and Nüsslin F 2006 Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities *Radiother. Oncol.* **78** 27–35

Bishop C M 2006 Pattern recognition and machine learning *Information Science and Statistics* (New York: Springer)

Bodnar T, Mazur S and Podgórski K 2016 Singular inverse Wishart distribution and its application to portfolio theory *J. Multivariate Anal.* **143** 314–26

Bondar L, Intven M, Burbach J P M, Budiarto E, Kleijnen J P, Philippens M, van Asselen B, Seravalli E, Reerink O and Raaymakers B 2014 Statistical modeling of CTV motion and deformation for IMRT of early-stage rectal cancer *Int. J. Radiat. Oncol.*Biol.*Phys.* **90** 664–72

Budiarto E, Keijzer M, Storchi P R, Hoogeman M S, Bondar L, Mutanga T F, de Boer H C J and Heemink A W 2011 A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases *Phys. Med. Biol.* **56** 1045–61

Cook R D and Forzani L 2011 On the mean and variance of the generalized inverse of a singular Wishart matrix *Electron. J. Stat.* **5** 146–58

Fontenla E, Pelizzari C A, Roeske J C and Chen G T Y 2001 Using serial imaging data to model variabilities in organ position and shape during radiotherapy *Phys. Med. Biol.* **46** 2317–36

Fujikoshi Y, Ulyanov V V and Shimizu R 2010 *Multivariate Statistics : High-Dimensional and Large-Sample Approximations* (Hoboken, UNITED STATES: John Wiley & Sons, Incorporated)

Gelman A, Carlin J B, Stern H S and Rubin D B 1995 *Bayesian Data Analysis* (Chapman and Hall/CRC)

Herschtal A, Foroudi F, Greer P B, Eade T N, Hindson B R and Kron T 2012 Finding the optimal statistical model to describe target motion during radiotherapy deliverya Bayesian approach *Phys. Med. Biol.* **57** 2743–55

Hysing L B, Ekanger C, Zolnay A, Helle S I, Rasi M, Heijmen B J M, Sikora M, Söhn M, Muren L P and Thörnqvist S 2018 Statistical motion modelling for robust evaluation of clinically delivered accumulated dose distributions after curative radiotherapy of locally advanced prostate cancer *Radiother. Oncol.: J. Eur. Soc. Therapeutic Radiol. Oncol.* **128** 327–35

Hysing L B, Söhn M, Muren L P and Alber M 2011 A coverage probability based method to estimate patient-specific small bowel planning volumes for use in radiotherapy *Radiother. Oncol: J. Eur. Soc. Therapeutic Radiol. Oncol.* **100** 407–11

Lam K L, Ten Haken R K, Litzenberg D, Balter J M and Pollock S M 2005 An application of Bayesian statistical methods to adaptive radiotherapy *Phys. Med. Biol.* **50** 3849–58

Lindegaard J C, Assenholt M, Ramlov A, Fokdal L U, Alber M and Tanderup K 2017 Early clinical outcome of coverage probability based treatment planning for simultaneous integrated boost of nodes in locally advanced cervical cancer *Acta Oncol. (Stockholm, Sweden)* **56** 1479–86

Magallon-Baro A, Loi M, Milder M T W, Granton P V, Zolnay A G, Nuyttens J J and Hoogeman M S 2019 Modeling daily changes in organ-at-risk anatomy in a cohort of pancreatic cancer patients *Radiother. Oncol.* **134** 127–34

Owens C A *et al* 2022 Development and validation of a population-based anatomical colorectal model for radiation dosimetry in late effects studies of survivors of childhood cancer *Radiother. Oncol.* **176** 118–126

Price G J and Moore C J 2007 A method to calculate coverage probability from uncertainties in radiotherapy via a statistical shape model *Phys. Med. Biol.* **52** 1947–65

