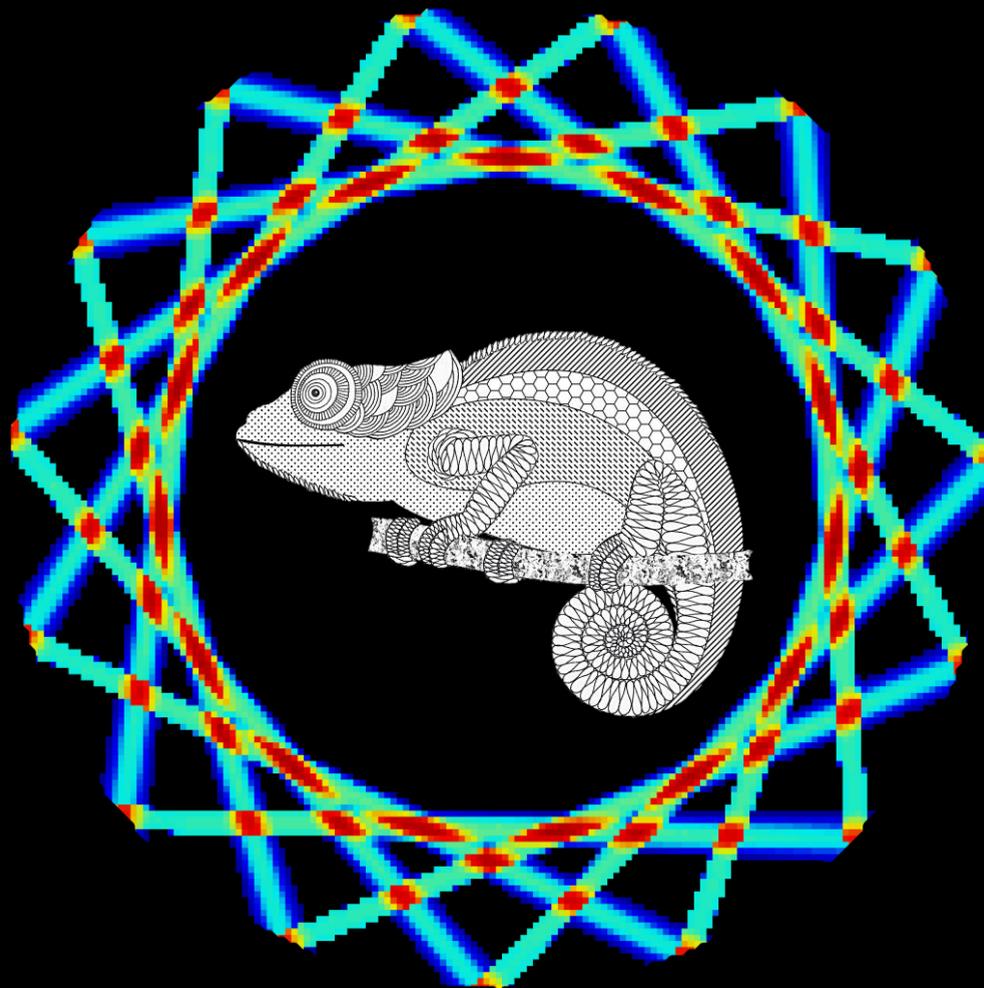


Automated Online-Adaptive Intensity-Modulated Proton Therapy



Thyrza Z. Jagt

Automated Online-Adaptive Intensity-Modulated Proton Therapy

Thyrza Z. Jagt

Copyright © Thyrza Z. Jagt, 2020

Printed by: Proefschriftmaken

ISBN: 978-94-6423-014-7

Automated Online-Adaptive Intensity-Modulated Proton Therapy

Geautomatiseerde online-adaptieve
intensiteitsgemoduleerde protontherapie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 25 november 2020 om 09:30 uur

door

Thyrza Zeralda Jagt
geboren te Lelystad

Promotiecommissie

Promotor: Prof.dr. M.S. Hoogeman

Overige leden: Prof.dr. L. Incrocci
Prof.dr.ir. J.-J. Sonke
Prof.dr. U.A. van der Heide

Copromotor: Dr.ir. S. Breedveld

*Medicine needs to be personalized.
Tailored to each disease, each patient, each moment.
Able to adapt to its surroundings.*

Like a chameleon

Table of contents

1	Introduction	1
2	Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations in prostate cancer	11
3	An automated planning strategy for near real-time adaptive proton therapy in prostate cancer	35
4	Plan-library supported automated replanning for online-adaptive intensity-modulated proton therapy of cervical cancer	55
5	Online-adaptive versus robust IMPT for prostate cancer: how much can we gain?	69
6	Evaluation of alternative parameter settings for dose restoration and full plan adaptation in IMPT for prostate cancer	83
7	Robust contour propagation using deep learning and image registration for online-adaptive proton therapy of prostate cancer	101
8	Correlations between the shifts in prompt gamma emission profiles and the changes in daily target coverage during simulated pencil-beam scanning proton therapy	127
9	Discussion	145
A	Appendices	157
	References	177
	Summary	193
	Samenvatting	199
	PhD portfolio	205
	List of Publications	207

Chapter

1

Introduction

1.1 INTENSITY-MODULATED PROTON THERAPY

Cancer is still a leading cause of death worldwide, with more than 3.7 million new patients and 1.9 million deaths each year in Europe (World Health Organization). A widely used treatment modality is radiation therapy. Radiation therapy, or radiotherapy, is a form of cancer treatment in which ionizing radiation is used to eradicate the tumor cells by damaging their DNA. Prior to the radiation treatment of a patient, a personalized treatment plan is constructed. A treatment plan consists of the individualized treatment unit settings and a simulated dose distribution resulting from these settings, projected on a 3-dimensional CT scan (planning CT scan). Due to unavoidable dose received by healthy tissues surrounding the tumor (organs at risk), severe side effects may be induced, which can have a long-lasting negative impact on the patients' quality of life. The goal during the construction of the treatment plan is to find the optimal balance between delivering an adequate dose to the target volume (including the tumor) and sparing of the organs at risk (OARs).

Intensity-modulated proton therapy (IMPT) is a type of radiotherapy in which the patient is irradiated using high-energy protons. Groups of protons, so-called pencil-beams or spots, are aimed at the tumor from different directions. Protons are positively charged particles, depositing most of their dose at the end of their range in a so-called Bragg peak. This is illustrated in Figure 1.1. The main advantage of this localized dose deposition is a better sparing of the healthy tissue surrounding the tumor. Better sparing can result in fewer side effects, thereby limiting the impact on the patients' quality of life.

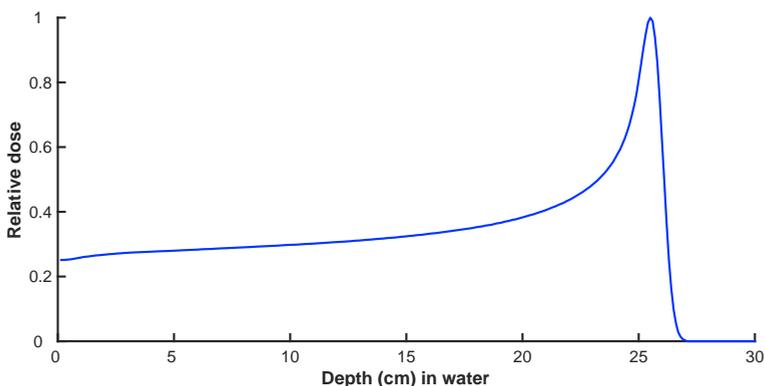


Figure 1.1: Bragg peak of a 200 MeV proton beam: integral dose as a function of depth in water.

1.2 1.2 UNCERTAINTIES IN IMPT

Although IMPT can deposit the dose very locally due to its characteristic Bragg peak, this same characteristic makes it very sensitive to variations in the daily anatomy [1–3], as illustrated in Figure 1.2. The depth of each Bragg peak within the body is dependent on the energy of the pencil-beam and to a large part on the electron density of the tissues the pencil-beam encounters along its path. Due to anatomical variations the densities along the pencil-beam paths can change, altering the depths of the individual Bragg peaks and changing the overall shape of the dose distribution. This possibly results in local over- and under-dosage. When considering the anatomical variations, one can distinguish between inter-fraction variation which occurs between the treatment fractions, and intra-fraction variation which occurs during a single treatment fraction. An example of inter-fraction variation is changes in bowel filling. Intra-fractionally the anatomy can change for example by breathing. In addition, gradual changes, e.g. caused by tumor regression or weight loss, may also impact the location of the Bragg peaks.

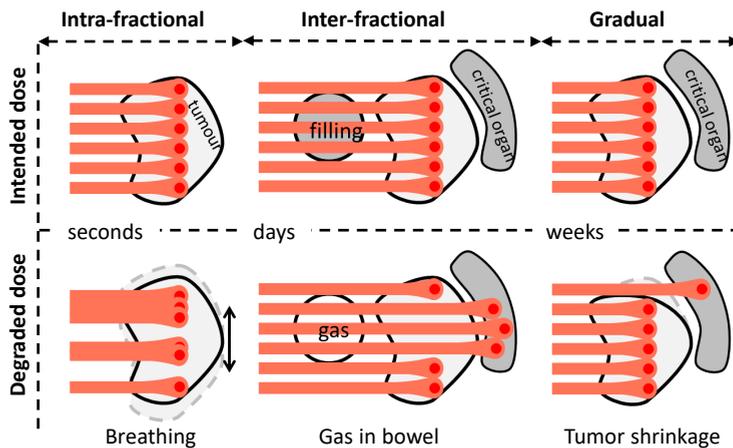


Figure 1.2: Illustration showing the effect of different anatomical variations on IMPT. Each pencil-beam is individually affected by the variations, leading to an overall change in shape and intensity of the dose distribution. Courtesy of M.S. Hoogeman.

In this thesis we focused on inter-fraction variations in the pelvis. The inter-fraction variations that have been investigated include variations in shape and location of female and male pelvic organs and the resulting density changes along the pencil-beam paths. An example of the impact these changes have on a dose distribution is shown in Figure 1.3 for a prostate cancer case.

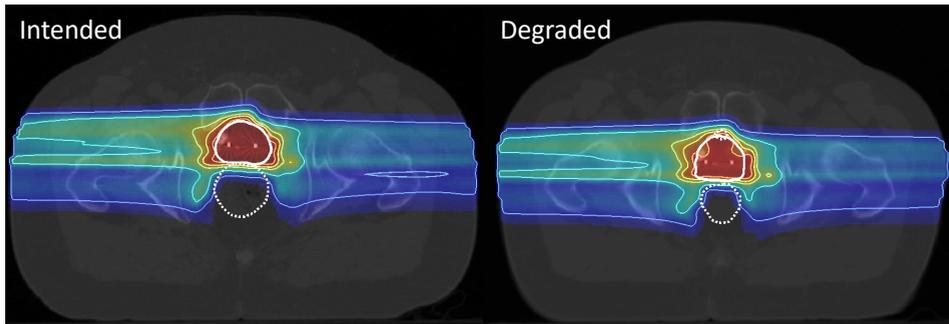


Figure 1.3: Illustration showing the dosimetric effects of anatomical variations on IMPT. Targets are depicted by solid lines, OARs by dotted lines.

1.3 ADAPTIVE PROTON THERAPY

To mitigate the effect of the anatomical variations on the dose distribution, several motion management strategies can be considered. The simplest strategy aims to account for the anatomical variations during the initial treatment planning, by enlarging the tumor region with a margin and generating the treatment plan using this enlarged target volume (so called planning target volume, PTV). In IMPT, however, the dose distribution may not be conserved after an anatomical change, making this approach inadequate [4]. A more advanced strategy is to minimize the impact of anatomical variations by explicitly including those as error scenarios in the optimization of the treatment unit settings [5, 6]. This approach of robust treatment planning [7–9] has proven to be more effective than the PTV margin approach. Still, the more the treatment plan is made robust, the less healthy tissue can be spared [10].

A third approach to account for inter-fraction variations is to adapt the treatment plan online to fit the daily anatomy, i.e. adaptive planning. It is expected that this approach results in a smaller treated volume and improved sparing of healthy tissues.

A possible workflow for adaptive proton therapy is described in the project description of ADAPTNOW – High-Precision Cancer Treatment by Online-Adaptive Proton Therapy. In this workflow, one would prior to each treatment fraction:

1. use an in-room CT scanner to generate an image of the daily anatomy, followed by automated delineation of the tumor and OARs,
2. automatically move the patient to the treatment unit, in which time in the background the computations for treatment plan adaptation are performed, and

3. treat the patient while using prompt gamma profiles to monitor the delivered dose for comparison against the planned dose as part of treatment delivery quality assurance (QA).

The described workflow is illustrated in Figure 1.4. Ideally the automated delineation and the automated plan adaptation should be completed within 30 seconds, in which time frame the patient can be moved from the in-room CT scanner to the gantry. Limiting the delineation and adaptation to this time frame limits possible intra-fraction variation, ensuring that the anatomy to which the plan gets adapted matches the anatomy at start of the dose delivery. Short adaptation times will furthermore ensure that the fraction time is not prolonged, adding to patient comfort.

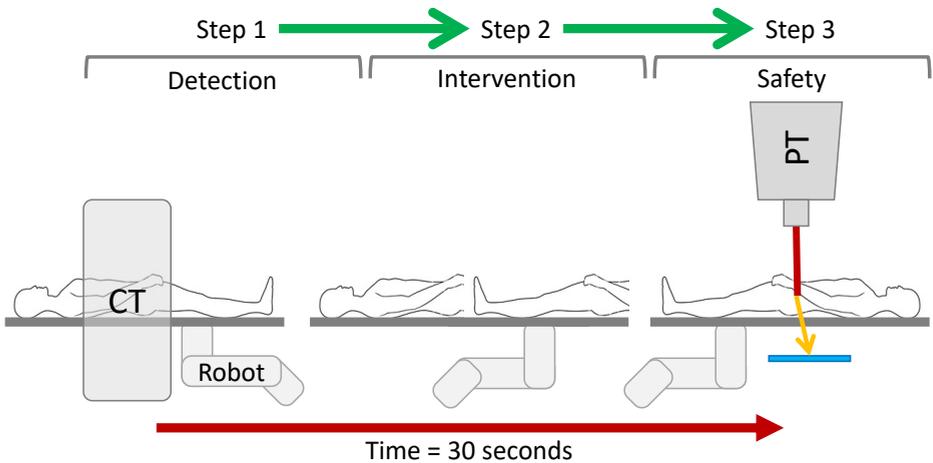


Figure 1.4: Illustration of the proposed online-adaptive workflow. A CT scan is made of the patient, after which the robotic manipulator moves the treatment couch from the CT to the treatment position. During the movement, the treatment plan is adapted to the daily anatomy. Courtesy of M.S. Hoogeman.

In this thesis we focused on the methods required for the second step of the workflow, i.e. the adaptation of the treatment plan (Chapters 2 to 6). We have developed and investigated different adaptive planning strategies focusing on daily target coverage, organ sparing and adaptation times.

Chapters 7 and 8 describe the work regarding the automated delineation method (step 1) and treatment delivery QA using prompt gamma profiles (step 3).

1.4 THIS THESIS

The aim of this thesis was to investigate feasibility of online-adaptive IMPT. To this end, we used high-risk prostate cancer and locally advanced cervical cancer as model system. These tumor sites were selected as they both show challenging inter-fraction variations. For high-risk prostate cancer, daily tissue density variations occur due to changes in bowel, bladder and rectum filling. The changes in rectum filling also cause variation in the position and orientation of the prostate. Another challenge lies in the motion of the nodal target volume, which is independent from that of the prostate. In locally advanced cervical cancer the target region can show large day-to-day variation in shape due to changes in rectum, bladder and sigmoid filling. Proving that adaptation is feasible for these challenging tumor sites will therefore imply feasibility for other sites. To achieve this goal the following research questions were answered in this thesis.

1.4.1 Can the dosimetric effects of density variations along the pencil-beam paths be reversed to restore the prior treatment plan?

An IMPT treatment plan can be described by the lateral location, energy and intensity of the individual pencil-beams. Density variations along the pencil-beam paths affect the depth and shape of the Bragg peaks, resulting in a distorted dose distribution. This suggests that the dose distribution can be restored to its initial state by restoring the depth of the Bragg peaks. This can be done by adjusting their energy. In Chapter 2, we describe a simple dose restoration strategy based on this idea. The method starts by adapting the pencil-beams' energy of the prior plan to the correct water equivalent path lengths. After this, a fast re-optimization of the spot intensities is performed using a fast quadratic solver, which minimizes residual dose differences between the prior plan and the restored plan. The developed method was tested on 80 repeat CT scans of 10 prostate cancer patients.

1.4.2 What is the benefit of using a prior plan as a warm-start for full plan adaptation instead of applying dose restoration?

The simple dose restoration described in Chapter 2 restores a prior treatment plan by accounting for density variations along the pencil-beam paths while ignoring changes in the shape and location of the organs and targets. The restoration method consequently does not allow the restored treatment plan to yield a 'better' plan than the original, or even to restore tumor dose in cases with high anatomical variations. In Chapter 3 we investigated whether it would be possible to take the complete daily anatomy into account by expanding the dose restoration method, and what the ad-

ded value of this would be. Taking the simple restoration strategy as a starting point, Chapter 3 describes an automated adaptation strategy which accounts for density variations as well as changes in organ shapes and locations. After adapting the energies of the prior plan as in Chapter 2, this adaptation method continues to add new spots to the optimization and performs a fast multi-criteria optimization using the Reference Point Method (RPM). The result is a Pareto optimal treatment plan for the daily anatomy, with the same trade-offs as were made in the prior treatment plan. The developed method was tested on 88 repeat CT scans of 11 prostate cancer patients. Results were amongst others compared to the restoration method of Chapter 2 to determine the added value of this full adaptation.

1.4.3 Can the addition of a plan-library improve the automated adaptation method for tumor sites showing large day-to-day variations?

Large day-to-day variations in shape and location of the target are seen in locally advanced cervical cancer. In photon beam radiotherapy, the use of a patient-specific plan-library has been clinically introduced for this reason. Prior to each fraction, a daily image is used to select the best fitting treatment plan from the plan-library. For IMPT however, it has been shown that this approach is not always sufficient [11]. In Chapter 4, a patient-specific plan-library was therefore combined with the automated adaptation method from Chapter 3, allowing for a prior plan to be selected from the plan-library at each fraction. To investigate the added value of the plan-library in the plan adaptation, the results were compared to applying the adaptation method with a single prior plan for all fractions. The comparison was done using the repeat CT scans of six cervical cancer patients.

1.4.4 What is the benefit of online-adaptive IMPT compared to robust treatment planning?

In Chapters 2 – 4 the feasibility of the simple dose restoration method and the full plan adaptation method has been demonstrated. In Chapter 5 we investigated the gain of these methods compared to a non-adaptive robust treatment planning approach. To this end, we first derived the robustness settings and safety margins required to obtain adequate target coverage in the repeat CT scans of the prostate dataset. We then adapted non-robust prior plans using the simple dose restoration method from Chapter 2 and the full adaptation method from Chapter 3. For each fraction the adapted treatment plans were compared to the recomputed robust treatment plans in terms of target coverage and OAR sparing.

1.4.5 What is the effect of varying the parameter settings in the online-adaptive methods?

The simple dose restoration method and the full plan adaptation method both require a prior treatment plan to start the adaptation. The amount of motion mitigation that is included in the prior plan can affect the results of the adaptation. Full plan adaptation furthermore requires several parameter settings, namely the sample size describing how many spots are to be added in each iteration, the stopping criteria describing how many iterations are to be performed and the optimization approach to be used for the spot-intensity optimization. In the previous chapters choices regarding these settings have been made, but the full effect of these choices has not yet been investigated. In Chapter 6 we therefore systematically varied the parameter settings to evaluate the effect on the output of the two adaptive methods. The effect of all variations was evaluated in terms of dosimetric results and calculation times on 88 repeat CT scans of 11 prostate cancer patients.

1.4.6 Can a combination of deep-learning and image registration be applied for contour propagation for daily CT scans?

To run the full plan adaptation method contours of the daily CT scans are required. The first step of the proposed online-adaptive workflow is thus to obtain contours of the daily CT scan. In Chapter 7 we describe an automated contour propagation approach which combines deformable image registration (DIR) with deep-learning. The approach starts with automatically segmenting the bladder of a daily CT scan using a deep-learning network. In the second step, possible gas pockets in the rectum and intestines are detected and inpainted (filled) with a realistic content using a Generative Adversarial Network (GAN). Finally, using the corrected image and the obtained bladder segmentation, DIR is applied to propagate the manual contours of the planning CT scan onto the daily CT scan. The method was trained and evaluated on CT scans of prostate cancer patients. Evaluation was done both geometrically and dosimetrically.

1.4.7 Can prompt gamma ray emission profiles be used to monitor dosimetric changes with respect to the planned dose distribution during IMPT?

The final step of the proposed online-adaptive workflow is to monitor treatment delivery using prompt gamma (PG) ray profiles as part of treatment delivery QA. PG rays result from nuclear interactions between the incoming protons and the patients' tissue, and emission profiles can be measured outside the patient. PG ray emission profiles have been shown to correlate with the depth-dose profile of the primary pro-

ton beam [12]. In Chapter 8 we investigated whether PG ray emission profiles can be used to detect changes in target coverage due to inter-fraction anatomical variations with respect to the planning CT scan. Using Monte Carlo, dose delivery on repeat CT scans of prostate cancer patients was simulated and PG ray emission profiles were obtained. Correlations were evaluated between the observed dosimetric changes and the changes in PG emission profiles.

Chapter

2

Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations in prostate cancer

Physics in Medicine and Biology, Vol. 62, pp 4254-4272, 2017,

doi: <https://dx.doi.org/10.1088/1361-6560/aa5c12>

Thyrza Z. Jagt¹
Sebastiaan Breedveld¹
Steven van de Water¹
Ben J.M. Heijmen¹
Mischa S. Hoogeman^{1,2}

¹ Department of Radiation Oncology, Erasmus MC Cancer Institute,
Rotterdam, The Netherlands

² Department of Medical Physics & Informatics, HollandPTC,
Delft, The Netherlands

Abstract

Purpose: Proton therapy is very sensitive to daily density changes along the pencil-beam paths. The purpose of this study is to develop and evaluate an automated method for adaptation of IMPT plans to compensate for these daily tissue density variations.

Methods and materials: A two-step restoration method for 'densities-of-the-day' was created: (1) restoration of spot positions (Bragg peaks) by adapting the energy of each pencil-beam to the new water equivalent path length; and (2) re-optimization of pencil-beam weights by minimizing the dosimetric difference with the planned dose distribution, using a fast and exact quadratic solver. The method was developed and evaluated using 8 – 10 repeat CT scans of 10 prostate cancer patients.

Results: Experiments demonstrated that giving a high weight to the PTV in the re-optimization resulted in clinically acceptable restorations. For all scans we obtained $V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$. For the bladder, the differences between the restored and the intended treatment plan were below +2 Gy and +2%-point. The rectum differences were below +2 Gy and +2%-point for 90% of the scans. In the remaining scans the rectum was filled with air, which partly overlapped with the PTV. The air cavity distorted the Bragg peak resulting in less favorable rectum doses.

2.1 INTRODUCTION

Intensity-modulated proton therapy (IMPT) allows for highly localized dose delivery, but is also sensitive to inter-fraction variations in the location of the Bragg peak [1, 2]. Such variations can be induced by variations in the tissue density along the pencil-beam path for example caused by changes in organ filling or by relative movements of organs, and may cause large discrepancies between the planned and delivered dose distribution [13]. A strategy to prevent this passively is the generation of robust treatment plans [14]. This strategy can, however, lead to increased doses to organs at risk (OARs) [10]. Ideally, adapting for the variations at time of treatment should be sub-minute after imaging. This is currently not feasible using a normal treatment planning workflow, where a full treatment plan is generated from scratch.

This study is part of a project which aims to reduce the time for a re-optimization of the treatment plan by greatly simplifying the optimization problem. The approach is to create in the treatment planning phase an individualized library-of-plans for possible patient anatomies capturing relatively large inter-fraction organ motion. The library can be derived from the patients' planning CT scans or a population-based statistical model describing anatomical variations [15, 16]. Just prior to each treatment fraction and based on in-room volumetric imaging, the treatment plan that best fits the anatomy-of-the day will be selected for delivery. Density changes along the beam paths will in general still occur, and need to be corrected, which is the topic of this study. Because generating fully optimized treatment plans for prostate cancer patients as described below takes on average about 25 min, a full optimization is not possible. Therefore we focus on correction for density changes only. The aim of this study is to develop and evaluate a re-optimization method that quickly and automatically restores a proton therapy dose distribution that has been distorted by density changes along the path of the beams. Applying this restoration method right before treatment is a step towards online-adaptive IMPT. The use of the restoration method is not exclusive to library-of-plans strategies but can also be applied to a static treatment plan to mitigate the impact of daily density changes. In this latter case, it would be assumed that (moderate) inter-fraction organ motion is accounted for by a margin around the clinical target volume (CTV).

Zhang *et al.* [17] also investigated a procedure to restore the planned dose to the prostate. In this procedure, the energy of every proton beam is adjusted according to the new water equivalent path length (WEPL) calculated from the daily CT scan. The intensities of the proton pencil-beams remained as planned. The method was tested on a phantom prostate patient, for which they assumed that the prostate would only shift rigidly inside the phantom from fraction to fraction. Two treatment techniques

were evaluated; one using distal edge tracking (DET), a type of IMPT placing spots only at the distal edge of the target, and one using 3D IMPT (which in this paper is abbreviated to IMPT), where spots are placed in the whole target volume. Effectiveness of this method was evaluated by comparing the adapted and non-adapted dose distributions for both treatment types. Good restorations were achieved for the DET plans, but the method did not work for IMPT plans, which is currently considered as the state-of-the-art treatment technique.

In our restoration method we also start with WEPL correction, but we proceed with a re-optimization of the pencil-beam weights. Evaluations were performed using CT scans of 10 patients. Four re-optimization methods have been compared to find the one resulting in the best restorations. For all patients we checked whether these restorations indeed resulted in clinically acceptable dose distributions and recorded the re-optimization time.

2.2 METHODS AND MATERIALS

Patient data

For the 10 study patients we had a planning CT scan taken with contrast and 8-10 repeat CT scans without contrast available, which were acquired during the course of a fractionated photon radiotherapy treatment. To avoid that the results would be perturbed by artificial density changes caused by the contrast we ignored the planning CT scan and used the first repeat CT scan as planning CT scan in this study. A total of 80 repeat scans were used. In each scan, the prostate, seminal vesicles, and lymph nodes were delineated as target structures. The delineated OARs in the planning CT were the rectum, bladder, small and large intestines, and the femoral heads.

Treatment planning

Dose was prescribed according to a simultaneously integrated boost scheme in which the high-dose PTV (prostate + 4 mm margin) was assigned 74 Gy and the low-dose PTV (lymph nodes and seminal vesicles + 7 mm margin) 55 Gy, to be delivered using two laterally opposed beams. We selected this treatment group for evaluation as a theoretical benefit of proton therapy has been demonstrated for the treatment of larger volumes associated with advanced-stage disease [18]. Note that in this study we used for each patient, instead of a library-of-plans, a static treatment plan with tight CTV-to-PTV margins that were supposed to account for inter-fraction geometrical errors due to internal organ motion, but not to account for density changes along the paths of the pencil-beams. The latter will be accounted for by the dose restoration

method proposed in this study. The PTV-intermediate is a 15 mm transition region between the expanded prostate and the expanded lymph nodes and seminal vesicles and was added to obtain the desired dose fall-off. The PTV-low consists of the expanded seminal vesicles and lymph nodes, excluding the transition region. To achieve the dose fall-off outside the target areas, conformity rings were created (see Figure 2.1).

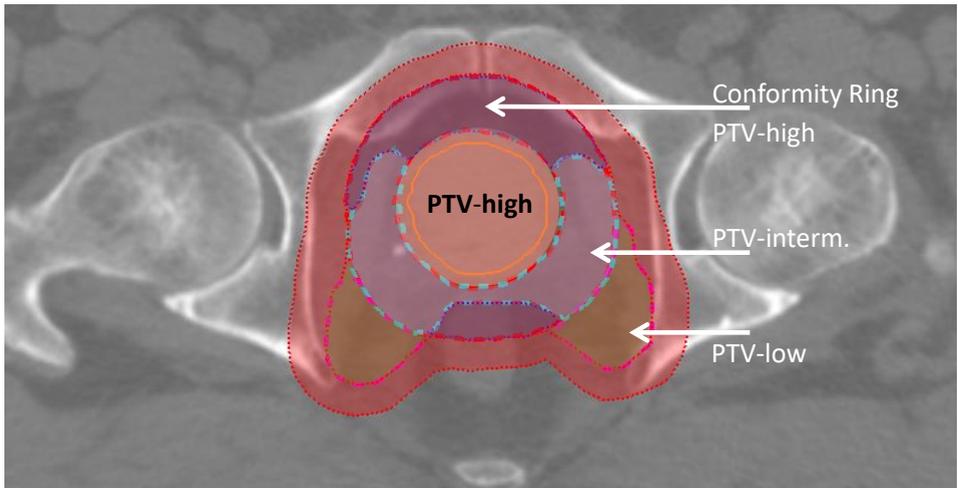


Figure 2.1: The PTV-high is an expansion of the prostate. The PTV-intermediate is a 15 mm transition region between the high-dose and low-dose PTV. The PTV-low consists of the expanded seminal vesicles and lymph nodes, excluding the transition region. The PTV-full consists of the PTV-high with a 15 mm expansion and PTV-low. The conformity ring around the PTV-high is the PTV-full excluding the PTV-low. The red area represents the 0 – 10 mm conformity ring of PTV-full.

All IMPT plans were generated using ‘Erasmus-iCycle’, our in-house developed treatment planning system for fully automated plan generation [19, 20], combined with the ‘Astroid’ dose engine [21]. Erasmus-iCycle uses a multi-criteria optimization to generate a clinically desirable Pareto optimal treatment plan on the basis of a wishlist consisting of hard constraints and objectives (see Table 2.1). This wishlist is created by physicians and is often used for the entire patient group (i.e. all prostate cancer patients). Constraints are never violated in the plan generation. Based on their assigned priorities, the objective functions are minimized sequentially. The achieved objective value is set as an additional hard constraint that has to be respected during the minimization of the lower priority objective functions (lexicographic optimization). More details on Erasmus-iCycle can be found in [19, 20, 22–24]. The wishlist used for plan generation in this study is shown in Table 2.1, combined with Figure 2.1. More details about the use of a wishlist are given in [22]. Generating treatment

plans using Erasmus-iCycle with this wishlist takes on average about 25 min.

To investigate the performance of the dose restoration method developed in this study, the intended treatment plans were generated without including patient setup and range robustness in the optimization. If the restoration method works well, the degree of robustness included in the treatment plan can be reduced, as the coverage loss due to density changes can be mitigated by our method of dose restoration. Reducing the degree of robustness, is expected to reduce the dose in OARs [10].

Table 2.1: The 'wishlist' with planning constraints and objectives used for automated IMPT plan generation. Constraints will always be met. The priority numbers of the objectives indicate the order in which objectives are to be optimized. A low number corresponds to a high priority. The PTV structures are shown in Figure 2.1. The objectives with priorities 4 – 8 were assigned a limit of 1 Gy in order to obtain very low dose values while at the same time not imposing an impossible goal.

Constraints	Structure	Type	Limit
	PTV-high	Minimum	0.97×74 Gy
	PTV-intermediate	Minimum	0.99×74 Gy
	PTV-low	Minimum	0.99×55 Gy
Objectives	Structure	Type	Limit
1	PTV-high	Maximum	1.07×74 Gy
1	PTV-intermediate	Maximum	1.07×74 Gy
1	PTV-low	Maximum	1.07×55 Gy
2	Conformity ring PTV-high	Maximum	1.07×74 Gy
2	Conformity ring PTV-full 0 – 10 mm	Maximum	1.07×55 Gy
2	Conformity ring PTV-full 10 – 15 mm	Maximum	0.90×55 Gy
3	Femoral heads	Maximum	50 Gy
4	Rectum	Mean	1 Gy
5	Small and large intestines	Mean	1 Gy
6	Bladder	Mean	1 Gy
7	Femoral heads	Mean	1 Gy
8	All conformity rings	Mean	1 Gy
8	All conformity rings	Maximum	1 Gy
9	Total spot-weight	Sum	1 Gp

Abbreviations: PTV = planning target volume; Gp = Gigaprotons

Dose restoration

The proposed restoration method assumes that a repeat CT scan acquired just prior to dose delivery is available and that the prostate is aligned to the treatment beams by a couch translation using implanted intra-prostatic markers. The restoration method takes this image-guidance procedure into account by aligning each repeat CT scan

to the planning CT using the implanted markers. Furthermore, we take as a starting point that repeat CT scans do not have automatically or manually delineated contours, meaning that only the structures projected from the planning CT to the repeat CT scans, i.e. the projected planning structures, are available for the re-optimization (see Figure 2.2, middle).



Figure 2.2: This study uses three combinations of CT scans and contour sets. Left: a planning CT scan with structures contoured in the planning CT scan. Middle: a repeat CT scan with structures projected from the planning CT to the repeat CT scan. Right: a repeat CT scan with structures contoured in the repeat CT scan.

The proposed restoration method restores the dose for all voxels of the structures that are used in the full optimization and hence are mentioned in the wishlist (see Table 2.1). For these voxels the dose deposition matrices are required as these matrices hold the dosimetric effect of every pencil-beam to every selected voxel. Multiplied with the pencil-beam weights this obtains the dose distribution in these selected voxels. For the planning CT with the planning contours, the matrices are already initialized due to the full optimization, using the energies chosen during optimization. As the dose deposition matrices depend on the path towards the voxels and the energies of the pencil-beams, they need to be recalculated for the new paths based on the repeat CT scan with the projected planning contours. When the pencil-beam energies are changed during restoration, the matrices need to be recalculated once more. Voxels of structures that are not included in the wishlist, i.e. which are not used in the full optimization, are therefore not included in the restoration to limit the calculation time. The order of importance of the structures of the planning CT scan, i.e. the planning structures, can be used to adjust the weight or importance factor of specific voxels in the re-optimization in order to improve the results. The advantage of this methods is that it does not require a time-consuming contouring step and can immediately commence after the alignment of the repeat CT scan to the planning CT scan. The definition of these importance factors is given in the next section.

The proposed restoration method consists of two steps. In the first step the spot positions (Bragg peaks) are restored by adapting the energy of each pencil-beam such

that the coordinates of the Bragg peaks in the planning CT scan and in the repeat scan are equal. Pencil-beam directions remain unchanged. Figure 2.3 shows a schematic representation of this procedure. The result of this step is the energy-restored dose distribution.

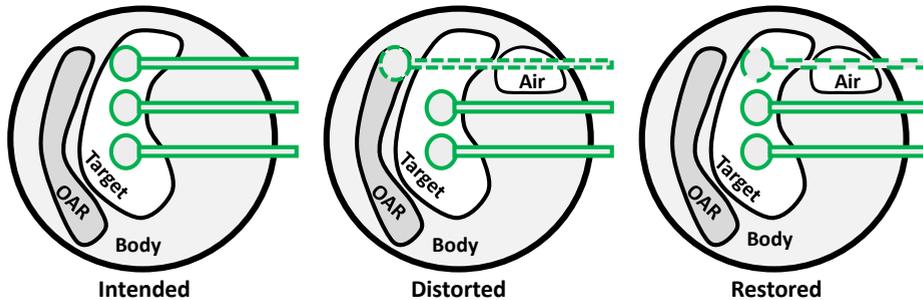


Figure 2.3: Restoring spot positions. Left: the spot positions as intended. Middle: an air cavity causes a displacement and deformation of the upper spot. Right: The energy of the pencil-beam has been adapted to restore the spot position. Note that the restoration of spot positions does not adapt for changes in shape and location of the target. If the target shows large geometric changes, the energy-restored spots will not necessarily end up in the target.

Restoring the Bragg peak changes its intensity and shape. Hereto we require to re-optimize the pencil-beam weights to match the intended dose as much as possible. The change in shape depends on the structures and air cavities along the pencil-beam path. Figure 2.4 illustrates the change in shape when a pencil-beam moves fully (middle) or partly (bottom) through an air cavity. We will refer to the changed Bragg peaks as distorted Bragg peaks.

Due to the changes in intensity and shape due to the energy-restoration, the dose deposition matrices need to be recalculated prior to the pencil-beam re-optimization. Instead of a full multi-criteria optimization as used for generation of the intended dose distributions with Erasmus-iCycle, the differences between the actual and intended dose distribution are used to define a quadratic objective function. This objective function contains all structures that are included as constraints or objectives in the wishlist. This re-optimization method uses the BOXCQP algorithm [25, 26].

The quadratic objective function is given by

$$s(\mathbf{f}) = (\mathbf{A}\mathbf{f} - \mathbf{d}^{\text{int}})^T W (\mathbf{A}\mathbf{f} - \mathbf{d}^{\text{int}}) + \kappa S. \quad (2.1)$$

Here $\mathbf{A}\mathbf{f}$ is the actual dose, calculated as the product of the dose deposition matrix \mathbf{A} and the spot-weight vector \mathbf{f} . At the start of a pencil-beam weight re-optimization, $\mathbf{A}\mathbf{f}$ is the energy-restored dose (see above), \mathbf{d}^{int} is the intended dose. W is a diagonal

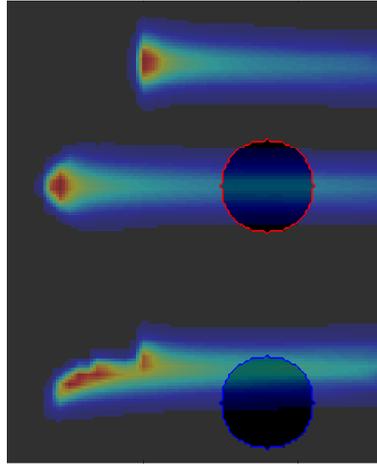


Figure 2.4: When a pencil-beam moves through an air cavity, the shape of the Bragg peak changes.

matrix containing importance factors of the voxels (see below). κS is a smoothing term that is also further explained below.

The quadratic objective function can be written in canonical form as

$$s(\mathbf{f}) = \frac{1}{2} \mathbf{f}^T H \mathbf{f} + \mathbf{f}^T \mathbf{b} + c, \quad (2.2)$$

where

$$H = A^T W A + \kappa S, \quad \mathbf{b} = A^T \mathbf{q}, \quad c = \frac{1}{2} (\mathbf{d}^{\text{int}})^T (W \mathbf{d}^{\text{int}}),$$

$$W = \sum_v \eta_v, \quad \mathbf{q} = -W \mathbf{d}^{\text{int}}.$$

More information on these equations can be found in [26]. The smoothing term κS was introduced to keep the Hessian H positive definite at all times. Without this term the Hessian is not positive definite when for instance the same dose can be achieved in two different ways. This can happen if two similar proton spots are included in the treatment plan. A simple approach which changes the solution minimally is to take κ small ($\mathcal{O}(10^{-4})$) and S the identity matrix.

The BOXCQP algorithm searches for the optimal spot-weight vector \mathbf{f} by minimization of the function $s(\mathbf{f})$.

Assignment of voxel importance factors in $s(\mathbf{f})$ (Equations 2.1 and 2.2)

Four different approaches for assigning importance factors to the voxels were evaluated. Table 2.2 contains the details of the different approaches. Approaches B –

D could be applied with 1 – 5 iterations (denoted as B₁ – B₅, C₁ – C₅ and D₁ – D₅), yielding a total of sixteen different re-optimization methods.

Table 2.2: Overview of the investigated approaches for assignment of voxel importance factors.

Method A (1 iteration)	All voxels in the structures in the wishlist (Table 2.1) have importance factor 1 throughout the re-optimization; W is the identity matrix.
Method B (1 – 5 iterations)	In the first iteration all voxels in the targets have importance factor 1000. The other voxels have factor 1. In each subsequent iteration the dose distribution is evaluated. Target voxels receiving either too little or too much dose, i.e. less than 95% or more than 107% of the prescribed dose, will get their factor doubled.
Method C (1 – 5 iterations)	All voxels in the targets have importance factor 1000. In the remaining structures the dose is evaluated. All voxels in the structure receiving the highest dose get a factor 500. In every iteration the next structure receiving the highest dose also gets factor 500. Each structure can only be selected once.
Method D (1 – 5 iterations)	In every iteration the difference between the intended dose and the actual dose is determined for each structure. The structure with the highest mean difference will get a factor 1000 for every voxel. In every iteration a new structure with factor 1000 is added.

Evaluation and comparison of intended and restored plans

All intended and restored treatment plans were evaluated by visual inspection of the dose distributions, the DVHs of the target volumes and OARs, and the clinical constraints. Visual inspection of the restored dose distribution was used to check for hotspots. For the PTV and CTV structures (see Table 2.1), we report the $V_{95\%}$ and $V_{107\%}$. The rectum was evaluated using the D_{mean} , $V_{45 \text{ Gy}}$, $V_{60 \text{ Gy}}$ and $V_{75 \text{ Gy}}$, and the bladder using the D_{mean} , $V_{45 \text{ Gy}}$ and $V_{65 \text{ Gy}}$.

For evaluation of the restored dose distribution, both the projected planning structures and the actual repeat structures were used. First we evaluated the restored dose distribution on the projected PTV and the actual repeat CTV. Secondly, the restored dose distribution was evaluated on the projected OARs and the actual repeat OARs. Besides the evaluation of the obtained treatment plans, the calculation times of the restoration methods were compared.

2.3 RESULTS

Distortion for the projected planning structures due to density changes

Figure 2.5 shows boxplots depicting the differences of the distorted dose distribution in the repeat CT scans minus the intended dose distribution of the planning CT scans for all 80 repeat CT scans. The differences are calculated for the projected planning structures showing the starting point for dose restoration. For all scans the target coverage deteriorates due to the density changes, whereas the OARs dose remain similar on average.

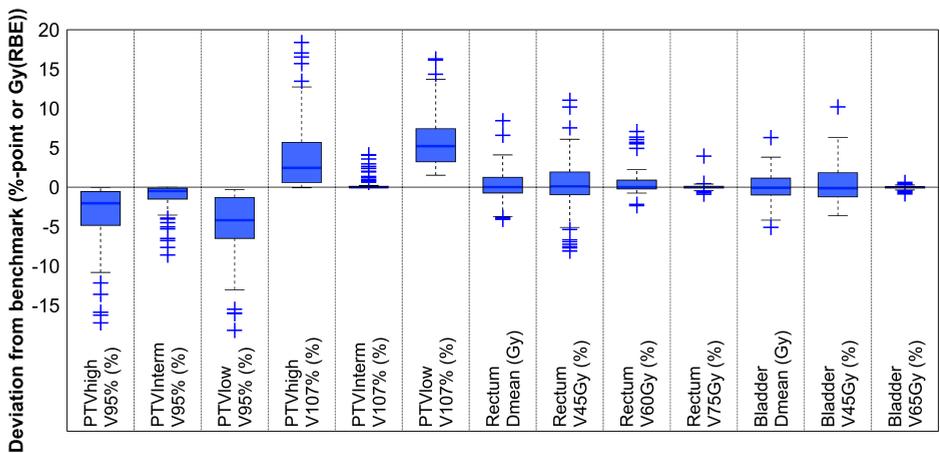


Figure 2.5: Boxplots depicting the difference in dosimetric parameters of the distorted minus the intended dose distributions for all 80 scans based on the projected planning structures. Each boxplot indicates the median and the 25th and 75th percentiles of the obtained differences. The dashed lines depict the remaining differences which are not outliers. Values are defined outliers if they are more than 1.5 times the distance between the 25th and 75th quartiles away from the quartiles. The plus marks indicate the outliers.

Results for projected planning PTV of all restoration methods

The intended treatment plans were optimized to ensure that at least 98% of the volume of the PTV structures given in the wishlist (Table 2.1 and 2.1) receives 95% of the prescribed dose and no more than 2% of the volume receives more than 107%. For the result of the restoration method to be clinically acceptable, we required that these objectives should still be met for 98% of the scans (i.e. for 98% of the scans $V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$). Table 2.3 shows the percentage of the scans for which $V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$ for the intended and distorted treatment plans as well as for each restoration method.

It can be seen that only in methods B₄ and B₅ (in which a higher importance factor is given to certain voxels in the targets) the constraints are met for all PTV structures for at least 98% of the scans. Method B₅ shows the best results and is therefore preferred at this point. It can be seen that for methods B₁ – B₂ and C₁ – C₅ (in which also the OARs get a higher importance factor) most constraints are met for at least 95% of the scans, but this is not the case for the V_{107%} of the low-dose PTV region. The results of methods A and D (with respectively no higher importance factors and higher importance factors for the voxels with the highest difference from the intended dose) meet the requirements for only a few patients, which means that these methods will be neglected in further analyses.

Results for actual repeat CTV structures of restoration methods B and C

Table 2.4 shows the results of the restoration methods B and C for the actual repeat CTV structures. Note that the restoration was done based on all voxels of the projected PTV and OAR structures. We required that for the CTV_{prostate} at least 98% of the volume obtains at least 95% of 74 Gy, and at most 2% of the volume receives 107% of 74 Gy. The CTV_{lymph nodes} and CTV_{seminal vesicles} both fall into the PTV-intermediate and the PTV-low (Figure 2.1). They should therefore receive at least 95% of 55 Gy and no more than 107% of 74 Gy.

To verify that the margin applied was sufficient to account for shape and position variations of the CTV structures, we first measured these parameters for the actual repeat structures without recalculating the dose distributions (without distortion due to density changes). For all scans the objectives are met, showing that the margins are indeed sufficient to account for the shape and position variations.

In Table 2.4, the percentages of repeat CT scans that meet the target constraints before and after restoration are listed. It shows that as expected target coverage was compromised due to density changes in the repeat CT scans (Distorted). It can be seen that starting restoration methods B or C from a static treatment plan with CTV-PTV margins results in a sufficient and acceptable target coverage for over 92.5% of the scans. When looking at the V_{107%}, methods B give better results than methods C, where method B₅ obtains the best results. Both methods B and C however obtain acceptable results when compared to the distorted results.

Results for projected planning OAR structures of restoration methods B and C

The results of restoration methods B₁ – C₅ for the projected planning OAR structures are shown in Figures 2.6 and 2.7. The results can be compared with Figure 2.5, which is showing the results of the distorted dose distribution for the projected contours.

Table 2.3: Percentages of the 80 dose distributions that meet the target constraints for the investigated restoration methods based on the projected planning structures.

	V _{95%} ≥ 98% PTV-high	V _{95%} ≥ 98% PTV- interm.	V _{95%} ≥ 98% PTV-low	V _{107%} ≤ 2% PTV-high	V _{107%} ≤ 2% PTV- interm.	V _{107%} ≤ 2% PTV-low
Intended	100	100	100	100	100	100
Distorted	45.0	77.5	31.3	45.0	91.3	0.0
Energy- restored	77.5	95.0	58.8	33.8	95.0	0.0
A	100	98.8	93.8	90.0	100	6.3
B ₁	100	100	100	98.8	100	58.8
B ₂	100	100	100	98.8	100	83.8
B ₃	100	100	100	100	100	96.3
B ₄	100	100	100	100	100	98.8
B ₅	100	100	100	100	100	100
C ₁	100	100	100	96.3	100	53.8
C ₂	100	100	100	95.0	100	46.3
C ₃	100	100	100	95.0	100	38.8
C ₄	100	100	100	95.0	100	26.3
C ₅	100	100	96.3	95.0	100	26.3
D ₁	83.8	93.8	36.3	65.0	100	0.0
D ₂	78.8	92.5	17.5	50.0	100	1.3
D ₃	78.8	93.8	10.0	53.8	97.5	1.3
D ₄	83.8	97.5	10.0	55.0	100	1.3
D ₅	85.0	98.8	16.3	60.0	100	0.0

Figure 2.6 shows that for the rectum the differences of the restored minus intended dose distributions are very small for most patients, with a total of 21 outliers for method B₅. Only 14 of these values were positive, meaning that the intended dosimetric parameter value was lower and hence better than the restored value. Using method C₅ decreased the total number of outliers to 11, of which only 9 were positive. Figure 2.6 shows that even though defined as outliers, some of these difference values are still very low. When using method B₅ only 8 scans show differences larger than or equal to +2 Gy for the D_{mean} and +2%-point for the V_{45 Gy}, V_{60 Gy} and V_{75 Gy}. Using method C₅ none of the scans obtain difference values larger than +2 Gy and +2%-point.

Figure 2.7 shows that on average the differences for all dosimetric parameters of the bladder are larger for the results of methods B. However, for both methods B and C the differences for the bladder are very small. Most scans differ less than +1 Gy for the D_{mean}, and +1%-point for the V_{45 Gy} and V_{60 Gy}. The outliers reach maximum differences of approximately +1.6 Gy and +1.6%-point.

Table 2.4: Percentages of the 80 dose distributions that meet the target constraints for the investigated restoration methods based on the actual repeat structures.

	$V_{95\%} \geq 98\%$ CTV _{prost}	$V_{95\%} \geq 98\%$ CTV _{ln}	$V_{95\%} \geq 98\%$ CTV _{sv}	$V_{107\%} \leq 2\%$ CTV _{prost}	$V_{107\%} \leq 2\%$ CTV _{ln}	$V_{107\%} \leq 2\%$ CTV _{sn}
Distorted	70.0	66.3	80.0	40.0	100	83.8
Energy-restored	87.5	92.5	90.0	35.0	100	95.0
B ₁	96.3	92.5	95.0	86.3	100	96.3
B ₂	96.3	92.5	95.0	87.5	100	96.3
B ₃	96.3	92.5	95.0	88.8	100	96.3
B ₄	96.3	92.5	95.0	92.5	100	96.3
B ₅	96.3	92.5	95.0	92.5	100	96.3
C ₁	96.3	92.5	96.3	81.3	100	96.3
C ₂	96.3	92.5	96.3	82.5	100	96.3
C ₃	96.3	92.5	96.3	83.8	100	96.3
C ₄	96.3	92.5	96.3	82.5	100	96.3
C ₅	96.3	92.5	96.3	81.3	100	96.3

Abbreviations: CTV_{prost} = CTV_{prostate}, CTV_{ln} = CTV_{lymph nodes}, CTV_{sv} = CTV_{seminal vesicles}

When comparing the results of methods B to the results of methods C, it can be seen that similar values are obtained. The largest differences are seen in the $V_{45 \text{ Gy}}$ of the rectum, where for method B₅ seven scans have a difference larger than +2%-point. Two of these scans also have a difference larger than 2 Gy for the D_{mean} . A scan of another patient has a difference larger than +2%-point for the $V_{60 \text{ Gy}}$ of the rectum. For these 8 scans, the PTV overlaps with a gas pocket in the rectum of the repeat. Closer inspection revealed that the air cavity distorted the Bragg peak resulting in less favorable rectum doses. To verify this assumption, we performed a full multi-criteria optimization using this repeat CT scan and the projected planning structures and compared the rectum dose values to the ones obtained from restoration. We found that in the fully optimized treatment plan the rectum D_{mean} was 28.1 Gy and the rectum $V_{45 \text{ Gy}}$ 31.9%. The values obtained from the restoration were 28.1 Gy and 32.1% for respectively the D_{mean} and the $V_{45 \text{ Gy}}$. As the differences between these values are very small we conclude that our assumption is correct and the high dose values are indeed caused by a distorted Bragg peak.

When using methods C, the rectum was restored better, but a larger part of the projected PTV received dose values higher than 107% of the prescribed dose (Table 2.3). It should be noted that the dose of the distorted Bragg peak was partly calculated inside the air cavity. Although this dose was contributing the rectum dose in the DVH calculation, in reality this dose will not be deposited in rectal tissue.

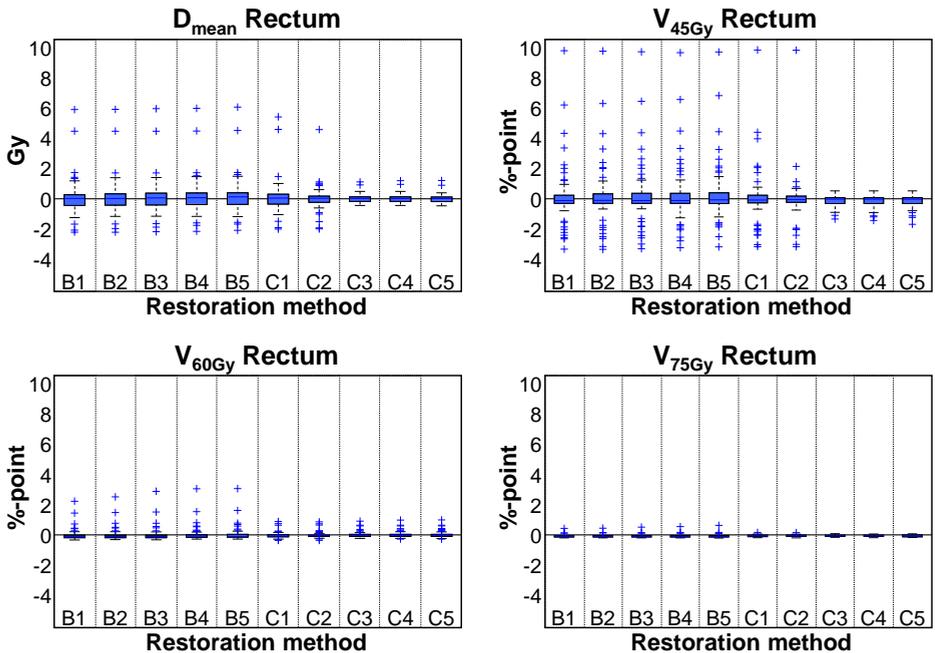


Figure 2.6: Boxplots showing differences of the restored minus the intended rectum dose parameters for all 80 scans for restoration methods B₁ – C₅, based on the projected planning structures. Positive values point at higher values for the restored dose distribution. Each boxplot indicates the median and the 25th and 75th percentiles of the obtained differences. The dashed lines depict the remaining differences which are not outliers. Values are defined outliers if they are more than 1.5 times the distance between the 25th and 75th quartiles away from the quartiles. The plus marks indicate the outliers.

Results for actual repeat OAR structures of restoration methods B and C

The results in Table 2.4 indicate that the target coverage can be effectively restored when evaluating on the actual repeat contours. Figure 2.8 shows the difference of the restored dose distributions for methods B₅ and C₅ minus the distorted dose distributions, i.e. without any restoration, for the actual repeat rectum (top) and bladder (bottom).

When looking at the results for the rectum (top Figure 2.8) it can be seen that both methods B₅ and C₅ have median differences close to zero when comparing to the distorted dose distribution. In all plots it can be seen that the largest outliers have negative difference values, meaning that for those scans the restored dose distribution obtains lower dose values in the rectum than the distorted dose distribution. When looking at the V_{75 Gy} after restoration, over 70% of the scans show differences equal to or smaller than zero. As can be seen in the boxplots, the remaining scans obtain very

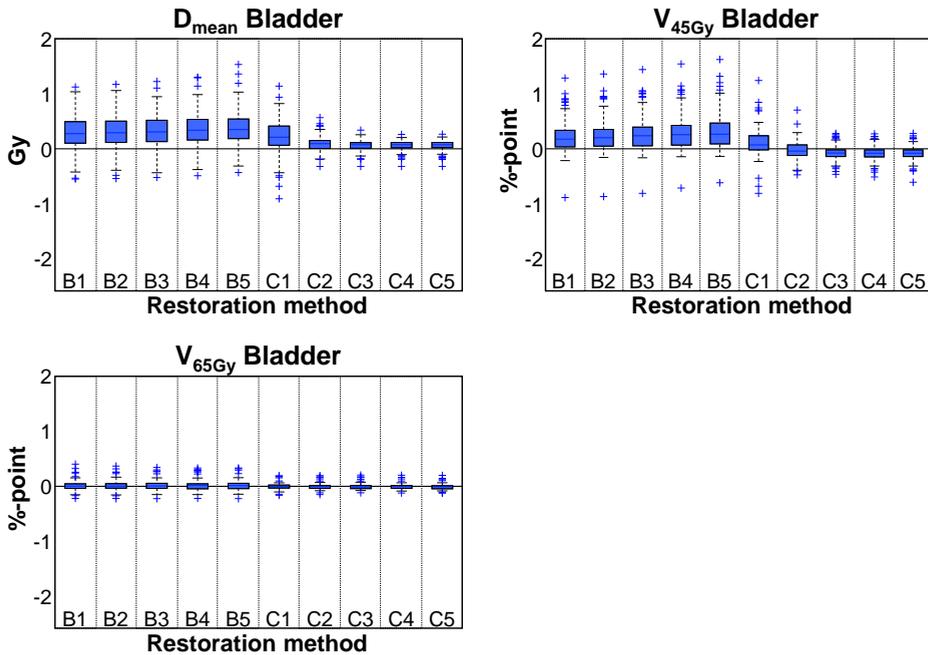


Figure 2.7: Boxplots showing differences of the restored minus the intended bladder dose parameters for all 80 scans for restoration methods B₁ – C₅, based on the projected planning structures. Positive values point at higher values for the restored dose distribution. Each boxplot indicates the median and the 25th and 75th percentiles of the obtained differences. The dashed lines depict the remaining differences which are not outliers. Values are defined outliers if they are more than 1.5 times the distance between the 25th and 75th quartiles away from the quartiles. The plus marks indicate the outliers.

low difference values. For the D_{mean} , $V_{60\text{ Gy}}$ and $V_{75\text{ Gy}}$ all difference values remain below +4 Gy and +4%-point. For the $V_{45\text{ Gy}}$ there are 7 and 6 scans in respectively method B₅ and C₅ with a difference value larger than +4%-point.

For the bladder (bottom Figure 2.8) we see that the differences between distorted and restored are very small for the $V_{65\text{ Gy}}$. For the D_{mean} and $V_{45\text{ Gy}}$ the differences are larger, though over 92% of the scans obtain difference values below +4 Gy and +4%-point.

Though the dosimetric parameter values of the OARs can for some scans increase after restoration, this loss remains smaller than the gain that is found for the target structures (Table 2.4).

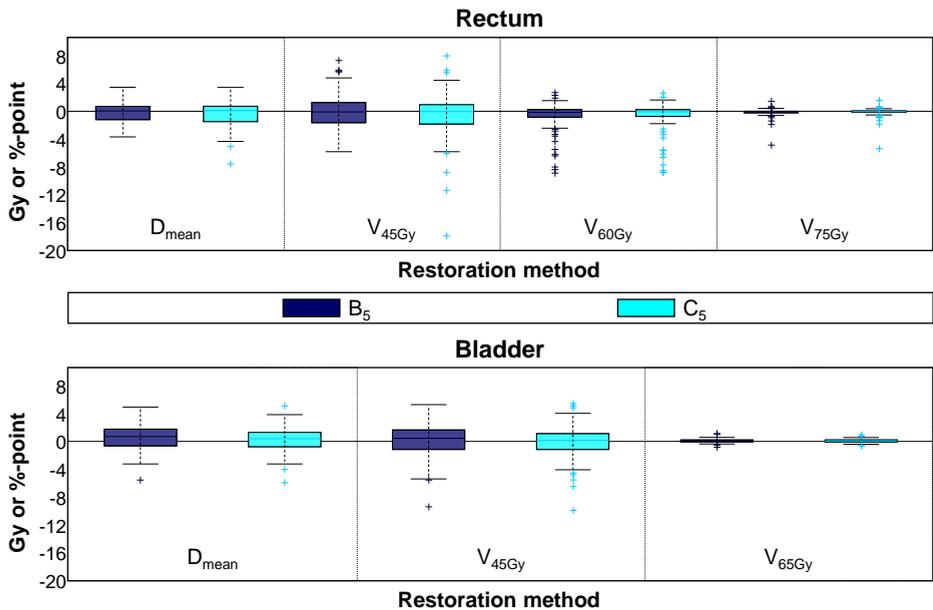


Figure 2.8: Boxplots showing differences of the restored minus the distorted dose parameters of the rectum (top) and bladder (bottom), for all 80 scans for restoration methods B₅ and C₅, based on the actual repeat structures. Positive values point at higher values for the restored dose distribution. Each boxplot indicates the median and the 25th and 75th percentiles of the obtained differences. The dashed lines depict the remaining differences which are not outliers. Values are defined outliers if they are more than 1.5 times the distance between the 25th and 75th quartiles away from the quartiles. The plus marks indicate the outliers.

Calculation times

The time needed for the energy adaptation, i.e. the restoration of the spot positions is independent of the restoration method and was on average 5.4 s (3.5 – 10.6). The re-optimization time includes the creation of the quadratic objective function, adapting the weight matrix and performing the minimization. Table 2.5 shows the re-optimization times for methods B₁ – B₅ and C₁ – C₅.

The mentioned calculation times do not include loading of the CT scans, the original plans and the dose calculations. The most time-consuming and limiting operation was the calculation of the dose deposition matrix A (mean 4.3 min (range 2.4 – 9.6)), which occurs once between the spot restoration (i.e. energy adaptation) and the weight re-optimization. Optimization of the dose calculation speed was not part of this study.

Table 2.5: Calculation times for the different B and C restoration methods. The mean is taken over all 80 scans.

	Weight re-optimization (seconds)	
	Mean	Range
B ₁	0.7	0.4 – 1.7
B ₂	1.5	0.9 – 3.6
B ₃	2.2	1.3 – 4.5
B ₄	3.1	1.8 – 5.9
B ₅	3.8	2.2 – 7.7
C ₁	0.9	0.4 – 2.5
C ₂	1.8	0.9 – 4.0
C ₃	2.7	1.3 – 5.6
C ₄	3.7	1.6 – 7.5
C ₅	4.5	2.3 – 9.0

2.4 DISCUSSION

In this study several re-optimization methods were compared, all aimed at restoring the dose distribution that was distorted due to density changes. All restoration methods were designed for near real-time performance enabling online-adaptive proton therapy. The goal of the restoration was to get dose distributions as close as possible to the intended dose distributions in the structures used for treatment planning. We found that the restoration method that best restores the dose in the target structures is B₅, which focuses on the target voxels. In every iteration, the target voxels that received either too much or too little dose were given a higher importance factor in the re-optimization. Using this method, all 80 scans had a restored dose distribution with a $V_{95\%} \geq 98\%$ and a $V_{107\%} \leq 2\%$ for the projected PTV structures used in the wish-list (Table 2.1). When using method B₅ the dosimetric parameters of the projected planning OARs showed on average very small differences from the intended values ($\leq +1$ Gy and $\leq +1\%$ -point). Eight outliers were found with differences larger than $+2$ Gy and $+2\%$ -point. These outliers can all be explained by an air cavity partly overlapping the PTV. The air cavity negatively affected the shape of the Bragg peak (see Figure 2.4), leading to a higher dose to the rectum after the restoration of the distorted dose distribution. For the worst outlier we generated a fully optimized treatment plan based on the repeat CT scan. The full optimization did not improve the dose to the rectum compared to the restoration, suggesting that the worsened rectum dose is due to the changed properties of the Bragg peak and that the restoration is close to the optimal result, i.e. a full re-optimization. An advantage of our method is that it can be applied using only the structures as contoured in the planning CT which means as soon as the daily CT scan has been aligned to the planning CT scan the restoration

can start.

Besides method B₅ method C₅ also performed well. Although slightly better results for the target structures were obtained with methods B, methods C achieved lower dose values to the OARs. Using the methods C, the target coverage was slightly compromised, obtaining $V_{95\%}$ values of less than 98% and $V_{107\%}$ values of more than 2%. For example, for the projected PTV-low 67 scans had a $V_{107\%}$ higher than 2% when using restoration method C₅. However, the $V_{107\%}$ was limited to 5.7%, which still may be considered clinically acceptable. As shown in Table 2.5, the calculation time is similar for both methods. When comparing the use of 1 iteration with the use of more iterations, we found that the increased computation time using more iterations is negligible. However, using more iterations obtained fewer hotspots in the targets when using method B (as shown in Table 2.3) and lower dose values to the OARs when using method C (Figures 2.6 and 2.7).

In addition to the projected planning PTV and OAR structures, we also evaluated the restored dose distributions for the actual CTV and OAR structures in the repeat CT scans. We found that for the coverage of the CTV structures of the prostate, lymph nodes and seminal vesicles, the number of patients receiving acceptable $V_{95\%}$ and $V_{107\%}$ values for the targets, increased when applying a restoration method (Table 2.4). Note that the intended treatment plan on the planning CT was optimized on a PTV volume, i.e. the actual target expanded by a margin, already anticipating some changes in shape and location. Performing a restoration on this PTV allowed for the CTV_{prostate} to be sufficiently irradiated at each treatment day without having to include robustness in the optimization a priori. Similar as to the evaluation on the PTV, the best results were obtained when using method B₅. For the OARs we compared the results of the distorted dose distribution, i.e. without restoration to the results of the dose distribution obtained with methods B₅ and C₅ (Figure 2.8). We found that the volumes receiving a high dose were reduced a little, and only small differences were found in the mean dose of the organs. Overall we can conclude that performing the restoration has no negative effect on the dosimetric values of the OARs.

Taken together, our findings prove the principle that clinically acceptable restorations for density changes can be obtained for prostate cancer patients within 10 seconds, when excluding the calculation of the dose deposition matrices. The calculation of these matrices currently takes several minutes. We believe that with some improvements of the dose engine this calculation time can be significantly reduced.

Though many more re-optimization methods are possible, as well as methods of updating the weight matrix W between iterations (see Equation (2.1)), only four main methods (A – D) were considered during this investigation. Method A was selected

to see the effect of minimal effort; by not using any importance factors, the method is very general and very fast. Methods B and C, in which we focus on the targets and OARs, were selected on the basic principle that in treatment planning the goal is always to get a high dose in the targets and a low dose in the OARs. Method D, in which higher importance factors are given to structures with higher differences, was inspired by our re-optimization method which aims to minimize the differences between the intended and the achieved dose.

In this study we analyzed the mean dose to OARs. To test whether the method also works for more serially responding organs, we applied a maximum dose objective to the rectum in the generation of the intended treatment plan. The results of applying restoration methods A, B₅, C₅ and D₅ on these cases were similar to the results of the previously discussed prostate cancer cases. Methods A and D yielded coldspots and hotspots, while methods B₅ and C₅ obtained acceptable results. Other approaches have not been investigated, as clinically sufficient results were already obtained using the methods developed and evaluated in this study. However, it is possible that for other treatment sites other restoration methods are more suitable.

Looking at the results of the restorations the difficulty seems to be in the restoration of the PTV-low as projected on the repeat CT scan. A possible explanation is that less degrees of freedom are available for the optimizer to restore the dose distribution as the dose to each lymph node is mostly delivered by one of the beams.

The developed restoration method aims to return to a clinically acceptable treatment plan which has already been through some level of quality assurance (QA). One could therefore assume that returning to this plan yields acceptable results. Some level of QA is however still required, as errors are always possible. This should involve a check on indicators that identify successful restorations. These indicators are for example PTV coverage and the difference between the intended and restored dose distribution e.g. using gamma evaluation if gamma analysis can be performed sufficiently fast. Also flagging large changes in spot-weights and WEPL corrections will be important as indicator of unsuccessful restorations. Furthermore, online QA of dose delivery errors is also required, as pre-treatment patient-specific QA cannot be performed in the online-adaptive setting. This is being developed in a project closely linked to this work, which aims at developing proton range verification for online QA using prompt-gamma imaging.

In generating the intended treatment plans, CTV-to-PTV margins were used to adapt for inter- and intra-fraction motion of the structures. With these margins, the developed dose restoration method based on the projected contours has shown to obtain clinically acceptable restorations for prostate cancer patients. Evaluating on

the actual contours obtained sufficient CTV coverage for most CT scans (Table 2.4) and acceptable values for the OARs (Figure 2.8). The intended treatment plans were generated using normal margins and no additional robustness settings. This shows that the re-optimization method can quickly adapt for changes, even if the treatment plan is not robust. If greater daily shape changes are expected, for instance in cervical cancer, the method will still work, but the PTV or internal target volume becomes very large. Therefore the aim will then be to first reduce the PTV, which can be achieved by implementing a plan-of-the-day approach based on a pretreatment established library-of-plans. After selecting the most tight fitting plan, our dose restoration method can be used to correct for density changes. We believe that will also work for other tumors in the pelvic region such as bladder cancer. To determine with how much the PTV can be reduced and whether this works for tumor sites outside the pelvic region needs further research.

In prostate cancer inter-fraction and intra-fraction variations in position and shape of the target volume and density changes along the proton beam paths can contribute to loss of coverage. Various studies have shown that the contribution of intra-fraction motion is much less than inter-fraction motion if treatment times are kept sufficiently short. The speed of the dose restoration is therefore of great importance. Residual shape changes of the target volume caused by intra-fraction motion can be accounted for by adding a small extra margin. If robustness is included in the initial plan we expect that due to the restoration of the spot positions the robustness will to some extent be preserved. The re-optimization of the spot-weights however might reduce the amount of preserved robustness. To what extent the robustness will be preserved and whether it is necessary to include robustness to tackle intra-fraction density changes requires further investigation.

The treatment plans generated in this study used two laterally opposing beams traversing through the hip bones. Rotational variation of the hip bone gives rise to density variations along the pencil-beam paths. On top of this, for the aligned scans shifts of the lymph nodes in the direction of the beam of at most 5 mm were detected, as well as anatomical changes of the seminal vesicles and prostate below 5 mm and 3 mm respectively. The small changes in the anatomy of the prostate itself can be explained by the scan alignment on the intra-prostatic markers. Andersen *et al.* [27] investigated plan robustness for different beam angles for prostate cancer patients. In their study they found that for the lymph nodes a low WEPL variation was found for beam angles around 40° and around 150°–160° for the left and corresponding angles for the right lymph nodes. Our method starts by correcting the WEPL, obtaining clinically acceptable restorations for all target structures. As the detected movement of

the lymph nodes has the same dosimetric effect as movement of the hip bones, we can assume that the method will also successfully adapt for the changing positions of the hip bones.

Zhang *et al.* [17] described a dose-adaptation method using only an energy adaptation. For the original DET treatment plan the adaptation obtained restorations which, when averaged over five shift datasets, differed less than 1% for the prostate $D_{98\%}$, $D_{50\%}$ and $D_{2\%}$. For the IMPT plan the restorations did not show sufficient improvements. These results were obtained with a phantom prostate patient in which only the prostate and femoral heads were delineated. In our study we used real patient CT scans and the rectum and bladder were used for evaluation. For all patients the targets $V_{95\%}$ and $V_{107\%}$ were restored to clinically acceptable values. Zhang *et al.* [17] did not report on hotspots in the restored dose distributions. In the present study the restored dose distributions with respect to hotspots were evaluated visually. Though the restored dose distributions were less homogeneous than the planned distributions, no clinically significant hotspots were observed. This, however, does not guarantee they will never occur. Our restoration method could be improved by including a quality-check and intervention system to prevent adverse effects on the dose distribution.

Bert *et al.* [28] created a method that adapts the pencil-beam positions as well as the beam energy (WEPL) during the treatment. To our knowledge they however did not change the pencil-beam weights.

The difference between the method of Bert *et al.* [28] and the method described in this work is in the steps that the methods use. Bert *et al.* adapt the pencil-beam positions and the beam energies, while this method adapts the beam energies and the beam weights. The difference between the two methods can be explained by the difference in application; Bert *et al.* compensate for intra-fraction target motion while we compensate for inter-fraction density changes.

For fractionated treatments it can be assumed that the impact of density changes are to some extent averaged out. However, this cannot be guaranteed for hypo-fractionated treatments. This method may therefore help to safely implement hypo-fractionated IMPT treatments by reducing the impact of the density changes before each treatment fraction. Another advantage of this restoration method is that it can replace the use of a rectal balloon. These balloons are sometimes inserted in prostate cancer patients in order to reduce the density changes and prevent large air cavities. These balloons however have to be inserted at each treatment day and can be a discomfort to the patients. In some cases the balloons are not even tolerated [29]. In this light our proposed restoration method may be an attractive alternative.

2.5 CONCLUSION

The impact of density variations on the pencil-beam paths in IMPT can be reduced by performing an automated dose restoration procedure consisting of a WEPL correction of the pencil-beams, followed by a re-optimization of the pencil-beam weights. Only performing the WEPL correction does not yield clinically acceptable results. The fast performance of the restoration method paves the way to future near-real time online-adaptive proton therapy.

2.6 ACKNOWLEDGMENTS AND CONFLICTS OF INTERESTS

The CT-data with contours were collected at Haukeland University Hospital, Bergen, Norway and were provided to us by responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing.

This study was financially supported by ZonMw, the Netherlands Organization for Health Research and Development, grant number 104003012 and by Varian Medical Systems. Erasmus MC Cancer Institute also has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc, Sunnyvale, USA.

Chapter

3

An automated planning strategy for near real-time adaptive proton therapy in prostate cancer

Physics in Medicine and Biology, Vol. 63, 135017 (11pp), 2018,

doi: <https://dx.doi.org/10.1088/1361-6560/aacaa7>

Thyrza Z. Jagt¹
Sebastiaan Breedveld¹
Rens van Haveren¹
Ben J.M. Heijmen¹
Mischa S. Hoogeman^{1,2}

¹ Department of Radiation Oncology, Erasmus MC Cancer Institute,
Rotterdam, The Netherlands

² Department of Medical Physics & Informatics, HollandPTC,
Delft, The Netherlands

Abstract

Purpose: Proton therapy plans are very sensitive to anatomical changes such as density changes along the pencil-beam paths and changes in organ shape and location.

Previously, we developed a restoration method which compensates for density changes along the pencil-beam paths but which is unable to adapt for anatomical changes. This study's purpose is to develop and evaluate an automated method for adaptation of IMPT plans in near real-time to the anatomy of the day.

Methods and materials: We developed an automated treatment plan adaptation method using (1) a restoration of spot positions (Bragg peaks) by adapting the energies to the new water equivalent path lengths; and (2) a spot addition to fully cover the target of the day, followed by a fast reference point method optimization of the spot-weights resulting in a Pareto optimal plan for the daily anatomy. The method was developed and evaluated using 8 – 10 repeat CT scans of 11 prostate cancer patients, prescribing 55 Gy(RBE) (seminal vesicles and lymph nodes) with a boost to 74 Gy(RBE) (prostate).

Results: Applying the automated adaptation method resulted in a clinically acceptable target coverage ($V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$) for 96% of the scans after a single iteration of adding 2500 spots. The other scans obtained target coverages with $V_{95\%} \geq 98\%$ and $2 < V_{107\%} \leq 5\%$. When using two spot-addition iterations, all scans obtained clinically acceptable results. Compared to the restoration method the adaptation lowered the mean dose to rectum and bladder with median values of 6.2 Gy(RBE) and 4.7 Gy(RBE) respectively. The largest improvements were obtained for $V_{45 \text{ Gy(RBE)}}$ for both rectum and bladder, with median differences of 10.3%-point and 10.8%-point respectively, and maximum differences up to 22%-point. The two adaptation steps took on average 7.3 seconds and 1.7 minutes respectively. No user interaction was needed, making this fast and fully automated method a first step towards online-adaptive proton therapy.

3.1 INTRODUCTION

The advantage of intensity-modulated proton therapy (IMPT) is that it deposits the dose very locally as most of the dose is deposited in the Bragg peak. This same characteristic however also makes the treatment modality prone to uncertainties, as the final position of the Bragg peak is very sensitive to variations in daily anatomy [1, 2]. These daily variations include density changes along the pencil-beam paths and variations in shape and location of all organs.

In previous work, we developed an automated restoration method that can adapt a treatment plan to compensate for density variations along the pencil-beam paths by restoring a dose distribution to the optimized distribution of the planning CT [30]. This restoration method starts by restoring the individual spots of a prior treatment plan generated on the planning CT scan to the correct water equivalent path length (WEPL) by adapting the spot energies. This step is then followed by a fast optimization of the spot-weights, which minimizes the difference between the obtained dose distribution and the intended dose distribution on a voxel base. To steer the optimization, additional weight is added to those voxels which do not meet the dose criteria. To apply this method, the planning contours were propagated onto the repeat CT scans, leaving only density variations along the pencil-beam paths to be restored. Combining the WEPL correction with the spot-weight optimization can restore the distorted dose distributions to their intended form. Both steps are fully automated, requiring no user interaction. A limitation however, is that this method does not compensate for variations in target shape and therefore substantial internal margins are still required to account for internal target motion. Additionally, the previous method does not optimize the dose to the new shape and position of the OARs. Another downside of the previous method is that the WEPL correction adds new energy layers in between the planned ones thereby increasing treatment delivery time.

The purpose of this study is to develop and evaluate a method to automatically adapt an IMPT plan in near real-time to the anatomy of the day. This will allow the use of tight dose distributions around the target. Our method simultaneously aims for target dose restoration, as well as minimizing the dose to the OARs, while maintaining the multi-criteria trade-offs of the treatment plan generated for the original planning CT scan. We hypothesize that the adaptation method will hence improve on the previous method as it will not only take into account the inter-fraction density changes but also the inter-fraction changes in organ shape and location. As the goal is to apply this method just prior to treatment, it should be fast. Therefore, a full multi-criteria optimization is unfeasible due to its long calculation time (about 25 min for prostate).

In addition, the method should be completely automated, meaning that it should be able to perform the required adaptations without any user interaction.

In our proposed method we start with a WEPL correction with a fixed number of energy layers, limiting the number of energy layers included in the problem. After this, we add new spots and optimize all spot-weights using the reference point method (RPM). The results of the automated adaptation method were compared with four other strategies, ranging from non-adapted to a full multi-criteria optimization.

3.2 METHODS AND MATERIALS

Patient data

This study included 11 prostate cancer patients, each having 8-10 available repeat CT scans. The original planning CT scans were generated using contrast fluid, making dose calculation on these scans inaccurate. These scans were therefore excluded. Instead, the first repeat CT scan of every patient was used as planning CT and the remaining in total 88 CT scans were used as repeat CT scans. The prostate, seminal vesicles and lymph nodes were defined as target structures, and the rectum, bladder, small and large intestines, and the femoral heads were defined as organs at risk (OARs). Delineations of the target volumes and OARs were available in all repeat CT scans.

Treatment planning volumes and prescription

In this study, the aim was to develop an automated method for fast online adaptation of the treatment plan prior to each fraction. Applying this method mitigated the effects of inter-fraction organ motion, but not of intra-fraction uncertainties. Another uncertainty that was not accounted for is the delineation uncertainty, and in the case of online-adaptive proton therapy especially, the auto-segmentation uncertainty. Therefore, to account for both uncertainties, the high-dose CTV (prostate) in each CT scan was expanded with a small margin of 2 mm and the low-dose CTV (seminal vesicles and lymph nodes) with 3.5 mm. Here the low-dose CTV was given a larger margin primarily to account for the intra-fraction motion of the seminal vesicles, as well as the larger uncertainty in the auto-segmentation of these structures [31]. Similar to our previous work [30], dose was prescribed according to a simultaneously integrated boost scheme in which the high-dose PTV was assigned 74 Gy(RBE) and the low-dose PTV 55 Gy(RBE), to be delivered using two laterally opposed beams.

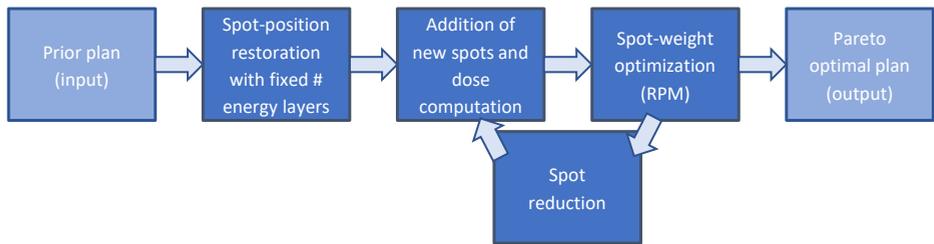


Figure 3.1: The workflow of the automated adaptive treatment planning method. The adaptation starts with a restoration of the spot positions with a fixed number of energy layers, in which all Bragg peaks are restored to their new water equivalent path length. After this the method enters a loop in which iteratively new spots are added, the intensities are optimized and the non-contributing spots are deleted. The output is a Pareto optimal treatment plan.

The automated adaptive treatment planning method

Figure 3.1 illustrates the steps of the automated adaptive treatment planning method. In short, we perform a restoration of the spot positions by adapting the energy of each spot to the new water equivalent path length, followed by the addition of new spots to fully cover the target volume. We then run a fast RPM optimization of the spot-weights resulting in a Pareto optimal plan for the daily anatomy. In the following sections we describe in more detail each step of the method.

Initialization of the automated adaptive treatment planning method

To initialize the automated adaptive treatment planning method, an IMPT treatment plan was created on the planning CT scan. We used the spots from these prior plans as a basis for each fraction. To ensure sufficient spot coverage for most target deformations in the repeat CT scans an additional spot-margin was added to the PTV regions, being an additional +5 mm for the high-dose PTV and +6.5 mm for the low-dose PTV. These plans were thus generated with total margins of +7 mm for the prostate and +10 mm for the lymph nodes and seminal vesicles. By adding a limited number of new spots in the re-optimization, we expect to add sufficient freedom in dose-shaping to compensate for all changes in anatomy between the fractions. Important to note is that these prior treatment plans are not intended for dose delivery. They are only used as a basis for daily re-optimization.

As in our previous work [30], several artificial structures were defined in the optimization. To steer the dose fall-off between the high-dose PTV and the low-dose PTV, a 15 mm transition region called the PTV-intermediate was defined between them. The PTV-low consists of the expanded seminal vesicles and lymph nodes, excluding the transition region, as shown in Figure 3.2.

Initial treatment plans were generated using ‘Erasmus-iCycle’, our in-house developed treatment planning system for fully automated plan generation [19, 20], combined with the ‘Astroid’ dose engine [21]. These multi-criteria optimizations were done according to the wishlist shown in Table 3.1 to generate clinically acceptable Pareto optimal plans in a fully automated way. The constraints in the wishlist, which in this case enforce a minimum dose to the targets, will never be violated, whilst the objective functions are minimized sequentially according to their priorities. Such an optimization takes on average about 25 min for prostate cancer patients. More details on Erasmus-iCycle and the use of a wishlist can be found in [19, 20, 22–24].

Table 3.1: The wishlist with planning constraints and objectives used for automated IMPT plan generation for prostate cancer patients. Constraints will always be met. The priority numbers of the objectives indicate the order in which objectives are to be optimized. A low number corresponds to a high priority. The objectives with priorities four to eight were assigned a limit of 1 Gy(RBE) in order to obtain very low dose values while at the same time not imposing an impossible goal. This wishlist is adapted from previous work [30].

Constraints	Structure	Type	Limit
	PTV-high	Minimum	0.97×74 Gy(RBE)
	PTV-intermediate	Minimum	0.99×74 Gy(RBE)
	PTV-low	Minimum	0.99×55 Gy(RBE)
Objectives			
Priority	Structure	Type	Limit
1	PTV-high	Maximum	1.07×74 Gy(RBE)
1	PTV-intermediate	Maximum	1.07×74 Gy(RBE)
1	PTV-low	Maximum	1.07×55 Gy(RBE)
2	Conformity ring PTV-high	Maximum	1.07×74 Gy(RBE)
2	Conformity ring PTV-full 0 – 10 mm	Maximum	1.07×55 Gy(RBE)
2	Conformity ring PTV-full 10 – 15 mm	Maximum	0.90×55 Gy(RBE)
3	Femoral heads	Maximum	50 Gy(RBE)
4	Rectum	Mean	1 Gy(RBE)
5	Small and large intestines	Mean	1 Gy(RBE)
6	Bladder	Mean	1 Gy(RBE)
7	Femoral heads	Mean	1 Gy(RBE)
8	All conformity rings	Mean	1 Gy(RBE)
8	All conformity rings	Maximum	1 Gy(RBE)
9	Total spot-weight	Sum	1 Gp

Abbreviations: PTV = planning target volume; Gp = Gigaprotons

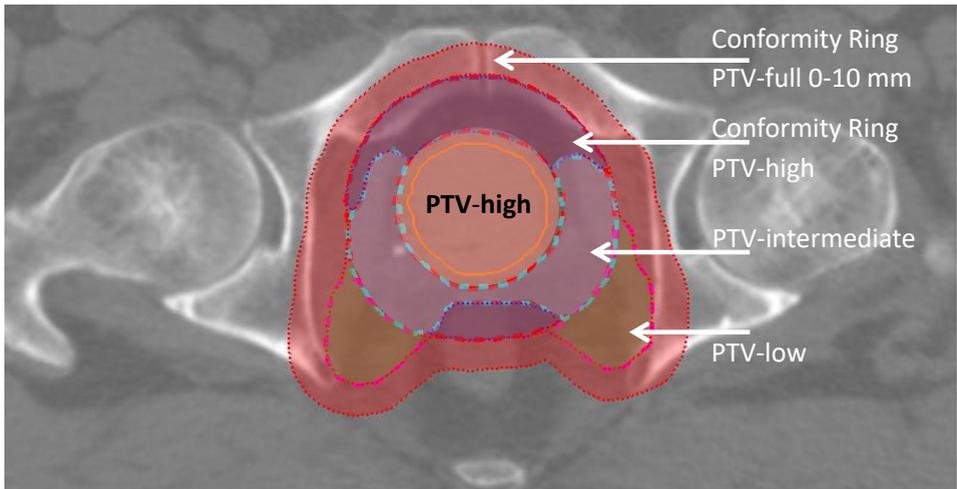


Figure 3.2: The PTV-high is an expansion of the prostate. The PTV-intermediate is a 15 mm transition region between the high-dose and low-dose PTV. The PTV-low consists of the expanded seminal vesicles and lymph nodes, excluding the transition region. The PTV-full consists of the PTV-high with a 15 mm expansion and PTV-low. The conformity ring around the PTV-high is the PTV-full excluding the PTV-low. The red area represents the 0 – 10 mm conformity ring of PTV-full. This Figure is adapted and re-used from previous work [30].

Spot-position restoration with a fixed number of energy layers

In previous work [30] we developed a method to restore spot positions to the new WEPL. In this method the spots could be restored to any available energy, causing the number of energy layers after spot-restoration to be higher than before, resulting in longer treatment delivery. To limit the number of energy layers after spot-restoration, the adaptive treatment planning method proposed in this study will set an upper limit to the number of energy layers. To this end, the energy of each spot can only be restored to one of the available energies that were used in the prior plan.

Spot addition and dose computation

As the restoration of spot positions does not adapt for changes in shape and location of the target, the energy restoration followed by a spot-weight optimization will not automatically result in an adequate coverage of the target. For this reason random new spots which will deposit dose to the target region are (iteratively) added to the weight optimization. New energy layers will be allowed in this step as long as the number of available energy layers is below a certain threshold.

Due to the changes in position and shape due to the energy-restoration and the addition of new spots, the dose deposition matrices need to be recalculated prior to the

optimization of the spot-weights. After this, the intensities will be optimized using the reference point method (RPM) (see below), which will optimize the intensities in a single optimization, while respecting the planning multi-criteria trade-offs made in the prior plan. The combination of adding new spots and doing the weight re-optimization can be repeated. To limit the number of spots in the final plan, a spot reduction technique is applied in this process to eliminate the non-contributing spots [23].

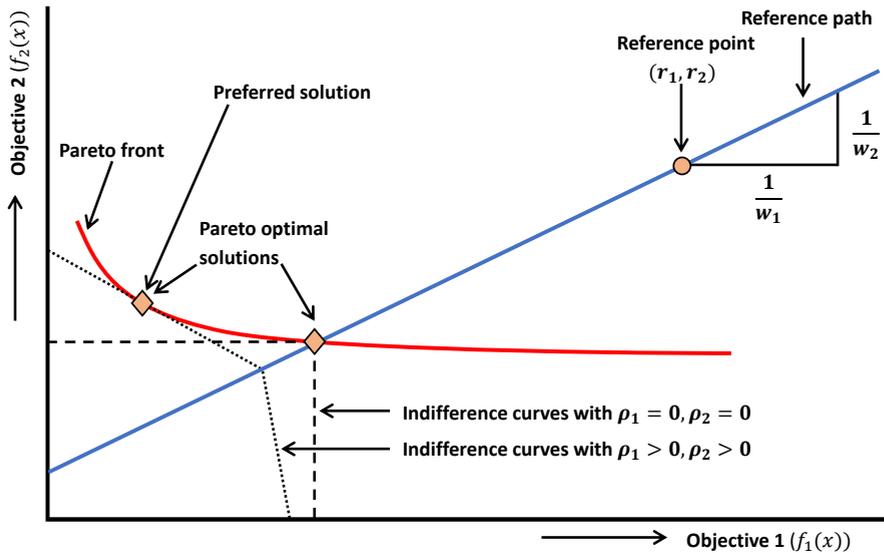


Figure 3.3: A visual explanation of the RPM for two objectives. The dashed lines are the indifference curves with $\rho = 0$, the dotted lines are the indifference curves with $\rho > 0$.

The reference point method

The RPM is a way of modeling a multi-criteria problem. The objective is to define the modeling parameters such that solving this model results in clinically favorable trade-offs between the objectives. A visual representation of the RPM is given in Figure 3.3. The principal idea of the RPM is to use the achieved clinical objective values of the prior plan as reference point, for guidance towards a Pareto optimal solution for the new anatomy.

More specifically, the RPM needs a reference point $r = (r_1, r_2, \dots, r_n)$, which has a value for all n prioritized objectives $f_i(x)$ for $i \in \{1, 2, \dots, n\}$. In the optimization the goal is to get all the objective values f_i close to their reference values r_i simultaneously, and

lower these further if possible, while still meeting the hard constraints $h(x)$. From the reference point a reference path can then be described going through the reference point and the ideal but unachievable goal, i.e. a value of zero for all objectives in case of minimization. The reference path is parameterized by z , which is minimized in the optimization. A lower value of z corresponds to a lower point on the reference path. The method minimizes the value of z in a single optimization to the Pareto front by following the reference point down. The shape of the reference path is modeled by the constraint $w_i(f_i(x) - r_i) \leq z$.

In this case when minimizing all objectives f_i , the RPM can be written as

$$\begin{aligned} & \text{minimize} && z \\ & \text{subject to} && w_i(f_i(x) - r_i) \leq z \quad i = 1, \dots, n \\ & && h(x) \leq 0. \end{aligned} \tag{3.1}$$

In the optimization problem given in equation (3.1) the trade-offs can be steered using the weights w_i . However, it is not always possible to find general values of w_i for which the optimization results in good results for multiple patients, due to overly simple indifference curves. For this reason trade-off parameters ρ_i are added for every objective, which define the angles of the indifference curves (see Figure 3.3). Though these trade-off parameters are usually used to ensure that the generated solution is not in the outer parts of the Pareto front, we will use them to optionally steer the RPM to a clinically more relevant solution. Including the trade-off parameters can be done by adding an additional term to the optimization problem (3.2):

$$\begin{aligned} & \text{minimize} && z + \sum_{i=1}^n \rho_i w_i (f_i(x) - r_i) \\ & \text{subject to} && w_i(f_i(x) - r_i) \leq z \quad i = 1, \dots, n \\ & && h(x) \leq 0. \end{aligned} \tag{3.2}$$

This problem is then solved by a general mathematical solver [32]. The RPM has been applied successfully in photon intensity-modulated radiotherapy [33, 34]. For more details on the RPM we refer the reader to [35] and [36].

RPM-parameter tuning

As shown above the RPM contains several parameters, namely the reference point and for each objective the weight w_i and the trade-off parameter ρ_i . For the reference point we chose to use the objective values that were obtained from the prior treatment plan that was based on the planning CT scan. This means that a unique reference point was calculated for each patient. The tuning of the weight and trade-off parameters was performed on a subset of six out of the 88 CT scans. Those six

CT scans were selected by applying the adaptation on all 88 scans with $\rho_i = 0$, $w_{14} = 1 \times 10^{-4}$ (for f_{14} (Total spot-weight)) and $w_i = 1$ for all other objectives, and then comparing with the outcome of the restoration method. The six selected CT scans were from six different patients, where two scans improved when compared to the restoration method, two scans obtained similar results, and two scans obtained worse results. The comparison was done based on the dosimetric parameter values on targets and OARs. The parameters were tuned such that all six scans met the target criteria. We then selected the parameters for which the lowest OAR doses were obtained.

Optimization and spot addition strategy

The number of iterations in which new spots are added randomly and the number of new spots has to be set. These settings involve a trade-off between improved plan quality (if more re-optimization iterations and spots are allowed) and longer optimization times. To limit optimization time the number of iterations was set to a maximum of two and the number of new spots to 2500 based on the trade-off between calculation time per iteration and the number of required iterations [20]. New energy layers were added as long as the number of energy layers was less than 20 for a certain beam direction. The upper limit of the number of energy layers was set to 200 for all beams together.

Comparison and benchmarking of the adaptive strategies

In this study we compared the novel RPM adaptive treatment planning approach to the spot-restoration method reported previously (spot-position restoration with weight optimization) [30]. In addition, the RPM adaptive method was benchmarked against a full multi-criteria optimization. This benchmark treatment plan was generated using 'Erasmus-iCycle' according to the wishlist shown in Table 3.1, with the tight margins to account for intra-fraction and automatic delineation uncertainties. We also recalculated the dose distribution as if the prior plan (generated for the planning CT scan) would have been delivered on the repeat CT scans. This is only to illustrate the impact of density variations and anatomical changes if the treatment plan is not optimized to the anatomy in the repeat CT scan. Table 3.2 gives an overview of the methods which were included in the evaluation.

Evaluation of the methods

For each repeat CT scan, the dose distributions of all methods (prior, spot-restored, RPM adaptive and benchmark) were checked to see whether they fulfilled the planning criteria ($V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$ for the PTVs). In addition, we checked all

Table 3.2: An overview of the different treatment plans that are compared.

Method	Explanation
Prior plan only	Treatment plan optimized from scratch using Erasmus-iCycle for the planning CT scan using large margins. For evaluation purposes, the dose distribution was recalculated for each aligned repeat CT scan as if the prior plan would have been delivered to that scan.
Spot-restoration	Prior treatment plan adapted by performing a full WEPL correction on the spots of the prior plan, followed by a spot-weight optimization [30].
RPM adaptive	Prior treatment plan adapted for each repeat CT scan by an energy layer constrained WEPL correction followed by one or two iterations of spot addition and RPM optimization.
Benchmark	Treatment plan optimized from scratch using Erasmus-iCycle for each repeat CT scan using tight margins. Currently the best achievable plan if no time constraints would apply. This plan was included as benchmark of plan quality.

dose distributions visually on hot spots inside and outside the target volumes. For the PTV and CTV structures (see Table 1), we report the $V_{95\%}$ and $V_{107\%}$. The rectum was evaluated on the $V_{75 \text{ Gy(RBE)}}$, $V_{60 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$ and D_{mean} , and the bladder on the $V_{65 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$ and D_{mean} . As the number of energy layers could change by the adaptation of the treatment plan, which impacts the delivery time [23], we also report the number of energy layers of each treatment plan.

3.3 RESULTS

Parameter tuning

Based on the training set of six scans, we found that the best results were obtained with $\rho_3 = 0.5$ for f_3 (PTV-low), while the trade-off parameters for the other objectives could be kept zero. Regarding the weights, good results were obtained with $w_1 = 10$, $w_3 = 2$ and $w_{14} = 1 \times 10^{-4}$ for f_1 (PTV-high), f_3 (PTV-low) and f_{14} (Total spot-weight), respectively, while $w_i = 1$ for all other objectives.

Results for the targets

The automatically generated prior treatment plans all fulfilled the planning criteria that at least 98% of the wide-margin PTV on the planning CT scan is covered by 95% of the prescribed dose, and no more than 2% of this PTV receives more than 107%. Table 3.3 shows the percentage of repeat CT scans for which the different methods

met the planning criteria for the PTV structures in the repeat CT scans ($V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$). The results of the RPM adaptive treatment plans that are reported were obtained after a single iteration of adding new spots and applying RPM optimization. In this case 85 out of 88 scans met the PTV criteria. This is a substantial improvement with respect to the spot-restoration method, for which only 45 repeat CT scans met all planning criteria. Of the remaining three scans, two obtained a $V_{107\%}$ value of the PTV-high between 2% and 2.5% and the other scan obtained a PTV-low $V_{107\%}$ value of 4.1%. After a second iteration of adding new spots and RPM optimization, the planning criteria were also met for these three scans.

Table 3.3: Percentages of the 88 dose distributions that meet the target constraints for the PTV structures for the investigated methods. The RPM adaptive treatment plans are obtained using a single iteration of adding new spots.

	Prior (%)	Spot-restoration (%)	RPM adaptive (%)	Benchmark (%)
$V_{95\%} \geq 98\%$ PTV-high	78.4	97.7	100	100
$V_{95\%} \geq 98\%$ PTV-intermediate	95.5	95.5	100	100
$V_{95\%} \geq 98\%$ PTV-low	72.7	96.6	100	100
$V_{107\%} \leq 2\%$ PTV-high	48.9	88.6	97.7	100
$V_{107\%} \leq 2\%$ PTV-intermediate	87.5	96.6	100	100
$V_{107\%} \leq 2\%$ PTV-low	1.1	58.0	98.9	100
Dose distributions meeting all criteria	1.1	51.1	96.6	100

Similar results were obtained for the CTV structures (see Table A.1.1 in appendix A.1).

Results for the OARs

Figure 3.4 shows scatterplots of the dosimetric parameters of the rectum and bladder, comparing the spot-restoration method with the RPM adaptive method. For all scans, the rectum dose was lower for the RPM adaptive method.

For the bladder mean dose there were six repeat CT scans for which spot-restoration performed better than RPM adaptive. The maximum difference was 3.4 Gy(RBE), which could be reduced to 1.1 Gy(RBE) by applying a second iteration of adding spots followed by RPM optimization.

Compared to the spot-restoration method, the RPM adaptive method reduced the mean bladder and rectum dose by 6.2 Gy(RBE) and 4.7 Gy(RBE), respectively. The largest reductions were obtained for the bladder and rectum $V_{45 \text{ Gy(RBE)}}$, with median

reductions of 10.3%-point¹ and 10.8%-point, respectively and a maximum reduction of 22%-point.

Figure 3.5 shows boxplots of the differences between the spot-restoration method and the benchmark and between the RPM adaptive method (one and two iterations) and the benchmark. The RPM adaptive results were much closer to the benchmark than the spot-restoration results. In particular, for the high-dose DVH parameters of the rectum, the RPM adaptive method yielded results close to the benchmark. For the mean dose to the rectum and bladder respectively, the benchmark further reduced the dose with 6.7 Gy(RBE) and 9.1 Gy(RBE) with respect to the RPM adaptive method. Small improvements were obtained by adding a second iteration of spot-addition followed by another RPM optimization.

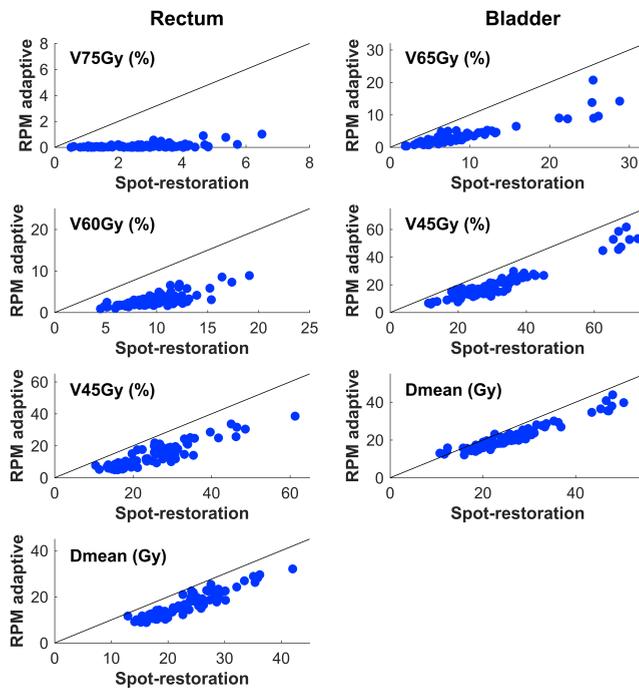


Figure 3.4: Scatterplots depicting the differences in dosimetric parameters of the rectum and bladder between the spot-restoration method and the RPM adaptive method with 1 iteration. Dots below the line represent plans with lower dose values for the RPM adaptive treatment plans.

¹This value provides the absolute difference. For example, if a value increases from 10% to 15%, it is a relative increase of 50%, but an absolute increase of 5%-point.

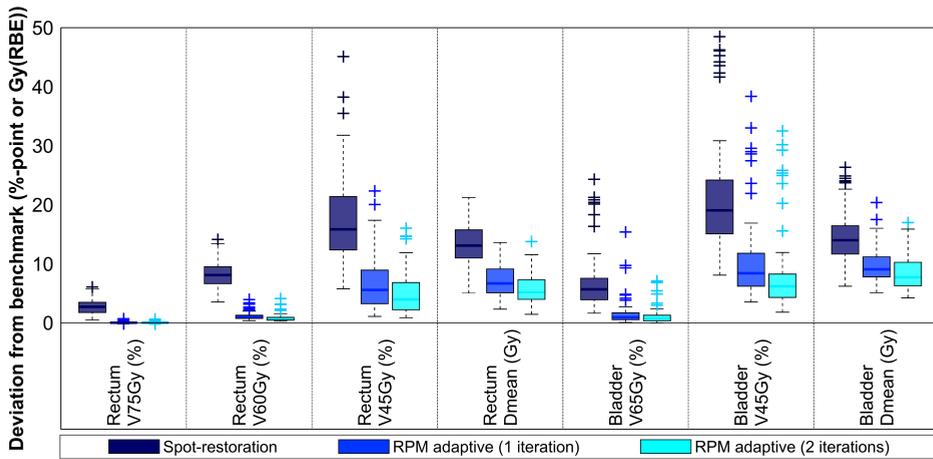


Figure 3.5: Boxplots depicting the differences in dosimetric parameters of the rectum and bladder between the spot-restoration method and the benchmark, and between the RPM adaptive method and the benchmark for 1 and 2 iterations. Differences of zero indicate that the value is equal to that of the benchmark. Larger differences indicate that the value is larger than the benchmark value, leaving room for improvement. Both improvements from spot-restored to RPM adaptive with 1 iteration and from RPM adaptive with 1 iteration to 2 iterations were statistically significant for all dosimetric parameters (Wilcoxon signed-rank test, 5% significance level, $p < 0.01$).

Number of energy layers

Table 3.4 shows the mean number of energy layers in the treatment plans. As in the spot-restoration method the number of energy layers was not constrained, it was significantly higher than in the RPM adaptive method and in the benchmark. Note that for the RPM adaptive method the maximum number of energy layers was below 180, although the maximum was set to 200.

Table 3.4: The number of energy layers (mean and range) for all dose distributions.