Ramlov A, Assenholt M S, Jensen M F, Grønborg C, Nout R, Alber M, Fokdal L, Tanderup K and Lindegaard J C 2017 Clinical implementation of coverage probability planning for nodal boosting in locally advanced cervical cancer *Radiother. Oncol.* **123** 158–63

Rigaud B *et al* 2019 Statistical shape model to generate a planning library for cervical adaptive radiotherapy *IEEE Trans. Med. Imaging* **38** 406–16

Rios R, De Crevoisier R, Ospina J D, Commandeur F, Lafond C, Simon A, Haigron P, Espinosa J and Acosta O 2017 Population model of bladder motion and deformation based on dominant eigenmodes and mixed-effects models in prostate cancer radiotherapy *Med. Image Anal.* **38** 133–49

Rørtveit Ø L, Hysing L B, Stordal A S and Pilskog S 2021 Reducing systematic errors due to deformation of organs at risk in radiotherapy *Med. Phys.* **48** 6578–87

Sanguineti G, Faiella A, Farneti A, D'Urso P, Fuga V, Olivieri M, Giannarelli D, Marzi S, Iaccarino G and Landoni V 2020 Refinement & validation of rectal wall dose volume objectives for prostate hypofractionation in 20 fractions *Clin. Transl. Radiat. Oncol.* **21** 91–7

Sobotta B, Söhn M and Alber M 2010 Robust optimization based upon statistical theory *Med. Phys.* **37** 4019–28

Söhn M, Birkner M, Yan D and Alber M 2005 Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation *Phys. Med. Biol.* **50** 5893–908

Söhn M, Sobotta B and Alber M 2012 Dosimetric treatment course simulation based on a statistical model of deformable organ motion *Phys. Med. Biol.* **57** 3693–709

Stroom J C, de Boer H C, Huizenga H and Visser A G 1999 Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability *Int. J. Radiat. Oncol. Biol. Phys.* **43** 905–19

Szeto Y Z, Witte M G, van Herk M and Sonke J J 2017 A population based statistical model for daily geometric variations in the thorax *Radiother. Oncol.* **123** 99–105

Thörnqvist S, Hysing L B, Zolnay A G, Söhn M, Hoogeman M S, Muren L P, Bentzen L and Heijmen B J M 2013a Treatment simulations with a statistical deformable motion model to evaluate margins for multiple targets in radiotherapy for high-risk prostate cancer *Radiother. Oncol.* **109** 344–9

Thörnqvist S, Hysing L B, Zolnay A G, Söhn M, Hoogeman M S, Muren L P and Heijmen B J M 2013b Adaptive radiotherapy in locally advanced prostate cancer using a statistical deformable motion model *Acta Oncol. (Stockholm, Sweden)* **52** 1423–9

Timmerman M E and Kiers H A L 2003 Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences *Psychometrika* **68** 105–21

Tipping M E and Bishop C M 1999 Probabilistic principal component analysis *J. R. Stat. Soc.* B **61** 611–22

Unkelbach J *et al* 2018 Robust radiotherapy planning *Phys. Med. Biol.* **63** 22TR02

van Herk M, Remeijer P, Rasch C and Lebesque J V 2000 The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy *Int. J. Radiat. Oncol.* **Biol.* **Phys.* **47** 1121–35

Yan D, Vicini F, Wong J and Martinez A 1997 Adaptive radiation therapy *Phys. Med. Biol.* **42** 123–32

Paper III

InsidePolyhedron - Fast Point-In-Polyhedron test on a grid

Øyvind Lunde Rørtveit

Submitted to the Journal of Open Research Software

Paper IV

Calculating cumulative dose and uncertainty for OARs of re-irradiation patients using a Bayesian motion model

Øyvind Lunde Rørtveit, Liv Bolstad Hysing, Andreas Størksen Stordal, Christian Ekanger, Sara Pilskog

Manuscript in preparation

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230867846 (print)
9788230859681 (PDF)