	Mean	Range
Spot-restored	371	323 – 400
RPM adaptive with one iteration	135	82 – 179
RPM adaptive with two iterations	119	79 – 160
Benchmark	72	56 – 84

Calculation times

The calculations were performed on a dual Intel Xeon E5-2690 server. The time needed for the first step of the RPM adaptive method, i.e. the spot-restoration, was on average 7.3 s. The second step including a single spot addition and one RPM optimization took on average 1.7 min (1.2 – 2.0). If a second iteration of adding spots was performed, the total average calculation time of the second step increased to 3.4 min (2.5 – 4.0). In contrast, a benchmark plan took on average 25 min. A detailed overview of the running time of the different steps is shown in Table 3.5.

Table 3.5: An overview of the running time of every step.

		Average time	Range
First step	Spot-position restoration with a fixed #energy layers	7.3 s	5.7 – 9.0
	Dose calculation	1.1 min	0.9 – 1.2
Second step one iteration	Spot addition and dose computation (1x)	1.1 min	0.7 – 1.4
	RPM optimization (1x)	0.6 min	0.5 – 0.8
Second step two iterations	Spot addition and dose computation (2x)	2.2 min	1.4 – 2.5
	RPM optimization (2x)	1.2 min	1.0 – 1.8
Benchmark	Erasmus-iCycle full multi-criteria optimization	25 min	

3.4 DISCUSSION

In this study we developed and evaluated an automated planning method for near real-time adaptive IMPT in prostate cancer that accounts for daily density changes along the pencil-beam paths and shape changes of target and organs at risk. The method obtained clinically acceptable treatment plans for 85 out of 88 scans after a single iteration of adding spots, and for all 88 scans after two iterations. Large improvements were seen for the rectum and the bladder with respect to a previously published spot-restoration method ([30], Table 3.2). The average optimization time was limited to 2.9 or 4.6 min, for one and two iterations, respectively.

Figures 3.4 and 3.5 show that when compared to the spot-restoration method, the novel RPM adaptive method yields a substantial reduction in doses to the OARs, especially in the high-dose region of the rectum. This can be explained by the fact that this method optimizes the spot positions and weights for the daily anatomy, subject to all objectives that were included in the wishlist. The spot-restoration method on

the other hand only restores the dose distribution of the initial treatment plan and ignores the new position and shape of the OARs. The use of the RPM also ensures that a Pareto optimal treatment plan is generated with similar trade-offs to those used in the prior treatment plan.

Another improvement with respect to our previous spot-restoration method is that in the RPM adaptive method we restricted the number of possible energy layers to 200. This was done to lower fraction duration, as changing between energy layers takes time [23]. Higher limits did not further improve the results. Even with the limit of 200 the number of selected energy layers remained well below this limit. The second iteration of the RPM adaptive method reduced the mean number of energy layers from 135 to 119. This is due to the spot reduction which is included between the first and second iteration. Spots with a low contribution are removed from the problem as well as energy layers that no longer contribute any spots. In the second iteration spots can only be added to the remaining existing energy layers.

The results of the benchmark plans show that the adapted treatment plans can be further improved. The differences between benchmark and RPM adaptive treatment plans can be explained by the use of resampling in the generation of the benchmark treatment plan [20]. For the benchmark treatment plan new spot positions and energies were iteratively added and removed to find an optimal set of spots. In the RPM adaptive plan, however, we allowed the addition of only one or two sets of spots in order to restrict optimization time. Using two iterations of adding spots instead of one resulted in lower doses to OARs, but at the cost of increased optimization times. Another explanation of the difference between the benchmark and adapted treatment plans lies in the use of the reference point. The reference point was taken from the treatment plan optimized on the planning CT using the large CTV-to-PTV margins including additional spot-margins. This reference point might not be optimal for the tight margin plan on each of the repeat CT scan. Note, that the benchmark plans uses lexicographic optimization, which does not need a reference point. A possible improvement could be to make another prior plan using small margins, such that the objective values to determine the reference point are more representative.

The prior plan is also used to generate a set of spots as starting point for the adaptive treatment planning. To account for changes in target shape and position, an additional spot-margin was used. Together with the spot addition step this yielded acceptable results. Future research is needed to find the optimal balance between the spot-margin and the number of spots added during the adaptive treatment planning. The efficiency of the spot addition step could also be improved by aiming new spots at for example cold spots in the dose distribution of the PTV.

As the addition of new spots is time-consuming, we also tested the method without this step, i.e. only performing the spot-restoration with a fixed number of energy layers, followed by an RPM optimization.

Without adding new spots the treatment plans were often not clinically acceptable because of hot spots in the PTV. These hot spots reflected an insufficient spot set by the RPM optimization. We found that in most cases (>96%) only adding new spots once was sufficient. This means that in only about 4% of the CT scans a longer optimization time is needed. An automatic check can be included to decide whether the second iteration is needed.

A limitation of the current study is that we did not perform an exhaustive search to find the optimal set of parameter values (e.g. the weights and trade-off parameters) for the proposed method. Because of the limited training set, we took the approach of minimally changing the parameters, while obtaining clinically acceptable results. Future research will focus on the optimization of the parameters in terms of plan quality and optimization speed. Furthermore, it should be noted that even though with these parameters acceptable results were obtained for this dataset, there is no guarantee that they will work for other sets.

The method proposed in this study assumes that manual or automatic delineations are available for all repeat CT scans. For the method to be used for adaptive proton therapy, this would mean that delineations would have to be obtained within seconds after the daily CT scan has been acquired. As manually delineating all organs can be time consuming, in practice this will only be possible with automated delineations. Work has been done in this area, amongst others by [31]. Using a fast and automatic deformable image registration (DIR) method, automatic delineations could be generated within 13 s with a success rate of about 80%. The deformation field of the DIR could also be used to adapt the spot positions after the WEPL correction to the new anatomy. A disadvantage of this approach is that the dose proximal to the Bragg peak (at the plateau region) will also be displaced, which may not be optimal for the proximally located target volumes and/or OARs.

The method in this study assumes that daily CT images are obtained. It should however be noted that most daily imaging is done using cone beam CT. Thus far these images are mostly used in photon therapy as setup correction, verification or plan-library selection. For proton therapy, the quality of the cone beam CT is insufficient for accurate dose computation (see e.g. [37]). To enable online adaptation, various proton therapy centers, have installed an in-room CT scanner for this purpose. In the future, improved cone beam CT reconstruction algorithms may overcome the hurdle

of dose computation. In case of absence of a CT scanner, a two-step approach is proposed by Veiga *et al.* [38], where the planning CT scan is deformed using DIR to the CBCT scan. The dose computation is then performed on the deformed CT scan.

To account for both the intra-fraction motion and the uncertainty in the auto-segmentation, the high-dose CTV (prostate) and low-dose CTV (lymph nodes and seminal vesicles) in each CT scan were expanded with a tight margin. In this study we used a 2 mm margin for the high-dose CTV and a 3.5 mm margin for the low-dose CTV. Whether these margins are sufficient needs to be further investigated. Note that the selected values for the margins do not affect the results of the method, as the same values were taken into account in all steps of the method. To account for range uncertainties, equation (3.2) could be modeled using robust objectives f_i . By including robustness into the optimization we could also account for the intra-fraction motion, decreasing the need of CTV-PTV margins. Future research is required to see if this would give better dosimetric results than using a margin and to what extent it would impact optimization time.

Due to the time constraints a physical quality assurance (QA) of the RPM adapted plan cannot be performed prior to delivery. Alternatively, various QA methods can be combined to achieve an equivalent safety level. One is to perform an independent dose calculation of the adapted plan prior to delivery. By using a different dose algorithm, possible failures can be detected. During the fraction, dose delivery can be checked by measuring prompt-gamma emission profiles and comparing these to the profiles derived from the planning CT scan [39]. Another possibility is to recalculate the dose distribution after the treatment fraction, based on the log-file information of the proton therapy delivery system. The information obtained with these methods can then be used to improve the adaptation for the next treatment fraction.

The current time to run the RPM adaptive method is about 3 min on average when applying one iteration of adding spots, including all calculations for the new dose deposition matrices. Though this might not be suitable for online-adaptive proton therapy yet, it is already a great improvement compared to the time it took to generate the benchmark treatment plan (25 min on average).

Most time is consumed in the addition of new spots and the recalculation of the dose matrices. This recalculation is first done after the spot-restoration step, taking on average 1.1 min. In the adaptation step the addition of 2500 spots and the calculation of the dose deposition matrices of these spots again takes 1.1 min on average. The optimization time took only 37 s. Therefore future work should focus on making dose calculation faster. This can be done by making dose calculations algorithms faster,

using faster computer hardware, or both. Various groups are working on dose calculation algorithms including both pencil-beam algorithms and Monte Carlo, which are faster than the ‘Astroid’ dose engine [21] used in our study. An example is the work of Silva *et al.* [40, 41], in which an analytical dose engine running on a GPU is described. They reported dose calculation times of less than half a second for the treatment of a skull base tumor. Other groups, amongst others [42] and [43], work on fast GPU-based Monte Carlo dose engines. They reported processing times in the order of seconds per million particles, while obtaining dosimetric results similar to those of CPU-based simulations.

3.5 CONCLUSIONS

IMPT treatment plans can be re-optimized for new daily anatomies by performing an automated adaptation method consisting of a WEPL correction of the spots, followed by a combination of adding new spots and doing a spot-weight optimization. The high acceptance rate, speed, and automation of the adaptive planning method demonstrates the feasibility of daily adapted treatment plans that maximally conform the dose to the target.

3.6 ACKNOWLEDGMENTS AND CONFLICTS OF INTERESTS

The CT-data with contours were collected at Haukeland University Hospital, Bergen, Norway and were provided to us by responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing.

This study was financially supported by ZonMw, the Netherlands Organization for Health Research and Development, grant number 104003012 and by Varian Medical Systems. Erasmus MC Cancer Institute also has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc., Sunnyvale, USA.

Chapter

4

Plan-library supported automated replanning for online-adaptive intensity-modulated proton therapy of cervical cancer

Acta Oncologica, Vol. 58, pp 1440-5, 2019,

doi: <https://dx.doi.org/10.1080/0284186X.2019.1627414>

Thyrza Z. Jagt¹, Sebastiaan Breedveld¹
Rens van Haveren¹, Remi A. Nout²
Eleftheria Astreinidou², Ben J.M. Heijmen¹
Mischa S. Hoogeman^{1,3}

¹ Department of Radiation Oncology, Erasmus MC Cancer Institute,
Rotterdam, The Netherlands

² Department of Radiation Oncology, Leiden University Medical Center,
Leiden, The Netherlands

³ Department of Medical Physics & Informatics, HollandPTC,
Delft, The Netherlands

Abstract

Purpose: Intensity-modulated proton therapy is sensitive to inter-fraction variations, including density changes along the pencil-beam paths and variations in organ shape and location. Large day-to-day variations are seen for cervical cancer patients. The purpose of this study was to develop and evaluate a novel method for online selection of a plan from a patient-specific library of prior plans for different anatomies, and adapt it for the daily anatomy.

Methods and materials: The patient-specific library of prior plans accounting for altered target geometries was generated using a pretreatment established target motion model. Each fraction, the best fitting prior plan was selected. This prior plan was adapted using (1) a restoration of spot-positions (Bragg peaks) by adapting the energies to the new water equivalent path lengths; and (2) a spot addition to fully cover the target of the day, followed by a fast optimization of the spot-weights with the reference point method (RPM) to obtain a Pareto-optimal plan for the daily anatomy. Spot addition and spot-weight optimization could be repeated iteratively. The patient cohort consisted of six patients with in total 23 repeat CT scans with a prescribed dose of 45 Gy(RBE) to the primary tumor and the nodal CTV. Using a 1-plan-library (one prior plan based on all motion in the motion model) was compared to choosing from a 2-plan-library (two prior plans based on part of the motion).

Results: Applying the prior-plan adaptation method with one iteration of adding spots resulted in clinically acceptable target coverage ($V_{95\%} \geq 95\%$ and $V_{107\%} \leq 2\%$) for 37/46 plans using the 1-plan-library and 41/46 plans for the 2-plan-library. When adding spots twice, the 2-plan-library approach could obtain acceptable coverage for all scans, while the 1-plan-library approach showed $V_{107\%} > 2\%$ for 3/46 plans. Similar OAR results were obtained.

Conclusion: The automated prior-plan adaptation method can successfully adapt for the large day-to-day variations observed in cervical cancer patients.

4.1 INTRODUCTION

Highly localized dose deposition is possible in intensity-modulated proton therapy (IMPT) using the characteristic Bragg peak. At the same time, this treatment modality is sensitive to inter-fraction variations, including density changes along the pencil-beam paths and variations in organ shape and location [1, 2].

Large day-to-day variations can be seen in the shape and position of the cervix-uterus, mostly due to changes in filling of bladder, rectum and sigmoid. Displacements of the tip of the uterus of more than 3 cm between an empty-bladder and a full-bladder anatomy are common. In photon beam radiotherapy, a plan-of-the-day approach has been clinically implemented in several centers, in which a daily image is used to select the best fitting treatment plan from a plan-library [15, 16].

For cervical cancer IMPT, such an approach has been investigated by Schoot *et al.* [11]. The cervix-uterus positions of a full- and empty-bladder CT scan were used to create an internal target volume (ITV) encompassing all possible positions. This ITV was divided into subITVs with which a patient specific plan-library was generated. All library plans were robustly optimized using 8 mm setup errors and 3% range errors. For each simulated fraction, the library plan encompassing the daily CTV was selected, and recalculated on the daily anatomy without further (re-)optimization. Despite the generous robustness settings, the selected plan resulted in inadequate CTV coverage in about 10% of the repeat CT scans, due to ‘substantial deviating anatomy compared to the pretreatment derived full range ITV’ [11]. This shows that when the daily anatomy greatly deviates from the pretreatment observed motion, using a plan-library with robust treatment plans is insufficient to guarantee target coverage.

In this study we therefore propose to automatically adapt the treatment plan that is selected from the plan-library using our automated plan adaptation method developed for prostate cancer in previous work [44]. In this proposed prior plan strategy, the plan selected from a library of prior plans is adapted by an energy adaptation of the pencil-beams, followed by adding spots and a weight optimization using the reference point method (RPM) using automatically tuned RPM-parameters. Outcomes were compared to forward dose calculation of the prior plans on the repeat CT scans (no replanning), and to full, time-consuming multi-criteria optimizations for the daily scans (benchmark). To investigate the effect of using a prior plan as a warm-start for optimization, outcomes were also compared to a time-constrained non-prior- plan strategy in which a new plan is generated from scratch for the daily scans.

4.2 METHODS AND MATERIALS

Patient data

This study included data of six patients with locally advanced cervical cancer selected from an institutional review board approved prospective study designed to investigate inter-fraction motion in cervical cancer patients. For each patient, a full- and empty-bladder CT scan was acquired pretreatment as well as four weekly repeat CT scans, resulting in a total of 23 repeat CT scans. More detailed background of the patient data can be found in the Supplementary Materials (Appendix A.2).

Treatment planning volumes and prescription

The goal was to obtain clinically acceptable daily treatment plans for the repeat CT scans. In the foreseen online-adaptive workflow, we assume that the structures are daily segmented automatically or with minimal user interaction. To account for intra-fraction uncertainties and inevitable uncertainties in the structure segmentation of the adaptive workflow, the daily targets were expanded with a margin: a PTV_{OAPT} ($PTV_{Online-Adaptive\ Proton\ Therapy}$) was created by adding a 5 mm margin around the primal CTV and a 2 mm margin around the nodal CTV [45, 46]. Prescribed dose to the PTV_{OAPT} was set to 45 Gy(RBE), using an RBE of 1.1, which was delivered by four beams (0° , 90° , 180° and 270°).

The automated adaptive treatment planning method

The proposed strategy starts by selecting the best prior plan from a plan-library. The spot-positions are then restored by adapting the energy of each spot to the new water equivalent path length (WEPL). To adapt for changes in shape and location of the target, 3000 new spots are added prior to the optimization with the RPM. The combination of the spot addition and spot-weight optimization can be repeated. In this study, we evaluated using the spot-position restoration in combination with zero (i.e. only optimize the restored spots), one and two iterations of adding spots and spot-weight optimization. Figure A.2.1 in the Supplementary Materials (Appendix A.2), illustrates the workflow of the proposed strategy. A detailed description of this approach is given in [44]. The two extensions of the existing method, namely the library of prior plans and RPM-parameter tuning, are discussed below.

Plan-library generation for the prior-plan strategy

Due to potentially large day-to-day variations in the shape and location of the cervix-uterus, prior plans generated solely on either the full- or empty-bladder CT scan will

likely result in insufficient spot coverage for the observed target deformations in the repeat CT scans. For this reason, an in-house, non-rigid registration was used to derive for each patient a motion model describing the cervix-uterus shape for every possible bladder volume [15]. Using this pretreatment established motion model we created a ‘Complete ITV’ including all observed motion, as well as a ‘Full’ and ‘Empty’ subITV, focusing on parts of the observed motion. The ‘Empty ITV’ ranges from the cervix-uterus corresponding to the empty-bladder to the cervix-uterus corresponding to a ‘half-full-bladder’ structure and the ‘Full ITV’ ranges from the cervix-uterus corresponding to this half-full-bladder to the cervix-uterus corresponding to the full-bladder. Figure A.2.2 in the Supplementary Materials (Appendix A.2) shows an example of the three ITV structures in the sagittal view.

We investigated two library types for the prior-plan strategy:

- 1-plan-library: One prior plan, based on the ‘Complete ITV’.
- 2-plan-library: Two prior plans, based on the ‘Full’ and ‘Empty’ subITVs.

All prior treatment plans were generated based on a PTV_{Prior} which encompassed the Complete ITV or Full/Empty subITV enlarged with a 5 mm margin and the nodal CTV enlarged with a 4 mm margin. Anatomical differences not accounted for by the PTV_{Prior} are expected to be handled by adding new spots during replanning.

The prior treatment plans were generated using ‘Erasmus-iCycle’, our in-house developed treatment planning system for fully automated plan generation, combined with the ‘Astroid’ dose engine. The optimization iteratively adds and removes spots to the target, without time restrictions, see [19–24] and the Supplementary Materials (Appendix A.2) for more details. It is important to note that these prior plans were not intended as the definitive treatment plan, but serve as a warm-start for daily replanning.

Library plan selection strategies

In the case of a 2-plan-library, a selection had to be made between the two prior plans in the library. Traditionally this is done by comparing the bladder volume to a half-full-bladder structure. In this work we selected based on the daily anatomy and the restored spot-positions, without assuming the cervix-uterus motion to be linked to bladder filling. The percentages of the total spots of the library plans that ended up in the daily target region after spot-position restoration were compared. If the difference was more than 1%-point, the plan with the highest percentage was selected as prior. If not, both plans fit the daily anatomy equally well. In that case, the plan with

the most spots ending up in the daily target region after restoration was selected as prior.

RPM-parameter tuning

The RPM is used in this study to automatically optimize the spot-weights in a single optimization. The output is a Pareto-optimal solution, with objective trade-offs in line with the original (i.e. prior) plan. To get these trade-offs, the required RPM-parameters were automatically tuned. As the results of the prior-plan adaptation method might depend on the RPM-parameters, three-fold cross validation was applied. For each fold, two different patients were used for parameter tuning. The planning strategies using the found parameters were tested on the other four patients of each fold. Evaluation was done on all folds simultaneously: i.e., on 46 plans (two plans for each scan). More information on the RPM, the RPM-parameter tuning and the individual folds can be found in the Supplementary Materials (Appendix A.2) and [33–36, 44, 47].

Comparison and evaluation of the methods

In this study, we benchmarked the results of the prior-plan strategy against fully multi-criteria optimized plans. These benchmark plans were generated for each fraction on the PTV_{OAPT} with the same approach as was used for the prior plans (above).

Besides the time-consuming full multi-criteria benchmark optimization, we investigated a replanning strategy that does not require a prior plan. New spots are placed in the target region, which are then optimized using the RPM. Two approaches for the non-prior-plan strategy were investigated:

- **New-Spots-E3:** New spots were positioned in a regular grid, using a 5 mm lateral spacing and an energy spacing three times the longitudinal width of the Bragg peak (at 80% of the peak height).
- **Sampled-New-Spots-3x:** New spots were iteratively added as was done for the benchmark and prior plans. To limit the calculation times, the optimization was stopped after three iterations.

We compared the prior-plan strategy to the non-prior-plan strategy to see whether the use of a prior plan as a warm-start is beneficial for either plan quality or calculation time. More details on the non-prior-plan strategy approaches can be found in the Supplementary Materials (Appendix A.2). Other strategies, energy spacings and number of iterations are also reported there.

Table 4.1 gives an overview of the different methods which were included in the evaluation.

For each repeat CT scan, the dose distributions of all strategies (forward calculation of the prior, prior-plan strategy, non-prior-plan strategy and benchmark) were checked to see whether they fulfilled the planning criteria ($V_{95\%} \geq 95\%$ and $V_{107\%} \leq 2\%$) for the PTV_{OAPT} . In addition, all dose distributions were visually checked for hotspots inside and outside the target volumes.

For the PTV, we report the $V_{95\%}$ and $V_{107\%}$. For rectum, bladder and bowelbag, we report the $V_{30 \text{ Gy(RBE)}}$, D_{mean} and D_{max} . For the sigmoid, femoral heads and whole body (patient) we report the D_{max} .

All calculations were performed on a dual Intel Xeon E5-2690 server.

Table 4.1: Overview of the different treatment plans that are compared.

Method	Explanation
No replanning: Forward dose calculation of prior plan on daily CT	Prior treatment plan selected from a plan-library, recalculated for each aligned repeat CT scan as if it would have been delivered to that scan. Note that as these prior plans were not intended for treatment, the results are only shown to illustrate that replanning is required.
1-plan-library/2-plan-library: Prior-plan strategy	Prior treatment plan selected from a plan-library, adapted for each repeat CT scan by an energy layer constrained WEPL correction followed by zero, one or two (0x, 1x, 2x) iterations of spot addition (adding 3000 spots per iteration) and RPM optimization.
Sampled-New-Spots-3x/New-Spots-E3: Non-prior-plan strategy	Treatment plan generated by placing only new spots in the target region and using the RPM to optimize the spot intensities on the PTV_{OAPT} for each repeat CT scan. Spots were either positioned in a regular grid, or randomly selected from a very fine regular grid using a limited number of iterations.
Benchmark	Treatment plan optimized from scratch using Erasmus-iCycle on the PTV_{OAPT} for each repeat CT scan. Currently the best achievable plan if no time constraints would apply. This plan was included as a benchmark of obtainable plan quality.

4.3 RESULTS

Results for the targets

All prior plans achieved the $V_{95\%}$ and $V_{107\%}$ requirements for the respective PTV_{Prior} volumes. Table 4.2 shows for each treatment strategy the number of plans that met the target demands. It can be seen that forward calculation of the prior plans, i.e., without replanning, always resulted in inadequate target coverage. Replanning using the prior-plan approaches without the addition of new spots (0x) achieved sufficient $V_{95\%}$ values, but too high $V_{107\%}$ values. Adding spots once (1x) yielded acceptable target coverage for more than 80% of the plans. Acceptable coverage was only obtained for all plans with the 2-plan-library-2x approach.

For the non-prior-plan strategy, Table 4.2 shows that while using a fine regular grid (New-Spots-E3) always resulted in acceptable target coverage, iteratively sampling new spots (Sampled-New-Spots-3x) achieved the demands in only 84% of the plans.

Table 4.2: For each treatment strategy, the number of plans that meet the prescribed target demands.

	$V_{95\%} \geq 95\%$ & $V_{107\%} \leq 2\%$	Calculation times (min.) mean (min – max)
1-plan-library No replanning	0/46	–
2-plan-library No replanning	0/46	–
1-plan-library-0x	2/46	1.9 (1.6 – 2.4)
2-plan-library-0x	0/46	2.1 (1.7 – 2.5)
1-plan-library-1x	37/46	4.2 (3.2 – 5.2)
2-plan-library-1x	41/46	4.2 (3.4 – 5.4)
1-plan-library-2x	43/46	6.4 (5.1 – 8.3)
2-plan-library-2x	46/46	6.6 (5.3 – 8.4)
Sampled-New-Spots-3x	39/46	7.1 (5.7 – 8.4)
New-Spots-E3	46/46	40.7 (25.0 – 78.4)
Benchmark	46/46	56.4 (25.3 – 85.1)

Results for the OARs

In Figure 4.1, the OAR results obtained using the 1-plan-library-2x approach and the 2-plan-library-2x approach are compared to the OAR results of the benchmark plans. The highest prioritized criteria (D_{max}) deteriorated less than 5 Gy(RBE) compared

to the benchmark plans, where some resulted in even lower doses (bowelbag, sigmoid D_{max}). For the 2-plan-library-2x approach, the largest deviation (+14%-point) was seen for the rectum $V_{30 Gy(RBE)}$; obtaining a value of 63%, where the benchmark plan had a value of 49%. In Figure 4.2, the OAR results of the best approach of the prior-plan strategy (2-plan-library-2x) and the two approaches of the non-prior-plan strategy are compared to benchmark. Similar OAR results were obtained for the 2-plan-library-2x and the New-Spots-E3 approaches, while the Sampled-New-Spots-3x approach showed slightly larger deviations from benchmark.

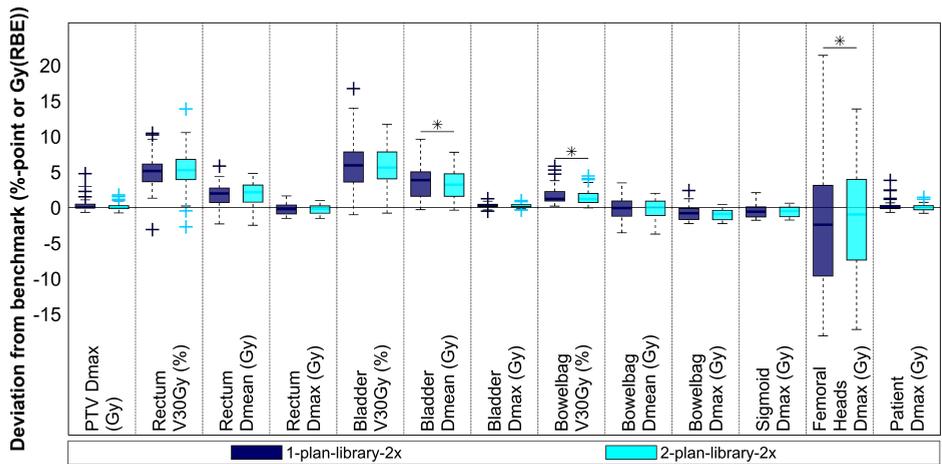


Figure 4.1: Boxplots depicting the OAR differences between the prior-plan strategies with two iterations of adding spots and benchmark. Results are shown for the prior-plan strategy using the 1-plan-library approach and the 2-plan-library approach. Negative deviations depict scans for which the OAR value is lower in the RPM plan than in benchmark. Statistically significant differences (Wilcoxon signed-rank test, 1% significance level, $p < 0.01$) are indicated by asterisks.

Calculation times

Generating the library of prior plans took on average 1.5 h per plan, including dose calculation. Table 4.2 shows the total calculation times required for all treatment strategies, excluding the final dose calculation. In the prior-plan strategy, the spot-position restoration step took on average 5.9 seconds (range 4.4 – 7.4) per restored plan. After restoration, the dose deposition matrix was recalculated in on average 1.2 min (1.0 – 1.4). Without the addition of spots, the RPM spot-weight optimization took on average 28.0 s (19.7 – 54.8). Adding new spots and calculating their dose deposition matrices was completed in on average 1.4 min per iteration (0.9 – 2.5). With the addition of new spots, the average calculation time of the RPM spot-weight optimization increased to 1 min (0.6 – 1.7) per iteration.

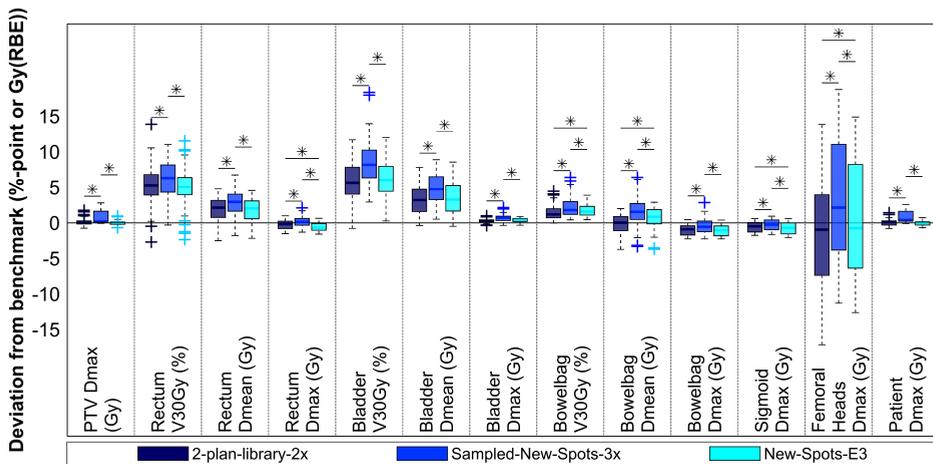


Figure 4.2: Boxplots depicting the OAR deviations from benchmark for the best prior-plan strategy approach and the non-prior-plan strategy approaches. The prior-plan strategy is shown using the 2-plan-library-2x approach, the non-prior strategy is shown using the Sampled-New-Spots-3x approach and the New-Spots-E3 approach. Negative deviations depict scans for which the OAR value is lower in the RPM plan than in benchmark. Statistically significant differences (Wilcoxon signed-rank test, 1% significance level, $p < 0.01$) are indicated by asterisks.

4.4 DISCUSSION

In this study, we combined a plan-library approach with a previously developed RPM adaptive method in a prior-plan strategy. The combination of selecting a prior plan and adding new pencil-beams could account for density changes along the pencil-beam paths and large inter-fraction shape changes of targets and OARs. Clinically acceptable treatment plans were obtained for all plans when using the 2-plan-library-2x approach. One iteration of spot addition was already sufficient for more than 80% of the plans.

Plans were considered acceptable if they achieved $V_{95\%} \geq 95\%$ and $V_{107\%} \leq 2\%$. As all discussed strategies were completely automated, these demands were strictly checked, even though slight deviations might be clinically acceptable. The latter can be incorporated by automatically notifying the user when the plan is within a prescribed bandwidth of the demands.

Applying the prior treatment plans without replanning resulted in inadequate target coverage, while in [11], most scans obtained acceptable target coverage. The differences can be explained by the fact that our prior plans were not intended for actual dose delivery, but only as a warm-start for daily replanning. For this reason, no robustness was used in the optimization of the prior plans. In the Supplementary

Materials (Appendix A.2), we show that robustness against uncertainties in stopping power prediction can be added at the cost of a limited increase in optimization time.

Comparing the 1-plan-library with the 2-plan-library approaches, Table 4.2 shows that the 2-plan-library performed slightly better, but the differences in success rate in terms of target coverage were small. Using a 2-plan-library-2x approach always resulted in clinically acceptable target coverage. Similar OAR results were obtained with the 1-plan-library-2x and the 2-plan-library-2x approaches (Figure 4.1).

We showed that a non-prior-plan strategy in which no warm-start is used can result in clinically acceptable treatment plans when sufficient spots are included. Figure 4.2 and Table 4.2 show that the plans from New-Spots-E3 obtain similar OAR results as the 2-plan-library-2x approach, while requiring over six times longer calculation times. This makes this option infeasible for online replanning. Although the New-Spots-Sampling-3x approach requires similar calculation times as the proposed prior-plan method with two iterations of spot addition, only 39/46 plans fulfilled the target demands (Table 4.2). To obtain good results for all plans would require more iterations of spot addition, again indicating that without a warm-start (i.e., a non-prior-plan approach) calculation times increase.

Several approaches of adaptive IMPT have been reported in the literature. An offline approach was proposed by Kurz *et al.*, in which a new plan is generated to serve as an update for the next fraction [48]. One hour was required for deformable image registration (DIR), optimization and dose calculation. Adaptation reduced over-dosage in the targets and partially improved OAR sparing. Bernatowicz *et al.* compared dose restoration methods using new spots (no prior plan) aiming to restore a given reference dose distribution [49]. Without restoration less than 45% of the repeat CT scans achieved adequate target coverage; with restoration this improved to 100%. The difference is that our method optimizes the dose distribution to the daily anatomy, while their restoration methods intend to only restore a prior dose distribution. Botas *et al.* [50] developed online-adaptation approaches based on cone beam CTs (CBCTs) in which only spots from the prior plan were used. A spot restoration was applied using DIR and, if necessary, this was followed by a weight tuning. Applying only a spot restoration was found to be insufficient; combined with a weight tuning acceptable results were obtained. Calculations were done using GPU-based Monte Carlo.

For conventional radiotherapy, the combination of a restoration followed by a weight optimization was reported by Ahunbay *et al.* [51]. Segment aperture morphing combined with segment weight optimization showed to improve target coverage and OAR sparing. Adaptation was possible in 10 min. Recently adaptive planning methods for the MR-Linac have been described by Winkel *et al.* [52]. Two main categories

of the adaptation methods are described as 'adapt to position' and 'adapt to shape'. Our replanning method would belong to the latter. Other centers have implemented a stereotactic MR guided adaptive workflow (SMART) [53–56]. Generating online-adaptive plans when target and OAR aims were not met resulted in adequate target coverage and better OAR sparing.

In this study, we decided to ignore the simultaneously integrated boost that is recommended in the EMBRACE II protocol. Including the boost would not alter the workflow. If more spots are needed to cover the boost this could slightly increase the calculation times. For the prior plan selection approach, a threshold of 1%-point was used for the initial selection criterion. While this setting is admittedly ad hoc, it demonstrated good results. Also, the proposed replanning methods currently add 3000 new spots to the optimization per iteration, which is the same number as was used in the optimization of the benchmark and prior plans. It is possible that using a different number in the replanning could result in acceptable results after fewer iterations. Fewer iterations might also be achieved by using a different beam setup (i.e., not 0°, 90°, 180°, and 270°), as other beam setups could be more robust against the daily anatomical variations. Finally, this proof of principle study was conducted on a small dataset consisting of six patients. Further investigation based on more data is necessary.

In conclusion, large day-to-day variations such as seen in cervical cancer radiotherapy can be accounted for in IMPT by applying a fast and automated prior-plan adaptation method. Selecting a prior plan from a plan-library, adapting its pencil-beams to the new WEPL, adding new spots and optimizing the spot-weights resulted in clinically acceptable treatment plans on daily anatomies. The use of a library of prior plans significantly reduced the optimization times to obtain clinically acceptable treatment plans.

4.5 DISCLOSURE STATEMENT

Erasmus MC Cancer Institute has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc., Sunnyvale, USA.

4.6 FUNDING

This study was financially supported by ZonMw, the Netherlands Organization for Health Research and Development, grant number [104003012] and by Varian Medical Systems.

Chapter

5

Online-adaptive versus robust IMPT for prostate cancer: how much can we gain?

Radiotherapy and Oncology, Vol. 151, pp 228-33, 2020

doi: <https://doi.org/10.1016/j.radonc.2020.07.054>

Thyrza Z. Jagt¹
Sebastiaan Breedveld¹
Rens van Haveren¹
Ben J.M. Heijmen¹
Mischa S. Hoogeman^{1,2}

¹ Department of Radiation Oncology, Erasmus MC Cancer Institute,
Rotterdam, The Netherlands

² Department of Medical Physics & Informatics, HollandPTC,
Delft, The Netherlands

Abstract

Purpose: Intensity-modulated proton therapy (IMPT) is highly sensitive to anatomical variations which can cause inadequate target coverage during treatment. Available mitigation techniques include robust treatment planning and online-adaptive IMPT. This study compares a robust planning strategy to two online-adaptive IMPT strategies to determine the benefit of online adaptation.

Methods and materials: We derived the robustness settings and safety margins needed to yield adequate target coverage ($V_{95\%} \geq 98\%$) for $> 90\%$ of 11 patients in a prostate cancer cohort (88 repeat CTs). For each patient, we also adapted a non-robust prior plan using a simple restoration and a full adaptation method. The restoration uses energy-adaptation followed by a fast spot-intensity re-optimization. The full adaptation uses energy-adaptation followed by the addition of new spots and a range-robust spot-intensity optimization. Dose was prescribed as 55 Gy(RBE) to the low-dose target (lymph nodes and seminal vesicles) with a boost to 74 Gy(RBE) to the high-dose target (prostate). Daily patient set-up was simulated using implanted intra-prostatic markers.

Results: Margins of 4 and 8 mm around the high- and low-dose target regions, a 6 mm setup error and a 3% range error were found to obtain adequate target coverage for all repeat CTs of 10/11 patients (94.3% of all 88 repeat CTs). Both online-adaptive strategies yielded $V_{95\%} \geq 98\%$ and better OAR sparing in 11/11 patients. Median OAR improvements up to 11%-point and 16%-point were observed when moving from robust planning to respectively restoration and full adaptation.

Conclusion: Both full plan adaptation and simple dose restoration can increase OAR sparing besides better conforming to the target criteria compared to robust treatment planning.

5.1 INTRODUCTION

Due to its characteristic Bragg Peak, intensity-modulated proton therapy (IMPT) can deliver dose locally, avoiding low dose baths and improving dose conformality. These Bragg Peaks however also make IMPT sensitive to anatomical variations such as changes in density, organ-shape and location [1–3]. Two mitigation strategies accounting for such uncertainties are robust treatment planning and online-adaptive IMPT. Robust treatment planning is a passive strategy which preemptively includes errors scenarios in the optimization possibly combined with safety margins to account for anatomical variations [7–9]. Conversely, online-adaptive IMPT is an active strategy taking the optimized plan and adapting it to better fit the daily anatomy and undo the effects of density variations prior to each fraction [30, 44, 48–50, 57, 58].

Making a treatment plan more robust inevitably results in increased doses to healthy tissues [10]. Conversely, online-adaptive planning aims at maintaining an adequate target volume coverage, while minimizing the dose to the organs at risk (OARs) for each fraction. In previous work, we developed online-adaptive treatment planning methods which are feasible for clinical implementation. Starting with the development of a dose restoration method [30], we could restore the initial dose distribution from a dose distribution distorted due to differences in density. Subsequently, we extended this into a full, but sufficiently fast, automated plan adaptation method to adapt the plan to the daily shape and position of the target volume and OARs [44, 57]. We demonstrated that both methods can achieve acceptable target coverage for (most of) the fractions and simultaneously yield OAR doses close to what can be achieved with a fully optimized treatment plan generated without time constraints [30, 44, 57].

So far, however, the proposed adaptive treatments have not been compared to non-adaptive treatments for which robust treatment planning is used to mitigate uncertainties in the daily patient anatomy. Such a comparison is complex, because although individual uncertainties such as intra-fraction motion and positioning variations have been described in literature, information on how to combine these in robustness settings and safety margins in robust optimization is still lacking. A comparison is nevertheless recommendable to establish the value of online-adaptive treatment approaches and to determine whether the benefits outweigh the costs.

In this work, we therefore evaluated the dosimetric benefit of the two developed methods for online-adaptive IMPT by comparing them to a robust treatment planning approach. To this end, we first derived the robustness settings and magnitude of the safety margins needed to yield adequate target volume coverage in a set of pro-

state cancer patients with repeat CT scans. Secondly, for each fraction we compared the online-adaptive approaches to the recomputed robust treatment plans in terms of target coverage and OAR dose.

5.2 METHODS AND MATERIALS

Patient data

This study included data of 11 prostate cancer patients, with 8-10 available repeat CT scans per patient selected from a phase II dose-escalation trial approved by the western Norway regional committee for medical and health research ethics (2006-15727). The original planning CT scans were excluded, as these were generated using contrast fluid, making dose calculation inaccurate. Taking instead the first repeat CT scan as planning CT scan (pCT), 88 repeat CT scans (rCTs) remained for evaluation. From here on, pCT refers to the first repeat CT scan being used as planning CT.

Treatment planning volumes and prescription

Dose was prescribed according to a simultaneously-integrated boost scheme comprising a high-dose region of 74 Gy(RBE) and a low-dose region of 55 Gy(RBE), to be delivered in 37 fractions, using an RBE of 1.1. An intermediate target dose-region, generated as the 15 mm transition between the high- and low-dose regions, was assigned a dose between 55 and 74 Gy(RBE) to steer dose fall-off. On each scan two clinical target volumes (CTVs) were delineated. For the high-dose region, a CTV_{High} was defined as the prostate, a CTV_{Low} was defined for the low-dose region as the lymph nodes and seminal vesicles. From here on we will denote the combination of the CTV_{High} and the CTV_{Low} as CTV. The rectum, bladder, small and large intestines, and the femoral heads were defined as OARs. Target delineations were available in all rCTs, OAR delineations in most. For scans missing the delineations of the intestines or femoral heads the pCT delineations were projected onto the rCT. Dose was to be delivered with two laterally opposed beams.

All rCTs were aligned to the corresponding pCT by a translation based on implanted intra-prostatic markers.

Both adaptation strategies require a prior treatment plan generated on the pCT to start the adaptation. These prior plans were generated using the PTV_{Prior} structures, which were generated by enlarging the CTV_{High} of the pCT by 7 mm, and the CTV_{Low} by 10 mm. Relatively large margins were selected to ensure sufficient spot coverage for most target deformations seen in the rCTs, as was done in previous work [44].

Mitigation strategies

Three mitigation strategies were compared in this study, all aiming for a clinically acceptable dose in all treatment fractions.

- *Strategy A – Robust treatment planning:* On each pCT a robust treatment plan was generated. To account for internal organ motion with respect to the daily alignment on the markers, targets were expanded by safety margins creating internal target volumes (ITVs). Subsequently, robust optimization was applied using the ‘minimax’ worst-case approach [7–9], including range robustness to account for uncertainties in the conversion from Hounsfield units to proton stopping power and including setup robustness to account for patient shifts relative to the daily alignment. Nine error scenarios were optimized simultaneously (nominal, \pm setup, \pm range). We derived the robustness settings and magnitude of the safety margins required for this dataset to ensure adequate coverage in all target regions of all rCTs for at least 90% of the patients. This was done by systematically increasing the margins (0 – 8 mm in steps of 2 mm) and the setup error (2 – 8 mm in steps of 2 mm), while evaluating the effect on the rCTs. The range error, related to uncertainties in the stopping power prediction, was fixed at 3%. For more details see appendix A.3. CTV coverage of the rCTs was evaluated by a forward dose calculation of the robust treatment plan on each rCT.
- *Strategy B – Plan restoration:* For each rCT the dose distribution of the prior treatment plan, optimized on the pCT, was restored. This was done using the delineations of the pCT projected onto the rCT. The restoration method uses energy-adaptation followed by a fast spot-intensity re-optimization focusing on the targets. Details on this method can be found in [30]. Evaluation was done on the CTV structures of the rCTs.
- *Strategy C – Full plan adaptation:* For each rCT, the prior plan optimized on the pCT was used as a warm-start for adaptation. The method starts with an energy-adaptation, followed by adding 2500 new spots and a spot-intensity optimization using the reference point method (RPM). To account for uncertainties in stopping power prediction the optimization is robust to a \pm 3% range. Adaptation is done based on the available contours in the rCTs. To account for inevitable segmentation errors as well as intra-fraction motion uncertainties, the CTV contours were expanded by small margins creating PTV_{OAPT} structures (Online-Adaptive Proton Therapy). As was done in previous work, a 2 mm margin was added around the CTV_{High} of the rCT and a 3.5 mm margin was added

around the CTV_{Low} [44]. Parameter tuning for this strategy was done using three-fold cross validation, where one third of the patients (selected randomly) was used for tuning and the remaining two thirds for testing. Evaluation was done on all folds simultaneously, i.e. 176 plans (two per scan). A brief explanation of the RPM and the tuning is shown in appendices A.5 and A.6. Details on the full adaptation method can be found in [44, 57]. Evaluation was done on the PTV_{OAPT} structures of the rCTs.

Note that evaluation of the three strategies is done on different target definitions, i.e. the daily CTVs for robust planning (A) and simple dose restoration (B) and the PTV_{OAPT} for full plan adaptation (C). This was done to include segmentation errors that are inevitable in an online-adaptive approach, thereby avoiding a too optimistic evaluation for strategy C.

The prior and robust treatment plans were generated using our in-house developed multi-criteria treatment planning system ‘Erasmus-iCycle’ combined with the ‘Astroid’ dose engine. All plans were optimized to obtain clinically acceptable target coverage defined as $V_{95\%} \geq 98\%$ while simultaneously aiming for $V_{107\%} \leq 2\%$ for their respective PTV and ITV. Here $V_{95\%}$ and $V_{107\%}$ are the percentages of the volumes receiving respectively 95% and 107% of the prescribed dose. Dose to the OARs was minimized according to the objectives shown in Table A.4.1 in the appendices. More details can be found in [19–24].

Figure 5.1 summarizes the compared methods.

Comparison and evaluation of the methods

For each rCT, the dose distributions obtained with the three strategies were checked visually and whether they fulfilled the targets planning criteria. We report the targets $V_{95\%}$, $V_{107\%}$ and $V_{110\%}$. In case of hotspots we also report the $D_{2\%}$ and D_{max} . For the rectum, we report the $V_{75 \text{ Gy(RBE)}}$, $V_{60 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$, D_{mean} and $D_{2\%}$ and for the bladder the $V_{65 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$, D_{mean} and $D_{2\%}$. For the whole body (patient) we report the $V_{10 \text{ Gy(RBE)}}$ and $D_{2\%}$. Here $V_{x \text{ Gy(RBE)}}$ is the percentage of the volume receiving $x \text{ Gy(RBE)}$, D_{mean} is the average dose and D_{max} is the maximum dose.

All calculations were performed on a dual Intel Xeon E5-2690 server.

Statistical analysis

Wilcoxon signed-rank tests were performed using MATLAB (Mathworks version 2017a) to evaluate the differences between the strategies. A p -value < 0.05 was considered to be statistically significant.

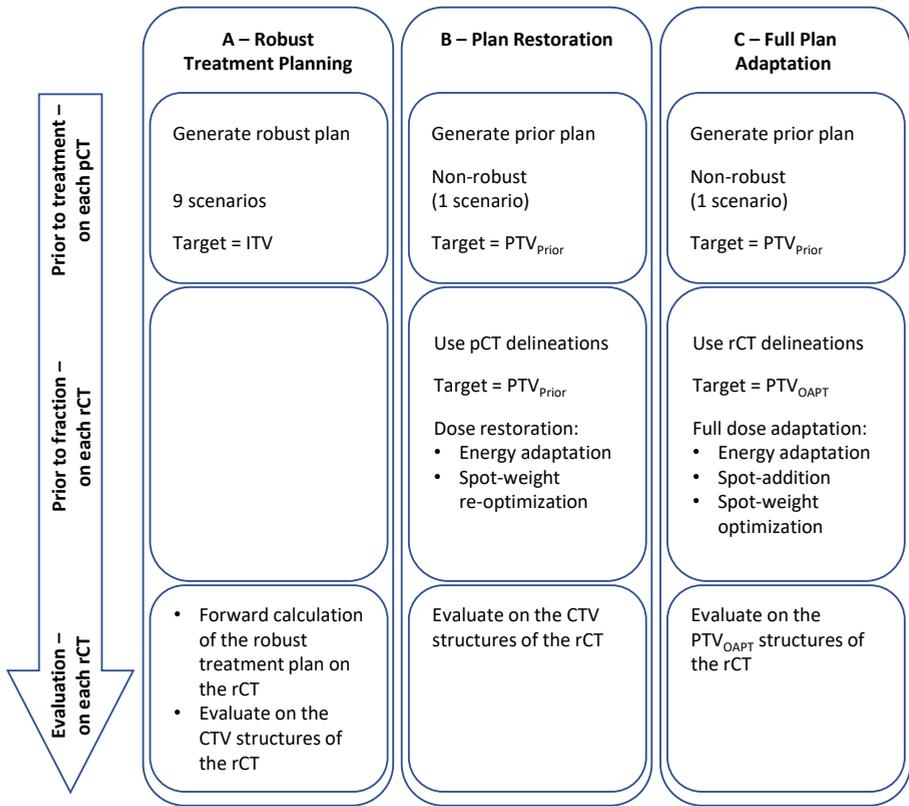


Figure 5.1: Summary of the compared strategies.

5.3 RESULTS

For robust treatment planning (A), expanding the CTV_{High} and CTV_{Low} with a 4 mm and 8 mm safety margin, respectively and applying a range error of 3% and a setup error of 6 mm to the targets during robust optimization yielded adequate target coverage ($V_{95\%} \geq 98\%$ for all target regions) for all rCTs in 10/11 patients. The other patient had $98\% > V_{95\%} \geq 95.5\%$ for the CTV_{Low} for 3/8 rCTs.

Applying the robust treatment plans on the rCTs resulted in a population-mean $V_{107\%}$ of the CTV_{Low} of 44.8% (19.5% – 60.9%) and a population-mean $V_{110\%}$ of 19.9% (5.6% – 37.6%). $D_{2\%}$ values up to 65.8 Gy(RBE) and D_{max} values up to 75.1 Gy(RBE) were obtained (respectively 119.6% and 136.5% of 55 Gy(RBE)). These high values are due to the proximity of the ITV_{High} and ITV_{Low} , as during robust optimization the dose in the ITV_{Low} is increased to achieve adequate ITV_{High} coverage in the error scenarios.

For the CTV_{High} all scans obtained $V_{107\%} \leq 2\%$ and $V_{110\%} = 0\%$. No combination of margins and robustness was found obtaining sufficient coverage for all target regions for all rCTs of all patients.

Applying plan restoration (B) yielded $V_{95\%} \geq 98\%$ for all scans. For the CTV_{High} all scans obtained $V_{107\%} \leq 2\%$, but for the CTV_{Low} 21/88 scans obtained $V_{107\%} > 2\%$, with values up to 3.7%. $D_{2\%}$ values up to 59.6 Gy(RBE) and D_{\max} values up to 71.7 Gy(RBE) were obtained (respectively 108.4% and 130.4% of 55 Gy(RBE)). All 88 scans obtained $V_{110\%} = 0\%$ for the CTV_{High}, but $0\% < V_{110\%} < 1.6\%$ for the CTV_{Low}.

Applying the full plan adaptation method (C) yielded $V_{95\%} \geq 98\%$ for all plans for the PTV_{OAPT_High} and PTV_{OAPT_Low}. For the PTV_{OAPT_High} 54/176 plans obtained $V_{107\%} > 2\%$ (up to 37.3%) and 9/176 plans $V_{110\%} > 0\%$ (up to 15.2%). $D_{2\%}$ values up to 84.0 Gy(RBE) and D_{\max} values up to 85.3 Gy(RBE) (respectively 113.5% and 115.3% of 74 Gy(RBE)) were obtained. For the PTV_{OAPT_Low} 20/176 plans obtained $V_{107\%} > 2\%$ (up to 13.7%), with $D_{2\%}$ and D_{\max} values up to 59.9 Gy(RBE) (108.9% of 55 Gy(RBE)). All plans obtained $V_{110\%} = 0\%$.

In terms of OAR sparing the adaptive strategies (B and C) outperformed strategy A for all patients. Figure 5.2 shows boxplots depicting the obtained OAR values for the three strategies. Largest differences between the strategies were observed for the $V_{45 \text{ Gy(RBE)}}$ of both rectum and bladder. For the rectum the median value improved with 11.1%-point when moving from robust treatment planning to plan restoration (A to B) and 16.3%-point when moving to full plan adaptation (A to C). For the bladder these improvements were respectively 6.9%-point and 9.9%-point. For the high dose criteria ($V_{75 \text{ Gy(RBE)}}$, $D_{2\%}$ and D_{\max}) smaller differences between the strategies were observed. For all evaluation criteria of the OARs the differences between robust treatment planning (A) and plan restoration (B), as well as the differences between plan restoration (B) and full plan adaptation (C) were statistically significant.

Figure 5.3 shows an example of a slice of the dose distributions obtained for one of the rCTs using the three different strategies. It can be seen that the high-dose region is largest for robust treatment planning (A) and smallest for plan adaptation (C).

Plan restoration (B) took on average 1.7 minutes (1.4 – 2.1) and full plan adaptation (C) took on average 6.6 minutes (5.0 – 9.8). These times include the adaptation steps and intermediate dose calculations, but exclude initialization and final dose calculation, thus reflecting the additional time required compared to recalculation of a static plan on the rCT (strategy A). For both methods, the initialization consumed on average 1 minute. The final dose calculation takes on average 3.9 minutes for plan restoration (2.1 – 5.5) and 7.0 minutes for full adaptation (3.1 – 11.3).

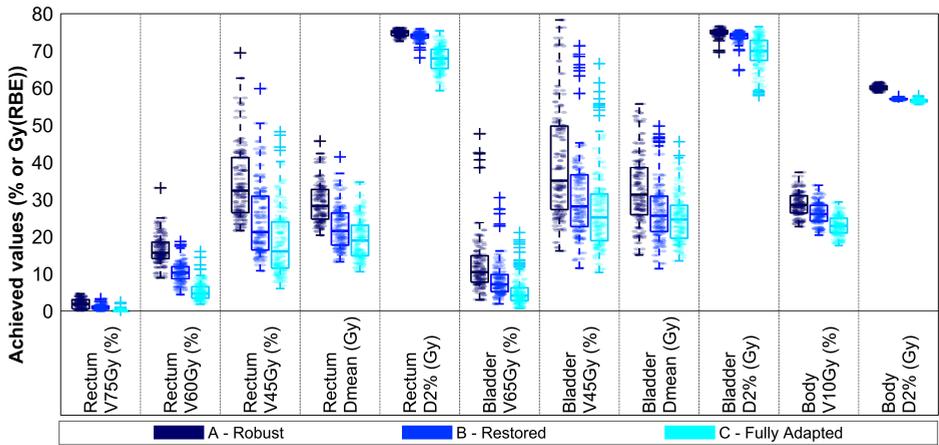


Figure 5.2: Boxplots depicting the obtained dosimetric parameter values for the three strategies.

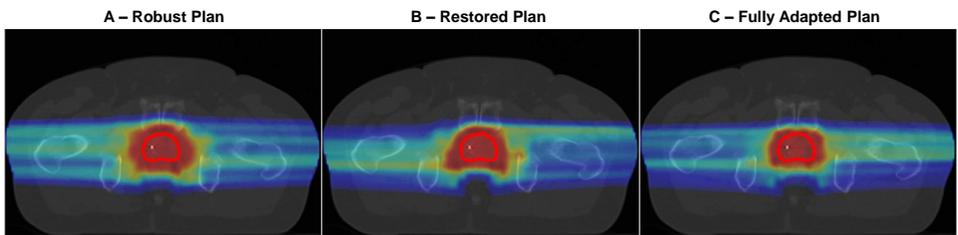


Figure 5.3: An example of the dose distributions obtained using the different strategies for one repeat CT scan. The red contour indicates the daily CTV_{High} .

5.4 DISCUSSION

In this study, robust treatment planning combined with safety margins (A) was compared to two adaptive strategies (B and C). Plan restoration and full plan adaptation both achieved $V_{95\%} \geq 98\%$ for all rCTs, while robust treatment planning did not meet this criteria for one patient. Applying adaptive treatment planning always resulted in lower OAR doses than robust treatment planning, with largest median improvements observed for the rectum (up to 16.3%-point). Several studies have shown OAR dose to be correlated to toxicity [59], so lowering these with online-adaptive IMPT can potentially reduce the expected toxicities compared to a strategy that fully relies on robust treatment planning.

For the robust treatment planning approach we derived required margins and robustness settings to achieve adequate target coverage for all rCTs in at least 90% of the patients. We obtained five combinations of margins and robustness settings all

yielding adequate target coverage for all rCTs of 10/11 patients. For this study, we selected the combination with the smallest margins and robustness settings. It should be noted that these settings are specific to the investigated dataset and the number of robustness scenarios and have not been validated on other datasets. The observed benefit of adaptive planning likely depends on the dataset and robustness settings that are used.

In this study both targets were robustly optimized using the same values for setup and range robustness to stay close to clinical practice. More research is needed to determine whether OAR doses could be reduced by applying target-specific robustness settings. The effect of fractionation has not been considered in this study. Fractionation can potentially average out underdosage or overdosage in the target volume. Hence, the evaluation criteria applied in this study for the three strategies might be too conservative. However, while uncertainties in photon radiotherapy mostly result in dose deviations around the target-edges, IMPT can result in underdosage in the center of the target. Whether an underdosage in the center of the target volume can be effectively compensated by an overdosage in another fraction is unclear. Besides this, treatments are increasingly delivered in fewer fractions reducing the averaging effect [60].

For the adaptation methods prior plans including large margins were used, as these have shown to be effective in previous work [44]. Changing these prior margins and changing the PTV_{OAPT} margins may influence the observed gain of adaptive planning. Furthermore, due to the used optimization method, the plans obtained through simple dose restoration are not explicitly made robust against range errors caused by uncertainties in stopping power prediction. While setup errors should be negligible in the daily adaptive workflow, range errors arising from Hounsfield Unit to proton stopping power conversion remain present. For full plan adaptation, we have therefore included range robustness in the spot-weight optimization. The tuning of the RPM-parameters however has been done without including range robustness. This could be an explanation for the elevated $V_{107\%}$ values obtained with full plan adaptation. Including range robustness in the tuning could possibly reduce these, although from a clinical perspective these values are acceptable.

In this study the three methods have been compared for a dataset of high-risk prostate cancer patients. This treatment group is interesting for online-adaptive planning due to the challenges which are related to the size of the target volume, its location in the pelvic region, and the differential motion between the low-dose and high-dose target volumes. Investigating the benefit of online-adaptive planning in other

treatment sites such as head and neck cancer and locally advanced cervical cancer and lung cancer is part of ongoing and future research.

All treatment plans were made using two laterally opposed beams. While more complex beam geometries might improve all three methods, finding such a geometry requires further research. For all methods the CTs were aligned based on intra-prostatic markers. This approach may differ between centres. The accuracy of alignment might influence the required setup robustness and margins. However, as the alignment was the same for all methods, no effect on the comparison is expected.

Intra-fraction variations have not been addressed in this study. However, we included a small margin (2.0/3.5 mm) to account for the extra intra-fractional motion potentially occurring between the start of the full adaptation and beam delivery and to account for segmentation uncertainties. Intra-fraction motion during beam delivery was ignored for all three methods, but could easily be included by expanding the PTV or ITV or increasing robustness. Whether the included margins are sufficient or whether larger margins or more robustness is required was outside the scope of this study and should be investigated before clinical implementation.

General challenges of introducing online-adaptive IMPT into the clinic include adaptation time, user interaction time, the need for daily delineations and plan quality assurance (QA). Considering the adaptation time, the fully automated process now takes on average 2.9 minutes for dose restoration, and 7.5 minutes for full adaptation. As anatomical variations could occur during this time span, calculation times should be further reduced. The intermediate dose calculations are the most time consuming. Dose calculation time can be shortened considerably for example by parallelization and running the calculations on a GPU, as shown by Silva *et al.* [40, 41] and Matter *et al.* [58]. This was however outside the scope of this study. Regarding the user interaction time, as both investigated adaptive strategies are fully automated, user interaction is only required once in advance to tune the parameters for an entire patient population, and once prior to each fraction to verify and approve the adaptation. The latter can be automated as well by automatically computing relevant dosimetric parameter values of the adapted plan and checking these against predefined limits. Prior to adaptation however the delineations of the rCT should be generated. When done manually, this step requires time-consuming user interaction. This can be largely avoided by (partly) generating the daily delineations automatically. For prostate cancer patients, an auto-propagation method combining deep-learning with deformable image registration has for example been developed with which already 80% of the automatically propagated pCT contours onto the rCT could be used without manual corrections [61]. Without deep-learning, contour propagation was used in for ex-

ample the work on adaptive planning by Kurz *et al.* and Botas *et al.* [48, 50]. Additional uncertainties in automatically generated contours can be accounted for by adding a margin to the targets as was done in the present study. It should be noted that daily delineations are only needed for full plan adaptation, as plan restoration uses the pCT contours. Another challenge lies in daily plan QA, for which little to no time is available in the daily adaptive workflow. This can be solved using alternatives such as a redundant dose calculation, online dose monitoring using prompt gamma emission profiles [62], and using machine log files [63, 64] to verify the correct delivery of the treatment plan.

In conclusion, having demonstrated that plan adaptation in IMPT can reduce dose to OARs compared to robust treatment planning within a clinically acceptable time frame, we consider it auspicious to start exploring clinical implementation of online-adaptive strategies in IMPT.

5.5 ACKNOWLEDGMENTS

The CT-data with contours were collected at Haukeland University Hospital, Bergen, Norway and were provided to us by responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing. This study was financially supported by ZonMw, the Netherlands Organization for Health Research and Development, grant number 104003012 and by Varian Medical Systems.

5.6 CONFLICTS OF INTERESTS

Varian Medical Systems has in part financed this research. Erasmus MC Cancer Institute also has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc., Sunnyvale, USA.

Chapter

6

Evaluation of alternative parameter settings for dose restoration and full plan adaptation in IMPT for prostate cancer

Abstract

Purpose: Intensity-modulated proton therapy is highly sensitive to anatomical variations which can cause inadequate target coverage during treatment. A dose restoration method and a full plan adaptation method have been developed in earlier work, both requiring several parameter settings. This study evaluates the previously selected settings by systematically comparing them to alternative settings.

Methods and materials: The dose restoration method takes a prior plan and uses an energy-adaptation followed by a fast spot-intensity re-optimization to restore the plan to its initial state. The full adaptation method uses an energy-adaptation followed by the addition of new spots and a spot-intensity optimization using the reference point method to fit the new anatomy. We varied the following parameter settings: 1) The amount of motion mitigation included in the prior plan; 2) the spot-addition sample size; 3) the spot-addition stopping criterion and 4) the spot-intensity optimization approach. The first variation was evaluated for both methods, the others only for the full plan adaptation. Evaluations were done on 88 repeat CT scans of 11 prostate cancer patients. Dose was prescribed as 55 Gy(RBE) to the low-dose target (lymph nodes and seminal vesicles) with a boost to 74 Gy(RBE) to the high-dose target (prostate).

Results: For the dose restoration method, changing the prior plans yielded insufficient target coverage or increased OAR doses. Results of the full plan adaptation were most affected by varying the stopping criterion and the optimization approach. More spot-addition iterations and using the 2-phase-epsilon-constraint approach both resulted in similar target coverage as the default settings, while achieving lower OAR doses. Both variations however increased calculation times, making them infeasible for online-adaptation.

Conclusion: We recommend maintaining the default setting for the dose restoration approach, while for full plan adaptation more OAR sparing was proved possible. To achieve this a more advanced tuning of the RPM-parameters is recommended.

6.1 INTRODUCTION

Intensity-modulated proton therapy (IMPT) is highly sensitive to anatomical variations, which can cause insufficient target coverage during treatment [1–3]. An approach to account for these anatomical variations, ensuring adequate target coverage in all treatment fractions while keeping the doses to the organs at risk (OARs) acceptable, is to use adaptive treatment planning. In previous work we have developed and evaluated two automated online-adaptive approaches, being a straightforward dose restoration approach and a full plan adaptation method.

The dose restoration aims to restore a prior treatment plan to its intended dose distribution, by adjusting each pencil-beam to the water equivalent path length of the current daily anatomy. The full adaptation approach expands this by also adapting the plan for daily variations in the target shape and location. This is done by adding new pencil-beams to the optimization. Both approaches require a prior treatment plan to start the adaptation. Full adaptation requires several parameter settings, namely: 1) sample size, prescribing how many spots are added in each iteration; 2) stopping criterion, prescribing how many iterations are performed; 3) the amount of motion mitigation accounted for in the prior plan and 4) the optimization method used for the spot-intensity optimization.

In our previously published research, we have only partly presented the effect of the choices made in that context. For future clinical implementation, it is important to verify and validate those settings. In this paper, we will therefore revisit the previously chosen parameter settings by systematically investigating and describing the effect of varying the parameter settings. In this study we address the following parameters:

1. the prior plan,
2. the sample size,
3. the stopping criterion, and
4. the spot-intensity optimization approach.

6.2 METHODS AND MATERIALS

Patient data

For this study a patient cohort of 11 prostate cancer patients was used, including 8 – 10 repeat CT scans (rCTs) per patient. The original planning CT scans (pCTs) were excluded as they were generated using contrast fluid. We instead took the first rCT as

pCT, keeping a total of 88 rCTs for evaluation. All patients gave consent after which they were enrolled in a phase II dose-escalation trial delivered with moderately hypofractionated IMRT at Haukeland University Hospital, Norway. Enrollment started in 2007. The trial had been approved by the local ethical committee REC West (the western Norway regional committee for medical and health research ethics; 2006-15727).

Treatment planning volumes and prescription

Dose was prescribed according to a simultaneously-integrated boost scheme to be delivered with two laterally opposed beams. The high-dose region was assigned 74 Gy(RBE) and the low-dose region 55 Gy(RBE), using an RBE of 1.1. To steer dose fall-off, an intermediate dose-region was defined as the 15 mm transition region between the high-dose region and low-dose region. This intermediate target should receive at least 55 Gy(RBE) and at most 74 Gy(RBE). On each scan two clinical target volumes (CTVs) were delineated. For the high-dose region, a CTV_{High} was defined as the prostate, and for the low-dose region a CTV_{Low} was defined as the lymph nodes and seminal vesicles. The rectum, bladder, small and large intestines, and the femoral heads were defined as OARs. Delineations of the target volumes, bladder and rectum were available in all rCTs. For some rCTs no manual delineations were available for the intestines and/or the femoral heads. For those scans, the missing delineations were obtained by propagating the delineations of the pCT onto the rCT.

Online-adaptive methods and default configuration

Treatment plan optimization of the prior plans

Both online-adaptive methods require a prior treatment plan to start the adaptation. All prior treatment plans were generated using 'Erasmus-iCycle', our in-house developed treatment planning system (TPS) for fully automated plan generation [19, 20], combined with the 'Astroid' dose engine [21]. The optimization applies the 2-phase-epsilon-constraint method (2pec), which performs a prioritized optimization of the objectives, obtaining a Pareto optimal plan. Plans were generated on the pCTs using the pCT delineations. PTV_{Wide} structures were generated by expanding the CTV_{High} and CTV_{Low} structures of the pCT with margins of 7 and 10 mm, respectively. These large default margins were selected to ensure sufficient spot coverage for most target deformations seen in the rCTs [44]. The prior plans were optimized to obtain clinically acceptable target coverage for the PTV_{Wide} structures ($V_{95\%} \geq 98\%$) while at the same time aiming for $V_{107\%} \leq 2\%$. Here $V_{95\%}$ and $V_{107\%}$ are the percentages of the volumes receiving 95% and 107% of the prescribed dose, respectively. Dose was fur-

ther minimized according to the objectives shown in Table A.4.1 of the appendices. More details on Erasmus-iCycle and the 2pec method can be found in [19, 20, 22–24].

The dose restoration method

The first method is a straightforward adaptation approach, which for each treatment fraction takes a prior plan optimized on the pCT and structure-wise restores that dose distribution on the new daily anatomy. Instead of requiring daily contours of the targets and OARs, the method uses the contours of the pCT projected onto the rCT. To restore the dose distribution to its initial state, the restoration method uses an energy-adaptation followed by a fast spot-intensity re-optimization with focus on the target volumes. Target coverage of the restored plans was evaluated on the CTV structures of the rCT. Further details on the method can be found in [30].

The full plan adaptation method

The second method is a more elaborate approach, which for each treatment fraction takes the prior plan optimized on the corresponding pCT and ‘adapts’ this, to fit the new daily anatomy. Using daily contours, the method uses the prior treatment plan as a warm-start to generate a new plan for the treatment fraction. To account for inevitable uncertainties in the daily delineations as well as some intra-fraction uncertainties, the full adaptation aims to obtain adequate target coverage on PTV_{OAPT} structures (Online-Adaptive Proton Therapy). These structures were created by expanding the daily CTV_{High} and CTV_{Low} structures by 2 and 3.5 mm, respectively [44]. Adaptation starts with an energy-adaptation followed by the addition of 2500 new spots and a spot-intensity optimization based on the rCT delineations, using the reference point method (RPM). The RPM is an optimizer which obtains a Pareto optimal treatment plan in a single optimization [33–36, 47]. Parameter tuning for the RPM was done using training plans generated using the 2pec optimization approach implemented in our in-house TPS. Tuning aims to minimize the differences between the original plans and the plans obtained when optimizing the training spots with the RPM, thereby training the RPM to make similar trade-offs as the 2pec optimizer. The used tuning protocol can be found in Table A.6.1 of the appendices. Tuning was done using three-fold cross validation, where one third of the patients was used for tuning and the remaining two thirds for testing. Evaluation was done on all folds simultaneously, i.e. 176 plans (two per scan). Target coverage was evaluated on the PTV_{OAPT} structures. More details on the adaptation method can be found in [44] and chapter 5.

For all PTV definitions, the PTV_{Low} consists of the expanded CTV_{Low} excluding the 15 mm transition region around the PTV_{High} . From here on we will denote the combination of the CTV_{High} and the CTV_{Low} as the CTV.

Variations on the default parameter settings

In this study we have investigated the effects of varying four of the default parameter settings. While altering one of these settings, all others were kept at the default values, as shown in Table 6.1.

1. *Parameter 1 – The prior treatment plan:* Both adaptation methods require a prior plan to start. In previous work we have generated these prior plans with the inclusion of additional margins around the target structures. These margins were added to ensure sufficient spot coverage for most target deformations seen in the rCTs. For the first setting we investigated what happens if we do not include such additional margins. Instead of using these ‘wide’ prior plans generated based on the PTV_{Wide} structures, ‘small’ prior plans were generated based on the PTV_{OAPT} structures of the pCTs. Second, we investigated what happens if we account for even more possible deformations. To this end, ‘robust’ prior plans were generated based on the CTV structures of the pCTs. Safety margins of 4 and 8 mm were added to the high- and low-dose regions, respectively (chapter 5), and robust optimization was applied using the ‘minimax’ worst-case approach [7–9]. Nine scenarios were optimized simultaneously (nominal, \pm setup, \pm range). A setup robustness of 6 mm and a range robustness of 3% were used (chapter 5).
2. *Parameter 2 – The sample size:* In the full adaptation method new spots are iteratively added to the target region. Previously, we found for the prostate cancer patient dataset used in this study that with a sample size of 2500, adequate target coverage could be obtained for all rCTs when applying a single iteration of spot addition and spot-intensity optimization [44]. For this parameter, we investigated the effect of varying the sample size on calculation time, target coverage and OAR sparing. The sample size was varied from 500 to 4500 in steps of 1000.
3. *Parameter 3 – Stopping criterion:* For the prostate cancer dataset used in this study adequate target coverage was obtained for all rCTs using only a single iteration [44]. By varying the stopping criterion we evaluated the effect of allowing the full adaptation to continue adding and optimizing spots until the improvements of the cost functions are all below 3%. With this approach no time constraints were given, thereby allowing for an evaluation of the final trade-offs made by the RPM optimization.
4. *Parameter 4 – Optimization approach:* To evaluate the effect of the RPM (which is a simpler model), we replaced this optimization by the slower 2pec optim-

ization. The 2pec method was applied using a single iteration of adding 2500 spots and optimizing the weights as well as running the adaptation until no improvement larger than 3% was observed.

A summary of the variations on the parameter settings we have investigated is provided in Table 6.1.

Table 6.1: A summary of the varied parameter settings, including default values and investigated variations.

Parameter	Default	Variations
Prior plan	Wide prior plans, optimized non-robustly on the PTV_{Wide} structures.	1. Small prior plans, optimized non-robustly on the PTV_{OAPT} structures, 2. Robust prior plans, optimized robustly on expanded structures and using 6 mm setup robustness and 3% range robustness.
Sample size	2500 spots	Investigate 500, 1500, 3500 and 4500.
Stopping criterion	Stop after 1 iteration of spot addition and optimization.	Continue adding new spots until no improvement of $> 3\%$ is observed for the cost functions.
Optimization approach	RPM	2pec

Evaluation of the investigated variations

For all variations of the parameter settings, we report the $V_{95\%}$ success rates (percentage of scans achieving $V_{95\%} \geq 98\%$ for all targets) and the $V_{107\%}$ success rates (percentage of scans achieving $V_{107\%} \leq 2\%$ for all targets). To compare sparing of the healthy tissues we evaluated dose to the rectum, bladder and whole body. For all investigated variations we show the rectum $V_{75 \text{ Gy(RBE)}}$, $V_{60 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$, D_{mean} and $D_{2\%}$, the bladder $V_{65 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$, D_{mean} and $D_{2\%}$, and the whole body $V_{10 \text{ Gy(RBE)}}$, and $D_{2\%}$. Here $V_x \text{ Gy(RBE)}$ is the percentage of the volume receiving $x \text{ Gy(RBE)}$, D_{mean} is the average dose and $D_{2\%}$ is the dose delivered to 2% of the volume.

All calculations were performed on a dual Intel Xeon E5-2690 server.

Statistical analysis

Wilcoxon signed-rank tests were performed using MATLAB (Mathworks version 2017a) to evaluate the differences in OAR dose between the varied parameter values and the

default parameter value. A p -value < 0.05 was considered to be statistically significant.

6.3 RESULTS

Parameter 1a: Changing the prior plan in the restoration method

All prior plans achieved adequate target coverage for all target regions. Table 6.2 shows the number of energy layers and pencil-beams in the different prior plans.

Table 6.2: Average number of energy layers and pencil-beams in the different prior plans.

	Small prior	Wide prior (default)	Robust prior
# energy layers	73 (59 – 80)	84 (65 – 96)	84 (65 – 95)
# spots	1268 (1129 – 1358)	1697 (1508 – 1856)	1682 (1560 – 1838)

Applying plan restoration using the ‘small’, ‘wide’ (default) and ‘robust’ prior plan resulted in $V_{95\%}$ success rates of 64.8%, 100% and 100%, respectively. For the $V_{107\%}$ the success rates were 67.0%, 76.1% and 0.0%, respectively. With the small prior $V_{107\%}$ values up to 5.4% were obtained for the high-dose region, and values up to 3.4% were obtained for the low-dose region. With the wide prior $V_{107\%}$ values up to 3.7% were obtained for the low-dose region. Using the robust prior resulted in $V_{107\%}$ values for the low-dose region between 30.5% and 53.8% (average value of 42.7%).

Figure 6.1 shows the achieved values of the evaluation criteria for the three methods, with the first two columns depicting the target coverage success rate. Largest differences were observed for the $V_{45 \text{ Gy(RBE)}}$ bladder and $V_{45 \text{ Gy(RBE)}}$ rectum. Using the small prior instead of the wide default prior resulted in a decrease in the median values of 22.6%-point and 20.6%-point, respectively. Changing from the wide prior to the robust prior further increased these median values with 9.7%-point and 7.7%-point for rectum and bladder, respectively. Overall, the differences in obtained values were larger between the small and wide prior than between the wide and robust prior. For all evaluation criteria of the OARs the differences between taking the small and default wide prior, as well as the differences between taking the wide and robust prior were statistically significant.

Parameter 1b: Changing the prior plan in the full adaptation method

Full adaptation resulted in a success rate of 100% for all three evaluated priors when adding 1 iteration of spots. Taking the wide and robust priors always resulted in $V_{107\%} \leq 2\%$. When using the small prior, 172/176 plans obtained a $V_{107\%} \leq 2\%$.

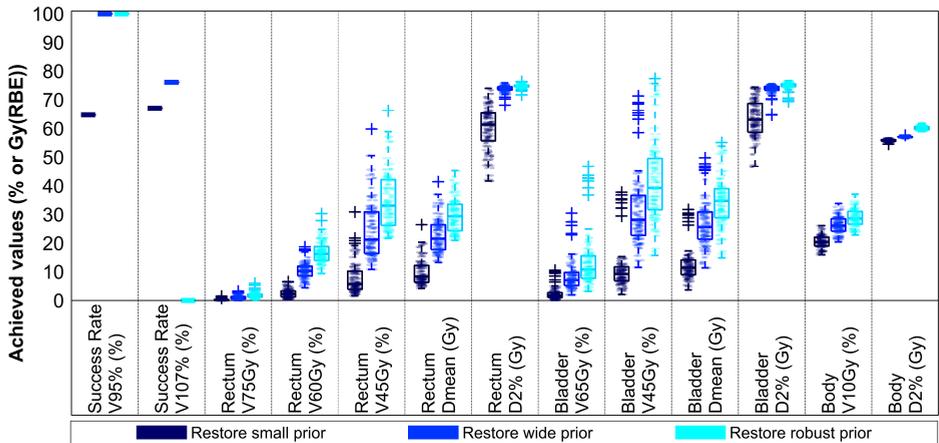


Figure 6.1: Boxplots depicting the obtained dosimetric parameter values for target and OARs using the dose restoration method with the three different priors. The first two columns show the percentage of scans that meet the target constraints for all target regions.

Two plans obtained $V_{107\%} > 2\%$ for the high-dose region (values 6.3% and 7.5%, respectively), the other two obtained $V_{107\%} > 2\%$ for the low-dose region (values 7.0% and 8.6%, respectively).

Figure 6.2 shows the achieved values of the evaluation criteria after full adaptation using the three different priors. The first columns depict the target coverage success rate.

Using the small prior plans mostly resulted in lower dosimetric parameter values than taking the wide (default) prior plan, with a maximum improvement of the median value of the bladder $V_{45 \text{ Gy(RBE)}}$ of 5.0%-point. The whole body $D_{2\%}$, however, was slightly lower in median value for the wide prior (55.4 Gy(RBE) compared to 55.6 Gy(RBE)). For all evaluation criteria of the OARs the differences between taking the small and wide prior were statistically significant

Taking the robust instead of the wide prior resulted in smaller differences, with increases (up to 3.2%-point in median value) in the volumes receiving lower dose values. For most evaluation criteria of the OARs the differences between taking the wide and robust prior were statistically significant. Non-significant differences were found for the rectum $V_{75 \text{ Gy(RBE)}}$, the rectum $V_{60 \text{ Gy(RBE)}}$ and the bladder $D_{2\%}$.

Parameter 2: Changing the sample size

Running the full adaptation with a sample size of 500 yielded adequate target coverage for all scans, and a $V_{107\%} \leq 2\%$ for 155/176 plans. For the high-dose region

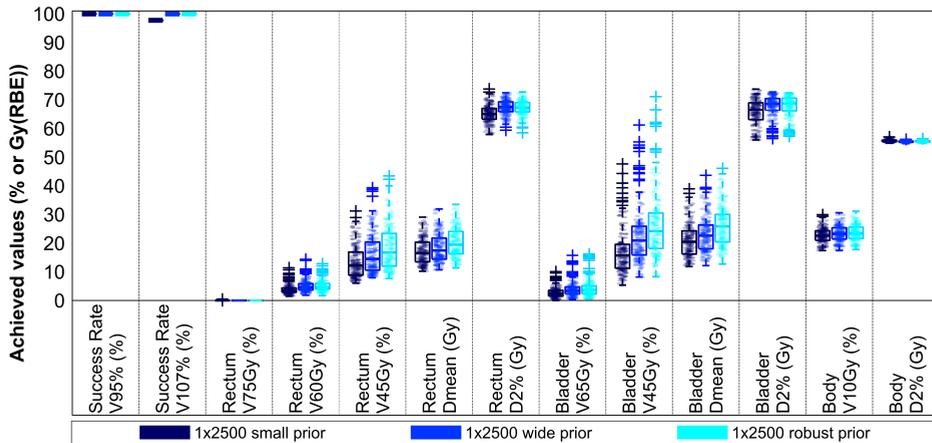


Figure 6.2: Boxplots depicting the obtained dosimetric parameter values using the full adaptation with the three different priors. Adaptation included one iteration of spot addition and spot-intensity optimization. The first two columns show the percentage of scans that met the target constraints for all target regions.

one plan obtained a $V_{107\%} > 2\%$, with a value of 4.2%. For the low-dose region, the 21 plans not achieving the goal obtained values between 5.0% and 66.8%, with an average value of 29.2%.

Running the full adaptation with a single iteration of spot addition using sample sizes of 1500, 2500, 3500, or 4500 yielded adequate target coverage for all scans, as well as yielding $V_{107\%} \leq 2\%$ for all scans.

Figure 6.3 shows the achieved values of the evaluation criteria after full adaptation with a single iteration of spot addition, using the different sample sizes. It can be seen that the high dose values were reduced when using a larger sample size, but that the mean dose of the bladder remained similar. The largest reduction in median value when increasing the sample size from 500 to 4500 was found for the bladder $V_{45\text{Gy(RBE)}}$ with an improvement of 3.3%-point.

The differences between taking a sample size of 500, 1500, 3500 or 4500 and taking the default sample size of 2500 were almost all statistically significant. The only non-significant differences were found for the bladder D_{mean} for a sample size of 500, the rectum $V_{75\text{Gy(RBE)}}$ for sample sizes of 1500, 3500 and 4500, and for the body $V_{10\text{Gy(RBE)}}$ for a sample size of 1500.

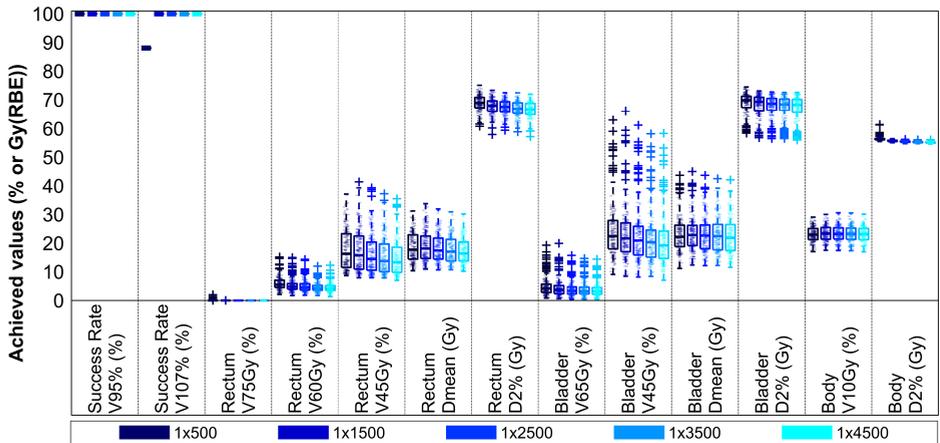


Figure 6.3: Boxplots depicting the obtained dosimetric parameter values using the full adaptation when using different sample sizes. Adaptation included one iteration of spot addition and spot-intensity optimization. The first two columns show the percentage of scans that meet the target constraints for all target regions.

Parameters 3 & 4 - Changing the stopping criterion and the optimization approach

Running the full adaptation method with a single iteration of spot addition and weight optimization resulted in adequate target coverage for all plans both when using the RPM as optimizer and when using the 2pec optimizer. All plans also achieved $V_{107\%} \leq 2\%$. The same was obtained when running the adaptation until no improvements larger than 3% were observed. The RPM required on average 6 iterations of spot addition and spot-weight optimization (range 3 – 12) to meet this stopping criterion. Using the 2pec optimizer required on average 7 iterations (range 4 – 10).

Figure 6.4 shows boxplots depicting the achieved values for the evaluation criteria when running the adaptation with a single iteration of spot addition for both optimizers (1×2500), as well as running the adaptation until no improvements larger than 3% were observed for both optimizers ($n \times 2500$).

For a single iteration the 2pec optimizer mostly resulted in lower evaluation criteria values than the RPM optimizer. Improvements in median values of up to 7.5% were observed for the $V_{45 \text{ Gy(RBE)}}$ and D_{mean} of both the rectum and bladder. A higher values for the 2pec optimizer was only observed for the whole body $D_{2\%}$.

Running the method until no improvement larger than 3% is obtained resulted in similar differences between the optimizers. The largest improvement when using 2pec instead of RPM was seen for the mean dose of the bladder, with a reduction of 13.1 Gy(RBE). Again, the whole body $D_{2\%}$ was slightly higher for the 2pec approach.

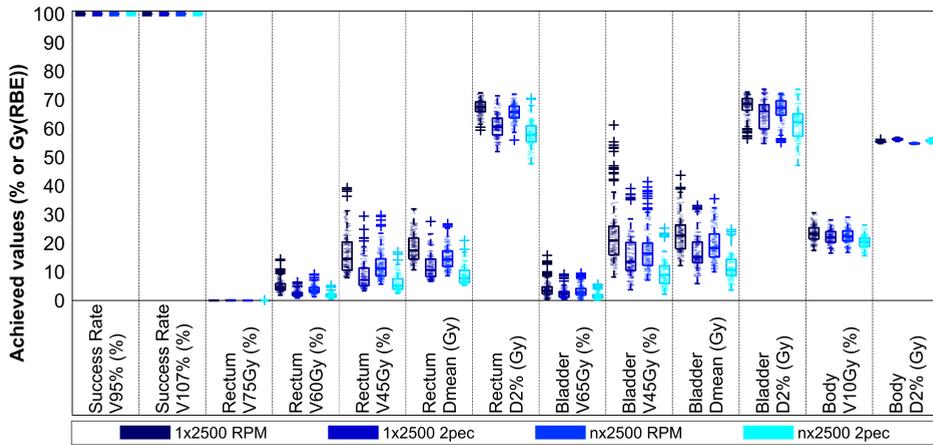


Figure 6.4: Boxplots depicting the obtained dosimetric parameter values when using either the 2pec optimizer or the RPM. Results are shown for when running the full adaptation with a single iteration of spot addition (1×2500), and for when running until no improvement larger than 3% is observed ($n \times 2500$). The first two columns show the percentage of scans that meet the target constraints for all target regions.

Comparing the results obtained with a single iteration and multiple iterations, it can be seen that applying the 2pec optimizer with a single iteration already achieved lower doses than applying the RPM optimizer with n iterations. Using the 2pec optimizer in more iterations further reduced the dosimetric parameter values.

Compared to the default setting of a single iteration using the RPM, all variations yielded statistically significant differences for the evaluation criteria of the OARs. The only exception was the difference for the rectum $V_{75 \text{ Gy(RBE)}}$ between a single and n iterations of using the RPM.

Calculation times

Table 6.3 shows the average calculation times of the different steps in the restoration method when using the three different prior treatment plans.

Table 6.4 provides the average calculation times of the different steps in the full adaptation when using the three different prior treatment plans, when using the different sample sizes in adapting the default prior plan, and when changing the stopping criterion and optimization approach. The times required for the energy-adaptation steps are omitted, as these can be found in Table 6.3.

Table 6.3: Average calculation times for the simple dose restoration.

Prior		Initialization <i>minutes</i> (range)	Energy- adaptation <i>seconds</i> (range)	Calculation of dose deposition matrix <i>minutes</i> (range)	Weight optimization <i>seconds</i> (range)
Default	Wide	1.1 (0.8 – 1.6)	6.8 (5.5 – 8.0)	1.3 (1.1 – 1.5)	7.5 (5.9 – 9.9)
Change the prior	Small	0.9 (0.7 – 1.1)	6.3 (5.1 – 7.8)	1.1 (0.9 – 1.2)	4.4 (3.5 – 5.6)
	Robust	0.9 (0.8 – 1.1)	6.3 (4.8 – 8.3)	1.3 (1.0 – 1.5)	7.1 (6.0 – 9.9)

Table 6.4: Average calculation times for the full adaptation method. The intermediate dose calculation to determine the dose deposition matrices of the new spots is included in the spot addition step.

	Prior and New spots	Beam setup <i>seconds</i> (range)	Spot addition <i>minutes</i> (range)	Optimization <i>minutes</i> (range)
Default	Wide + 1×2500	1.3 (0.9 – 5.8)	2.2 (1.3 – 5.1)	47.1 (39.6 – 57.4)
Change the prior	Small + 1×2500	1.2 (0.9 – 1.8)	2.1 (1.2 – 5.1)	39.0 (30.0 – 47.5)
	Robust + 1×2500	1.2 (0.8 – 1.8)	2.1 (1.2 – 5.1)	47.0 (40.2 – 59.6)
Change the sample size	Wide + 1×500	1.3 (0.9 – 4.3)	1.2 (0.8 – 1.8)	25.4 (20.7 – 31.2)
	Wide + 1×1500	1.2 (0.9 – 2.5)	1.6 (1.0 – 2.4)	34.7 (27.6 – 42.4)
	Wide + 1×3500	1.2 (0.8 – 1.9)	2.3 (1.2 – 4.1)	66.0 (55.3 – 80.0)
	Wide + 1×4500	1.2 (0.9 – 1.7)	2.6 (1.5 – 5.9)	90.1 (71.8 – 106.6)
Change the stopping criterion & optimization approach	Wide + 1×2500 2pec	1.2 (0.9 – 1.9)	2.2 (1.3 – 5.1)	187.2 (161.6 – 214.4)
	Wide + n×2500 RPM	1.2 (0.9 – 1.8)	14.3 (6.1 – 28.6)	1112.8 (372.9 – 3305.2)
	Wide + n×2500	1.3	14.5	5356.9
	2pec	(0.9 – 6.7)	(9.1 – 30.0)	(2598.1 – 10123.9)

6.4 DISCUSSION

In this study we have investigated and presented variations on the default parameter settings for the previously developed online adaptation methods, and compared the results to those obtained using the previously chosen default settings. We investig-

ated the effect of changing the amount of included motion mitigation in the prior plan for both the restoration and full adaptation method. For the full adaptation method, we also investigated the effect of changing the sample size, removing the time constraints, and changing the optimization approach.

Changing the prior treatment plan had a large influence on the results of the restoration method in terms of target coverage and OAR sparing. As this method aims to return to the prior dose distribution, this result was expected. Without additional motion mitigation in the prior treatment plan, i.e. in this case taking the small prior plan, resulted in inadequate target coverage in some of the dose restorations. Including too much motion mitigation, i.e. in this case taking the robust prior plan, led to a $V_{107\%} > 2\%$ for all plans in the low-dose region. Using the robust prior also resulted in higher values for the OARs than using the default prior. The high $V_{107\%}$ values can be explained by the proximity of the optimized high/intermediate-dose region to the evaluated low-dose region. High $V_{107\%}$ values were also observed for the robust treatment plans in the nominal scenario on the pCTs. Apparently, the extra dose is needed to obtain sufficient target coverage in the error scenarios. Calculation times only varied slightly between the different prior plans, where taking a prior plan with fewer energy layers and spots (Table 6.2) led to shorter dose calculation times and faster weight optimizations (Table 6.3).

For the full adaptation approach, smaller differences were observed. Using the different prior plans always resulted in adequate target coverage for all targets, although using the small prior resulted in $V_{107\%} > 2\%$ in 4/176 plans. With values below 10% and the high dose focused in the target regions, these values are however still acceptable from a clinical perspective. Considering the OAR doses, Figure 6.2 shows that including less motion mitigation in the prior plan leads to more OAR sparing. A possible explanation for this is that when taking the smaller prior plan, the prior spots are more focused in the target area. As the new spots only fill in the 'gaps' in the spot distribution of the target area, fewer spots will be located near the OARs. A similar result has been found in previous work using CT data of cervical cancer patients, where taking a plan-library of smaller plans yielded better results than taking a single large plan as prior [57]. This indicates that even with the new spots, the optimizer needs the original spots to obtain sufficient target coverage. Increasing the sample size or increasing the number of iterations could mitigate this effect. Changing the prior plan only influenced the running time of the RPM optimization (Table 6.4), again likely caused by the number of spots and energy layers.

We found that only a small number of new spots is needed to yield adequate target coverage for scans in this dataset. Increasing the sample size showed to have most

effect on dose conformity; reducing the volumes receiving high doses. Less effect was seen on the lower dose values and mean dose. Calculation times proved not to increase linearly with the sample size (Table 6.4). Adding a sample size of 500 twice took approximately the same time as adding a sample size of 2500 once.

Changing the optimizer from RPM to 2pec showed to have large effects on the generated treatment plans. Both using a single iteration of spot addition, and running until no improvement larger than 3% was observed, most dosimetric evaluation criteria obtained lower values when the 2pec was used. Only for the high doses the RPM achieved lower values. It should be noted though, that all plans that were generated using the RPM and 2pec are Pareto optimal. While Figure 6.4 seemingly shows that the RPM plans are of lower quality than the 2pec plans, evaluation of other dosimetric parameters show the opposite. For example, the maximum dose in OARs was lower when using the RPM compared to 2pec. Regarding the mean dose in OARs, we found that the RPM obtained a lower mean dose in the femoral heads, while the mean dose for rectum and bladder was higher. This shows that the trade-offs between the RPM and the 2pec optimizer were different. Another consequence of this is that the results of 2pec cannot be replicated by increasing the number of RPM iterations.

This result shows that the tuning of the RPM-parameters, which aims at forcing the RPM to make similar trade-offs as the 2pec optimizer, should be improved. Taking the same pencil-beam sets, such as tested with the single iteration, the RPM and 2pec optimizers still yield different trade-offs. Interestingly, when taking the spots selected within a 2pec optimization (such as the spots used during the tuning) the RPM does yield similar results to the 2pec optimizer. This indicates that the RPM-parameter tuning may have been focused too much on a specific pencil-beam set; being the set selected using the 2pec optimization. We conclude that the obtained parameters are not sufficiently robust to handle different pencil-beam sets. This may be improved by performing the RPM-parameter tuning on treatment plans with various pencil-beam sets. By including multiple plans per CT scan, we expect that the RPM-parameters obtained becomes more robust against differences between pencil-beam sets. Possibly the adaptation itself could also be added to the tuning, in which case the difference between an adapted plan based on the plan of a different scan and the fully optimized plan can be minimized as well.

Besides these changes, it should also be noted that the protocol used for tuning (Table A.6.1 in the appendices) might not yet be optimal and that training has been done on only a small number of scans. For this study we tuned using three-fold cross validation, training only on a third of the data. The training sets consisted of 30, 34 and 35 scans, respectively. For all variations it should furthermore be noted that the effect

they yield on the output plans are likely to be highly dependent on the investigated dataset.

Considering the running times, it can be seen in Table 6.4 that using the *2pec* approach takes longer than using the RPM. This was expected, as the *2pec* approach optimizes the objectives sequentially in a prioritized way, while the RPM only requires a single optimization. Running the method until no improvement larger than 3% is observed takes about 10 – 20 times longer. An explanation for this is that the optimization time depends on the number of pencil-beams, which increases with each spot-addition iteration.

While we have investigated the variations separately, they will probably be correlated in their effect on the adapted plans. For example, we have observed that when using full adaptation with a sample size of 2500, a smaller prior plan can be used while still yielding adequate treatment plans. We have also observed that when taking the default wide prior plan, the sample size can be reduced to 500 while still obtaining adequate target coverage. However, combining a sample size of 500 with a smaller prior plan is not expected to yield adequate results.

It should be noted that while this study has addressed many variations on the default parameter settings, not all possibilities have been investigated. An example of this is in the generation of the prior plans. While we have varied the amount of robustness included in the prior plans, all prior plans were made using our in-house TPS. The optimization in our in-house TPS features a spot-reduction loop to shorten delivery times [20]. On average the default prior plans included around 1700 pencil-beams (Table 6.2). It is possible that if less spot-reduction is performed in generating the prior plans, the need for new spots in the adaptation is reduced. While this could allow a reduction of the sample size in full plan adaptation, the impact on the overall calculation time would be small as the total number of spots in the adaptation would remain similar.

6.5 CONCLUSION

For the dose restoration approach, we recommend maintaining the default setting. This is based on the observation that changing the amount of motion mitigation accounted for in the prior plans can yield insufficient target coverage or increased OAR doses. Full plan adaptation appeared to be more robust against variations of the parameter settings, except for the choice of optimization method. To minimize the impact of the choice optimization method, a more advanced tuning of the RPM-parameters is recommended.

6.6 ACKNOWLEDGMENTS

The CT-data with contours were collected at Haukeland University Hospital, Bergen, Norway and were provided to us by responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing. This study was financially supported by ZonMw, the Netherlands Organization for Health Research and Development, grant number 104003012 and by Varian Medical Systems.

6.7 CONFLICTS OF INTERESTS

Varian Medical Systems has in part financed this research. Erasmus MC Cancer Institute also has research collaborations with Elekta AB, Stockholm, Sweden and Accuray Inc., Sunnyvale, USA.

Chapter

7

Robust contour propagation using deep learning and image registration for online-adaptive proton therapy of prostate cancer

Medical Physics, Vol. 46, pp 3329-43, 2019,

doi: <https://dx.doi.org/10.1002/mp.13620>

Mohamed S. Elmahdy¹, Thyrsa Z. Jagt², R. Th. Zinkstok³
Yuchuan Qiao¹, Rahil Shahzad¹
Hessam Sokooti¹, Sahar Yousefi¹
Luca Incrocci², C.A.M. Marijnen³
Mischa S. Hoogeman^{2,5}, Marius Staring^{1,3,5}

¹ Division of Image Processing, Department of Radiology,
Leiden University Medical Center, The Netherlands

² Department of Radiation Oncology, Erasmus MC Cancer Institute,
Rotterdam, The Netherlands

³ Department of Radiation Oncology, Leiden University Medical Center,
The Netherlands

⁴ Department of Medical Physics & Informatics, HollandPTC,
Delft, The Netherlands

⁵ Intelligent Systems Department, Delft University of Technology, The Netherlands

Abstract

Purpose: To develop and validate a robust and accurate registration pipeline for automatic contour propagation for online-adaptive intensity-modulated proton therapy (IMPT) of prostate cancer using *elastix* software and deep learning.

Methods and materials: A 3D convolutional neural network was trained for automatic bladder segmentation of the CT scans. The automatic bladder segmentation alongside the CT scan are jointly optimized to add explicit knowledge about the underlying anatomy to the registration algorithm. We included three datasets from different institutes and CT manufacturers. The first was used for training and testing the ConvNet, where the second and the third were used for evaluation of the proposed pipeline. The system performance was quantified geometrically using the dice similarity coefficient (DSC), the mean surface distance (MSD), and the 95% Hausdorff distance (HD). The propagated contours were validated clinically through generating the associated IMPT plans and compare it with the IMPT plans based on the manual delineations. Propagated contours were considered clinically acceptable if their treatment plans met the dosimetric coverage constraints on the manual contours.

Results: The bladder segmentation network achieved a DSC of 88% and 82% on the test datasets. The proposed registration pipeline achieved a MSD of 1.29 ± 0.39 , 1.48 ± 1.16 , and 1.49 ± 0.44 mm for the prostate, seminal vesicles, and lymph nodes, respectively on the second dataset and a MSD of 2.31 ± 1.92 and 1.76 ± 1.39 mm for the prostate and seminal vesicles on the third dataset. The automatically propagated contours met the dose coverage constraints in 86%, 91%, and 99% of the cases for the prostate, seminal vesicles, and lymph nodes, respectively. A conservative success rate (CSR) of 80% was obtained, compared to 65% when only using intensity-based registration.

Conclusion: The proposed registration pipeline obtained highly promising results for generating treatment plans adapted to the daily anatomy. With 80% of the automatically generated treatment plans directly usable without manual correction, a substantial improvement in system robustness was reached compared to a previous approach. The proposed method therefore facilitates more precise proton therapy of prostate cancer, potentially leading to fewer treatment related adverse side effects.

7.1 INTRODUCTION

Prostate cancer is one of the leading causes of mortality and the most common cancer among men. The National Cancer Society (NCS) estimates around 164,690 new cases and 24,430 deaths from prostate cancer in the United States only for 2018 [65]. Due to its slow progress, individuals could develop prostate cancer for many years without explicit signs. There are treatment options for prostate cancer including surgical removal of the prostate, hormone therapy, and radiotherapy. Intensity-modulated proton therapy (IMPT) is able to deliver a highly localized dose distribution to the target volume, while minimizing collateral damage to the surrounding healthy tissues [66]. IMPT is however more sensitive to daily changes than photon therapy, which may result in distortion of the delivered dose distribution [1, 17]. These changes could arise from anatomical variations in the shape and position of both target volumes and organs at risk (OARs) or a misalignment in the patient setup. In order to compensate for these changes, a margin is added to the clinical target volume (CTV) to generate the planning target volume (PTV) in addition to robust treatment planning. These margins result in extra dose to the OARs, leading to an increase in the treatment-related toxicities that may prevent dose escalation. Traditionally, motion-induced variations are minimized by implanting fiducial markers in the prostate, subsequently compensating for the daily prostate motion using online imaging [67]. However, such correction strategies are invasive and only capable of correcting for translational motion and limited amount of rotational motion [68]. Online imaging and replanning should be able to handle this problem without using fiducial markers [69]. These online CT scans have to be delineated first in order to update the treatment plan. Usually this task is done by radiation oncologists according to certain guidelines [70, 71]. However, intra and inter-observer inconsistency has been noted due to different preferences and experience among radiation oncologists [72, 73]. Typically, daily manual recontouring is not performed because it is time consuming and new anatomical variations may be introduced in the time it takes to delineate the scan [74]. Automatic recontouring algorithms can alleviate these issues, but robust methods are required, because otherwise still time consuming fallback strategies are needed.

Automatic recontouring could be accomplished effectively using deformable image registration (DIR) by deducing the correspondence between the daily CT scan and the planning CT scan. Using the generated deformation vector field (DVF), manual contours can be propagated from the planning CT scan to the daily CT scan. The automatically generated contours together with fast re-optimization of the treatment plan [44] could compensate for the daily variation and ensure the delivery of the pre-

scribed dose distribution at small margins and robustness settings. DIR is a crucial step towards developing online-adaptive IMPT alongside replanning and personalized dose quality assurance (QA). Currently, these steps are time consuming, thus severely limiting online procedures.

There are commercially available applications for automatic recontouring such as atlas based auto segmentation (ABAS), Mirada, and RayStation. These applications are, however, considered a black box for the end-users and therefore limit the parameter choices and tuning. Open source DIR packages provide a high level of flexibility with a concrete scientific evidence and reproducibility. Qiao *et al.* [75] reported a mean surface distance (MSD) of 1.36 ± 0.30 mm, 1.75 ± 0.84 mm, 1.49 ± 0.44 mm for the prostate, seminal vesicles, and lymph nodes, respectively for 18 patients using the open source `elastix` software. A clinical success rate of 69% was achieved, which means that 31% of the delineations have to be corrected, leading to increased costs and a suboptimal patient workflow. In 2011, Thor *et al.* deployed DIR to propagate the contours of the prostate and OARs from CT to cone beam CT [76]. The system achieved a mean dice similarity coefficient (DSC) of 0.80 for the prostate, 0.77 for the rectum, and 0.73 for the bladder with a relatively high variance. Moreover, the system was not qualitatively evaluated in terms of dosimetric coverage. Recently, Woerner *et al.* [77] investigated the error between different radiologists and both DIR and rigid registration in different body regions. They only reported the results for the prostate, which were 0.90, 0.99 mm, and 8.12 mm for the DSC, MSD, and Hausdorff distance (HD), respectively. Thörnqvist *et al.* [78] used two different demons-based registration algorithms, with one more conservative than the other. They achieved an average DSC of 0.88, 0.85, 0.89, 0.78 for the lymph nodes, prostate, bladder, and rectum, respectively.

In spite of the existence of quite accurate registration algorithms, they still suffer from a lack of robustness, which is a critical aspect for clinical application. Therefore, in this paper we focus on the robustness aspect of the registration pipeline. The main challenges in Qiao *et al.* [75] were the presence of gas pockets and large deformations surrounding the seminal vesicles, bladder, and rectum. Hence, we propose to tackle these challenges by inpainting the rectum gas pockets as well as embedding the bladder segmentation in the registration pipeline using deep learning to enhance the system's robustness. The proposed registration pipeline was evaluated geometrically and dosimetrically for generating clinically acceptable IMPT plans. Compared to our conference paper [79], we made several improvements, such as the inclusion of more datasets, dealing with gas pockets, data normalization, and multi-stage registration. Moreover, we carried out an extensive dosimetric validation for the auto-

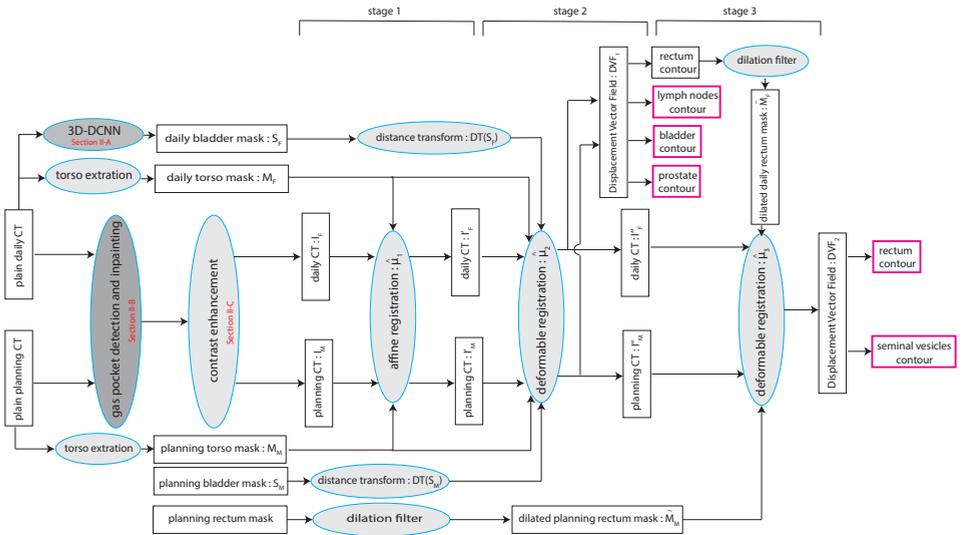


Figure 7.1: The proposed multi-stage registration process using elastix software and deep learning. The red boxes denote the contours finally used as output of the algorithm.

matically generated contours to verify its clinical viability.

7.2 METHODS

The prostate and seminal vesicles are positioned between the bladder and the rectum, therefore prostate motion is mainly influenced by the filling and motion of both the bladder and the rectum [80]. Hence, we hypothesize that embedding an explicit prior knowledge about the deformation of either organs to the intensity-based DIR method may improve the accuracy and robustness of the registration. Here, we considered the bladder because it has a well-defined shape that could be more easily delineated in a fully automatic manner than the rectum. Since the registration is intensity-based, the quality of the registration process is correlated to the quality of the input images. Hence, we introduced multiple data preprocessing steps to enhance the quality of the input images. These steps include rectum gas pocket detection and inpainting and contrast clipping as shown in Figure 7.1.

7.2.1 Bladder segmentation using deep learning

In this study, we automatically segment the bladder using a 3D U-net convolutional neural network (3D-CNN) similar to the architecture introduced in [81]. The network consists of encoding and decoding branches connected with skip connections as shown in Figure 7.2. In order to represent the volumetric information and tissue

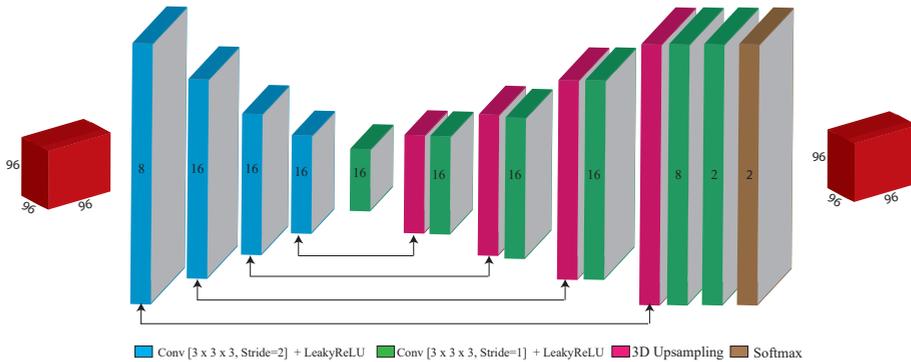


Figure 7.2: The architecture for the 3D-CNN network, where the numbers on the blocks denote the number of feature maps.

homogeneity of the CT volume, 3D convolution layers followed by non-linear leaky rectified linear units were used. The original maxpooling layers were replaced by strided convolution in both encoder and decoder branches. Negative dice similarity coefficient (DSC) [82] is deployed as a cost function and the network is trained using the Adam optimizer [83] with a fixed learning rate of 10^{-4} . The network has 64,320 trainable parameters which enables network inference of the entire CT image in approximately 2 seconds. The network was designed to output the same size as input, however the input size should be divisible by 16. Largest connected component analysis was applied as a post-processing step to eliminate irrelevant activations.

7.2.2 Gas pocket detection and inpainting

A problem that usually arises for intensity-based DIR of the pelvic region is the presence of gas pockets in the bowel and rectum. These pockets appear as dark areas surrounded by soft tissue. Usually the size and position of these pockets are not the same in the planning and the daily CT scans. In such situations, physical correspondence between images at different sessions does not exist because of the insertion or occlusion of image content. Only few studies addressed this issue in the literature. Gao *et al.* [84] proposed introducing a virtual gas pocket to the planning CT scan that follows the pocket in the daily CT scan. They tested it on 15 prostate cancer patients with distended rectum. Foskey *et al.* [85] proposed to deflate the pocket to a virtual point. In both papers, the authors assumed no gas pockets in the planning CT scan, which is not usually the case.

Recently, deep learning based algorithms have revolutionized the medical image analysis field [86]. One category of deep learning architectures is generative adversarial

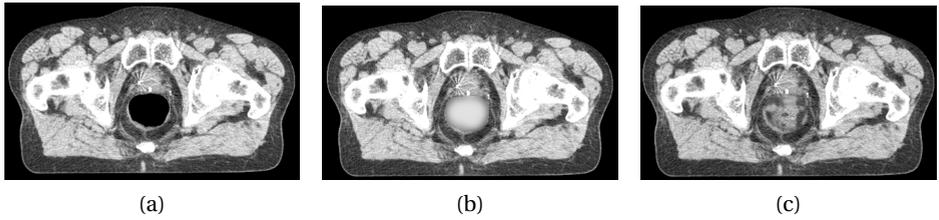


Figure 7.3: Different inpainting algorithms, where (a), (b), and (c) represent the original CT, the result from simple-inpainting, and the result from GAN-inpainting, respectively.

networks (GANs) introduced by Goodfellow *et al.* [87] in 2014. GANs have been growing since then in generating realistic natural and synthetic images. As for medical images, GANs have been used in image segmentation [88], synthesis [89], registration [90], and denoising [91]. Recently Yu *et al.* [92] proposed a 2D GAN network with a contextual attention model to restore and inpaint occluded regions in natural images. The network also blends the restored region with the surrounding texture to make it look more realistic. The proposed model has two successive networks for image generation in order to generate patches with fine quality. The first ‘generator’ network generates a coarse result through a dilated convolution network. This result is then fed to the second network. The second ‘discriminator’ network has two routes, one goes to a dilated convolution network while the other goes through a contextual attention model. Finally, the results from these two routes are concatenated and fed to a prediction network. This network has shown an improvement over a similar network proposed by Iizuka *et al.* [93]. In this paper, we retrained this network so that it can inpaint (fill) gas pockets of different shapes and sizes with a more sophisticated and realistic content rather than a fixed value. The same implementation and hyper parameters were used as in the original paper.

Alternatively, we also experimented with a simplified method for inpainting. Following the idea proposed by Rodriguez-Vila *et al.* [94] we fill the gas pockets with a fixed value and smooth the output to blend it with the surrounding tissues. A threshold of -200 is used to generate a binary mask of the gas pockets. This mask is then dilated with a kernel of size $7 \times 7 \times 1$ voxels (M) while the CT image is filled with a fixed HU number of 60 (the average HU number for faeces), and smoothed with a sigma of 4 mm ($I_{smoothed}$). Equation (7.1) shows the simple-inpainting process:

$$I_{out} = I_{input} \times (1 - M) + I_{smoothed} \times M \quad (7.1)$$

Figure 7.3 shows a comparison between gas pocket inpainting using the GAN network and simple-inpainting.

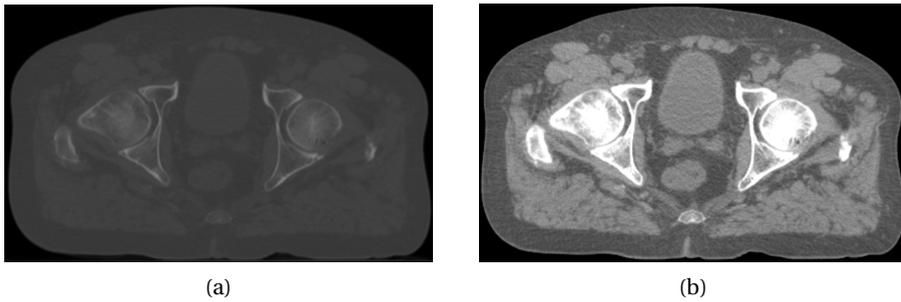


Figure 7.4: The effect of contrast clipping, where (a) and (b) represent the image before and after intensity clipping, respectively.

7.2.3 Contrast Enhancement

To enhance the soft tissue contrast, the CT intensity was clipped to the range of $[-300, 300]$. This clipping is similar to viewing the soft tissue with an appropriate window level. Moreover, such enhancement improves the registration convergence. Figure 7.4 shows the effect of intensity clipping.

7.2.4 Image Registration

For carrying out the DIR experiments, we used the open software package `elastix` [95]. For more details, see the website <http://elastix.isi.uu.nl>. All the experiments were performed on a cluster of workstations operated on the oracle grid engine (OGE), which has 500 nodes with a total of 800 cores. Testing time is reported using a PC with 16 GB memory, Windows 7 Professional 64 bit operation system and an Intel Xeon E51620 CPU with 4 cores at 3.6 GHz, utilizing only the CPU.

In this study, the planning CT scan (moving image) was aligned with the daily CT scan (fixed image) of each patient. The registrations were initialized based on the center-of-gravity of the bony anatomy defined by a Hounsfield number larger than 200. A mask of the body torso was generated using Pulmo software [96] to remove the effect of the CT table. The registration process is done in three stages. First, the moving and fixed images are registered using a single resolution affine transformation using 200 iterations as defined in Equation (7.2):

$$\hat{\mu}_1 = \arg \min_{\mu} C_1(I_F, I_M, M_F, M_M; T_{\mu_1}), \quad (7.2)$$

where I_F is the daily scan, I_M is the planning scan, M_F is the torso mask of the daily scan, M_M is the torso mask of the planning scan, and C_1 is the mutual information cost function. The affine transformation aligns the bones and large structures. Second, a deformable registration is applied to tackle the local deformations of the organs. In

this stage, the planning CT scan of each patient combined with the manual delineation of the bladder are considered the moving images, while the repeat CT scan of the same patient accompanied with the bladder segmentation resulting from the proposed 3D-CNN are the fixed images. Equation (7.3) defines the optimization problem for this stage:

$$\hat{\mu}_2 = \arg \min_{\mu} \{C_1(I_F, I_M, M_F, M_M, T_{\mu_1}; T_{\mu_2}) + \alpha C_2(DT(S_F), DT(S_M), T_{\mu_1}; T_{\mu_2})\}, \quad (7.3)$$

where C_2 is the mean squared difference (MSD) cost function, α is a weight for balancing these two cost functions, $DT(S_F)$ is the distance transform of the 3D-CNN bladder segmentation, and $DT(S_M)$ is the distance transform of the manual annotation of the planning scan. The distance transform (DT) of the bladder segmentations is used instead of the binary segmentations themselves, to ensure a smooth and stable optimization process. The generated deformation vector field (DVF) from this step is then used to propagate the contours of the prostate, lymph nodes, bladder, and rectum from the planning CT scan to the repeat CT scan. Because the seminal vesicle is a small irregular structure, which is highly affected by the deformation in the rectum, we introduce a third stage to focus the registration on the rectum and seminal vesicle region. In this stage, the rectum contour of the planning CT scan and the rectum contour of the daily CT scan (from the previous stage) are dilated with a kernel of $45 \times 45 \times 1$ voxels and used as a registration mask together with the fixed and moving CT scans. The contours of the rectum and seminal vesicles are then propagated using the generated DVF from the final stage. Equation (7.4) defines the optimization problem for this stage:

$$\hat{\mu}_3 = \arg \min_{\mu} C_1(I_F, I_M, \tilde{M}_F, \tilde{M}_M, T_{\mu_1}, T_{\mu_2}; T_{\mu_3}), \quad (7.4)$$

where \tilde{M}_M is the dilated rectum mask of the planning CT scan and \tilde{M}_F is the dilated rectum mask of the daily CT scan. A fast recursive implementation of the B-spline transformation was employed for DIR [97] in stage 2 and 3. Adaptive stochastic gradient descent was used for optimization [98] in all three stages. For the DIR stage we used a three level Gaussian pyramid with smoothing factors of 4, 2, and 1 mm. Figure 7.1 illustrates the proposed registration pipeline in detail.

7.3 EXPERIMENTS AND RESULTS

7.3.1 Dataset

This study includes three datasets representing three different institutes and CT scanners from three different vendors for patients who underwent intensity-modulated

Table 7.1: Details of the datasets reported in this study. LUMC, EMC, and HMC are abbreviations for Leiden University Medical Center (Netherlands), Erasmus Medical Center (Netherlands), and Haukeland Medical Center (Norway), respectively. SV and LN denote Seminal Vesicles, and Lymph Nodes, respectively.

Institute	Scanner	#Patients	#Scans/ patient	Image size	Voxel spacing (mm)	Manual delineations
LUMC	Toshiba	418	1	512x512x(68-240)	~1.0x1.0x3.0	bladder, rectum
EMC [99]	Siemens	14	4	512x512x(91-218)	~0.9x0.9x1.5	prostate, SV bladder, rectum
HMC [100]	GE	18	8-11	512x512x(90-180)	~0.9x0.9x2.0	prostate, SV, LN bladder, rectum

radiation therapy for prostate cancer. Table 7.1 shows detailed information about these datasets. The LUMC dataset was used to train and validate the neural network for segmenting the bladder (Section 7.2.1) as well as the inpainting network (Section 7.2.2), while the EMC and HMC dataset were used as independent test sets for the complete registration pipeline. Geometric evaluation was performed on both the EMC and HMC dataset. Eleven out of the eighteen HMC patients were considered for dosimetric evaluation due to the availability of not only the manual delineations for the target organs (prostate, seminal vesicles, lymph nodes) and OARs (bladder, rectum), but moreover the manual delineations of the bowels and femoral heads needed for planning.

7.3.2 Evaluation measures

The quality of the registration is quantified in terms of geometric aspects and dosimetric coverage. The geometric quality is measured by comparing the manual contours and the automatically propagated contours of the daily CT scan for the prostate, lymph nodes, seminal vesicles, rectum, and bladder. The dice similarity coefficient (DSC) measures the overlap between the segmentations, while the mean surface distance (MSD), and the 95% Hausdorff distance (HD) measure the residual distance between the contours in 3D space.

$$DSC = \frac{2|F \cap M|}{|F| + |M|}, \quad (7.5)$$

where F and M are the propagated contour and the ground truth contour, respectively.

$$MSD = \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n d(a_i, M) + \frac{1}{m} \sum_{i=1}^m d(b_i, F) \right), \quad (7.6)$$

$$HD = \max \left\{ \max_i \{d(a_i, M)\}, \max_j \{d(b_j, F)\} \right\}, \quad (7.7)$$

where $\{a_1, a_2, \dots, a_n\}$ and $\{b_1, b_2, \dots, b_m\}$ are the surface mesh points of the fixed and moving contours, respectively and $d(a_i, M) = \min_j \|b_j - a_i\|$. The geometrical success rate, as a marker for geometric robustness, is defined as the percentage of registrations with MSD < 2 mm (slice thickness): $\gamma = \frac{n}{N} \{\text{MSD} < 2 \text{ mm}\}$, where (N) is the total number of registrations performed.

IMPT plans were generated for 11 patients from the HMC dataset using both the manual and the automatic delineations. The plans were then evaluated on the manual delineations to investigate the clinical effect of the error between these two delineations. Erasmus-iCycle, an in-house developed treatment planning optimization system, [19, 20, 22–24] together with the Astroid dose engine were used to generate the IMPT plans. Erasmus-iCycle uses a multi-criteria optimization to generate a clinically desirable Pareto optimal treatment plan on the basis of a wishlist consisting of hard constraints and objectives. A small margin of 2 mm around the prostate and 3.5 mm around the lymph nodes and seminal vesicles is used to compensate for the marginal error of the propagated contours and to account for intra-observer variations in the manual contouring. These margins alone can not account for variations in shape and location of the target volumes. Dose was prescribed according to a simultaneously-integrated boost scheme in which the high-dose PTV (prostate + 2 mm margin) was assigned 74 Gy and the low-dose PTV (seminal vesicles and lymph nodes + 3.5 mm margin) 55 Gy, to be delivered using two laterally opposed beams. In order to avoid under-dosage, the optimization ensures that at least 98% of the target volumes receive at least 95% of the prescribed dose ($V_{95\%} \geq 98\%$). To avoid over-dosage the optimization ensures that less than 2% of the target volumes receive more than 107% of the highest prescribed dose ($V_{107\%} \leq 2\%$). To achieve a clinically acceptable result, automatically generated treatment plans from the propagated contours should still fulfill these goals. Hence, IMPT plans from the propagated contours are evaluated based on the manual contours. The clinical success rate, as a marker for geometric robustness, is defined as the percentage of registrations for which the prostate directly meets the dose treatment criteria: $\eta = \frac{n}{N} \{V_{95\%} \geq 98\%\}$. conservative success rate (CSR) is a more conservative measure of clinical success when all target volumes (the prostate, seminal vesicles and lymph nodes) meet this dosimetric criterion. For dosimetric coverage calculation $N = 99$.

7.3.3 Network training and performance

We implemented the 3D-CNN and GAN-inpainting networks using Tensorflow [101]. For training these networks, we used the LUMC dataset. This dataset was a sufficiently large dataset to be able to train the neural networks. Since the LUMC dataset

only had one CT scan per patient, it was not used for registration evaluation. From the 418 LUMC patients, 350 patients were used for network training, and 68 patients for validation. The trained network was then applied without modification to the CT scans in the EMC and HMC datasets. In order to account for the variations in voxel size between datasets and scans, all scans were resampled to a fixed voxel size of $1.0 \times 1.0 \times 2.0$ mm. For the 3D-CNN, 100,000 patches of size $96 \times 96 \times 96$ voxels were randomly extracted from the training volumes, making sure they are equally distributed between foreground and background. For the GAN-inpainting network, all the slices with gas pockets were eliminated from training. Moreover, all slices were resampled to a pixel size of 1.0×1.0 mm and centrally cropped to 256×256 pixels so that more patches could fit into memory as well as it would be beneficial for the network to learn the most relevant contextual information to the rectum. Randomly selected windows of size 64×64 pixels were occluded in order to train the network to inpaint these regions with a realistic content. Both the 3D-CNN and the 2D-GAN-inpainting network were trained for 100,000 iterations on the raw CT patches without any preprocessing except for resampling. All the experiments were carried out using an NVIDIA GTX1080 Ti with 11 GB of GPU memory. The 3D-CNN bladder segmentation network obtained a DSC of $85.4\% \pm 1.4\%$ on the validation scans. Moreover, the network was tested on the EMC and HMC datasets and achieved an average DSC of $82.3\% \pm 1.5\%$ and $87.9\% \pm 1.2\%$, respectively. Using a single GPU, the average inference time of the segmentation and inpainting networks were approximately 2 seconds and 3 seconds per volume depending on the number of slices per volume. Figure 7.5 shows examples of the network output.

7.3.4 Parameter optimization and preprocessing analysis

For a fair comparison, the same registration parameters as in [75] were used. For the weight α that balances the contribution of the bladder segmentation in the cost function (7.3), we investigated multiple settings based on initial experiments on EMC and HMC datasets. The weight was set for the coarse (first) resolution only and was set to zero for the other two resolutions, in order to avoid overfitting issues. Here we compared four settings for α : 0.2, 0.1, 0.05, and 0.01. For this experiment we did not use inpainting. The results are shown in Table 7.2 for the HMC dataset where "Affine" refers to the affine registration defined in Equation (7.2), which is considered a reference method. The weights 0.05 and 0.20 yielded very similar performance. We opted for a weight of 0.05 to avoid overfitting on the bladder. Since the target areas (prostate, lymph nodes, and seminal vesicles) obtained slightly better accuracy for a lower weight and these are important for radiotherapy planning, we selected 0.05. For the EMC dataset a similar experiment gave a weight of 0.01 (not reported).

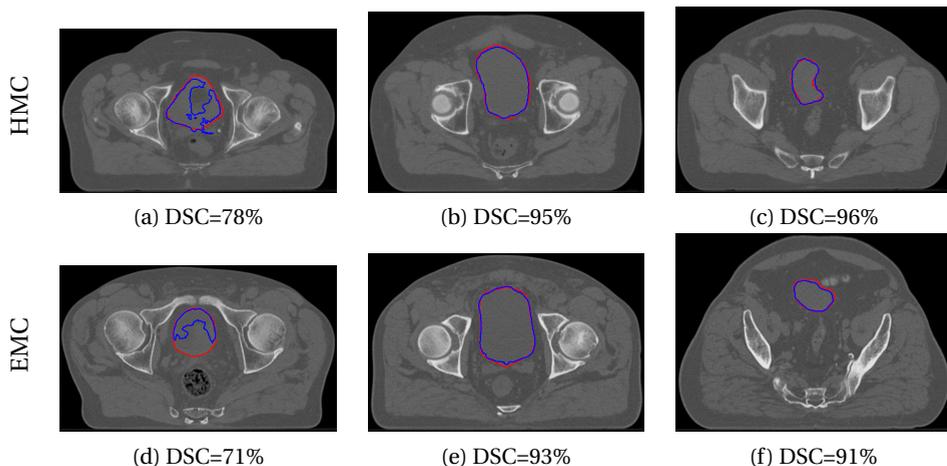


Figure 7.5: Examples of the automatic bladder segmentation using the 3D-CNN alongside the DSC of the volume. First and second rows represent samples from HMC and EMC, respectively. (a) and (d) are suboptimal results and the rest are good results. The red line represents the ground truth and the blue line is the network output.

Table 7.2: MSD (mm) of the target volumes and OARs of the HMC dataset for different registration and weight settings after the third stage of registration. Registrations using 100 and 500 iterations were both tested.

Method	α	Prostate	Seminal vesicles	Lymph nodes	Rectum	Bladder
		$\mu \pm \sigma$				
Affine, 200		1.63 ± 0.74	2.92 ± 1.74	1.23 ± 0.49	3.89 ± 1.62	4.37 ± 2.11
B-spline, 100	0.20	1.55 ± 0.90	1.70 ± 0.74	1.63 ± 0.58	2.70 ± 1.12	1.85 ± 1.85
	0.10	1.53 ± 0.82	1.72 ± 0.73	1.58 ± 0.50	2.72 ± 1.11	1.85 ± 1.71
	0.05	1.50 ± 0.75	1.74 ± 0.79	1.55 ± 0.46	2.75 ± 1.16	1.86 ± 1.56
	0.01	1.41 ± 0.36	1.75 ± 0.86	1.57 ± 0.38	2.76 ± 1.15	1.98 ± 1.19
B-spline, 500	0.20	1.49 ± 0.90	1.76 ± 0.80	1.65 ± 0.64	2.87 ± 1.39	1.74 ± 1.63
	0.10	1.45 ± 0.77	1.77 ± 0.93	1.59 ± 0.52	2.78 ± 1.19	1.77 ± 1.58
	0.05	1.43 ± 0.77	1.78 ± 0.90	1.55 ± 0.47	2.79 ± 1.19	1.81 ± 1.57
	0.01	1.36 ± 0.47	1.76 ± 0.82	1.56 ± 0.48	2.81 ± 1.18	1.84 ± 1.24

Therefore, for the remainder of the paper these weights have been used.

In order to investigate the difference between simple-inpainting and GAN-inpainting, we run the registration on HMC dataset using both techniques as shown in Table 7.3. The results shows a very similar performance for simple-inpainting and GAN-inpainting. Hence, the simple-inpainting is used for gas pocket inpainting for the remainder of the paper.

From the aforementioned experiments and analysis (Table 7.2 and 7.3), we noticed a similar performance between 100 and 500 iterations and in order to reduce the

Table 7.3: MSD (mm) of the target volumes and OARs for different registration settings and inpainting methods at $\alpha = 0.05$. Registrations using 100 and 500 iterations were both tested.

# It.	Inpainting Method	Prostate	Seminal vesicles	Lymph nodes	Rectum	Bladder
		$\mu \pm \sigma$				
100	Simple	1.29 ± 0.39	1.48 ± 1.16	1.49 ± 0.44	2.39 ± 1.92	1.72 ± 1.17
	GAN	1.29 ± 0.41	1.70 ± 2.12	1.49 ± 0.44	2.65 ± 2.17	1.71 ± 1.16
500	Simple	1.28 ± 0.42	1.36 ± 0.40	1.49 ± 0.44	2.19 ± 1.03	1.67 ± 1.22
	GAN	1.28 ± 0.42	1.36 ± 0.38	1.48 ± 0.45	2.33 ± 0.95	1.67 ± 1.22

registration time, we considered only the results from 100 iterations for the final experiments.

7.3.5 Registration performance

Since the LUMC dataset did not have any follow-up scans, we only consider the EMC and HMC datasets for evaluating the registration performance. Figure 7.6 shows example results of the automatically propagated contours. We compared the proposed method with the intensity-based registration approach of Qiao *et al.* [75]. For the HMC data we directly compare with the results reported in [75], as the same dataset was used. For the EMC data we applied their algorithm, and compare with our results. The DSC overlap of the proposed algorithm is presented in Table 7.4. For the HMC dataset, the prostate, lymph nodes, and bladder performed similarly for the proposed method and Qiao *et al.*, while the seminal vesicles and rectum showed substantial improvements. The median DSC values of the prostate, seminal vesicles, lymph nodes, rectum, and bladder were 0.88, 0.70, 0.89, 0.78, and 0.91, respectively for Qiao *et al.*, while they were 0.89, 0.73, 0.89, 0.85, and 0.94, respectively for the proposed method. For the EMC dataset, the proposed algorithm showed consistent improvement for the seminal vesicles, rectum, and bladder. The median DSC values of the prostate, seminal vesicles, rectum, and bladder were 0.91, 0.80, 0.76, and 0.86, respectively for Qiao *et al.* and 0.89, 0.81, 0.81, and 0.90, respectively for the proposed method. For the MSD results shown in Table 7.5, the proposed method outperformed Qiao *et al.* for all the target areas and OARs. The MSD of most of the targets and the OARs was less than one voxel (2 mm). The geometrical success rate was 97%, 93%, and 87% for the prostate, seminal vesicles, and lymph nodes, respectively for the HMC dataset and 67% and 71% for the prostate and seminal vesicles for the EMC dataset. Table 7.6 shows the 95%HD, yielding a significant improvement for the proposed method over Qiao *et al.* on the HMC dataset, but less improvement for the EMC dataset. Moreover, Qiao *et al.* and the proposed method show a significant improvement from the affine method except for the lymph nodes. Figure 7.7 shows scatterplots depicting the ef-

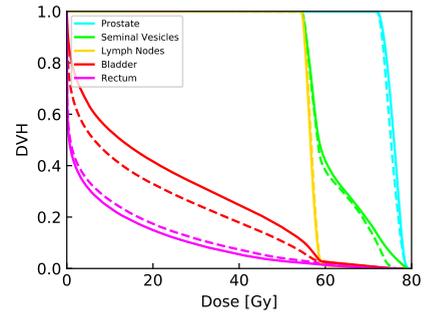
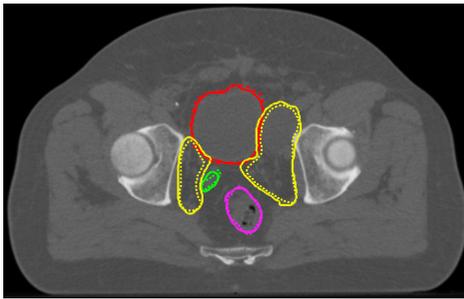
Table 7.4: DSC value of the target volumes and the OARs of the HMC and EMC datasets for different registration methods. † represents a significant difference (at $p = 0.05$) between Qiao *et al.* and the proposed algorithm.

	Method	# It.	Prostate	Seminal vesicles	Lymph nodes	Rectum	Bladder
			$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
HMC	Affine	200	0.84 ± 0.11	0.46 ± 0.26	0.90 ± 0.08	0.71 ± 0.10	0.77 ± 0.11
	Qiao <i>et al.</i>	100	0.87 ± 0.08	0.65 ± 0.18	0.88 ± 0.07	0.77 ± 0.09	0.88 ± 0.11
	Proposed	100	0.87 ± 0.08	$0.70 \pm 0.13^\dagger$	0.87 ± 0.07	$0.82 \pm 0.12^\dagger$	0.89 ± 0.12
EMC	Affine	200	0.78 ± 0.20	0.49 ± 0.32	-	0.62 ± 0.18	0.66 ± 0.25
	Qiao <i>et al.</i>	100	0.87 ± 0.13	0.70 ± 0.26	-	0.72 ± 0.16	0.78 ± 0.22
	Proposed	100	0.87 ± 0.12	$0.75 \pm 0.18^\dagger$	-	$0.78 \pm 0.15^\dagger$	$0.83 \pm 0.17^\dagger$

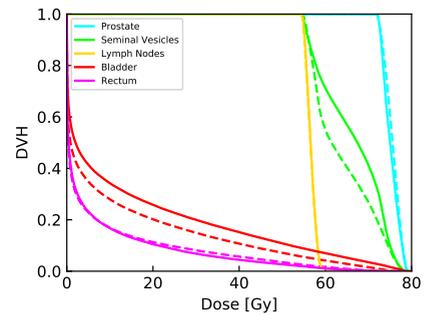
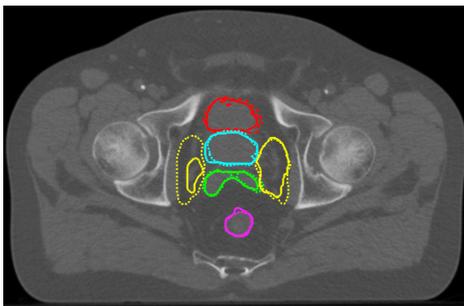
Table 7.5: MSD (mm) of the target volumes and the OARs of the HMC and EMC datasets for different registration methods. † represents a significant difference (at $p = 0.05$) between Qiao *et al.* and the proposed algorithm.

	Method	# It.	Prostate	Seminal vesicles	Lymph nodes	Rectum	Bladder
			$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
HMC	Affine	200	1.70 ± 0.96	3.02 ± 1.96	1.26 ± 0.51	3.92 ± 1.59	4.47 ± 2.27
	Qiao <i>et al.</i>	100	1.40 ± 0.47	1.85 ± 1.26	1.51 ± 0.44	3.13 ± 1.38	2.38 ± 1.79
	Proposed	100	1.29 ± 0.39	1.48 ± 1.16	1.49 ± 0.44	$2.39 \pm 1.92^\dagger$	$1.72 \pm 1.17^\dagger$
EMC	Affine	200	2.82 ± 3.18	4.42 ± 6.03	-	4.63 ± 3.01	8.03 ± 6.46
	Qiao <i>et al.</i>	100	1.41 ± 0.76	2.24 ± 3.14	-	3.21 ± 1.85	5.42 ± 5.84
	Proposed	100	1.54 ± 0.67	$1.67 \pm 1.38^\dagger$	-	$2.67 \pm 1.76^\dagger$	$3.89 \pm 4.00^\dagger$

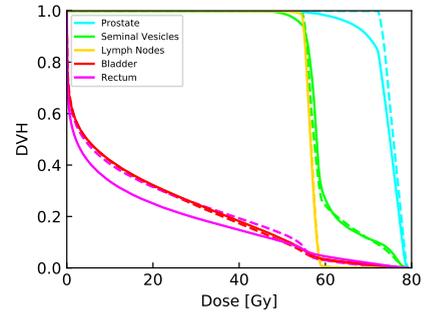
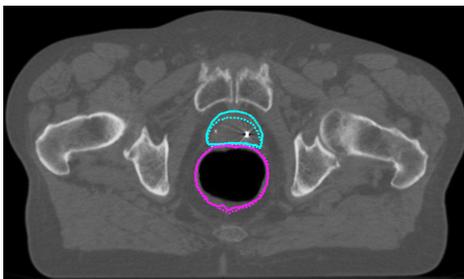
fect of the bladder distension (volume difference between planning and daily CT) on the mean surface distance (MSD) of different target organs of the HMC dataset. The figure shows that the MSD of the proposed method is less than the slice thickness (2 mm) for most of the cases, and that there is little correlation between registration performance and bladder distensibility. Figure 7.8 shows the comparison of the registration performance between Qiao *et al.* (intensity only) and the proposed method (intensity and bladder segmentation), both using 100 iterations for the HMC dataset. The comparison illustrates the performance in terms of DSC, MSD, and 95%HD for the target volumes and OARs. The figure shows a similar pattern between the proposed method using the manually annotated contours of the bladder and the contours from the 3D-CNN network. This pattern emphasizes that the proposed method achieved the upper limit of the system. The average runtime for the proposed pipeline is 98.3 seconds for each registration at 100 iterations.



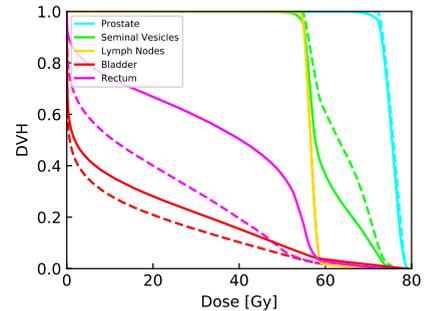
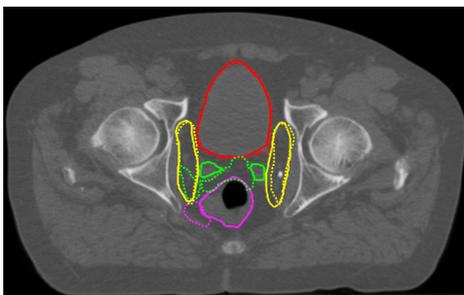
(a) The MSD of the prostate, lymph nodes, and seminal vesicles is 0.8, 1.6, and 1.0 mm, respectively.



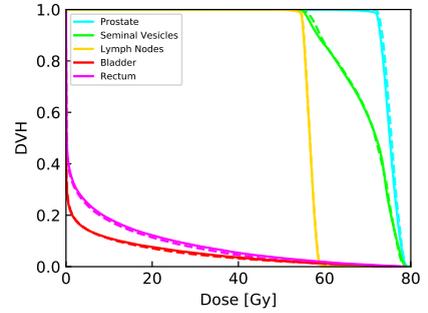
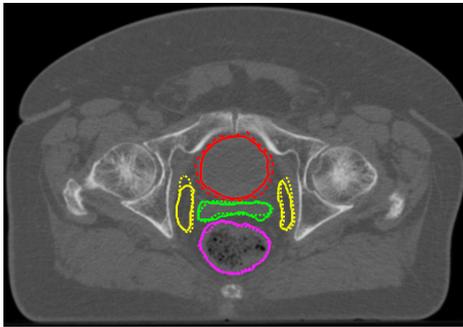
(b) The MSD of the prostate, lymph nodes, and seminal vesicles is 1.4, 1.8, and 1.2 mm, respectively.



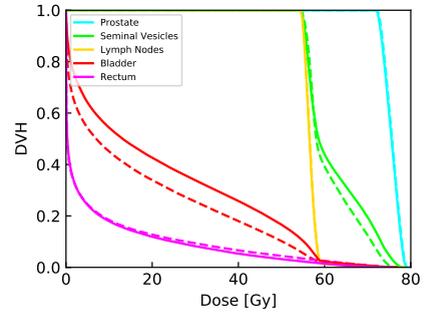
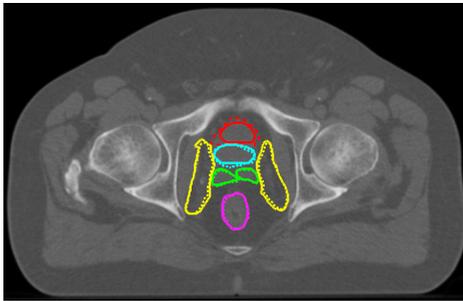
(c) The MSD of the prostate, lymph nodes, and seminal vesicles is 2.1, 1.6, and 1.5 mm, respectively.



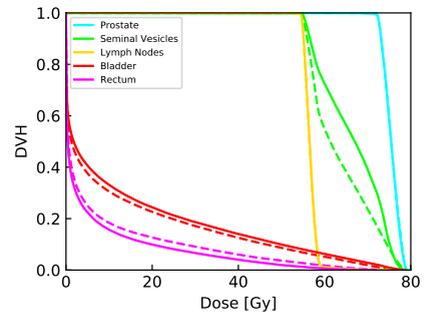
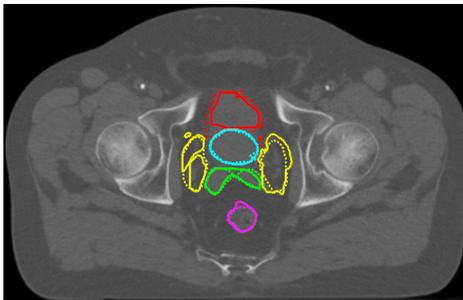
(d) The MSD of the prostate, lymph nodes, and seminal vesicles is 1.5, 1.6, and 11.0 mm, respectively.



(e) The MSD of the prostate, lymph nodes, and seminal vesicles is 1.3, 1.3, and 1.1 mm, respectively.



(f) The MSD of the prostate, lymph nodes, and seminal vesicles is 0.9, 1.4, and 0.9 mm, respectively.



(g) The MSD of the prostate, lymph nodes, and seminal vesicles is 1.3, 1.1, and 1.1 mm, respectively.

Figure 7.6: Examples from the automatic contours propagation of the HMC dataset and the corresponding dose volume histograms evaluated on the manual contours. The solid line represents the manual contouring results while the dotted line is the automatically propagated one.

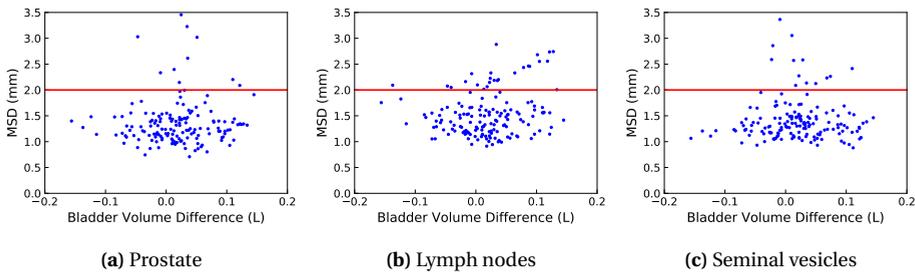


Figure 7.7: Scatterplots showing the effect of the bladder volume change between planning and daily scans of the HMC dataset on the performance of the proposed method in terms of MSD. Red line represents the slice thickness.

Table 7.6: %95HD (mm) of the target volumes and the OARs of the HMC and EMC datasets for different registration methods. † represents a significant difference (at $p = 0.05$) between Qiao *et al.* and the proposed algorithm.

			Prostate	Seminal vesicles	Lymph nodes	Rectum	Bladder
		Method	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
HMC		Affine	3.97 ± 1.96	6.61 ± 3.70	3.12 ± 1.27	11.8 ± 5.98	12.5 ± 7.06
		Qiao <i>et al.</i>	3.31 ± 1.16	4.59 ± 2.95	3.73 ± 1.02	10.4 ± 5.99	7.41 ± 6.85
		Proposed	3.07 ± 1.30	$3.82 \pm 3.19^\dagger$	3.74 ± 1.02	$8.66 \pm 6.92^\dagger$	$5.11 \pm 4.38^\dagger$
EMC		Affine	5.98 ± 6.19	8.11 ± 7.66	-	13.2 ± 6.88	21.3 ± 16.3
		Qiao <i>et al.</i>	3.65 ± 2.31	4.80 ± 5.09	-	11.3 ± 6.77	16.5 ± 17.2
		Proposed	3.93 ± 2.24	4.92 ± 5.13	-	10.4 ± 7.77	$11.5 \pm 12.5^\dagger$

7.3.6 Dosimetric performance

Figure 7.6 shows the Dose Volume Histogram (DVH) of the target organs and OARs for some examples. The clinical constraints in terms of $V_{95\%}$ and $V_{107\%}$ were calculated for the prostate, seminal vesicles, and lymph nodes based on the manual contours. In order to monitor the accumulated dose for the OARs, we calculated $V_{45\text{ Gy}}$, $V_{60\text{ Gy}}$, $V_{75\text{ Gy}}$, and D_{mean} for the rectum, as well as $V_{45\text{ Gy}}$, $V_{65\text{ Gy}}$, and D_{mean} for the bladder. Here D_{mean} is the structure’s average dose and $V_{x\text{ Gy}}$ is the percentage of volume receiving a dose of x Gy. Table 7.7 shows a comparison between the propagated contours from Qiao *et al.* and the proposed algorithm in terms of the percentage of scans that achieved the clinical criteria of $V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$. The table shows a significant improvement for the seminal vesicles, which is a small and difficult target organ, while the performance of the prostate and lymph nodes was very similar. The boxplots in Figure 7.9 illustrate the difference between the dosimetric parameter values of the manual delineations, calculated by using either the treatment plan based on the automated delineations or the manual delineations. We can see that the difference for all dosimetric parameters of all the target organs and OARs is almost 0%

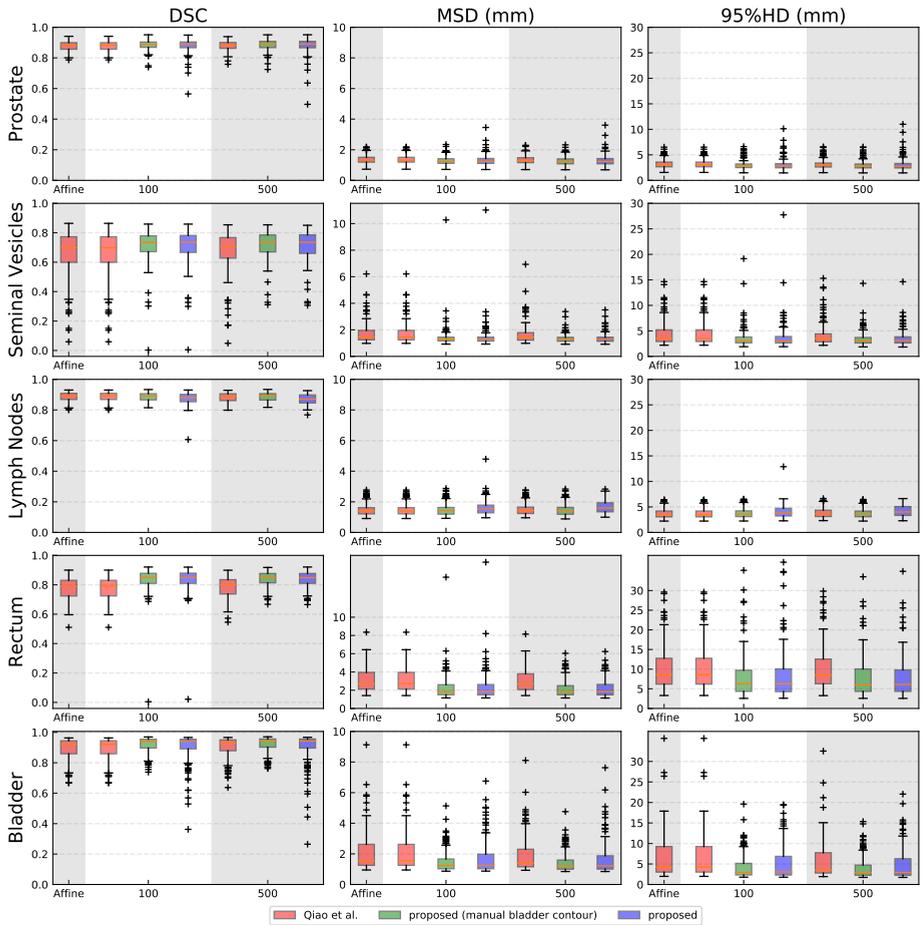


Figure 7.8: Boxplot comparison between Qiao *et al.* and the proposed algorithm for image registration on the HMC dataset versus the number of iterations. The columns show the DSC, MSD, and 95%HD from left to right. Prostate, seminal vesicles, lymph nodes, rectum, and bladder are shown from top to bottom rows, respectively. The red box is the method from Qiao *et al.*, the blue box is the proposed method, while the green box is an upper bound of the proposed method using manual daily contours.

or Gy except for the lymph nodes, which is approximately 1%.

7.4 DISCUSSION

In this study, we developed and evaluated an automatic contour propagation pipeline using DIR, while considering the robustness, accuracy, and clinical acceptance rate for the target organs and the OARs of prostate cancer. Online-adaptive IMPT is a cru-

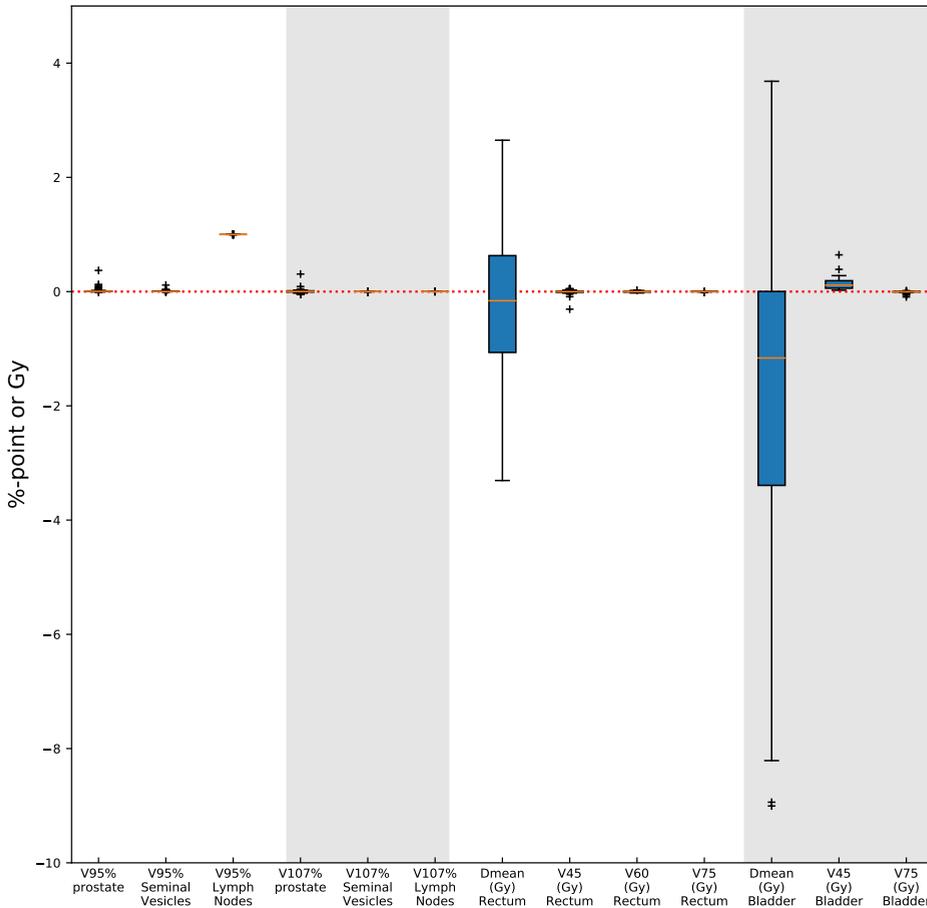


Figure 7.9: Boxplots depicting the difference in dosimetric parameters of the manual delineations, calculated by using either the treatment plan based on the automated delineations or the manual delineations for 99 scans of the HMC dataset.

cial step towards treatment with small margins for target organs. In this study we used margins of 2 mm for the prostate and 3.5 mm for the seminal vesicles and lymph nodes, respectively. Such small margins are only viable when online and daily replanning is performed. This replanning procedure should be accurate as well as robust to avoid any subsequent adverse side effects. The automatically propagated contours were validated geometrically on the EMC and HMC datasets as well as dosimetrically on the HMC dataset in order to investigate whether or not the propagated contours meet the clinical acceptance criteria for dose coverage. DSC, MSD, and 95%HD were chosen for geometric validation while $V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$ were used for dosimetric coverage validation. Here, $V_{95\%} \geq 98\%$ ensures that at least 98% of the

Table 7.7: Percentage of registrations that meets the dose constraints for different registration iterations. conservative success rate (CSR) refers to the percentage of registrations for which all target volumes (the prostate, seminal vesicles and lymph nodes) meet the dose constraints.

	$V_{95\%} \geq 98\%$				$V_{107\%} \leq 2\%$		
	Prostate	SV	LN	CSR	Prostate	SV	LN
Qiao <i>et al.</i>	83.8%	75.7%	97.9%	65%	100%	100%	100%
Proposed	85.8%	90.9%	98.9%	80%	100%	100%	100%

target volumes receive at least 95% of the prescribed dose and $V_{107\%} \leq 2\%$ ensures that less than 2% of the target volumes receive more than 107% of the highest prescribed dose.

In order to enhance the registration robustness, the segmentation of the bladder was introduced to steer the optimization. Since the registration process is partially driven by the bladder segmentation, this segmentation should be as accurate and robust as possible. Hence, we chose a 3D-CNN for bladder segmentation, and obtained a DSC of 87.9% and a Jaccard index of 80.2%, which is very comparable to the reported Jaccard index of 81.9% in [102], where the authors developed a CNN network alongside level-sets to segment the bladder in CT urography. Moreover, our proposed network outperformed the 2D-CNN network developed by Zhou *et al.* [103], where the authors reported a DSC of 72%. The high performance of the proposed network may be attributed to the use of a large receptive field as well as replacing the 2D convolutions with 3D convolutions, which helps the network to embed depth information.

Applying contrast clipping to the CT scans before registration was beneficial to the registration process, since the registration is intensity-based, which is consistent with the findings in [104]. Inpainting gas pockets in the rectum enhanced the registration of the rectum as well as the seminal vesicles. The presence of these pockets were challenging for the registration due to the physical non-correspondence between the daily and planning CT scans. Although the inpainting results from the GAN-inpainting network were more realistic than the simple-inpainting procedure, a similar performance with respect to the registration was obtained. Our explanation for this finding is that the mutual information similarity metric pays more attention to the overall intensity distribution and since the results from the simple-inpainting were blended and smoothed with respect to its neighbors, it produces a similar histogram distribution to the GAN-inpainting and subsequently gives a similar registration performance.

The initialization of the registration algorithm on the bony structures is a crucial step for optimal performance, which is consistent with the reported results in [75]. Moreover, masking out the couch using a torso mask removed its disrupting effect on the registration. Increasing the number of iterations had a minimal effect on the

registration performance while increasing the registration time. We found that the effect of adding a third registration step focussing on the rectal area, boosted the performance regarding the rectum and seminal vesicles while there was no detrimental effect for the prostate, lymph nodes, and bladder.

In this study, we focused on the generalizability and robustness of the registration represented by performance on different datasets and the number of failed registrations according to geometrical and dosimetric criteria. This target is achieved through several steps. First, inpainting the rectum gas pockets. Second, enhancing the CT image contrast by contrast clipping. Third, introducing the bladder segmentation with an optimized weights ($\alpha = 0.05$ and 0.01) to steer the optimization problem to a better local minimum while avoiding overfitting to the bladder. Fourth, using a third stage for registration to focus on the rectum and consequently the seminal vesicles by using a dilated mask for the rectum. Overall, these steps yielded a more robust registration and substantially decreased the number of registrations with insufficient quality, especially for the seminal vesicles, rectum, and bladder. Improving the MSD for the seminal vesicles, which is an important target volume, resulted in a more precise targeting with potential benefits in terms of local control (lower probability of recurrences). Moreover, both the rectum and the bladder improved in terms of MSD and 95%HD, thereby avoiding treatment-induced complications after the therapy, so a higher probability of better quality of life after treatment. For the bladder, 11 of the 18 registrations with an MSD larger than the top whisker in Figure 7.8, were belonging to two patients. For these two patients the 3D-CNN achieved an average DSC of 0.65, explaining the suboptimal performance of the proposed method on these cases. From the CT images no apparent reason for this was found. In terms of the geometric success rate defined by the number of registrations that achieved an MSD lower than 2 mm (slice thickness), the system achieved 97%, 93%, and 87% for the prostate, seminal vesicles, and lymph nodes, respectively. This compares to a success rate of 95%, 78%, and 86% for Qiao *et al.*, i.e. especially improving the performance for the seminal vesicles. Moreover, the proposed system showed robustness to the change in bladder distension between planning and daily CT scans as shown in Figure 7.7. The proposed registration method achieved quite similar results on the EMC and HMC datasets, except for the bladder. We suspect this is partially due to the difference in bladder segmentation performance of the neural network, which was 82% on the EMC data and 88% on the HMC data. It could also be related to the affine registration results for the EMC dataset (Table 7.5) being slightly less than HMC dataset. We visually checked the affine results and noticed that the field of view for some cases were cropped or zoomed. The average runtime for the proposed pipeline is 98.3 seconds for each registration at 100 iterations, comparing to 13.5 seconds reported by Qiao

et al. However, the pipeline could be further optimized and adapted for GPU acceleration.

For validating the clinical acceptance of the proposed algorithm, we considered $V_{95\%} \geq 98\%$, $V_{107\%} \leq 2\%$, and CSR for dosimetric coverage for 99 registrations. All the scans meet the $V_{107\%} \leq 2\%$ constraint. Fourteen out of the 99 registrations (14.1%) did not directly meet the $V_{95\%} \geq 98\%$ constraint for the prostate. After visual inspection of these failure cases, we found inconsistencies between the manual delineations for the planning and daily CT scans for 7 cases. These cases had a $V_{95\%}$ of $92.5\% \pm 0.1\%$, meaning that these cases were still close to be dosimetrically acceptable. The proposed algorithm improved the contouring quality and robustness especially for the seminal vesicles, which directly increased the percentage of acceptable scans from 75.5% to 90.9% for this important target organ. These success rates imply that the automatically generated contours have the potential to be employed for online-adaptive IMPT. Moreover, the typical 7 mm margins [78] may be replaced with smaller daily margins, which means delivering an effective dose with potentially less adverse effects.

The reported performance of the proposed pipeline could be further improved by correcting the inconsistency present in the manual contouring. Also, the weighting parameter α could be selected automatically by introducing it as a trainable parameter. Moreover, the current 3D-CNN was trained using CT scans without contrast material, and therefore is unlikely to perform well on scans acquired with contrast. In case the clinical protocol dictates contrast-enhanced CT acquisitions, the network could be easily retrained. We may further investigate the effect on segmentation performance of CT clipping as a preprocessing step for the 3D-CNN for bladder segmentation. We also consider developing an end-to-end neural network to jointly optimize the registration and segmentation tasks to further improve the system robustness and accuracy.

7.5 CONCLUSION

In this study we proposed a registration pipeline for automatic contour propagation for online-adaptive IMPT of prostate cancer using the open source package `elastix` software in combination with deep learning. The proposed pipeline achieved a geometrical success rate of 97%, 93%, and 87% for the prostate, seminal vesicles, and lymph nodes, respectively for the HMC dataset as well as 67% and 71% for the prostate and seminal vesicles, respectively for the EMC dataset. The HMC automatically propagated contours meet the dose coverage constraints in 86%, 91%, and 99% of cases for these targets. A conservative success rate (CSR) of 80% was achieved,

meaning that 80% of the automatically generated treatment plans can be directly used without manual correction. This recontouring showed a promise for generating daily treatment plans. Moreover, it showed a substantial improvement in the system robustness compared to a previous open source method, which means that more treatment plans can be directly used without manual correction, which is a crucial factor for enabling online daily adaptation and thus the use of relatively small treatment margins. Therefore, the proposed method could facilitate online-adaptive proton therapy of prostate cancer. The authors have no relevant conflicts of interest to disclose.

7.6 ACKNOWLEDGMENTS

This study was financially supported by Varian Medical Systems and ZonMw, the Netherlands Organization for Health Research and Development, grant number 104003012. The HMC dataset with contours were collected at Haukeland University Hospital, Bergen, Norway and were provided to us by responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing; they are gratefully acknowledged.

Chapter

8

Correlations between the shifts in prompt gamma emission profiles and the changes in daily target coverage during simulated pencil-beam scanning proton therapy

Physics in Medicine and Biology, Vol. 64, 085009 (10pp), 2019,

doi: <http://dx.doi.org/10.1088/1361-6560/ab145e>

Eelco Lens¹

Thyrza Z. Jagt²

Mischa S. Hoogeman^{2,3}

Dennis R. Schaart^{1,3}

¹ Department of Radiation Science and Technology, Delft University of Technology,
Delft, The Netherlands

² Department of Radiation Oncology, Erasmus MC Cancer Institute,
Rotterdam, The Netherlands

³ Department of Medical Physics & Informatics, HollandPTC,
Delft, The Netherlands

Abstract

Purpose: The aim of this study was to investigate the feasibility of using prompt gamma (PG) ray emission profiles to monitor changes in dose to the planning target volume (PTV) during pencil-beam scanning (PBS) proton therapy as a result of day-to-day variations in patient anatomy.

Methods and materials: For 11 prostate patients, we simulated treatment plan delivery using the patients daily anatomy as observed in the planning CT and 7 – 9 control CT scans, including the detected PG profiles resulting from the 5%, 10%, and 20% most intense proton pencil-beams. For each patient, we determined the changes in dosimetric parameters for the high- and low-dose PTVs between the simulations performed using the planning CT scan and the different control CT scans and correlated these to changes in the PG emission profiles.

Results: Changes in coverage of the high- and low-dose PTV correlated most strongly with the median and mean absolute PG emission profile shifts of the 5% most intense pencil-beams, respectively. With a mean Pearson correlation coefficient of -0.76 (SD: 0.17) for the high-dose PTV and of -0.60 (SD: 0.51) for the low-dose PTV.

Conclusion: We showed, as a proof of principle, that PG emission profiles obtained during PBS proton therapy could be used to detect changes in PTV coverage due to day-to-day anatomical variation.

8.1 INTRODUCTION

Proton therapy can be negatively affected by errors in proton range prediction and by day-to-day anatomical variations, both resulting in a variation in the Bragg peak location, this study focuses on the latter. The sensitivity of proton therapy accuracy to small daily anatomical variations, as well as the corresponding effects on the dose distribution, have been previously described [105]. The detection of secondary radiation exiting the patient has been proposed as a potential tool for day-to-day dose monitoring, so as to ensure patient safety and an effective treatment [12, 106–110].

Prompt gamma (PG) rays resulting from nuclear interactions between the incoming protons and the patients' tissue can be used for proton range monitoring during treatment [12]. PG emission profiles have been shown to correlate strongly with the depth-dose profile of the primary proton beam [12]. Detecting changes in the location of the fall-off region of the PG emission profiles to estimate the change in proton range has been shown to be feasible in both Monte Carlo simulations as well as in in-situ and in-vivo measurements [107, 111–113]. Different techniques have been proposed, such as using spectral or timing information [114–117] or neutron background suppression by applying time-of-flight discrimination [107].

PG emission profile measurements have been used in a clinical setting and in combination with patient data [112, 118–121]. However, information was provided on range shifts of individual pencil-beams or energy layers only, not on clinically relevant dosimetric parameters, which is needed to ensure daily target coverage and to increase treatment effectiveness.

If PG emission profiles are to be used for (near) real-time dose monitoring during proton therapy, the deviation between the detected and expected PG emission profiles should be determined. In addition, a translation from the simple detection of range shifts for a set of pencil-beams to the quantitative assessment of clinically relevant changes in the dose distribution should be established. To our knowledge, it has not yet been shown whether changes in PG emission profiles correlate with changes in clinically relevant dosimetric parameters such as target coverage.

The goal of this study is to investigate the feasibility of using detected PG emission profiles for the daily monitoring of dosimetric changes during pencil-beam scanning (PBS) proton therapy. We used Monte Carlo to simulate dose delivery on multiple control CT scans to determine the changes in dose to target volumes and in PG emission profiles, detected outside the patient, as a result of day-to-day variations in patient anatomy. We established the correlations between dosimetric changes and changes

in the PG emission profiles to determine what observables could potentially be used for day-to-day dose monitoring.

8.2 METHODS AND MATERIALS

Clinical dataset and treatment plans

Our dataset consisted of 11 prostate cancer patients, previously treated with conventional intensity-modulated radiotherapy (IMRT) without a bladder filling protocol or rectal balloon. For each patient we had a planning CT scan and 7 – 9 control CT scans (Appendix Table A.7.1), obtained during the course of treatment (99 CT scans in total). The available CT scans represented the day-to-day anatomical variations as were observed for this patient group in daily practice. These anatomical changes can result in density changes along the pencil-beam path affecting the dose distribution.

For each planning CT scan, we had a structure set delineated by an experienced radiation oncologist. We generated the high-dose planning target volume (PTV_{High}; prostate + 4 mm) and the low-dose PTV (PTV_{Low}; seminal vesicles + lymph nodes + 7 mm). In addition, there was the intermediate-dose PTV (PTV_{Inter}), which consisted of part of the PTV_{Low} that was situated in the 15 mm ring surrounding the PTV_{High} that was used to regulate the dose gradient around the PTV_{High} during plan optimization [30]. The PTV_{Inter} was excluded from the PTV_{Low}, thus these two volumes did not overlap. The PTV_{High} was prescribed 74 Gy in 37 fractions of 2 Gy, the PTV_{Low} and PTV_{Inter} were prescribed 55 Gy but the PTV_{Inter} will partially receive a higher dose since it is situated in the dose gradient surrounding the PTV_{High}. For each patient, a PBS proton treatment plan was made using the Erasmus-iCycle treatment planning system [20, 30] and the dose was calculated using the ‘Astroid’ dose engine [21]. This system generates Pareto optimal plans by using a multi-criteria optimization based on constraints and objectives. All planning constraints and objectives were in accordance with the work from Jagt *et al.* [30]. On average, a plan contained 1417 (range: 1247 – 1540) pencil-beams and a total of 2.8×10^{12} to 3.6×10^{12} protons (appendix Table A.7.1), divided over two opposing lateral fields. The plans were not robustly optimized.

Monte Carlo simulation of dose delivery and PG detection

Dose delivery using the proton therapy plans was simulated using the TOPAS (version 3.1.p1) (Geant4 10.3.p01 based) Monte Carlo code [122] and was performed on the Dutch national computing cluster LISA (SURFsara, Amsterdam, the Netherlands). We used the TOPAS default physics list, which has been shown to compare well to lists

that were proposed for proton therapy in earlier versions of Geant4 [123]. Geant4 has been shown to overestimate PG yields, while the PG emission profile length was accurately simulated [124].

We extracted the treatment parameters (e.g. pencil-beam energies and locations relative to the isocenter) from the treatment plans and used these to generate the initialization files required to perform the simulations. The pencil-beam weights were determined by dividing the number of monitor units (MUs) per pencil-beam by the number of MUs of the most intense pencil-beam within that treatment plan. For each plan, the most intense pencil-beam was simulated using 2.8×10^6 protons and the number of protons of the other pencil-beams were scaled by the corresponding beam weight. This resulted in an approximate downscaling of the actual number of planned protons by a factor of 5000 and on average $3.3 \times 10^5 - 5.3 \times 10^5$ simulated protons per pencil-beam per patient. The maximum number of protons was used to evenly distribute the available calculation time over all patients. The number of pencil-beams and the number of planned and simulated protons per treatment plan are given in appendix Table A.7.1.

For each patient, dose delivery was simulated on both the planning and control CT scans to obtain dose distributions and PG emission profiles corresponding to the anatomy represented on all CT scans. The extracted dosimetric values from the simulated planned dose distributions (i.e. treatment plans simulated on the corresponding planning CT scans) were used as reference values for each patient. This was done to negate any differences between the dose calculation engine of Erasmus-iCycle and TOPAS. The extracted values from the simulated dose distributions on the control CT scans were compared to the reference values to establish how the dose changed due to day-to-day anatomical variations.

In daily clinical practice, the patient is positioned based on a registration of the intra-prostatic markers. In this study, the isocenter of each CT, which was defined as the center of mass of the prostate, was placed at the origin of our coordinate system, mimicking perfect daily patient positioning. No positioning uncertainty of an envisioned PG emission detector was taken into account since we assumed a fixed detector position with respect to the isocenter of the treatment unit. The conversion from Hounsfield units to proton stopping power was performed using the Schneider (stoichiometric) conversion [125]. The used Hounsfield look-up table was not optimized for the used CT scanner and conversion uncertainties were not taken into account, but this should not influence the relative difference between the different simulations.

For each simulation, we scored the dose (i.e. dose to water) deposited within the CT volume on a 200×200 grid for each CT slice, resulting in slightly different voxel sizes per patient. However, this will not influence the results since no inter-patient comparison will be performed. All photons exiting the patient were scored on a cylindrical surface around the CT volume, coaxially to the treatment fields. For each photon, the position, direction, energy and which pencil-beam generated the photon were scored. This allowed us to analyze PG emission profiles per pencil-beam. The scoring cylinder encompassed the complete CT and would intersect with the patient in real life, but was used to increase the number of detected PG photons compared to using a scoring surface with an area similar to that of clinically used detectors [107]. The creation of the initialization files as well as all data analysis was performed in MATLAB (The MathWorks Inc., Natick, MA, USA).

Data analysis

The simulated dose files were linearly interpolated to the same grid as the planned dose distribution. All dose distributions were evaluated on the planning CT scan of that patient, which means that all dosimetric parameters were determined for the PTVs as defined on the planning CT scan. This ensured that observed dosimetric changes were solely caused by changes in the dose distributions resulting from density changes observed along the pencil-beam path. If the PTV, as delineated on the control CT scan, would be used, changes in dose could also be due to delineation variations and this would then not be observed in the prompt gamma emission profiles. The aim of this study was to show the feasibility of using PG emission profiles to determine changes in volumetric dosimetric parameters.

For each dose distribution, we created dose volume histograms (DVHs) for the different PTVs. We then calculated the mean dose (D_{mean}), median dose (D_{median}), the maximum dose received by $\geq 2\%$ of the volume ($D_{2\%}$), the minimum dose received by $\geq 98\%$ of the volume ($D_{98\%}$), and the volume percentage receiving $\geq 95\%$ of the prescribed dose ($V_{95\%}$).

To determine the PG emission profiles, we discriminated the scored photons based on energy (≥ 1 MeV) and angle of incidence ($87^\circ \leq \theta \leq 93^\circ$), as were used in other studies as well [118, 126]. The latter was done to select photons that were emitted close to perpendicular to the treatment beam, simulating a simple multi-slit collimator. We tallied the photons within 4 mm wide spatial bins resembling the pixel size of a realistic scintillation detector [107]. Next, we selected the PG emission profiles corresponding to the 5%, 10% and 20% most intense pencil-beams (i.e. highest number of protons) per plan, regardless of the pencil-beam locations or proton energy.

To reduce noise, each profile was filtered using a third-order median filter. The total number of protons in the selected pencil-beams corresponded to approximately 18%, 30% and 49% of the total number of protons in the treatment plans, respectively. The total numbers of simulated protons in the selected pencil-beams and the total number of PG photons used for analysis are given in appendix Table A.7.2. For each PG emission profile, we calculated five parameters that could each be used to compare the profiles from the simulated planned dose to those from the simulations on the control CT scans. For the first and second parameters, we automatically selected the point closest to the 50%-point of the fall-off region of the PG profile. Next, we selected 10 points in both directions and these 21 points were considered as the fall-off region of the PG profile. We fitted a sigmoid function to the fall-off region and used the 50%-point of the sigmoid curve (X_{50}) as a measure for the fall-off location of the profile (Figure 8.1) [118]. This point is known to correlate strongly with the Bragg peak location of the corresponding pencil-beam [12]. For each analyzed pencil-beam, we determined the difference as well as the absolute difference between the X_{50} of the simulation on the planning CT scan and the X_{50} of that same pencil-beam when simulated on the control CT scan (i.e. ΔX_{50} and $|\Delta X_{50}|$). For the third parameter, we summed the squared differences of the fall-off region between the PG profiles belonging to the same pencil-beams of the different simulations (i.e. local summed-squared differences). For the fourth and fifth parameters, we summed the squared differences and the chi-squared differences over the entire profiles (i.e. overall summed-squared and chi-squared differences). As a result, for each simulation on a control CT scan, we had three distributions of values for each of the five parameters, namely for the 5%, 10% and 20% most intense pencil-beams, respectively.

To investigate the feasibility of using these distributions for dose monitoring during treatment, we calculated the mean, median, and standard deviation (SD) of each distribution and determined, per patient, the Pearson correlation coefficients (r) between these values and the calculated dosimetric parameters. The Pearson correlation coefficient is a measure of the linear correlation strength between two variables and can range from $-1 \leq r \leq 1$, where a negative value of r stands for a negative correlation and a positive value for a positive correlation. Small errors in the sigmoid fits to the fall-off regions of the PG emission profiles due to low counting statistics will only have a minor influence on the results since for each simulation we averaged over ± 70 to ± 280 pencil-beams, depending on whether 5%, 10% or 20% of the pencil-beams was used.

No comparison was made between the five parameters used to determine the PG profile shifts and the individual Bragg peak shifts since we did not score the dose per

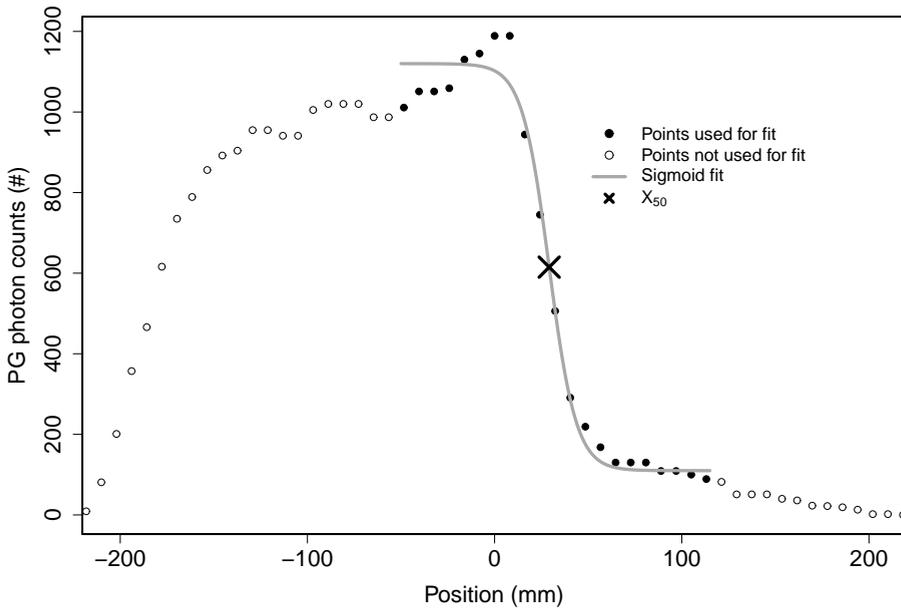


Figure 8.1: Prompt gamma emission profile corresponding to a simulated proton pencil-beam, with fitted sigmoid curve and X_{50} point.

pencil-beam. We were only interested in the PG profile shifts of a subset of pencil-beams and the clinically relevant dosimetric changes.

8.3 RESULTS

Figure 8.2 exemplarily shows the planned dose distribution, the simulated planned dose distribution, the dose distribution obtained from the simulation on control CT scan 4, and the same dose distribution shown on the planning CT scan, all for patient 1. The situations shown in Figures 8.2a, b and d were used for the calculation of the dosimetric parameters, the situation in Figure 8.2c is purely illustrative of what was simulated. In addition, Figure 8.3 shows the DVHs of patient 4, illustrating the agreement between the planned and simulated planned dose for this patient.

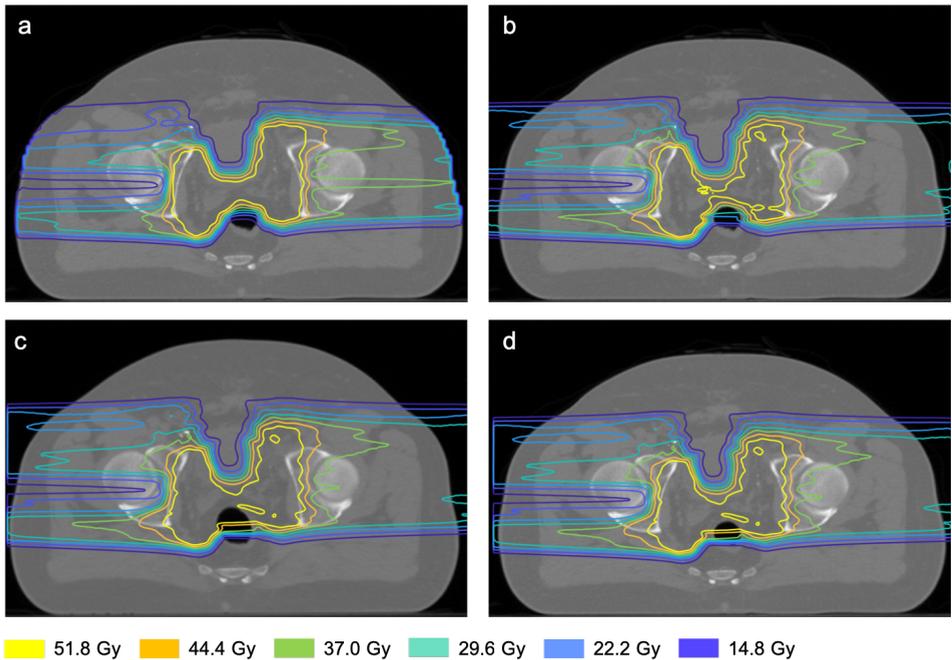


Figure 8.2: Illustration of (a) the planned dose, (b) the simulated planned dose, (c) the dose based on a simulation on control CT scan 4 and (d) the simulated dose of (c) projected on the planning CT scan, all for patient 1.

Correlations between changes in PG emission profiles and in dosimetric parameters

Figure 8.4 shows the ΔX_{50} and $|\Delta X_{50}|$ distributions when using the 5% most intense pencil-beams for patient 4. For this particular patient, the Pearson correlation coefficients between the change in $V_{95\%}$ of the PTV_{High} , PTV_{Inter} and PTV_{Low} (Figure 8.3) and the mean ΔX_{50} (Figure 8.4a) were 0.29, 0.10 and 0.20, respectively. When using the mean $|\Delta X_{50}|$ (Figure 8.4b), the coefficients were -0.97, -0.81, -0.99. In general, stronger correlations were observed when using the absolute shifts of the PG profiles (i.e. disregarding whether a profile shifts in one direction or the other). The calculated correlation coefficients over all patients when using $|\Delta X_{50}|$ are illustrated in Figure 8.5 in the form of boxplots. The strongest correlations were observed for the $V_{95\%}$ of the PTV_{High} and PTV_{Low} . This is because the deposited dose was optimized to be conformal to and uniform within these PTVs. Therefore, a disturbance in dose compromises this conformality or could create cold-spots, decreasing the $V_{95\%}$. The PTV_{Inter} is positioned completely in the dose gradient from 74 Gy to 55 Gy, thus receiving a non-uniform dose higher than the prescribed 55 Gy. Therefore, the $V_{95\%}$ is less sensitive to disturbances, hence the weak correlations. In general, the strongest

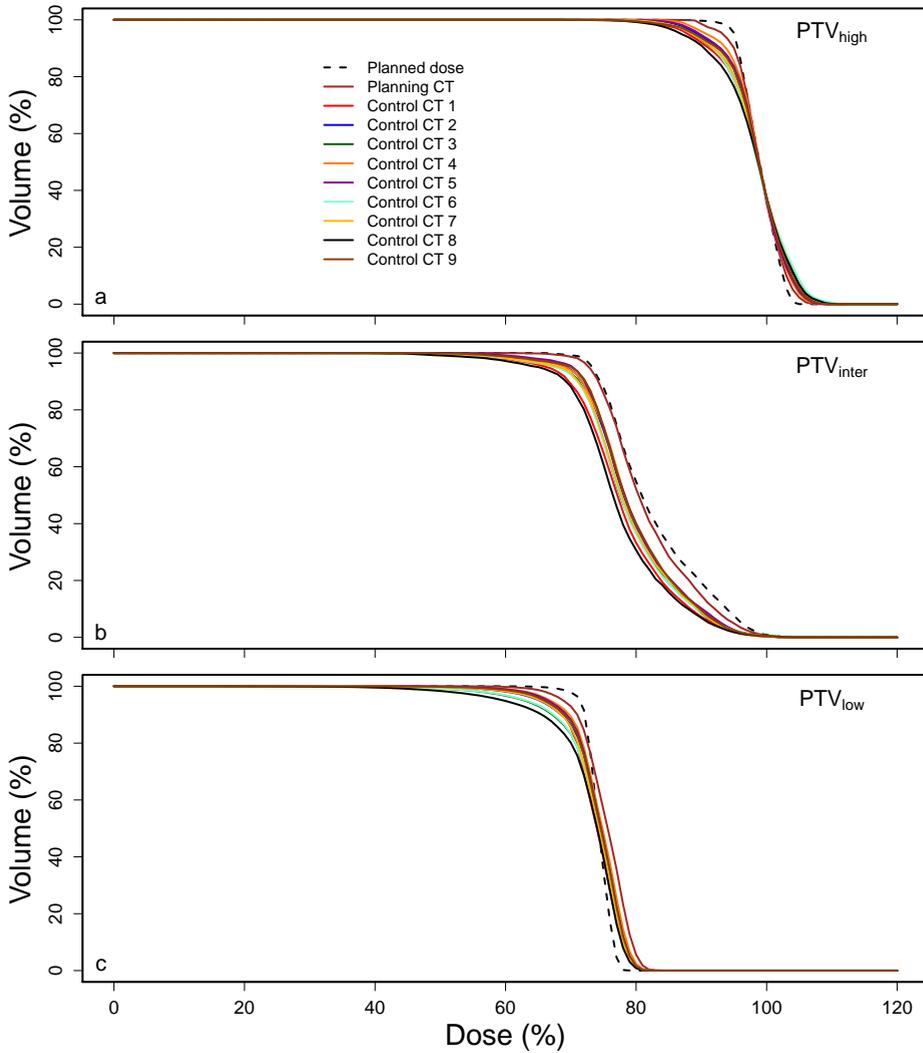


Figure 8.3: Dose volume histograms of (a) the high-dose PTV (PTV_{High}), (b) the intermediate-dose PTV (PTV_{Inter}) and (c) the low-dose PTV (PTV_{Low}) for all available dose distributions of patient 4.

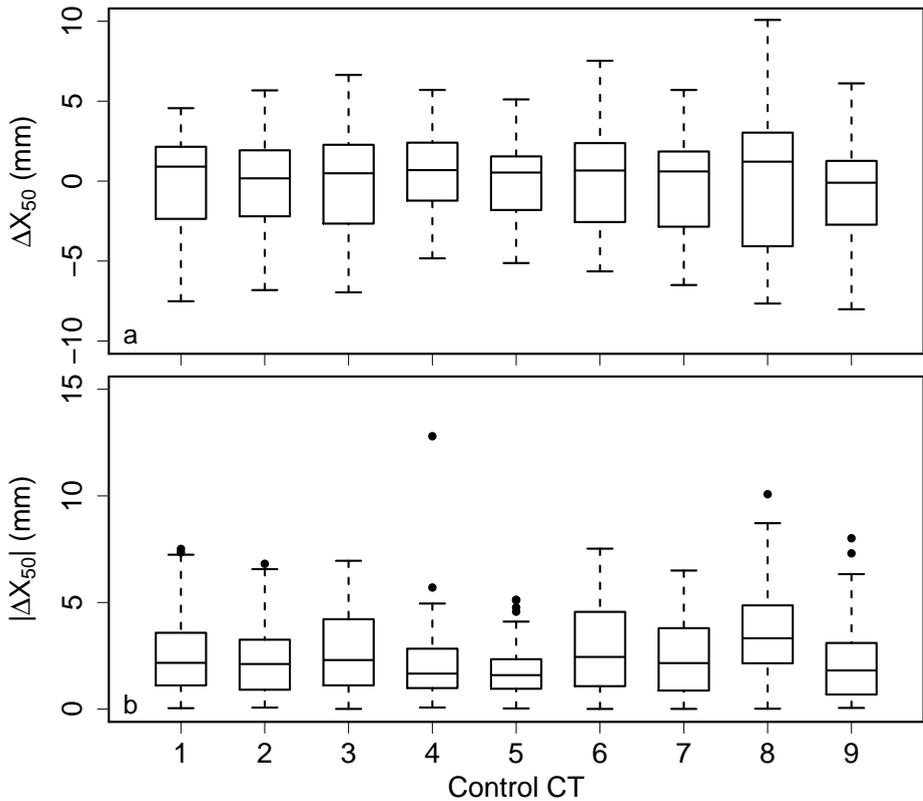


Figure 8.4: The (a) ΔX_{50} and (b) $|\Delta X_{50}|$ distributions when using the 5% most intense pencil-beams for patient 4.

correlations were observed when using the 5% most intense pencil-beams.

Figure 8.6 shows the linear fits used to calculate the correlation coefficients r and the corresponding p -values between the $V_{95\%}$ of the PTV_{High} and the median $|\Delta X_{50}|$ when using the 5% most intense pencil-beams, for all patients. We observed strong correlations in most patients, but the relation between the two parameters was highly variable between patients.

Using either the $|\Delta X_{50}|$, local summed-squared differences, overall summed-squared differences or the overall summed chi-squared differences yielded similar results, but the strongest correlations were observed when using the $|\Delta X_{50}|$. This section therefore focused on the $|\Delta X_{50}|$ distributions; the other results are shortly discussed in the Supplementary Materials (appendix Figures A.7.1 – A.7.4).

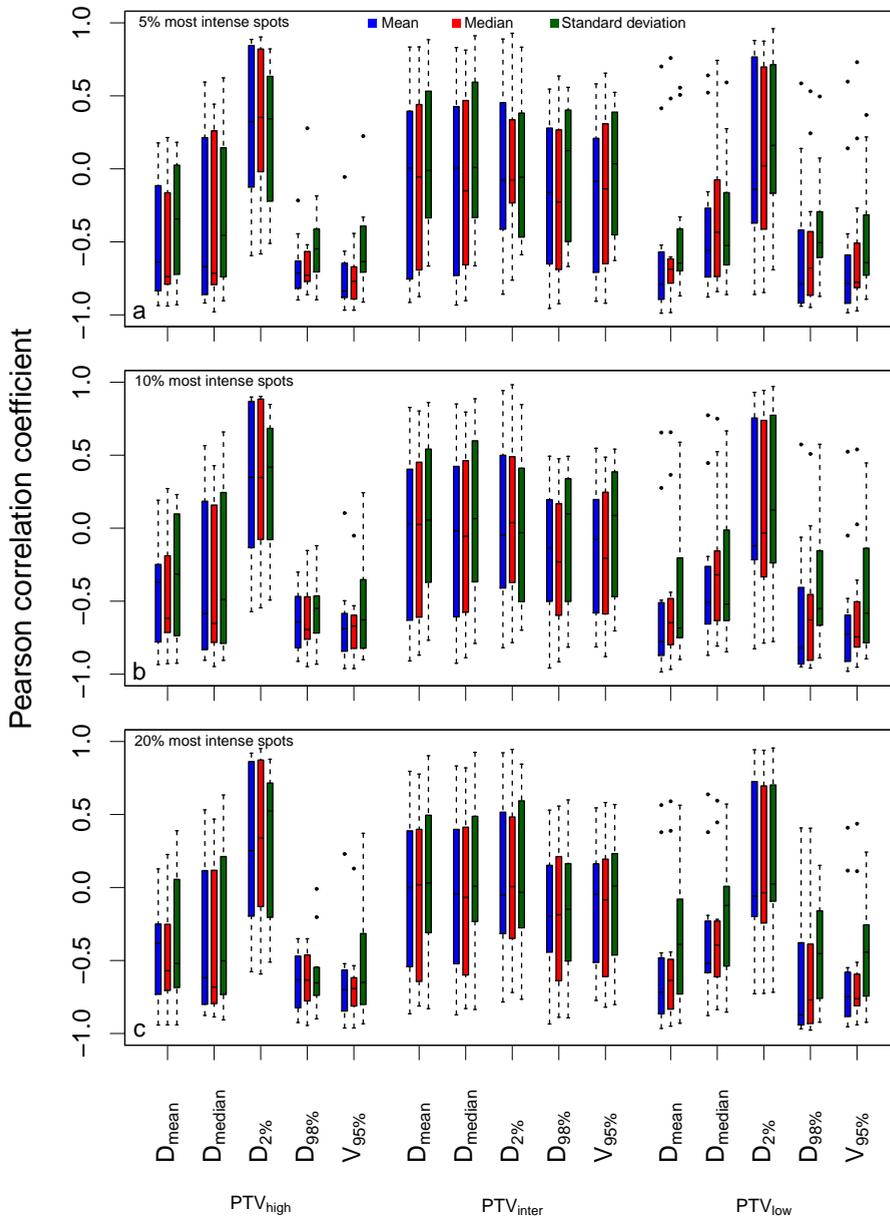


Figure 8.5: Distributions of Pearson correlation coefficients between the mean, median and standard deviation (color coded) of the $|\Delta X_{50}|$ distributions and the dosimetric parameters when using the (a) 5%, (b) 10% or (c) 20% most intense pencil-beams.

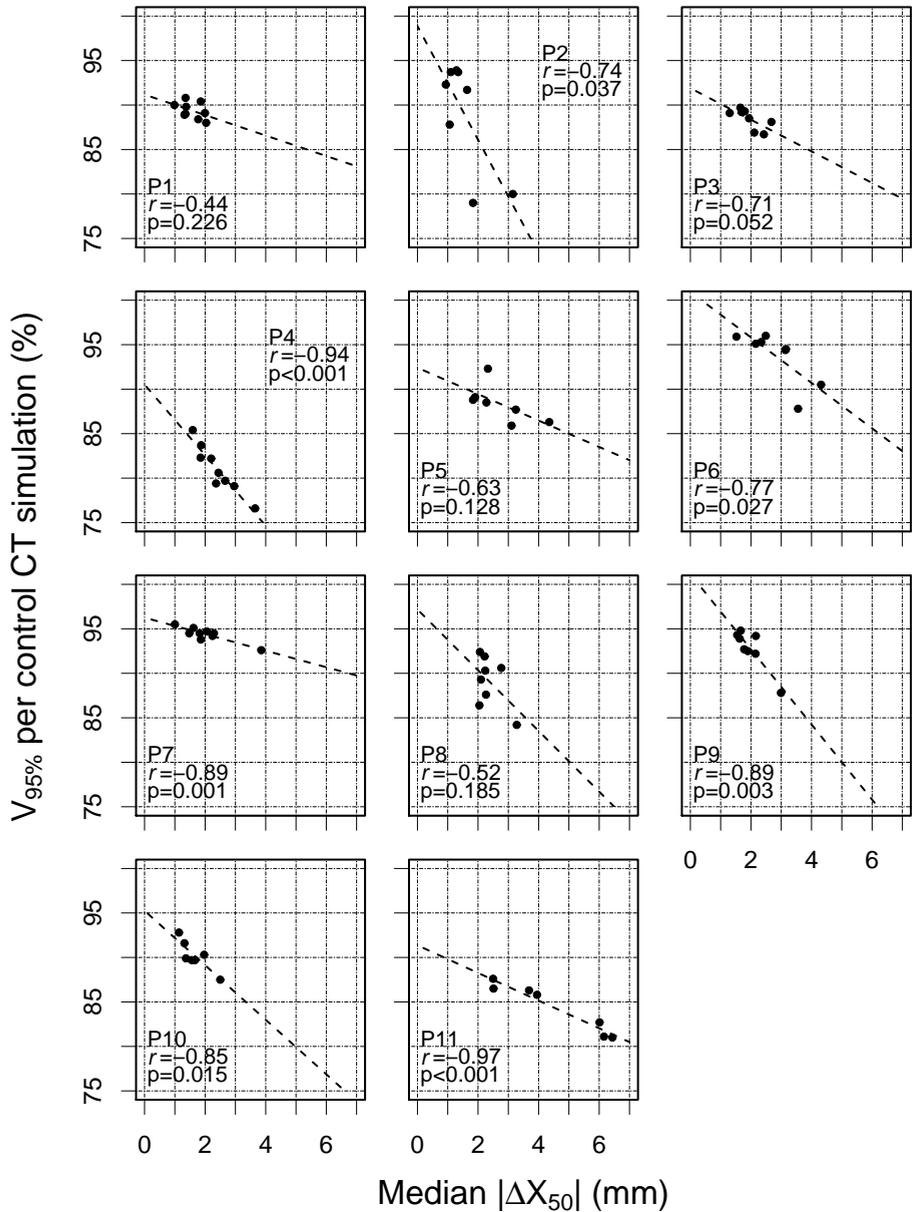


Figure 8.6: Scatterplots of the target coverage of the high dose PTV ($V_{95\%}$) and the median $|\Delta X_{50}|$ when using the 5% most intense pencil-beams for all patients (one patient per panel). Linear fits were performed to calculate the Pearson correlation coefficient (r) and corresponding p -values.

8.4 DISCUSSION

This study is the first to correlate the day-to-day changes in PG emission profiles with the day-to-day volumetric dosimetric changes of the PTVs for fractionated PBS proton therapy of prostate cancer patients. The observed correlations indicate that PG emission profiles could be used to detect daily volumetric dosimetric changes. The strength of the observed correlations and the relations between the changes in dose and PG emission profiles varied greatly. For example, weaker correlations were observed for the mean, median and max dose to the PTV_{High} and PTV_{Low} . This was to be expected since a shift of a subset of the pencil-beams within the volume does not necessarily change these parameters, while it does shift the PG emission profiles. Therefore, we did not report on the exact relations between the parameters extracted from the PG profiles and the dosimetric parameters. The large variation in correlation strength indicates that the calculated relations, which could be established through linear fitting, would vary greatly in predictive value and would have little added value. In addition, population-based relations could also not be determined due to the large variation between patients. This study was also limited by statistics in terms of the number of incident protons that were simulated and this inhibited using all pencil-beams for analysis because many would simply have a too low number of detected PG photons. The aim of the current work was to establish, as a proof of principle, that correlations between the changes in PG emission profiles and in dosimetric parameters exist.

The simulated target coverage was considerably lower compared to what was planned and would not be clinically acceptable. This was due to the difference in dose calculation technique and an imperfect translation of the plan parameters (e.g. exact angles of incidence), compromising the target coverage when compared to a fully optimized dose distribution. However, this did not influence the result since we only analyzed the relative differences.

If the proposed technique is to be used for dose monitoring in clinical practice, a set of reference PG emission profiles should be determined before the first treatment fraction, which could be done through Monte Carlo simulations. However, the full detector system (e.g. collimator and detector characteristics) should be considered in the predictive model. It would also be possible to determine the reference profiles from measurements during the first treatment fraction, but in this case no dose monitoring could be done during this fraction. In addition, if there is a large difference in patient anatomy during this first fraction, all subsequent fractions might be classified as incorrect.

The presented approach for dose monitoring using PG is not suitable for treatment plans that are re-optimized on a daily basis because there are no reference profiles to which the profiles of that day can be compared. It could e.g. be possible to calculate the new reference PG profiles from a re-optimized dose distribution by using a filtering approach similar to the approach described by Parodi *et al.* for dose monitoring using positron emission tomography techniques [127].

Only small differences were observed between the results when using the 5%, 10% or 20% most intense pencil-beams. It thus appears that relatively little information is added when increasing the number of pencil-beams used for analysis when using the current pencil-beam selection method (i.e. based on intensity). Other pencil-beam selection methods (e.g. based on their contribution to specific dosimetric parameters) could benefit from using a greater number of pencil-beams for analysis. Other factors that can have an even greater effect on the performance of the proposed method are the background signal caused by neutrons and including a realistic detector efficiency.

The pencil-beam selection method that was used in this study was purely based on achieving the best counting statistics for the PG emission profiles. Since the 20% most intense pencil-beams automatically included the 10% most intense pencil-beams and again this included the 5% most intense pencil-beams, the results from these three selections were correlated. Future research should focus on deriving parameters from the detected PG signals that have better specificity. For example, by selecting pencil-beams that specifically contribute to the dose to a selected structure rather than the most intense beams, and using these pencil-beams to determine dosimetric changes for that structure. Then, it might be possible to determine the exact relations between the PG emission profile shifts and the dosimetric parameters and to determine exact threshold values and action levels that can be used for dose monitoring in daily practice. This was not possible in the current work because of a too low number of simulated protons per pencil-beam and a change in pencil-beam selection could result in performing the analysis with PG emission profiles that had a poor contrast-to-noise ratio. In addition, devising methods to select pencil-beams that contribute to specific dosimetric parameters was outside the scope of this study.

This study has a number of known limitations such as: relatively low counting statistics, the use of a perfect and cylindrical detector surface, the rejection of neutrons at the detector surface, using only the PTV on the planning CT to calculate the dosimetric parameters, and not considering positioning uncertainties. Despite these limitations, the obtained results are a next step in the advancement of PG treatment monitoring research. The presented data shows that there is a correlation between the shifts of a

set of PG emission profiles and the change in volumetric dosimetric parameters. This is a first step in the clinical translation of the detection of range shifts of individual pencil-beams to changes in clinically relevant dosimetric parameters.

8.5 CONCLUSIONS

It appears feasible to use prompt gamma ray emission profiles, obtained during pencil-beam scanning proton therapy, to detect dosimetric changes of the PTVs resulting from day-to-day anatomical variations. Changes in PTV coverage correlate with changes in PG emission profiles, but more research is needed to establish the exact relation between the changes in PG emission profiles and in dosimetric parameters. PG based treatment monitoring could then be used to obtain real-time quantitative information on the dosimetric quality in a non-invasive manner.

8.6 ACKNOWLEDGMENTS AND CONFLICTS OF INTERESTS

This study was financially supported by ZonMw, the Netherlands Organization for Health Research and Development, grant number 104003012 and by Varian Medical Systems. The calculation time on the SURFsara Lisa system was granted by NWO Physical Sciences. The CT-data with contours were collected at Haukeland University Hospital, Bergen, Norway and were provided to us by responsible oncologist Svein Inge Helle and physicist Liv Bolstad Hysing.

Chapter

9

Discussion

In this thesis, we have worked towards online adaptation in intensity-modulated proton therapy (IMPT). We have developed and evaluated two fully automated adaptation methods, aiming to restore target coverage and increase organ at risk sparing in the treatment fractions. With the dose restoration method, we managed to reconstruct the prior dose distribution as made for the planning CT scan, increasing the success rate in terms of target coverage. Consequently, using this method will allow for less robustness to be included during treatment planning, which in turn can increase healthy tissue sparing. The full adaptation method allowed for the daily anatomy to be taken into account, thereby improving organ at risk (OAR) sparing. We have investigated the feasibility of the latter method for tumor sites showing large day-to-day variations, and we compared both the dose restoration and the full adaptation method to robust treatment planning.

In this chapter we will discuss the limitations of the methods developed throughout this thesis, followed by a discussion of the general developments and the future of (online-adaptive) IMPT.

9.1 LIMITATIONS AND IMPROVEMENTS OF THE ONLINE ADAPTATION METHODS

In the development of the adaptation methods described in this thesis, the focus has been on making the workflow fully automated. To this end, we aimed for a success rate of 100% in the adapted plans. This entails making them clinically acceptable without any user alterations, thereby reducing the need for review. To achieve this, we have been rather strict on certain plan characteristics.

In terms of target dose, the aim of adaptation was always to achieve an adequate target coverage, with OAR sparing having a lower priority. In the dose restoration method this aim was achieved by increasing the optimization-weight of the voxels not conforming to the planning criteria. In the full plan adaptation, minimum target dose was included as a hard constraint in the reference point method (RPM) optimization. In some instances, however, it could be that slightly lower target coverage is clinically desirable, as critical OARs could be better spared. If this holds for the entire treatment course, obviously the solution would be to alter the planning criteria in the RPM for this patient. For the plan restoration method, another solution could be to leave the prior plan unchanged, but put more focus on the optimization-weights of the OARs instead of the targets (see Chapter 2).

Besides the target coverage, we have invested in limiting the number of energy layers that are included in the adapted treatment plans. Limiting the number of energy layers reduces the treatment time, and as the running time of our dose calculation

[21] depends on the number of energy layers, the restrictions also reduce the calculation times of the (intermediate) dose calculations. Instead of using all feasible energy layers, we limited the number of energy layers by only allowing those energy layers that were used in the prior treatment plan. Additional energy layers were only allowed if these prior energy layers proved insufficient or too few. Although restricting the number of energy layers could yield an imperfect spot-restoration and reduce the number of degrees of freedom for spot addition, we have found that even with these restrictions adequate treatment plans were obtained. It should be noted that the arbitrary limit of 200 energy layers in the spot-addition step was actually almost never limiting for the prostate cancer database. For other treatment sites it could be limiting, so it should be evaluated whether a limit is actually necessary or should be removed altogether.

The RPM makes trade-offs between planning objectives, which can be tuned by the RPM-parameters. These parameters can be automatically obtained by training the RPM on a set of clinically desirable plans [47]. Depending on the tuning protocol and stopping criteria however, this training, though automated, can take in the order of days to complete for a training set of ten treatment plans. Furthermore, the evaluation of the parameter values is challenging, as their individual meaning cannot be directly linked to a specific planning objective. In Chapter 6, we showed that RPM-parameters trained on Erasmus-iCycle plans did not produce exactly the same trade-offs as in the training set. Therefore, as discussed in Chapter 6, there is room for improvement regarding the training and evaluation of the RPM-parameters. This could for example be achieved by performing the RPM-parameter tuning on treatment plans with various pencil-beam sets, avoiding overfitting on a specific type of pencil-beam set.

Possible improvements of the adaptation method could be in the handling of the restored spots and the addition of new spots. The energy-adaptation of the prior plan restores the prior spots to their initial position. As however different contours are used, these restored spots might not end up in the new target region. In the current method, such spots are still included in the optimization. As no real contribution of these spots is expected, they could be removed immediately after the energy-adaptation. Whether this will result in changes in plan quality and calculation times will have to be evaluated. For the new spots, the approach used for selecting the new spots could be altered. Presently the spot addition is done by sampling a number of spots using latin hypercube sampling. This sampling method aims at taking a homogeneous sample from all possible spot positions. Better results might be obtained when adding spots in a fine regular grid instead of sampling them. A downside

would be the increased calculation times, as more pencil-beams would be added. Another possibility could be to first determine where new spots are required. This is not straightforward, because a pencil beam not only contributes dose to the dose distribution in the Bragg peak, but also proximally to it.

9.2 APPLICATIONS OF THE ONLINE ADAPTATION METHODS

The methods developed in this thesis were created to adapt IMPT treatment plans on a daily basis. So far, the methods have only been evaluated on CT-data of a cohort of prostate and cervical cancer patients. The results demonstrate that the adaptation can be applied for relatively large target volumes exposing large day-to-day variation in shape and location and with significant daily density variations along the pencil-beam paths. This implies that the developed methods can also be applied to other treatment sites. For example, tumors in the head and neck region would be interesting to investigate as there are many organs at risk in close proximity. As was shown by van de Water *et al.* [10], reducing the amount of robustness can reduce the normal tissue complication probability (NTCP) value. The amount of required robustness can be reduced by accounting for density variations such as varying nasal cavity fillings using the dose restoration approach. Even more reduction of robustness can be achieved, by using the full plan adaptation approach to account for gradual changes such as weight loss and residual positional variations caused by spine bending.

Obviously, the daily variations in shape and location of the target volume and OARs do not only occur in IMPT, but may also happen in conventional radiotherapy. The dose distribution, however, is less affected as the static dose cloud approximation holds. Therefore, adding a margin around the target is usually sufficient. However, applying full plan adaptation on a daily basis can reduce margins in conventional radiotherapy as well, thereby improving OAR sparing. Plan adaptation using a restoration followed by a weight optimization was already described for prostate radiotherapy by Ahunbay *et al.* [51]. Instead of the energy-adaptation step they described a segment aperture morphing (SAM) technique. After this a segment weight optimization is performed which optimizes the segment weights of the individual segments. Similar techniques are also used in treatments with an MR-Linac, where a reference plan is either restored after an isocenter shift, or fully adapted to new daily contours (adapt to position and adapt to shape, respectively) [52]. While the restoration steps, such as the SAM and the energy-adaptation, are specific to the treatment modality, the subsequent weight optimization techniques can be used interchangeably. It would be interesting to see how the RPM optimization would perform when used for optimizing these segment weights.

Although the dose restoration method was described here as a method to restore a plan to its initial state, this method could also be applied to generate plans for a new anatomy. Skipping the energy adaptation step and instead positioning pencil-beams in a regular grid, any dose distribution could be given as input for the restoration method. The input could for example be a predicted dose distribution, based on the daily anatomy. Three dimensional predictions of photon dose distributions have already been generated by several groups using knowledge-based modeling and deep learning [128]. Taking such a dose prediction as input for a restoration method can result in a feasible treatment plan, as has already been shown by Fan *et al.* [129]. It should however be noted that while this could lead to clinically acceptable plans, there is no guarantee that the restored plans will be Pareto optimal.

The online adaptation methods developed in this thesis can also be used to generate new treatment plans, as already shown for IMRT by van Haveren *et al.* [33, 34]. In the Supplementary Materials of Chapter 4 (Appendix A.2) we have also evaluated the use of the full plan adaptation to generate plans from scratch. For an online setting, we found that the required calculation times were too long. Applying the method as a tool to generate new treatment plans in an offline setting would however be feasible.

9.3 THE FUTURE OF (ONLINE-ADAPTIVE) IMPT

9.3.1 Restoration, full plan adaptation or something else?

Considering the future of online-adaptive IMPT, the question arises whether daily plan adaptation is possible, and whether it is actually necessary.

In Chapter 2 we have seen that applying the dose restoration can return a dose distribution to its intended state. The restoration method aims for a similar dose distribution in each treatment fraction, by only adapting for density variations along the pencil-beam paths. Applying this dose restoration method to a prior plan effectively overcomes the limitation of the static dose cloud approximation in proton therapy. Therefore, by adding planning target volume (PTV) margins to the prior plan, it can be used in a similar way as is currently done in photon therapy where PTV margins are used to account for geometrical errors. As no new contours are required, applying this method on a daily basis is feasible, as long as daily volumetric imaging is available that can be used for accurate proton dose calculation.

In Chapters 3 – 5 we evaluated the more complex full plan adaptation approach. This method has the benefit of optimizing the treatment plan for the actual daily anatomy. This means that when using this method, the degree of included robustness/margins can be further reduced as the prior plan no longer has to account for all possible

shapes and locations of the target. Furthermore, using this approach we can adapt the dose to the OARs depending on their position with respect to the target, and also adapt for shape changes of the target and possible tumor shrinkage. However, for this method to become feasible, daily contours are required. Different methods of obtaining such contours are of course possible, including (combinations of) manual delineation, delineation propagation and full auto-delineation. Full use of auto-delineation is obviously preferable (for example as described in Chapter 7), where at most a manual check needs to be performed prior to the plan adaptation. If such methods, or other combinations, can be run in a limited time, then applying full plan adaptation is feasible as was demonstrated in this thesis.

In Chapter 5 we have shown that when compared to robust treatment planning, online-adaptive IMPT can lead to reductions in OAR doses in prostate cancer treatments. The amount of benefit will however depend on the treatment site and individual patient. To determine whether daily adaptation is truly necessary, an essential step is to evaluate whether these dose reductions are clinically relevant. Determining which patients or patient groups are likely to benefit from the plan adaptation is therefore required. In line with the Dutch model-based approach of patient selection for proton therapy (for more details see section 9.3.4), an interesting approach would be to trigger an adaptation based on a change in NTCP. Bijman *et al.* showed that some NTCP models are more affected by dose uncertainties than others [130]. This illustrates that, while less dose is preferable, the clinical impact of dose changes depends on other factors as well.

Besides determining whether a patient will benefit from plan adaptation, the optimal frequency of adaptation should be determined. For treatment sites with small day-to-day variations, but significant weekly changes, a weekly plan adaptation could be sufficient. Combinations of different mitigation approaches could also be considered, such as daily dose restoration combined with weekly plan adaptation. Important to note is that the daily plan adaptation as described throughout this thesis is based on daily acquired CT scans. If implemented in the clinical workflow, the additional dose that is given to the patient during these scans should be acknowledged [131]. Furthermore, in-room CT scanners are not widely available. Alternatively, other groups have shown that daily adaptation may also be achieved using cone beam CT (CBCT) scans [48, 50].

Prior to performing an actual comparison study, some patient groups and treatment schedules can already be expected to greatly benefit from daily adaptation. Treatments using hypo-fractionation will likely benefit from online plan adaptation, as there is less chance for random daily errors to be averaged out in other treatment

fractions and higher daily doses are delivered per fraction. Treatment sites for which density changes frequently occur, could also benefit from online adaptation. An example is tumors in the thoracic region. In terms of patient groups, online plan adaptation is an attractive option for pediatrics. Due to their longer survival, children are at a higher risk of secondary cancers caused by dose to the healthy tissues. Online plan adaptation can reduce these doses and thereby reduce the risks.

9.3.2 Towards clinical implementation of online-adaptive IMPT

In this thesis methods for online adaptation were developed and tested. The next step will be to move towards clinical implementation. The clinical implementation of automated online-adaptive proton therapy will be pursued in a research project within the HollandPTC-consortium, at the Erasmus MC Cancer Institute and the Technical University of Delft (the IMAGINATION project, 'Improved healthy-tissue sparing by automated daily online plan adaptation in proton therapy for head and neck cancer'). This project will also focus on automated quality assurance of adapted plans and online patient-specific quality assurance based on the actually delivered treatment.

The success of the implementation will critically depend on the used parameter settings for the adaptation (Chapter 6). For dose restoration for instance, additional research is needed to determine the settings for an optimal prior plan. Full plan adaptation requires daily contours as input. For some treatment sites, e.g. with large day-to-day variations, such as cervical cancer, obtaining new contours in a short time frame is currently still very difficult. For these type of treatment sites, dose restoration could be a better option to start with. For treatment sites with large day-to-day variations, the dose restoration can be combined with a library of prior plans, as was done for the full plan adaptation in Chapter 4. Chapter 6 has shown that taking a smaller prior plan for restoration can result in lower doses to the OARs. Based on this result, dose restoration is likely to benefit of using such a plan-library. This approach is pursued in the PROTECT ('Online-adaptive proton therapy for cervical cancer to reduce the impact on morbidity and the immune system') project for online-adaptive proton therapy in locally advanced cervical cancer.

Another important aspect is to establish how much the degree of robustness and/or margins can be reduced for those patients receiving daily adaptation. Adaptations in the presence of shrinking targets should be carried with restraint as remaining microscopic tumor could be missed, which could result in loss of tumor control and increased recurrence rates.

In the daily online-adaptive workflow, the adapted plan has to be verified to determine whether it can be delivered. For both methods the planning criteria should obviously be checked, to ensure that target coverage is adequate and whether constraints on healthy tissues are met. If not, the user should automatically be warned and preferably be provided with an alternative. For example, when a dose-limiting OAR has moved closer to the target in a certain fraction, a target under-dosage could be required in that fraction. An automatic warning of this could be sent to the user, and with a predefined alternative set of planning criteria, an alternative treatment plan could then be generated.

Besides verifying the clinical treatment planning constraints, it could be checked whether the adaptation actually performs better than (or similar to) the prior plan. For plan restoration, this can for example be done by computing the dosimetric parameter values of the restored plan and comparing these to the prior plan. As the aim was to restore the dose distribution, and the same contours are being used for evaluation, this comparison can provide a good indication whether the restored plan is acceptable. By defining thresholds for the differences per criteria, this check can then be fully automated. An automatic warning can be issued, allowing the user to for example try an alternative weighting. For the full plan adaptation, the evaluation is done on the new daily contours, meaning that a comparison with the prior values might not be conclusive. While it might provide an indication of what kind of values would be considered acceptable, a comparison with the distorted plan, i.e. the prior plan recalculated on the new anatomy, would be more meaningful. Parallel to the adaptation a forward calculation of the prior plan can be performed, and the dosimetric parameter values of the distorted plan can be automatically computed and compared to the adapted plan. For both online-adaptive methods, comparing to the prior (or distorted) plan can provide an indication of the plan quality compared to non-adaptive treatment. No information is however obtained on how good the plan is for the specific anatomy of the day. To include a check such as this, a knowledge-based approach could be implemented which gives a prediction of obtainable dose distributions for a specific anatomy [128].

Another important aspect of the clinical implementation of a daily online-adaptive workflow is patient-specific plan quality assurance (QA). With the limitation of available QA time, alternative solutions, preferably automated, will have to be investigated. Checking the viability of the adapted plan in terms of MU can be fully automated. Alternative QA methods could include real-time monitoring of the dose delivery and retrospective control through machine log files. In Chapter 8 we have shown that dosimetric changes caused by anatomical variations can be correlated to changes in

measured prompt-gamma profiles. The use of this methodology as an online monitoring measure, however, is still far from routine clinical use. QA through the use of machine log files on the other hand has been proposed as an alternative and may be more straightforward to implement [63, 64]. The drawback of this method is that it yields the results after the delivery of the treatment fraction. Secondary independent calculations of the proton therapy machine settings compared to the actual machine settings can be used to verify the treatment plan prior to delivery.

9.3.3 Future innovations of (online-adaptive) IMPT

When online-adaptive IMPT gets introduced into the clinic together with technical innovations improving speed, applications can go beyond daily plan adaptation. Where the work in this thesis focused solely on adapting for inter-fractional variations, the next step would be to also adapt for intra-fractional variations. This would be especially interesting in moving targets such as tumors in the chest and abdomen. To achieve this, the first step would be to reduce the running times, with in particular those of the intermediate dose calculations. Adaptation could then be done after each beam, or even continuously. A method for dose accumulation should be included to check whether the prescribed fraction dose has been achieved.

The methods we developed in this thesis are able to adapt for setup uncertainties and changing anatomies. To account for uncertainties in the conversion from Hounsfield unit to proton stopping power, the full plan adaptation can robustly optimize for range errors. These uncertainties could also be eliminated by applying transmission beam proton therapy. In this technique a single sufficiently high energy is selected, such that the proton beams do not stop in the patient. Currently a popular research field for this application is flash therapy, where treating tissue with a very high dose-rate induces the so-called ‘flash-effect’ [132–135]. So far obtaining these high dose-rates only seems possible when using the highest energy. As no Bragg peaks are used in these types of plans, range uncertainty and the effects of density changes are eliminated.

Daily treatment plans obtained from scratch are expected to become available in the future with further improvements on dose calculation and optimization speed. Running the calculations on a GPU and making certain assumptions and simplifications, Matter *et al.* [58] already showed that a simple treatment plan can be generated within 10 seconds. It is important to consider that the simplifications used do not compromise on treatment plan quality.

9.3.4 The future of (online-adaptive) IMPT in the Netherlands

When considering the Netherlands specifically, proton therapy is still a very new treatment modality, with the first patients only treated in 2018. As proton therapy is more expensive than conventional radiotherapy, the Dutch health authorities decided that currently only 2200 patients per year will get their treatment reimbursed. Besides a limited number of standard indications, which automatically qualify for proton therapy, patients will pre-dominantly be selected by the model-based approach [136, 137]. This model-based approach has been introduced to objectively select patients for whom a clinically relevant benefit of proton therapy can be expected. Prospective observational cohort studies are used to validate this approach.

For patients with a potential benefit of IMPT, two treatment plans are generated. One simulating a state-of-the-art photon therapy plan, and one simulating a state-of-the-art proton therapy plan. For these plans NTCP values are calculated for one or multiple side effects of the treatment. Patients for whom the proton NTCP value is considerably lower than the photon NTCP value (the threshold is dependent on the grade of the side effect), are then referred to proton therapy. The NTCP values are calculated using models based on parameters such as delivered dose to specific organs but also on patient and tumor characteristics. Generally speaking, the NTCP value increases with increasing dose. For some tumor sites such as brain, for which no validated NTCP models are available, the actual dose levels to critical organs are used for selection.

The NTCP models are derived from dose distributions calculated on the planning CT scan. Currently, in robust treatment planning, the results obtained for the nominal planning scenario are used to calculate the NTCP values. When using full plan adaptation, the plan will be adapted throughout the treatment, changing the dose at each fraction. That is why one may wonder whether the standard NTCP models used for patient selection are still valid for online-adaptive treatments. Therefore, it is necessary to validate the NTCP models separately for online-adaptive treatments and to develop methods to calculate the expected NTCP value before the start of the adaptive treatment.

The use of online-adaptive treatment planning can assist in the clinical validation of proton therapy. The main claim regarding proton therapy is that it can better spare healthy tissues than photon therapy at the same tumor control probability. For this to be validated, the claimed dose reduction to healthy tissues needs to be realized, and clinical relevance of the dose reduction needs to be demonstrated. This claim implicitly urges the need for the development and implementation of motion mitigation, which is known to reduce the OAR sparing ability of IMPT [10, 138]. In Chapter

5 we have observed that online plan adaptation can reduce the required motion mitigation and can increase OAR sparing compared to current passive motion mitigation strategies such as robust treatment planning. The claimed dose reduction compared to photon therapy will hence be easier to realize using online adaptation instead of using the current robust treatment planning approaches.

In conclusion, this thesis shows that online-adaptive IMPT is within reach and that the transition to clinical implementation of online-adaptive IMPT is the most logical step. Ultimately this development should lead to reduced radiation-induced side effects and to improved patients' quality of life.

Chapter

A

Appendices

A.1 SUPPLEMENTARY MATERIALS – AN AUTOMATED PLANNING STRATEGY FOR NEAR REAL-TIME ADAPTIVE PROTON THERAPY IN PROSTATE CANCER

Table A.1.1: Percentages of the 88 dose distributions that meet the target constraints for the CTV structures for the investigated methods. The RPM adaptive treatment plans are obtained using a single iteration of adding new spots. Scans that meet these criteria do not necessarily have the ability to account for intra-fraction motion.

	Prior (%)	Spot-restoration (%)	RPM adaptive (%)	Benchmark (%)
$V_{95\%} \geq 98\%$ CTV _{prostate}	78.4	97.7	100	100
$V_{95\%} \geq 98\%$ CTV _{seminal vesicles}	90.9	96.6	100	100
$V_{95\%} \geq 98\%$ CTV _{lymph nodes}	79.6	98.9	100	100
$V_{107\%} \leq 2\%$ CTV _{prostate}	63.6	92.1	94.3	100
$V_{107\%} \leq 2\%$ CTV _{seminal vesicles}	32.0	94.3	100	100
$V_{107\%} \leq 2\%$ CTV _{lymph nodes}	100	100	100	100
Dose distributions meeting all criteria	36.4	87.5	94.3	100

A.2 SUPPLEMENTARY MATERIALS – PLAN-LIBRARY SUPPORTED AUTOMATED REPLANNING FOR ONLINE-ADAPTIVE INTENSITY-MODULATED PROTON THERAPY OF CERVICAL CANCER

Patient data

All patients were treated at the Leiden University Medical Center between 8 April 2015 and 3 May 2017. Written informed consent was obtained for all patients. For each patient, a full- and empty-bladder CT scan was acquired as well as four weekly repeat CT scans, resulting in a total of 23 repeat CT scans. All scans were acquired in supine position. For one patient no empty-bladder CT scan was available. For this patient the repeat CT scan with the smallest bladder volume was used as the empty-bladder CT scan. The remaining 23 repeat CT scans were used as representations of the daily anatomies. The low-risk primary tumor related clinical target volume (CTV-TLR) and the lymph nodes (CTV-E) were defined as target structures according to the EMBRACE II protocol [139, 140], and the rectum, bladder, bowelbag (outer extension), sigmoid and the femoral heads were defined as organs at risk (OARs). Delineations of the target volumes and OARs were available in all repeat CT scans.

The automated adaptive treatment planning method

Figure A.2.1 shows the workflow of the prior-plan adaptation method investigated in this study.

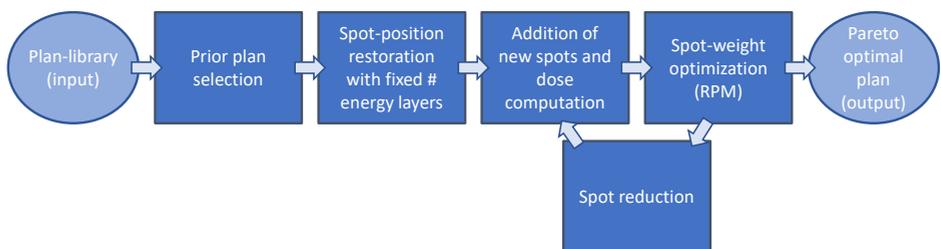


Figure A.2.1: The workflow of the automated prior-plan adaptive treatment planning method. A library of prior plans is generated for the patient. Each fraction, the best fitting prior plan is selected from the plan-library. A restoration of the spot-positions with a fixed number of energy layers follows, in which all Bragg peaks are restored to their new water equivalent path length. After this the method enters a loop in which new spots are added, the intensities are optimized and the non-contributing spots are deleted. The loop can be repeated; we evaluated running the loop zero, one or two times. The output is a Pareto-optimal treatment plan. Adapted from [44].

Plan-library generation for the prior-plan strategy

Figure A.2.2 shows examples of the ITV structures used for the library of prior plans, in the sagittal view.

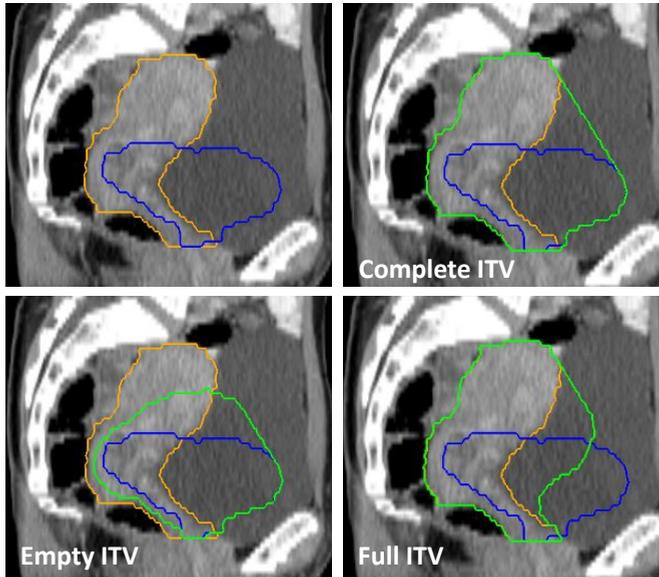


Figure A.2.2: An example of the ITV structures, shown in sagittal view. Top-left shows the cervix-uterus corresponding to the empty-bladder CT scan in blue, and the cervix-uterus corresponding to the full-bladder CT scan in orange. Top right shows the Complete ITV in green. Bottom left and right show respectively the Empty ITV and Full ITV in green.

Multi-criteria optimization for the benchmark and prior plans

Benchmark and prior plans were generated on their respective PTV_{OAPT} and PTV_{Prior} , using our in-house developed fully automated treatment planning system ‘Erasmus-iCycle’. Optimizations were done according to the wishlist shown in Table A.2.1. The full multi-criteria optimization includes iteratively sampling 3000 spots to the target region using latin hypercube sampling, optimizing the spot-weights and removing non-contributing spots. Latin hypercube sampling is a sampling method that takes the previously sampled spots into account when selecting the next spot, aiming at a homogeneous sampling. Spots are sampled from a virtual grid with a 1 mm resolution; a lateral spacing of 1 mm, with the energy spacing set to the longitudinal width of the Bragg peak (relative energy factor of 1).

Table A.2.1: The wishlist with planning constraints and objectives used for automated IMPT plan generation for cervical cancer patients. Constraints will always be met. The priority numbers of the objectives indicate the order in which objectives are to be optimized. A low number corresponds to a high priority.

Constraints	Structure	Type	Limit
	PTV	Minimum	$0.95 \times 45 \text{ Gy(RBE)}$
Objectives			
Priority	Structure	Type	Limit
1	PTV	Maximum	$1.07 \times 74 \text{ Gy(RBE)}$
1	Patient	Maximum	$1.07 \times 74 \text{ Gy(RBE)}$
1	PTV-low	Maximum	$1.07 \times 55 \text{ Gy(RBE)}$
2	Conformity ring PTV-high	Maximum	$1.07 \times 74 \text{ Gy(RBE)}$
2	Conformity ring PTV-full 0 – 10 mm	Maximum	$1.07 \times 55 \text{ Gy(RBE)}$
2	Conformity ring PTV-full 10 – 15 mm	Maximum	$0.90 \times 55 \text{ Gy(RBE)}$
3	Femoral heads	Maximum	50 Gy(RBE)
4	Rectum	Mean	1 Gy(RBE)
5	Small and large intestines	Mean	1 Gy(RBE)
6	Bladder	Mean	1 Gy(RBE)
7	Femoral heads	Mean	1 Gy(RBE)
8	All conformity rings	Mean	1 Gy(RBE)
8	All conformity rings	Maximum	1 Gy(RBE)
9	Total spot-weight	Sum	1 Gp

Abbreviations: PTV = planning target volume; Gp = Gigaprotons

RPM-parameter tuning

In this study the RPM-parameters were automatically tuned on a subset of the database. The parameters were obtained using the method presented in [47], which aims at minimizing the differences in plan objective values between the benchmark plans and the plans obtained by the RPM with the found parameters. Tuning was stopped when the differences complied the demands of the protocol shown in Table A.2.2.

As the results of the prior-plan adaptation method might depend on the RPM-parameters, a three-fold cross validation was applied. For each fold two different patients were used for parameter tuning. The planning strategies using the found parameters were evaluated on the other four patients of each fold.

To evaluate the RPM-parameters for the three tuning folds, RPM plans were generated on the scans not used for tuning, using the benchmark spots and the obtained RPM-parameters. Since both the benchmark plans and RPM plans are Pareto-optimal, differences are always observed. The preferences in the protocol define and limit these trade-offs, but might not be optimal; changing the protocol can affect the res-

ults of the method. For an acceptable configuration, generated plans should achieve adequate target coverage and the differences between the benchmark and RPM plans should be small. For this study we achieved most differences being smaller than 1 Gy(RBE) and 1%-point. Further improvement of this protocol was outside the scope of this study.

Results of the planning strategies for the individual folds

Table A.2.3 shows the number of scans that achieve the target demands for each planning strategy for the individual folds.

It can be seen that different success rates were obtained for the different folds. While folds 1 and 2 show acceptable target coverage for all scans after one iteration of spot addition, fold 3 requires a second iteration. Including more patients in the tuning would probably result in more stable results, and possibly in acceptable coverage for all scans after one iteration. When however only a small dataset is available for the RPM-parameter tuning, running an additional iteration of adding spots could be a solution. Another possibility is to check the target demands after each iteration of spot addition and weight optimization, only running another iteration when necessary. Figure A.2.3 shows the differences for the OARs between the plans obtained using the 2-plan-library approach with two iterations of spot addition and the benchmark plans for the three different folds. For all folds, the highest prioritized criteria (D_{max}) deteriorated less than 1 Gy(RBE) compared to the benchmark plans, where some resulted in even lower doses (bowelbag, sigmoid D_{max}). For lower prioritized criteria (D_{mean}), all folds obtained OAR values within 8 Gy(RBE) and 8%-point from benchmark.

Non-prior-plan strategy

Two approaches of the non-prior-plan strategy were investigated in this study. In the first approach, which places the spots in a regular grid, different spacings between the energy layers were investigated, while keeping a 5 mm lateral spacing. The spacing between energy layers was set to vary with the energy by making it relative to the longitudinal width of the Bragg peak (at 80% of the peak height). For evaluation, besides taking a relative energy factor of three (i.e. New-Spots-E3), relative energy factors of four and five were considered. New-Spots-E4 and New-Spots-E5, respectively. A larger relative energy factor corresponds to a larger energy spacing, resulting in fewer spots.

In the second approach, the spots are iteratively sampled from a fine grid as was done for the benchmark plans: with the energy layer spacing set to the longitudinal width

Table A.2.2: The protocol which was used in the RPM-parameter tuning. Includes the user preferences for the population-based differences between the benchmark plans and the RPM generated plans (benchmark – RPM).

Plan objective	Measure	Lower bound
PTV D_{\max}	Median	-0.5
	1 st quartile	-0.2
	5 th percentile	-0.1
PTV rings D_{\max} , D_{mean}	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.2
Patient D_{\max}	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.2
Rectum D_{\max} , D_{mean}	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.3
Rectum $V_{40 \text{ Gy(RBE)}}$, $V_{30 \text{ Gy(RBE)}}$	Median	-1
	1 st quartile	-0.75
	5 th percentile	-0.5
Bladder D_{\max} , D_{mean}	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.3
Bladder $V_{40 \text{ Gy(RBE)}}$, $V_{30 \text{ Gy(RBE)}}$	Median	-1
	1 st quartile	-0.75
	5 th percentile	-0.5
Bowelbag D_{\max} , D_{mean} , $V_{40 \text{ Gy(RBE)}}$, $V_{30 \text{ Gy(RBE)}}$	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.3
Sigmoid D_{\max} , D_{mean}	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.2
Femoral heads D_{\max} , D_{mean}	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.2
Total spot-weight	Median	-3
	1 st quartile	-2
	5 th percentile	-1

of the Bragg peak (relative energy factor of 1). Besides stopping this non-prior-plan approach after taking three iterations, we also investigated stopping after one and

Table A.2.3: For each treatment strategy, the number of evaluation scans that meet the prescribed target demands for the individual folds.

	Fold 1 $V_{95\%} \geq 95\% \ \& \ V_{107\%} \leq 2\%$	Fold 2 $V_{95\%} \geq 95\% \ \& \ V_{107\%} \leq 2\%$	Fold 3 $V_{95\%} \geq 95\% \ \& \ V_{107\%} \leq 2\%$
1-plan-library No replanning	0/15	0/16	0/15
2-plan-library No replanning	0/15	0/16	0/15
1-plan-library-0x	1/15	1/16	0/15
2-plan-library-0x	0/15	0/16	0/15
1-plan-library-1x	12/15	15/16	10/15
2-plan-library-1x	15/15	16/16	10/15
1-plan-library-2x	15/15	15/16	13/15
2-plan-library-2x	15/15	16/16	15/15
Sampled-New-Spots-3x	15/15	16/16	8/15
New-Spots-E3	15/15	16/16	15/15

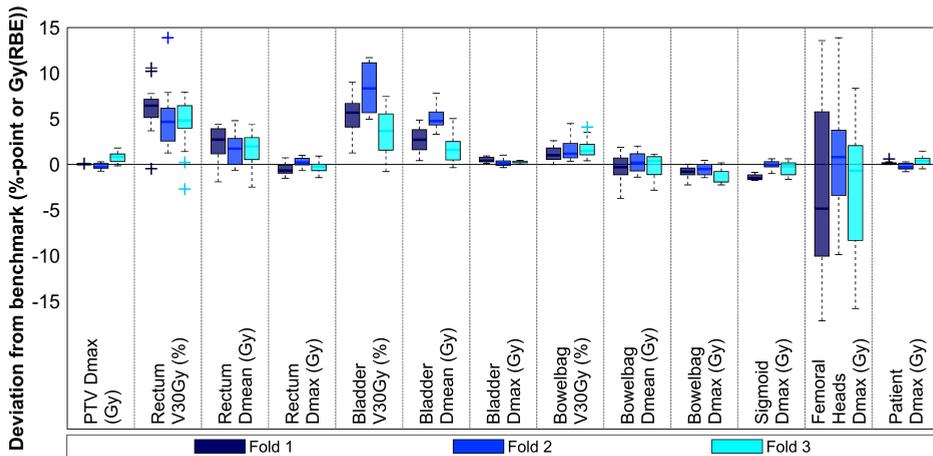


Figure A.2.3: Boxplots depicting the OAR differences between the 2-plan-library-2x and benchmark plans for the three evaluation folds. Negative deviations depict scans for which the OAR value is lower in the RPM plan than in benchmark.

two iterations. In between the iterations non-contributing spots are removed. The results of these approaches will be denoted by respectively Sampled-New-Spots-1x and Sampled-New-Spots-2x.

Table A.2.4 shows the number of plans which meet the prescribed target results for

all investigated non-prior-plan approaches. It can be seen that stopping after fewer iterations in the Sampled-New-Spots approach or taking a larger relative energy factor in the New-Spots approach results in shorter calculation times, but also fewer scans meeting the target demands. New-Spots-E4 however does meet the target demands for all plans.

Table A.2.4: For each non-prior-plan approach, the number of evaluation scans that meet the prescribed target demands. The last column shows the total calculation time excluding the final dose calculation.

	$V_{95\%} \geq 95\%$ & $V_{107\%} \leq 2\%$	Calculation times (min.) mean (min – max)
Sampled-New-Spots-1x	0/46	2.7 (2.0 – 3.9)
Sampled-New-Spots-2x	24/46	4.8 (3.7 – 5.6)
Sampled-New-Spots-3x	39/46	7.1 (5.7 – 8.4)
New-Spots-E5	14/46	10.3 (6.5 – 15.6)
New-Spots-E4	46/46	18.9 (11.3 – 32.9)
New-Spots-E3	46/46	40.7 (25.0 – 78.4)

In Figure A.2.4 the OAR results of the 2-plan-library-2x approach and the New-Spots-E4 and New-Spots-E3 approaches are compared to benchmark. It can be seen that although the New-Spots-E4 approach meets the target demands, the OAR results differ more from benchmark than those obtained with the New-Spots-E3 approach and the 2-plan-library-2x approach. Though faster than New-Spots-E3, the calculation times of the New-Spots-E4 approach are still more than twice the calculation times required for the 2-plan-library-2x approach.

Robustness in the prior-plan adaptation method

The proposed prior-plan adaptation method does not account for range errors arising from Hounsfield units to proton stopping power conversion. For this reason we also evaluated the strategy with the inclusion of 3% range robustness applied to the PTV_{OAPT} on plan quality and calculation speed. Robustness was included using a ‘minimax’ worst-case approach [7–9] optimizing three scenarios simultaneously (nominal, +3%, -3%). Results showed that range robustness could easily be incorporated in the prior-plan adaptation method, with a limited increase in calculation time compared to non-robust optimization. On average per iteration, the calculation time needed for the spot addition and calculation of the dose deposition matrices increased from 1.4 to 2.7 minutes (1.5 – 5.2), and the calculation time of the RPM optimization increased from 1 to 2 minutes (1.3 – 3.1).

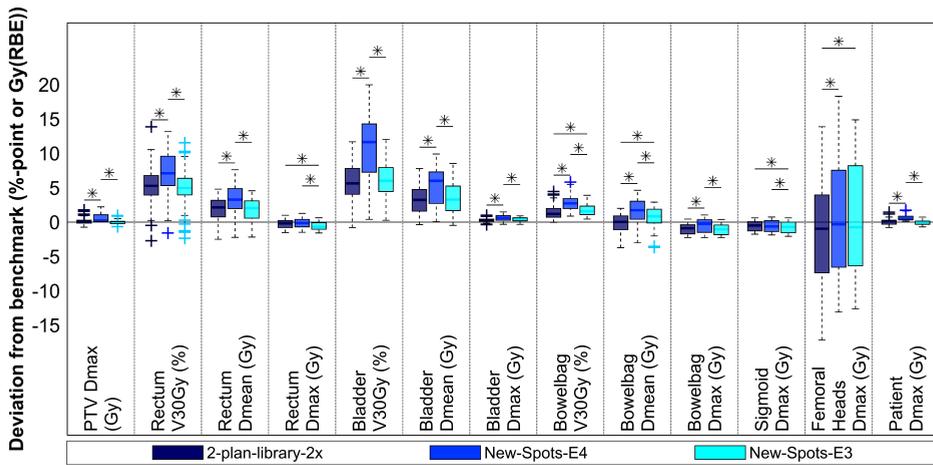


Figure A.2.4: Boxplots depicting the OAR deviations from benchmark for the best prior-plan strategy approach and the non-prior-plan strategy approaches. The prior-plan strategy is shown using the 2-plan-library-2x approach, the non-prior strategy is shown the using New-Spots-E4 approach and the New-Spots-E3 approach. Negative deviations depict scans for which the OAR value is lower in the RPM plan than in benchmark. Statistically significant differences (Wilcoxon signed-rank test, 1% significance level, $p < 0.01$) are indicated by asterisks.

A.3 SUPPLEMENTARY MATERIALS – ONLINE-ADAPTIVE VERSUS ROBUST IMPT FOR PROSTATE CANCER: HOW MUCH CAN WE GAIN?

Deriving the robustness settings

As part of this study we derived the robustness-parameters and the margins required for this dataset to ensure sufficient coverage in both target regions of all repeat CT scans (rCTs) for at least 90% of the patients (10/11). Obtaining the settings was done by methodically increasing the margins (0 – 8 mm in steps of 2 mm) and the setup error (2 – 8 mm in steps of 2 mm), while evaluating the effect on the rCTs, until at least 10/11 patients obtained adequate target coverage for all rCTs. We investigated the settings when applying the robustness to the target structures alone, and to both the target structures and one or two artificial rings around these target structures. The first ring was a 15 mm ring around the ITV_{High} , excluding the ITV_{Low} . The second ring was a 10 mm ring around the combined ITVs. In all 108 combinations, the range robustness was kept at 3%.

Five combinations yielding sufficient target coverage for all rCTs of 10/11 patients were found. We presented the settings achieving good results for 5/8 rCTs of the last patient. The other four combinations achieved good results for respectively 7/8 (2x), 3/8 and 2/8 rCTs, but also involved larger margins and applied the robustness to both the expanded target structures and artificial rings. The first three combinations applied the robustness on the expanded target structures and the first artificial ring, the last combination applied the robustness on the expanded target structures and both artificial rings. Consequently, we observed higher OAR doses with these settings.

A.4 WISHLIST FOR PROSTATE TREATMENT PLANS

Table A.4.1: The wishlist with planning constraints and objectives used for automated IMPT plan generation for prostate cancer patients. Constraints will always be met. The priority numbers of the objectives indicate the order in which objectives are to be optimized. A low number corresponds to a high priority. An artificial PTV-intermediate was generated as the 15 mm transition between the high-dose PTV and the low-dose PTV, included to steer the dose fall-off.

Constraints				
		Structure	Type	Limit
		PTV-high	Minimum	0.97×74 Gy
		PTV-intermediate	Minimum	0.99×74 Gy
		PTV-low	Minimum	0.99×55 Gy
Objectives				
	Priority	Structure	Type	Limit
f_1	1	PTV-high	Maximum	1.06×74 Gy(RBE)
f_2	1	PTV-intermediate	Maximum	1.06×74 Gy(RBE)
f_3	1	PTV-low	Maximum	1.06×55 Gy(RBE)
f_4	2	Conformity ring PTV-high	Maximum	1.07×74 Gy(RBE)
f_5	2	Conformity ring PTV-full 0 – 10 mm	Maximum	1.07×55 Gy(RBE)
f_6	2	Conformity ring PTV-full 10 – 15 mm	Maximum	0.90×55 Gy(RBE)
f_7	3	Rectum	Maximum	1.02×74 Gy(RBE)
f_8	3	Bladder	Maximum	1.02×74 Gy(RBE)
f_9	3	Body	Maximum	1.06×74 Gy(RBE)
f_{10}	3	Femoral heads	Maximum	50 Gy(RBE)
f_{11}	4	Rectum	Mean	0 Gy(RBE)
f_{12}	5	Small and large intestines	Mean	0 Gy(RBE)
f_{13}	6	Bladder	Mean	0 Gy(RBE)
f_{14}	7	Femoral heads	Mean	0 Gy(RBE)
f_{15}	8	All conformity rings	Mean	0 Gy(RBE)
f_{16}	8	All conformity rings	Maximum	0 Gy(RBE)
f_{17}	9	Total spot-weight	Sum	1 Gp

Abbreviations: PTV = planning target volume; Gp = Gigaprotons

A.5 REFERENCE POINT METHOD OPTIMIZATION

The reference point method (RPM) is used in this study to optimize the spot-weights in a single iteration during the full plan adaptation. Detailed information can be found in [33–36, 44]. The main idea of the RPM is to guide the optimization to a Pareto optimal solution by using a reference path combined with carefully tuned indifference curves of the objective functions. As a result, a single set of RPM-parameters (reference path and indifference curves) generates for each patient a Pareto optimal plan with clinically desired trade-offs, within a single optimization run.

A.6 RPM-PARAMETER TUNING PROTOCOL FOR PROSTATE TREATMENT PLANS

To run the reference point method, the desired trade-offs between objectives should be defined in the RPM-parameters. For this study the RPM-parameters were automatically tuned using a three-fold cross validation, where one third of the patients (selected randomly) was used for tuning and the remaining two thirds for testing. The tuning method that was used iteratively minimizes the differences in plan objective values between predefined benchmark plans and plans obtained using the RPM with the tuned parameters. The full tuning method is presented in [47] and has been applied before in [57]. Tuning was stopped after 20 iterations, or when the differences complied the demands of the tuning protocol shown in Table A.6.1.

Table A.6.1: The protocol which was used in the RPM-parameter tuning for the prostate cancer database. Includes the user preferences for the population-based differences between the benchmark plans and the RPM generated plans (benchmark – RPM).

Plan objective	Measure	Lower bound
PTV D_{max}	Median	-0.5
	1 st quartile	-0.3
	5 th percentile	-0.1
PTV conformity rings D_{max} , D_{mean}	Median	-1
	1 st quartile	-0.8
	5 th percentile	-0.5
Rectum D_{max} , D_{mean} , $V_{75 \text{ Gy(RBE)}}$ $V_{60 \text{ Gy(RBE)}}$, $V_{45 \text{ Gy(RBE)}}$	Median	-0.8
	1 st quartile	-0.5
	5 th percentile	-0.3
Bladder D_{max} , D_{mean} , $V_{65 \text{ Gy(RBE)}}$	Median	-0.8
	1 st quartile	-0.5
	5 th percentile	-0.3
Bladder $V_{45 \text{ Gy(RBE)}}$	Median	-1
	1 st quartile	-0.5
	5 th percentile	-0.3
Small and large intestines D_{mean}	Median	-0.8
	1 st quartile	-0.5
	5 th percentile	-0.3
Femoral heads D_{max} , D_{mean}	Median	-1.5
	1 st quartile	-1
	5 th percentile	-0.5
Total spot-weight	Median	-2
	1 st quartile	-1.5
	5 th percentile	-1

A.7 SUPPLEMENTARY MATERIALS – CORRELATIONS BETWEEN THE SHIFTS IN PROMPT GAMMA EMISSION PROFILES AND THE CHANGES IN DAILY TARGET COVERAGE DURING SIMULATED PENCIL-BEAM SCANNING PROTON THERAPY

Table A.7.1: Distribution of control CTs, number of protons, and pencil-beams per patient and treatment plan.

Pat	# Control CTs	# pencil-beams per plan	# P+ per plan	# P+ per simulated plan	Mean (SD) P+ per planned pencil-beam	Mean (SD) P+ per simulated pencil-beam
1	9	1521	3.4×10^{12}	6.6×10^8	$2.3 (2.1) \times 10^9$	$4.4 (4.0) \times 10^5$
2	8	1407	3.2×10^{12}	7.7×10^8	$2.3 (2.0) \times 10^9$	$5.5 (4.8) \times 10^5$
3	8	1438	3.0×10^{12}	6.3×10^8	$2.1 (2.0) \times 10^9$	$4.4 (4.1) \times 10^5$
4	9	1247	2.9×10^{12}	5.2×10^8	$2.3 (2.1) \times 10^9$	$4.2 (3.8) \times 10^5$
5	7	1361	2.8×10^{12}	5.2×10^8	$2.1 (2.0) \times 10^9$	$3.8 (3.6) \times 10^5$
6	8	1540	3.5×10^{12}	7.6×10^8	$2.2 (2.0) \times 10^9$	$4.9 (4.3) \times 10^5$
7	9	1476	3.2×10^{12}	7.8×10^8	$2.2 (1.9) \times 10^9$	$5.3 (4.6) \times 10^5$
8	8	1297	3.1×10^{12}	6.2×10^8	$2.4 (1.9) \times 10^9$	$4.8 (3.8) \times 10^5$
9	8	1409	3.4×10^{12}	4.6×10^8	$2.4 (2.2) \times 10^9$	$3.3 (3.0) \times 10^5$
10	7	1399	3.1×10^{12}	5.1×10^8	$2.2 (1.8) \times 10^9$	$3.6 (3.0) \times 10^5$
11	7	1496	3.6×10^{12}	7.5×10^8	$2.4 (2.1) \times 10^9$	$5.0 (4.4) \times 10^5$

Abbreviations: Pat = Patient, P+ = proton; SD = standard deviation

Results from ΔX_{50} , the local summed squared differences, overall summed squared differences and overall summed chi-squared differences

As mentioned in the Results section, the results when using ΔX_{50} , the local summed squared differences, overall summed squared differences and overall summed chi-squared differences are highly similar to the results obtained when using the $|\Delta X_{50}|$ distributions. Figures A.7.1, A.7.2, A.7.3 and A.7.4 are the equivalents of Fig. 8.4 and 8.5, but when using the local summed squared differences, overall summed squared differences and overall summed chi-squared differences, respectively.

Table A.7.2: Total number of protons in the pencil-beams used for analysis and the percentage of the total number of protons simulated in the complete treatment plan per patient. In addition, the total number of PG photons used for analysis, after applying the angle of incidence selection and averaged over the simulations on the planning and control CT scans.

Pat	5% most intense pencil-beams		10% most intense pencil-beams		20% most intense pencil-beams	
	Protons	PG photons	Protons	PG photons	Protons	PG photons
1	1.3×10^8 (19%)	1269293	2.1×10^8 (32%)	2051829	3.3×10^8 (50%)	3139781
2	1.4×10^8 (18%)	1381508	2.3×10^8 (30%)	2287554	3.8×10^8 (49%)	3552127
3	1.2×10^8 (19%)	1151009	2.0×10^8 (32%)	1947741	3.2×10^8 (51%)	3552127
4	9.6×10^7 (18%)	975779	1.6×10^8 (31%)	1595006	2.6×10^8 (50%)	2459828
5	9.8×10^7 (19%)	939255	1.7×10^8 (32%)	939255	2.7×10^8 (51%)	2352405
6	1.3×10^8 (18%)	1308651	2.3×10^8 (30%)	2161750	3.7×10^8 (49%)	3406027
7	1.4×10^8 (17%)	1293143	2.4×10^8 (30%)	2150725	3.8×10^8 (49%)	3432533
8	1.0×10^8 (16%)	1001286	1.7×10^8 (28%)	1656947	2.9×10^8 (47%)	2692195
9	8.4×10^7 (18%)	877776	1.4×10^8 (31%)	1424679	2.3×10^8 (50%)	2285463
10	8.8×10^7 (17%)	854818	1.5×10^8 (29%)	1373496	2.4×10^8 (47%)	2168823
11	1.3×10^8 (17%)	1369986	2.3×10^8 (30%)	2250886	3.7×10^8 (49%)	3565075

Abbreviations: Pat = patient, PG = prompt gamma

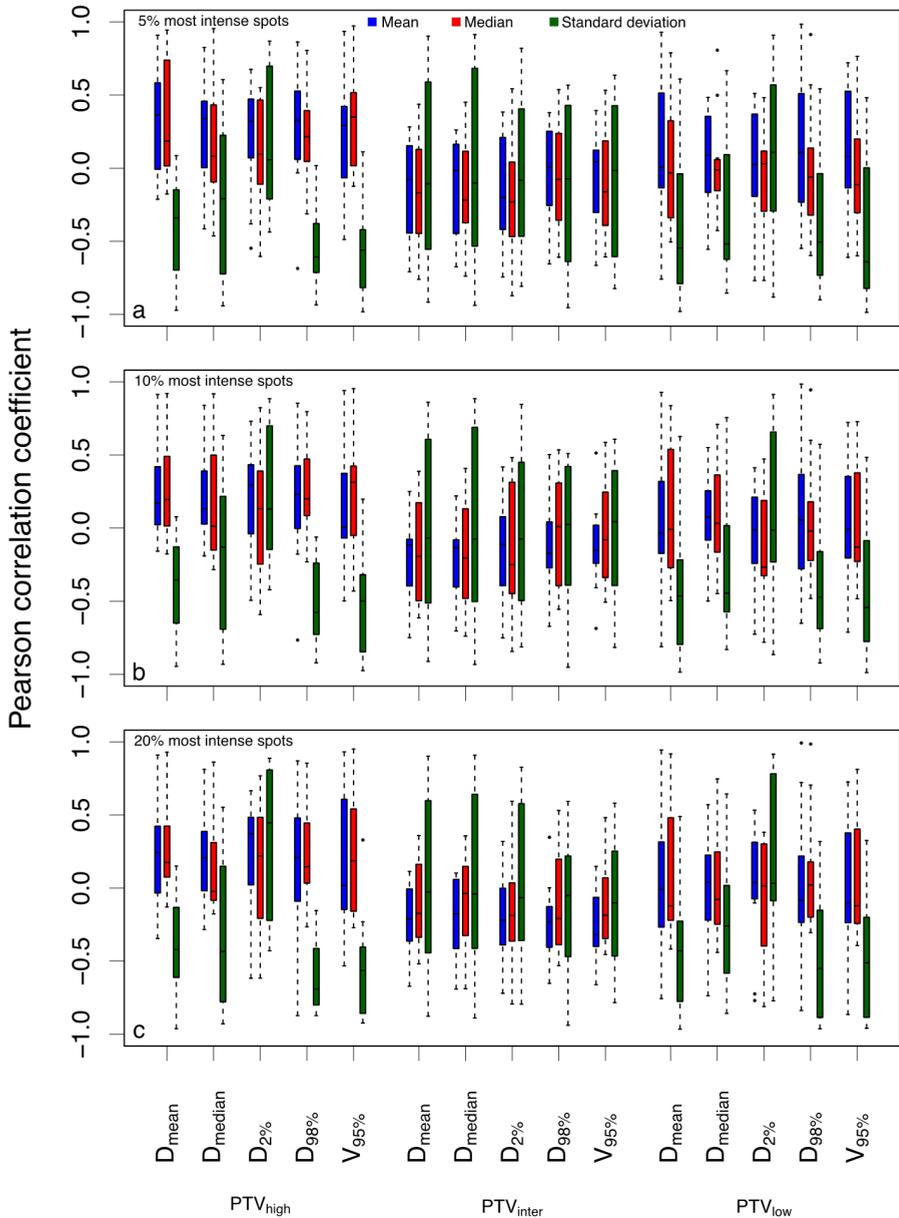


Figure A.7.1: Distributions of Pearson correlation coefficients between the mean, median and standard deviation (color coded) of the ΔX_{50} distributions and the dosimetric parameters when using the (a) 5%, (b) 10% or (c) 20% most intense pencil-beams.

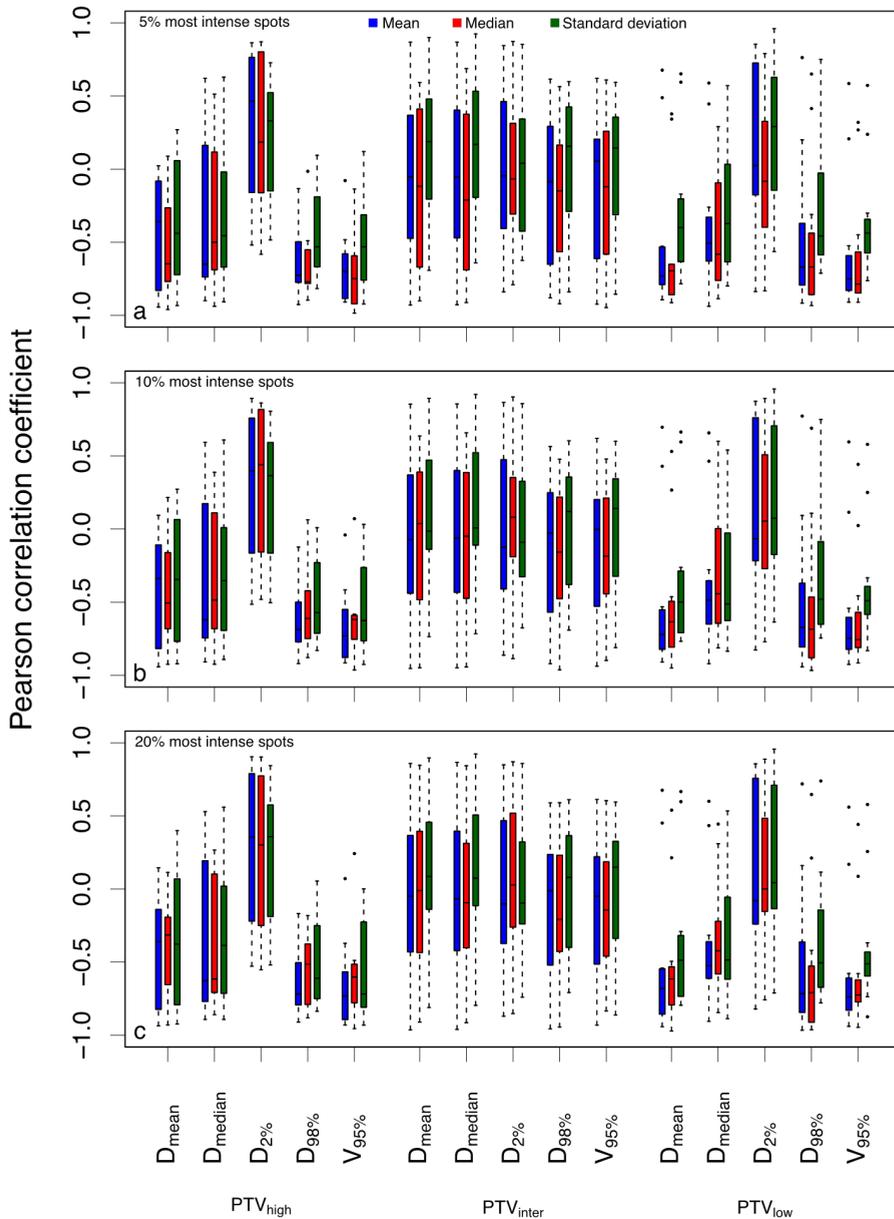


Figure A.7.2: Distributions of Pearson correlation coefficients between the mean, median and standard deviation (color coded) of the distributions of the summed local squared differences and the dosimetric parameters when using the (a) 5%, (b) 10% or (c) 20% most intense pencil beams.

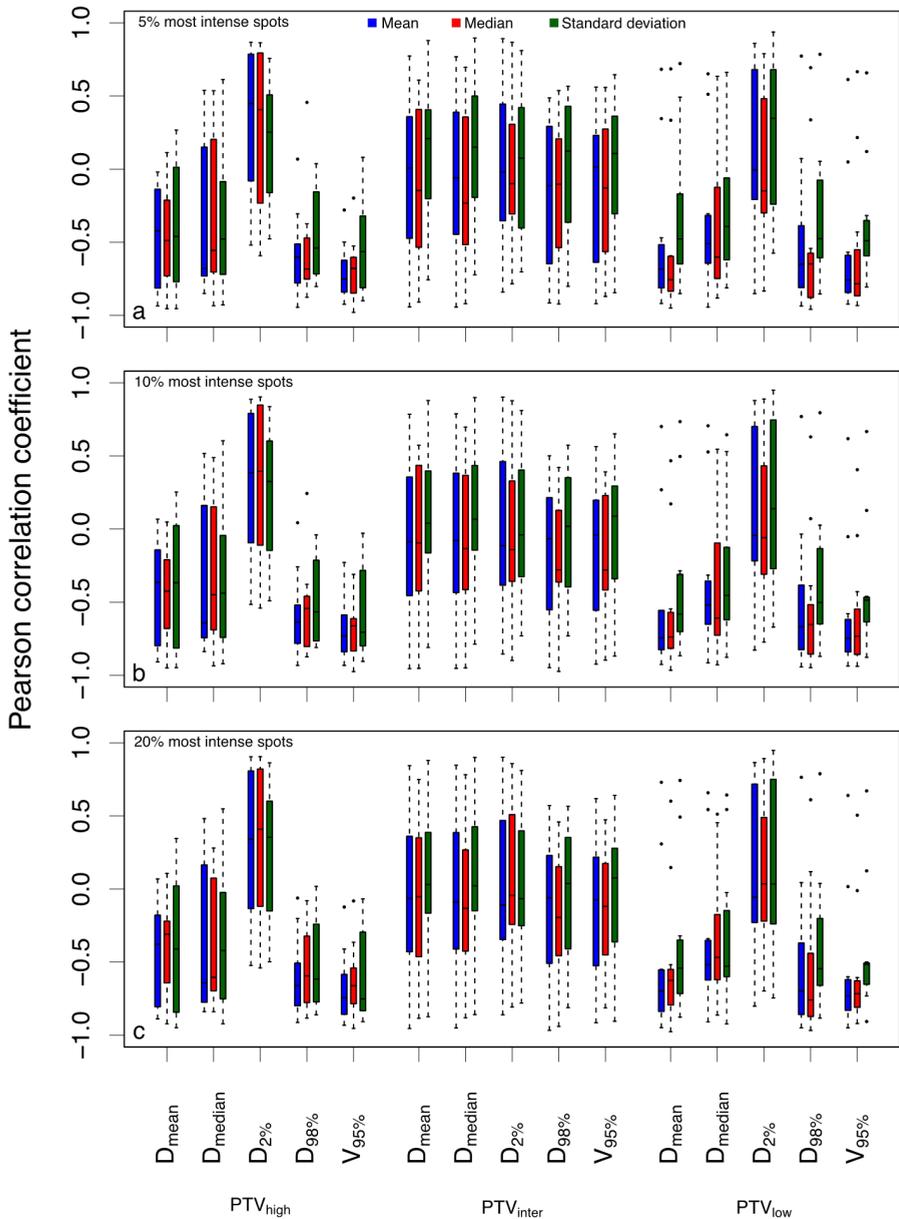


Figure A.7.3: Distributions of Pearson correlation coefficients between the mean, median and standard deviation (color coded) of the distributions of the summed overall squared differences and the dosimetric parameters when using the (a) 5%, (b) 10% or (c) 20% most intense pencil beams.

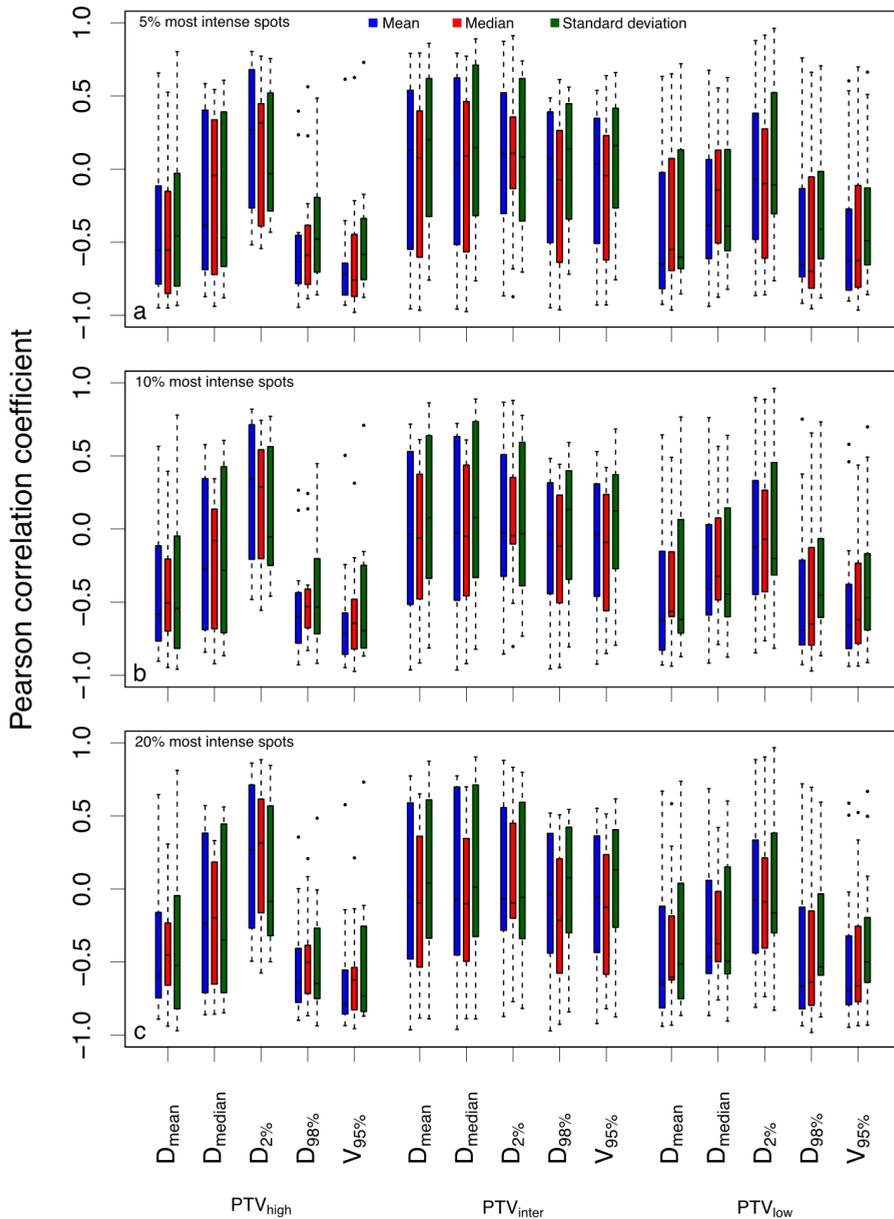


Figure A.7.4: Distributions of Pearson correlation coefficients between the mean, median and standard deviation (color coded) of the distributions of the summed local chi-squared differences and the dosimetric parameters when using the (a) 5%, (b) 10% or (c) 20% most intense pencil-beams.

References

- [1] A.J. Lomax. “Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties”. In: *Physics in Medicine and Biology* 53.4 (2008), pp. 1027–1042. DOI: [10.1088/0031-9155/53/4/014](https://doi.org/10.1088/0031-9155/53/4/014).
- [2] A.J. Lomax. “Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions”. In: *Physics in Medicine and Biology* 53.4 (2008), pp. 1043–1056. DOI: [10.1088/0031-9155/53/4/015](https://doi.org/10.1088/0031-9155/53/4/015).
- [3] S. Thörnqvist, L.P. Muren, L. Bentzen, et al. “Degradation of target coverage due to inter-fraction motion during intensity-modulated proton therapy of prostate and elective targets”. In: *Acta Oncologica* 52.3 (2013), pp. 521–527. DOI: [10.3109/0284186X.2012.752860](https://doi.org/10.3109/0284186X.2012.752860).
- [4] M. Engelsman and H.M. Kooy. “Target volume dose considerations in proton beam treatment planning for lung tumors”. In: *Medical Physics* 32.12 (2005), pp. 3549–3557. DOI: [10.1118/1.2126187](https://doi.org/10.1118/1.2126187).
- [5] W. Liu, S.J. Frank, X. Li, et al. “Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers”. In: *Medical Physics* 40.5 (2013), p. 051711. DOI: [10.1118/1.4801899](https://doi.org/10.1118/1.4801899).
- [6] W. Liu, S.J. Frank, X. Li, et al. “PTV-based IMPT optimization incorporating planning risk volumes vs robust optimization”. In: *Medical Physics* 40.2 (2013), pp. 021709–021709. DOI: [10.1118/1.4774363](https://doi.org/10.1118/1.4774363).
- [7] A. Fredriksson, A. Forsgren, and B. Hårdemark. “Minimax optimization for handling range and setup uncertainties in proton therapy”. In: *Medical Physics* 38.3 (2011), pp. 1672–1684. DOI: [10.1118/1.3556559](https://doi.org/10.1118/1.3556559).
- [8] W. Chen, J. Unkelbach, A. Trofimov, et al. “Including Robustness in Multi-criteria Optimization for Intensity Modulated Proton Therapy”. In: *Physics in Medicine and Biology* 57 (2012), pp. 591–608. DOI: [10.1088/0031-9155/57/3/591](https://doi.org/10.1088/0031-9155/57/3/591).
- [9] A. Fredriksson and R. Bokrantz. “A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning”. In: *Medical Physics* 41.8 Part 1 (2014), p. 081701. DOI: [10.1118/1.4883837](https://doi.org/10.1118/1.4883837).
- [10] S. van de Water, I. van Dam, D.R. Schaart, et al. “The price of robustness; impact of worst-case optimization on organ-at-risk dose and complication probabil-

- ity in intensity-modulated proton therapy for oropharyngeal cancer patients”. In: *Radiotherapy and Oncology* 120.1 (2016), pp. 56–62. DOI: [10.1016/j.radonc.2016.04.038](https://doi.org/10.1016/j.radonc.2016.04.038).
- [11] A.J.A.J. van de Schoot, P. de Boer, K.F. Crama, et al. “Dosimetric advantages of proton therapy compared with photon therapy using an adaptive strategy in cervical cancer”. In: *Acta Oncologica* 55.7 (2016), pp. 892–899. DOI: [10.3109/0284186X.2016.1139179](https://doi.org/10.3109/0284186X.2016.1139179).
- [12] C.H. Min, C.H. Kim, M.Y. Youn, et al. “Prompt gamma measurements for locating the dose falloff region in the proton therapy”. In: *Applied Physics Letters* 89.18 (2006), p. 183517. DOI: [10.1063/1.2378561](https://doi.org/10.1063/1.2378561).
- [13] A.C. Kraan, S. van de Water, D.N. Teguh, et al. “Dose Uncertainties in IMPT for Oropharyngeal Cancer in the Presence of Anatomical, Range, and Setup Errors”. In: *International Journal of Radiation Oncology Biology Physics* 87.5 (2013), pp. 888–896. DOI: [10.1016/j.ijrobp.2013.09.014](https://doi.org/10.1016/j.ijrobp.2013.09.014).
- [14] H. Li, X. Zhang, P. Park, et al. “Robust optimization in intensity-modulated proton therapy to account for anatomy changes in lung cancer patients”. In: *Radiotherapy and Oncology* 114.3 (2015), pp. 367–372. DOI: [10.1016/j.radonc.2015.01.017](https://doi.org/10.1016/j.radonc.2015.01.017).
- [15] M.L. Bondar, M.S. Hoogeman, J.W. Mens, et al. “Individualized Nonadaptive and Online-Adaptive Intensity-Modulated Radiotherapy Treatment Strategies for Cervical Cancer Patients Based on Pretreatment Acquired Variable Bladder Filling Computed Tomography Scans”. In: *International Journal of Radiation Oncology, Biology, Physics* 83.5 (2012), pp. 1617–1623. DOI: [10.1016/j.ijrobp.2011.10.011](https://doi.org/10.1016/j.ijrobp.2011.10.011).
- [16] S.T. Heijkoop, T.R. Langerak, S. Quint, et al. “Clinical Implementation of an Online Adaptive Plan-of-the-Day Protocol for Nonrigid Motion Management in Locally Advanced Cervical Cancer IMRT”. In: *International Journal of Radiation Oncology, Biology, Physics* 90.3 (2014), pp. 673–679. DOI: [10.1016/j.ijrobp.2014.06.046](https://doi.org/10.1016/j.ijrobp.2014.06.046).
- [17] M. Zhang, D.C. Westerly, and T.R. Mackie. “Introducing an on-line adaptive procedure for prostate image guided intensity modulate proton therapy”. In: *Physics in Medicine and Biology* 56.15 (2011), pp. 4947–4965. DOI: [10.1088/0031-9155/56/15/019](https://doi.org/10.1088/0031-9155/56/15/019).
- [18] L. Widesott, A. Pierelli, C. Fiorino, et al. “Helical Tomotherapy vs. Intensity-Modulated Proton Therapy for Whole Pelvis Irradiation in High-Risk Prostate Cancer Patients: Dosimetric, Normal Tissue Complication Probability, and Gen-

- eralized Equivalent Uniform Dose Analysis”. In: *International Journal of Radiation Oncology, Biology, Physics* 80.5 (2011), pp. 1589–1600. DOI: [10.1016/j.ijrobp.2010.10.005](https://doi.org/10.1016/j.ijrobp.2010.10.005).
- [19] S. Breedveld, P.R.M. Storchi, P.W.J. Voet, et al. “iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans”. In: *Medical Physics* 39.2 (2012), pp. 951–963. DOI: [10.1118/1.3676689](https://doi.org/10.1118/1.3676689).
- [20] S. van de Water, A.C. Kraan, S. Breedveld, et al. “Improved efficiency of multicriteria IMPT treatment planning using iterative resampling of randomly placed pencil beams”. In: *Physics in Medicine and Biology* 58.19 (2013), pp. 6969–6983. DOI: [10.1088/0031-9155/58/19/6969](https://doi.org/10.1088/0031-9155/58/19/6969).
- [21] H.M. Kooy, B.M. Clasié, H.M. Lu, et al. “A Case Study in Proton Pencil-Beam Scanning Delivery”. In: *International Journal of Radiation Oncology, Biology, Physics* 76.2 (2010), pp. 624–630. DOI: [10.1016/j.ijrobp.2009.06.065](https://doi.org/10.1016/j.ijrobp.2009.06.065).
- [22] S. Breedveld, P.R.M. Storchi, and B.J.M. Heijmen. “The equivalence of multicriteria methods for radiotherapy plan optimization”. In: *Physics in Medicine and Biology* 54.23 (2009), pp. 7199–7209. DOI: [10.1088/0031-9155/54/23/011](https://doi.org/10.1088/0031-9155/54/23/011).
- [23] S. van de Water, H.M. Kooy, B.J.M. Heijmen, et al. “Shortening Delivery Times of Intensity Modulated Proton Therapy by Reducing Proton Energy Layers During Treatment Plan Optimization”. In: *International Journal of Radiation Oncology, Biology, Physics* 92.2 (2015), pp. 460–468. DOI: [10.1016/j.ijrobp.2015.01.031](https://doi.org/10.1016/j.ijrobp.2015.01.031).
- [24] P.W.J. Voet, M.L.P. Dirx, S. Breedveld, et al. “Toward Fully Automated Multicriterial Plan Generation: A Prospective Clinical Study”. In: *International Journal of Radiation Oncology, Biology, Physics* 85.3 (2013), pp. 866–872. DOI: [10.1016/j.ijrobp.2012.04.015](https://doi.org/10.1016/j.ijrobp.2012.04.015).
- [25] C. Voglis and I. Lagaris. *BOXCQP: an algorithm for bound constrained convex quadratic problems*. Conference proceedings. 1st IC-SCCE Conf. Athens, 2004.
- [26] S. Breedveld, P.R.M. Storchi, M. Keijzer, et al. “Fast, multiple optimizations of quadratic dose objective functions in IMRT”. In: *Physics in Medicine and Biology* 51.14 (2006), pp. 3569–3579. DOI: [10.1088/0031-9155/51/14/019](https://doi.org/10.1088/0031-9155/51/14/019).
- [27] A.G. Andersen, O. Casares-Magaz, L.P. Muren, et al. “A method for evaluation of proton plan robustness towards inter-fractional motion applied to pelvic lymph node irradiation”. In: *Acta Oncologica* 54.9 (2015), pp. 1643–1650. DOI: [10.3109/0284186X.2015.1067720](https://doi.org/10.3109/0284186X.2015.1067720).

- [28] C. Bert, N. Saito, A. Schmidt, et al. “Target motion tracking with a scanned particle beam”. In: *Medical Physics* 34.12 (2007), pp. 4768–4771. DOI: 10.1118/1.2815934.
- [29] B.B. Ronson, L.T. Yonemoto, C.J. Rossi, et al. “Patient tolerance of rectal balloons in conformal radiation treatment of prostate cancer”. In: *International Journal of Radiation Oncology, Biology, Physics* 64.5 (2006), pp. 1367–1370. DOI: 10.1016/j.ijrobp.2005.11.001.
- [30] T.Z. Jagt, S. Breedveld, S. van de Water, et al. “Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations in prostate cancer”. In: *Physics in Medicine and Biology* 62.11 (2017), pp. 4254–4272. DOI: 10.1088/1361-6560/aa5c12.
- [31] Y. Qiao, T.Z. Jagt, M.S. Hoogeman, et al. “Evaluation of an Open Source Registration Package for Automatic Contour Propagation in Online Adaptive Intensity-Modulated Proton Therapy of Prostate Cancer”. In: *Frontiers in Oncology* 9 (2019), p. 1297. DOI: 10.3389/fonc.2019.01297.
- [32] S. Breedveld, B. van den Berg, and B.J.M. Heijmen. “An interior-point implementation developed and tuned for radiation therapy treatment planning”. In: *Computational Optimization and Applications* 68.2 (2017), pp. 209–242. DOI: 10.1007/s10589-017-9919-4.
- [33] R. van Haveren, W. Ogryczak, G.M. Verduijn, et al. “Fast and fuzzy multi-objective radiotherapy treatment plan generation for head and neck cancer patients with the lexicographic reference point method (LRPM)”. In: *Physics in Medicine and Biology* 62.11 (2017), pp. 4318–4332. DOI: 10.1088/1361-6560/62/11/4318.
- [34] R. van Haveren, S. Breedveld, M. Keijzer, et al. “Lexicographic extension of the reference point method applied in radiation therapy treatment planning”. In: *European Journal of Operational Research* 263 (2017). DOI: 10.1016/j.ejor.2017.04.062.
- [35] W. Ogryczak. “Preemptive Reference Point Method”. In: (1997). DOI: 10.1007/978-3-642-60667-0_16.
- [36] W. Ogryczak and B. Kozowski. “Reference point method with importance weighted ordered partial achievements”. In: *TOP* 19.2 (2011), pp. 380–401. DOI: 10.1007/s11750-009-0121-4.
- [37] G. Landry, C. Kurz, F. Kamp, et al. “SP-0032: Enabling proton dose calculations on CBCT images”. In: *Radiotherapy and Oncology* 127 (2018), S11–S12. DOI: 10.1016/S0167-8140(18)30342-6.

- [38] C. Veiga, J. Alshaikhi, R. Amos, et al. “Cone-Beam Computed Tomography and Deformable Registration-Based Dose of the Day Calculations for Adaptive Proton Therapy”. In: *International Journal of Particle Therapy* 2.2 (2015), pp. 404–414. DOI: [10.14338/IJPT-14-00024](https://doi.org/10.14338/IJPT-14-00024). 1.
- [39] E. Lens, T.Z. Jagt, M.S. Hoogeman, et al. “OC-0082: Using prompt gamma emission profiles to monitor day-to-day dosimetric changes in proton therapy”. In: *Radiotherapy and Oncology* 127 (2018), S40–S41. DOI: [10.1016/S0167-8140\(18\)30392-X](https://doi.org/10.1016/S0167-8140(18)30392-X).
- [40] J. da Silva, R. Ansorge, and R. Jena. “Sub-second pencil beam dose calculation on GPU for adaptive proton therapy”. In: *Physics in Medicine and Biology* 60.12 (2015), pp. 4777–4795. DOI: [10.1088/0031-9155/60/12/4777](https://doi.org/10.1088/0031-9155/60/12/4777).
- [41] J. da Silva, R. Ansorge, and R. Jena. “Fast Pencil Beam Dose Calculation for Proton Therapy Using a Double-Gaussian Beam Model”. In: *Frontiers in Oncology* 5 (2015), pp. 281–281. DOI: [10.3389/fonc.2015.00281](https://doi.org/10.3389/fonc.2015.00281).
- [42] D. Giantsoudi, J. Schuemann, X. Jia, et al. “Validation of a GPU-based Monte Carlo code (gPMC) for proton radiation therapy: clinical cases study”. In: *Physics in Medicine and Biology* 60.6 (2015), pp. 2257–2269. DOI: [10.1088/0031-9155/60/6/2257](https://doi.org/10.1088/0031-9155/60/6/2257).
- [43] H. Wan Chan Tseung, J. Ma, and C. Beltran. “A fast GPU-based Monte Carlo simulation of proton transport with detailed modeling of nonelastic interactions”. In: *Medical Physics* 42.6Part1 (2015), pp. 2967–2978. DOI: [10.1118/1.4921046](https://doi.org/10.1118/1.4921046).
- [44] T.Z. Jagt, S. Breedveld, R. van Haveren, et al. “An automated planning strategy for near real-time adaptive proton therapy in prostate cancer”. In: *Physics in Medicine and Biology* 63.13 (2018), p. 135017. DOI: [10.1088/1361-6560/aacaa7](https://doi.org/10.1088/1361-6560/aacaa7).
- [45] S.T. Heijkoop, T.R. Langerak, S. Quint, et al. “Quantification of intra-fraction changes during radiotherapy of cervical cancer assessed with pre- and post-fraction Cone Beam CT scans”. In: *Radiotherapy and Oncology* 117.3 (2015), pp. 536–541. DOI: [10.1016/j.radonc.2015.08.034](https://doi.org/10.1016/j.radonc.2015.08.034).
- [46] M.A.E. van de Sande, C.L. Creutzberg, S. van de Water, et al. “Which cervical and endometrial cancer patients will benefit most from intensity-modulated proton therapy?” In: *Radiotherapy and Oncology* 120.3 (2016), pp. 397–403. DOI: [10.1016/j.radonc.2016.06.016](https://doi.org/10.1016/j.radonc.2016.06.016).
- [47] R. van Haveren, B.J.M. Heijmen, and S. Breedveld. “Automatically configuring the reference point method for automated multi-objective treatment plan-

- ning". In: *Physics in Medicine and Biology* 64.3 (2019), p. 035002. DOI: 10.1088/1361-6560/aaf9fe.
- [48] C. Kurz, R. Nijhuis, M. Reiner, et al. "Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images". In: *Radiation Oncology (London, England)* 11 (2016), pp. 64–64. DOI: 10.1186/s13014-016-0641-7.
- [49] K. Bernatowicz, X. Geets, A. Barragan, et al. "Feasibility of online IMPT adaptation using fast, automatic and robust dose restoration". In: *Physics in Medicine and Biology* 63.8 (2018), p. 085018. DOI: 10.1088/1361-6560/aaba8c.
- [50] P. Botas, J. Kim, B. Winey, et al. "Online adaption approaches for intensity modulated proton therapy for head and neck patients based on cone beam CTs and Monte Carlo simulations". In: *Physics in Medicine and Biology* 64.1 (2018), p. 015004. DOI: 10.1088/1361-6560/aaf30b.
- [51] E.E. Ahunbay, C. Peng, S. Holmes, et al. "Online Adaptive Replanning Method for Prostate Radiotherapy". In: *International Journal of Radiation Oncology, Biology, Physics* 77.5 (2010), pp. 1561–1572. DOI: 10.1016/j.ijrobp.2009.10.013.
- [52] D. Winkel, G.H. Bol, P.S. Kroon, et al. "Adaptive radiotherapy: The Elekta Unity MR-linac concept". In: *Clinical and Translational Radiation Oncology* 18 (2019), pp. 54–59. DOI: 10.1016/j.ctro.2019.04.001.
- [53] O. Bohoudi, A.M.E. Bruynzeel, S. Senan, et al. "Fast and robust online adaptive planning in stereotactic MR-guided adaptive radiation therapy (SMART) for pancreatic cancer". In: *Radiotherapy and Oncology* 125.3 (2017), pp. 439–444. DOI: 10.1016/j.radonc.2017.07.028.
- [54] L. Henke, R. Kashani, C. Robinson, et al. "Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen". In: *Radiotherapy and Oncology* 126 (2017). DOI: 10.1016/j.radonc.2017.11.032.
- [55] M. Tyran, N. Jiang, M. Cao, et al. "Retrospective evaluation of decision-making for pancreatic stereotactic MR-guided adaptive radiotherapy". In: *Radiotherapy and Oncology* 129.2 (2018), pp. 319–325. DOI: 10.1016/j.radonc.2018.08.009.
- [56] T. Finazzi, M.A. Palacios, E.O.B. Spoelstra, et al. "Role of On-Table Plan Adaptation in MR-Guided Ablative Radiation Therapy for Central Lung Tumors". In: *International Journal of Radiation Oncology, Biology, Physics* 104.4 (2019), pp. 933–941. DOI: 10.1016/j.ijrobp.2019.03.035.

- [57] T.Z. Jagt, S. Breedveld, R. van Haveren, et al. “Plan-library supported automated replanning for online-adaptive intensity-modulated proton therapy of cervical cancer”. In: *Acta Oncologica* 58.10 (2019), pp. 1440–1445. DOI: [10.1080/0284186X.2019.1627414](https://doi.org/10.1080/0284186X.2019.1627414).
- [58] M. Matter, L. Nenoff, G. Meier, et al. “Intensity modulated proton therapy plan generation in under ten seconds”. In: *Acta Oncologica* 58 (2019), pp. 1–5. DOI: [10.1080/0284186X.2019.1630753](https://doi.org/10.1080/0284186X.2019.1630753).
- [59] V. Landoni, C. Fiorino, C. Cozzarini, et al. “Predicting toxicity in radiotherapy for prostate cancer”. In: *Physica Medica* 32.3 (2016), pp. 521–532. DOI: [10.1016/j.ejmp.2016.03.003](https://doi.org/10.1016/j.ejmp.2016.03.003).
- [60] A. Widmark, A. Gunnlaugsson, L. Beckman, et al. “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 (2019), pp. 385–395. DOI: [10.1016/S0140-6736\(19\)31131-6](https://doi.org/10.1016/S0140-6736(19)31131-6).
- [61] M.S. Elmahdy, T.Z. Jagt, R.T. Zinkstok, et al. “Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer”. In: *Medical Physics* 46.8 (2019), pp. 3329–3343. DOI: [10.1002/mp.13620](https://doi.org/10.1002/mp.13620).
- [62] E. Lens, T.Z. Jagt, M.S. Hoogeman, et al. “Correlations between the shifts in prompt gamma emission profiles and the changes in daily target coverage during simulated pencil beam scanning proton therapy”. In: *Physics in Medicine and Biology* 64.8 (2019), p. 085009. DOI: [10.1088/1361-6560/ab145e](https://doi.org/10.1088/1361-6560/ab145e).
- [63] C. Winterhalter, G. Meier, D. Oxley, et al. “Log file based Monte Carlo calculations for proton pencil beam scanning therapy”. In: *Physics in Medicine and Biology* 64.3 (2019), p. 035014. DOI: [10.1088/1361-6560/aaaf82d](https://doi.org/10.1088/1361-6560/aaaf82d).
- [64] J.E. Johnson, C. Beltran, H. Wan Chan Tseung, et al. “Highly efficient and sensitive patient-specific quality assurance for spot-scanned proton therapy”. In: *PLOS ONE* 14.2 (2019), e0212412. DOI: [10.1371/journal.pone.0212412](https://doi.org/10.1371/journal.pone.0212412).
- [65] National Cancer Society. *Cancer Stat Facts: Prostate Cancer*. URL: <https://seer.cancer.gov/statfacts/html/prost.html>.
- [66] H.M. Kooy and C. Grassberger. “Intensity modulated proton therapy”. In: *The British Journal of Radiology* 88.1051 (2015), p. 20150195. DOI: [10.1259/bjr.20150195](https://doi.org/10.1259/bjr.20150195).
- [67] F. Van den Heuvel, J. Fugazzi, E. Seppi, et al. “Clinical application of a repositioning scheme, using gold markers and electronic portal imaging”. In: *Ra-*

- diotherapy and Oncology* 79.1 (2006), pp. 94–100. DOI: 10.1016/j.radonc.2006.03.002.
- [68] M.S. Hoogeman, M. van Herk, J. de Bois, et al. “Strategies to reduce the systematic error due to tumor and rectum motion in radiotherapy of prostate cancer”. In: *Radiotherapy and Oncology* 74.2 (2005), pp. 177–185. DOI: 10.1016/j.radonc.2004.12.010.
- [69] E.K. Hansen, M.K. Bucci, J.M. Quivey, et al. “Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer”. In: *International Journal of Radiation Oncology, Biology, Physics* 64.2 (2006), pp. 355–362. DOI: 10.1016/j.ijrobp.2005.07.957.
- [70] D. Boehmer, P. Maingon, P. Poortmans, et al. “Guidelines for primary radiotherapy of patients with prostate cancer”. In: *Radiotherapy and Oncology* 79.3 (2006), pp. 259–269. DOI: 10.1016/j.radonc.2006.05.012.
- [71] C. Salembier, G. Villeirs, B. De Bari, et al. “ESTRO ACROP consensus guideline on CT- and MRI-based target volume delineation for primary radiation therapy of localized prostate cancer”. In: *Radiotherapy and Oncology* 127.1 (2018), pp. 49–61. DOI: 10.1016/j.radonc.2018.01.014.
- [72] N.K.G. Jensen, D. Mulder, M. Lock, et al. “Dynamic contrast enhanced CT aiding gross tumor volume delineation of liver tumors: An interobserver variability study”. In: *Radiotherapy and Oncology* 111.1 (2014), pp. 153–157. DOI: 10.1016/j.radonc.2014.01.026.
- [73] A.C. Riegel, J.G. Antone, H. Zhang, et al. “Deformable image registration and interobserver variation in contour propagation for radiation therapy planning”. In: *Journal of Applied Clinical Medical Physics* 17.3 (2016), pp. 347–357. DOI: 10.1120/jacmp.v17i3.6110.
- [74] P. Kupelian, T. Willoughby, A. Mahadevan, et al. “Multi-institutional clinical experience with the Calypso System in localization and continuous, real-time monitoring of the prostate gland during external radiotherapy”. In: *International Journal of Radiation Oncology, Biology, Physics* 67.4 (2007), pp. 1088–1098. DOI: 10.1016/j.ijrobp.2006.10.026.
- [75] Y. Qiao. “Fast Optimization Methods For Image Registration In Adaptive Radiation Therapy”. In: (2017). URL: <https://openaccess.leidenuniv.nl/bitstream/handle/1887/59448/05.pdf?sequence=8>.
- [76] M. Thor, J.B.B. Petersen, L. Bentzen, et al. “Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of pro-

- state cancer”. In: *Acta Oncologica* 50.6 (2011), pp. 918–925. DOI: [10.3109/0284186X.2011.577806](https://doi.org/10.3109/0284186X.2011.577806).
- [77] A.J. Woerner, M. Choi, M.M. Harkenrider, et al. “Evaluation of Deformable Image Registration–Based Contour Propagation From Planning CT to Cone-Beam CT”. In: *Technology in Cancer Research and Treatment* 16.6 (2017), pp. 801–810. DOI: [10.1177/1533034617697242](https://doi.org/10.1177/1533034617697242).
- [78] Sara Thörnqvist, Jørgen B. B. Petersen, Morten Høyer, et al. “Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration”. In: *Acta Oncologica* 49.7 (2010), pp. 1023–1032. DOI: [10.3109/0284186X.2010.503662](https://doi.org/10.3109/0284186X.2010.503662).
- [79] M.S. Elmahdy, T.Z. Jagt, S. Yousefi, et al. “Evaluation of Multi-metric Registration for Online Adaptive Proton Therapy of Prostate Cancer”. In: *Biomedical Image Registration*. 2018, pp. 94–104. DOI: [10.1007/978-3-319-92258-4_9](https://doi.org/10.1007/978-3-319-92258-4_9).
- [80] S. Mangar, J. Coffey, H. McNair, et al. “Prostate Radiotherapy: Evaluating the Effect of Bladder and Rectal Changes on Prostate Movement-A CT Study”. In: *Trends in Medical Research* 1 (2006), pp. 55–65. DOI: [10.3923/tmr.2006.55.65](https://doi.org/10.3923/tmr.2006.55.65).
- [81] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, et al. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. 2016, pp. 424–432. DOI: [10.1007/978-3-319-46723-8_49](https://doi.org/10.1007/978-3-319-46723-8_49).
- [82] F. Milletari, N. Navab, and S.A. Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. 2016, pp. 565–571. DOI: [10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79).
- [83] D. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. In: *International Conference on Learning Representations* (2014).
- [84] S. Gao, L. Zhang, H. Wang, et al. “A deformable image registration method to handle distended rectums in prostate cancer radiotherapy”. In: *Medical Physics* 33.9 (2006), pp. 3304–3312. DOI: [10.1118/1.2222077](https://doi.org/10.1118/1.2222077).
- [85] M. Foskey, B. Davis, L. Goyal, et al. “Large deformation three-dimensional image registration in image-guided radiation therapy”. In: *Physics in Medicine and Biology* 50.24 (2005), pp. 5869–5892. DOI: [10.1088/0031-9155/50/24/008](https://doi.org/10.1088/0031-9155/50/24/008).
- [86] G. Litjens, T. Kooi, B.E. Bejnordi, et al. “A survey on deep learning in medical image analysis”. In: *Medical Image Analysis* 42 (2017), pp. 60–88. DOI: [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005).

- [87] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al. “Generative Adversarial Networks”. In: *Advances in Neural Information Processing Systems* 3 (2014).
- [88] Y. Xue, T. Xu, H. Zhang, et al. “SegAN: Adversarial Network with Multi-scale L_1 Loss for Medical Image Segmentation”. In: *Neuroinformatics* 16 (2017). DOI: 10.1007/s12021-018-9377-x.
- [89] D. Nie, R. Trullo, J. Lian, et al. *Medical Image Synthesis with Context-Aware Generative Adversarial Networks*. Vol. 10435. 2017, pp. 417–425. DOI: 10.1007/978-3-319-66179-7_48.
- [90] Y. Hu, E. Gibson, N. Ghavami, et al. *Adversarial deformation regularization for training image registration neural networks*. 2018. DOI: 10.1007/978-3-030-00928-1_87.
- [91] Q. Yang, P. Yan, Y. Zhang, et al. “Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss”. In: *IEEE Transactions on Medical Imaging* PP (2017). DOI: 10.1109/TMI.2018.2827462.
- [92] J. Yu, Z. Lin, J. Yang, et al. *Generative Image Inpainting with Contextual Attention*. Journal Article. 2018. DOI: 10.1109/CVPR.2018.00577.
- [93] S. Iizuka, E. Simo-Serra, and H. Ishikawa. “Globally and locally consistent image completion”. In: *ACM Transactions on Graphics* 36 (2017), pp. 1–14. DOI: 10.1145/3072959.3073659.
- [94] B. Rodriguez-Vila, F. Garcia-Vicente, and E.J. Gomez. “Methodology for registration of distended rectums in pelvic CT studies”. In: *Medical Physics* 39.10 (2012), pp. 6351–6359. DOI: 10.1118/1.4754798.
- [95] S. Klein, M. Staring, K. Murphy, et al. “elastix: A Toolbox for Intensity-Based Medical Image Registration”. In: *IEEE Transactions on Medical Imaging* 29.1 (2010), pp. 196–205. DOI: 10.1109/TMI.2009.2035616.
- [96] M. Staring, M.E. Bakker, J. Stolk, et al. “Towards local progression estimation of pulmonary emphysema using CT”. In: *Medical Physics* 41.2 (2014), p. 021905. DOI: 10.1118/1.4851535.
- [97] W. Huizinga, S. Klein, and D. Poot. *Fast Multidimensional B-spline Interpolation Using Template Metaprogramming*. 2014. DOI: 10.1007/978-3-319-08554-8_2.
- [98] Y. Qiao, B. van Lew, B.P.F. Lelieveldt, et al. “Fast Automatic Step Size Estimation for Gradient Descent Optimization of Image Registration”. In: *IEEE Transactions on Medical Imaging* 35 (2015). DOI: 10.1109/TMI.2015.2476354.

- [99] G.J. van der Wielen, T.F. Mutanga, L. Incrocci, et al. “Deformation of Prostate and Seminal Vesicles Relative to Intraprostatic Fiducial Markers”. In: *International Journal of Radiation Oncology, Biology, Physics* 72.5 (2008), 1604–1611.e3. DOI: 10.1016/j.ijrobp.2008.07.023.
- [100] L.P. Muren, E. Wasbø, S.I. Helle, et al. “Intensity-Modulated Radiotherapy of Pelvic Lymph Nodes in Locally Advanced Prostate Cancer: Planning Procedures and Early Experiences”. In: *International Journal of Radiation Oncology, Biology, Physics* 71.4 (2008), pp. 1034–1041. DOI: 10.1016/j.ijrobp.2007.11.060.
- [101] M. Abadi, A. Agarwal, P. Barham, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. 2015.
- [102] K.H. Cha, L. Hadjiiski, R.K. Samala, et al. “Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets”. In: *Medical Physics* 43.4 (2016), pp. 1882–1896. DOI: 10.1118/1.4944498.
- [103] X. Zhou, T. Ito, R. Takayama, et al. *Three-Dimensional CT Image Segmentation by Combining 2D Fully Convolutional Network with 3D Majority Voting*. 2016, pp. 111–120. DOI: 10.1007/978-3-319-46976-8_12.
- [104] K. Men, J. Dai, and Y. Li. “Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks”. In: *Medical Physics* 44.12 (2017), pp. 6377–6389. DOI: 10.1002/mp.12602.
- [105] H. Paganetti. “Range uncertainties in proton therapy and the role of Monte Carlo simulations”. In: *Physics in Medicine and Biology* 57.11 (2012), R99–R117. DOI: 10.1088/0031-9155/57/11/r99.
- [106] K. Parodi, F. Ponisch, and W. Enghardt. “Experimental study on the feasibility of in-beam PET for accurate monitoring of proton therapy”. In: *IEEE Transactions on Nuclear Science* 52.3 (2005), pp. 778–786. DOI: 10.1109/TNS.2005.850950.
- [107] P. Cambraia Lopes, E. Clementel, P. Crespo, et al. “Time-resolved imaging of prompt-gamma rays for proton range verification using a knife-edge slit camera based on digital photon counters”. In: *Physics in Medicine and Biology* 60.15 (2015), pp. 6063–6085. DOI: 10.1088/0031-9155/60/15/6063.
- [108] P. Cambraia Lopes, J. Bauer, A. Salomon, et al. “First in situ TOF-PET study using digital photon counters for proton range verification”. In: *Physics in Medicine and Biology* 61.16 (2016), pp. 6203–6230. DOI: 10.1088/0031-9155/61/16/6203.

- [109] K. Parodi and J.C. Polf. “In vivo range verification in particle therapy”. In: *Medical Physics* 45.11 (2018), e1036–e1050. DOI: [10.1002/mp.12960](https://doi.org/10.1002/mp.12960).
- [110] J. Krimmer, D. Dauvergne, J.M. Létang, et al. “Prompt-gamma monitoring in hadrontherapy: A review”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), pp. 58–73. DOI: [10.1016/j.nima.2017.07.063](https://doi.org/10.1016/j.nima.2017.07.063).
- [111] M. Moteabbed, S. España, and H. Paganetti. “Monte Carlo patient study on the comparison of prompt gamma and PET imaging for range verification in proton therapy”. In: *Physics in Medicine and Biology* 56.4 (2011), pp. 1063–1082. DOI: [10.1088/0031-9155/56/4/012](https://doi.org/10.1088/0031-9155/56/4/012).
- [112] C. Richter, G. Pausch, S. Barczyk, et al. “First clinical application of a prompt gamma based in vivo proton range verification system”. In: *Radiotherapy and Oncology* 118.2 (2016), pp. 232–237. DOI: [10.1016/j.radonc.2016.01.004](https://doi.org/10.1016/j.radonc.2016.01.004).
- [113] L. Nenoff, M. Priegnitz, G. Janssens, et al. “Sensitivity of a prompt-gamma slit-camera to detect range shifts for proton treatment verification”. In: *Radiotherapy and Oncology* 125.3 (2017), pp. 534–540. DOI: [10.1016/j.radonc.2017.10.013](https://doi.org/10.1016/j.radonc.2017.10.013).
- [114] J.M. Verburg, H.A. Shih, and J. Seco. “Simulation of prompt gamma-ray emission during proton radiotherapy”. In: *Physics in Medicine and Biology* 57.17 (2012), pp. 5459–5472. DOI: [10.1088/0031-9155/57/17/5459](https://doi.org/10.1088/0031-9155/57/17/5459).
- [115] J.M. Verburg, K. Riley, T. Bortfeld, et al. “Energy- and time-resolved detection of prompt gamma-rays for proton range verification”. In: *Physics in Medicine and Biology* 58.20 (2013), pp. L37–L49. DOI: [10.1088/0031-9155/58/20/L37](https://doi.org/10.1088/0031-9155/58/20/L37).
- [116] J.M. Verburg and J. Seco. “Proton range verification through prompt gamma-ray spectroscopy”. In: *Physics in Medicine and Biology* 59.23 (2014), pp. 7089–7106. DOI: [10.1088/0031-9155/59/23/7089](https://doi.org/10.1088/0031-9155/59/23/7089).
- [117] C. Golnik, F. Hueso-González, A. Müller, et al. “Range assessment in particle therapy based on prompt-ray timing measurements”. In: *Physics in Medicine and Biology* 59.18 (2014), pp. 5399–5422. DOI: [10.1088/0031-9155/59/18/5399](https://doi.org/10.1088/0031-9155/59/18/5399).
- [118] F. Claassen-Janssen, G. Landry, P. Cambraia Lopes, et al. “Factors influencing the accuracy of beam range estimation in proton therapy using prompt gamma emission”. In: *Physics in Medicine and Biology* 59 (2014), pp. 4427–4441. DOI: [10.1088/0031-9155/59/15/4427](https://doi.org/10.1088/0031-9155/59/15/4427).
- [119] S. Schmid, G. Landry, C. Thieke, et al. “Monte Carlo study on the sensitivity of prompt gamma imaging to proton range variations due to interfractional

- changes in prostate cancer patients”. In: *Physics in Medicine and Biology* 60.24 (2015), pp. 9329–9347. DOI: 10.1088/0031-9155/60/24/9329.
- [120] G. Janssens, J. Smeets, F. Vander Stappen, et al. “Sensitivity study of prompt gamma imaging of scanned beam proton therapy in heterogeneous anatomies”. In: *Radiotherapy and Oncology* 118.3 (2016), pp. 562–567. DOI: 10.1016/j.radonc.2015.11.002.
- [121] Y. Xie, E.H. Bentefour, G. Janssens, et al. “Prompt Gamma Imaging for In vivo Range Verification of Pencil Beam Scanning Proton Therapy”. In: *International Journal of Radiation Oncology, Biology, Physics* 99.1 (2017), pp. 210–218. DOI: 10.1016/j.ijrobp.2017.04.027.
- [122] J. Perl, J. Shin, J. Schuemann, et al. “TOPAS: An innovative proton Monte Carlo platform for research and clinical applications”. In: *Medical Physics* 39 (2012), pp. 6818–6837. DOI: 10.1118/1.4758060.
- [123] C.Z. Jarlskog and H. Paganetti. “Physics Settings for Using the Geant4 Toolkit in Proton Therapy”. In: *IEEE Transactions on Nuclear Science* 55.3 (2008), pp. 1018–1025. DOI: 10.1109/TNS.2008.922816.
- [124] M. Pinto, D. Dauvergne, N. Freud, et al. “Assessment of Geant4 Prompt-Gamma Emission Yields in the Context of Proton Therapy Monitoring”. In: *Frontiers in Oncology* 6 (2016), p. 10. DOI: 10.3389/fonc.2016.00010.
- [125] W. Schneider, T. Bortfeld, and W. Schlegel. “Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions”. In: *Physics in Medicine and Biology* 45.2 (2000), pp. 459–478. DOI: 10.1088/0031-9155/45/2/314.
- [126] A.K. Biegun, E. Seravalli, P. Cambraia Lopes, et al. “Time-of-flight neutron rejection to improve prompt gamma imaging for proton range verification: a simulation study”. In: *Physics in Medicine and Biology* 57.20 (2012), pp. 6429–6444. DOI: 10.1088/0031-9155/57/20/6429.
- [127] K. Parodi and T. Bortfeld. “A filtering approach based on Gaussianpowerlaw convolutions for local PET verification of proton radiotherapy”. In: *Physics in Medicine and Biology* 51.8 (2006), pp. 1991–2009. DOI: 10.1088/0031-9155/51/8/003.
- [128] C. Wang, X. Zhu, J.C. Hong, et al. “Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future”. In: *Technology in cancer research and treatment* 18 (2019). DOI: 10.1177/1533033819873922.

- [129] J. Fan, J. Wang, Z. Chen, et al. “Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique”. In: *Medical Physics* 46.1 (2019), pp. 370–381. DOI: [10.1002/mp.13271](https://doi.org/10.1002/mp.13271).
- [130] R.G. Bijman, S. Breedveld, T. Arts, et al. “Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy”. In: *Acta Oncologica* 56.11 (2017), pp. 1444–1450. DOI: [10.1080/0284186X.2017.1355113](https://doi.org/10.1080/0284186X.2017.1355113).
- [131] R. Smith-Bindman, J. Lipson, R. Marcus, et al. “Radiation Dose Associated With Common Computed Tomography Examinations and the Associated Lifetime Attributable Risk of Cancer”. In: *Archives of Internal Medicine* 169.22 (2009), pp. 2078–2086. DOI: [10.1001/archinternmed.2009.427](https://doi.org/10.1001/archinternmed.2009.427).
- [132] A. Patriarca, C. Fouillade, M. Auger, et al. “Experimental Set-up for FLASH Proton Irradiation of Small Animals Using a Clinical System”. In: *International Journal of Radiation Oncology, Biology, Physics* 102.3 (2018), pp. 619–626. DOI: [10.1016/j.ijrobp.2018.06.403](https://doi.org/10.1016/j.ijrobp.2018.06.403).
- [133] M. Buonanno, V. Grilj, and D.J. Brenner. “Biological effects in normal cells exposed to FLASH dose rate protons”. In: *Radiotherapy and Oncology* 139 (2019), pp. 51–55. DOI: [10.1016/j.radonc.2019.02.009](https://doi.org/10.1016/j.radonc.2019.02.009).
- [134] E. Beyreuther, M. Brand, S. Hans, et al. “Feasibility of proton FLASH effect tested by zebrafish embryo irradiation”. In: *Radiotherapy and Oncology* 139 (2019), pp. 46–50. DOI: [10.1016/j.radonc.2019.06.024](https://doi.org/10.1016/j.radonc.2019.06.024).
- [135] S. van de Water, S. Safai, J.M. Schippers, et al. “Towards FLASH proton therapy: the impact of treatment planning and machine characteristics on achievable dose rates”. In: *Acta Oncologica* 58.10 (2019), pp. 1463–1469. DOI: [10.1080/0284186X.2019.1627416](https://doi.org/10.1080/0284186X.2019.1627416).
- [136] J.A. Langendijk, P. Lambin, D. De Ruysscher, et al. “Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach”. In: *Radiotherapy and Oncology* 107.3 (2013), pp. 267–273. DOI: [10.1016/j.radonc.2013.05.007](https://doi.org/10.1016/j.radonc.2013.05.007).
- [137] A. Jakobi, A. Bandurska-Luque, K. Stützer, et al. “Identification of Patient Benefit From Proton Therapy for Advanced Head and Neck Cancer Patients Based on Individual and Subgroup Normal Tissue Complication Probability Analysis”. In: *International Journal of Radiation Oncology, Biology, Physics* 92.5 (2015), pp. 1165–1174. DOI: [10.1016/j.ijrobp.2015.04.031](https://doi.org/10.1016/j.ijrobp.2015.04.031).
- [138] T. Arts, S. Breedveld, M.A. de Jong, et al. “The impact of treatment accuracy on proton therapy patient selection for oropharyngeal cancer patients”. In: *Radio-*

-
- therapy and Oncology* 125.3 (2017), pp. 520–525. DOI: 10.1016/j.radonc.2017.09.028.
- [139] *EMBRACE II Study - Title page and total protocol v15 151015*. URL: https://www.embracestudy.dk/UserUpload/PublicDocuments/Docs/EMBRACE_II_title_page_and_total_protocol_v15_151015.pdf.
- [140] R. Pötter, K. Tanderup, C. Kirisits, et al. “The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies”. In: *Clinical and Translational Radiation Oncology* 9 (2018), pp. 48–60. DOI: 10.1016/j.ctro.2018.01.001.

Summary

Intensity-modulated proton therapy (IMPT) is a type of radiotherapy in which the patient is irradiated using groups of high-energy protons called pencil-beams or spots. Protons are positively charged particles with have finite range in tissue. Most of their dose is deposited at the end of their range in a so-called Bragg peak, with no dose being deposited after the Bragg peak. Due to these characteristic Bragg peaks a very localized dose distribution can be generated, allowing for adequate target coverage and sparing of the healthy tissues (organs at risk). However, the characteristic Bragg peak also makes this treatment modality sensitive to daily anatomical variations such as density changes along the pencil-beam paths and changes in organ shape and location. These variations can cause the intended dose distribution to be distorted, possibly resulting in local over- and under-dosage.

To account for these daily anatomical changes, we have developed online adaptation strategies to ensure sufficient dose in each treatment fraction, while maintaining the benefit of healthy tissue sparing. With the development of these strategies, we have proven the feasibility of online-adaptive IMPT and obtained answers to seven research questions.

A treatment plan can be restored to its initial state by applying an energy-adaptation followed by a spot-weight re-optimization.

In **Chapter 2**, we developed a two-step dose restoration method that can restore a distorted dose distribution to its intended state by compensating for daily tissue density variations. The restoration strategy considers the daily repeat CT scan together with the contours of the planning CT scan projected onto the repeat CT scan. Taking a prior plan generated on the planning CT scan, it applies (1) a restoration of the spot positions (Bragg peaks) by adapting the energy of each pencil-beam to the new water equivalent path length; and (2) a re-optimization of pencil-beam weights by minimizing the dosimetric difference with the prior dose distribution, using a fast and exact quadratic solver. For 10 prostate cancer patients, with a total of 80 repeat CT scans, we evaluated different approaches of assigning importance factors in the spot-weight re-optimization. Results showed that giving a high weight to the planning target volume (PTV) in the re-optimization resulted in adequate restorations. All scans obtained clinically acceptable target coverage ($V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$) for the projected PTV. Dosimetric parameter values of the rectum and bladder

were mostly within 2 Gy and 2%-point of the intended values. Energy-adaptation took on average 5.4 seconds, and the spot-weight re-optimization took on average 3.8 seconds.

Full plan adaptation can increase both target coverage and healthy tissue sparing compared to dose restoration.

In **Chapter 3** we expanded the dose restoration method to a full plan adaptation method to be able to account for variations in shape and location of the target and organs at risk (OARs). This method considers the daily repeat CT scan combined with the contours of this daily CT scan. Starting with a prior treatment plan, the method uses (1) a restoration of spot positions as introduced in the dose restoration method of Chapter 2; and (2) a spot addition to fully cover the target of the day, followed by a fast reference point method (RPM) optimization of the spot-weights resulting in a Pareto optimal plan for the daily anatomy. The method was developed and evaluated using 88 repeat CT scans of 11 prostate cancer patients. For all scans we applied the simple dose restoration method and the new online adaptation method. Applying the simple dose restoration method resulted in a clinically acceptable coverage of the target of the day ($V_{95\%} \geq 98\%$ and $V_{107\%} \leq 2\%$) for 51.1% of the scans. Applying the automated online adaptation method resulted in a clinically acceptable target coverage for 96.6% of the scans after a single iteration of adding 2500 spots and for 100% of the scans after two iterations. Full plan adaptation resulted in lower doses to the OARs than dose restoration. The largest improvements between dose restoration and full plan adaptation with two iterations were obtained for $V_{45 \text{ Gy(RBE)}}$ for both rectum and bladder, with median differences of 10.3%-point and 10.8%-point, respectively (maximum 22%-point). The two steps of full plan adaptation took on average 7.3 seconds and 1.7 minutes, respectively.

For treatment sites showing large day-to-day variations the full plan adaptation approach yields better treatment plans when combined with a patient-specific plan-library.

The developed full plan adaptation method from Chapter 3 uses a prior treatment plan as a warm-start for adaptation to the daily anatomy. In **Chapter 4** we evaluated whether treatment sites showing large day-to-day variations could benefit from the addition of a patient-specific plan-library to the full plan adaptation method. This was evaluated on the CT data of a patient cohort consisting of six cervical cancer patients with in total 23 repeat CT scans, prescribing dose to the primary tumor and the nodal clinical target volume (CTV). Instead of generating a single prior plan per patient, a patient-specific library of prior plans accounting for altered target geometries was

generated using a pretreatment established target motion model. Each fraction, the best fitting prior plan was selected. The selected prior plan was adapted according to the method developed in Chapter 3. RPM-parameters were automatically tuned for this study. Three-fold cross validation, i.e. tuning three times on a third of the data and using the found parameters on the rest of the data, was used to remove any dependency of the adaptation outcome on the tuning data. Spot addition and spot-weight optimization were done once or twice. Using a 1-plan-library (one prior plan based on all motion in the motion model) was compared to selecting a plan from a 2-plan-library (two prior plans based on part of the motion). Applying full plan adaptation with one iteration of adding 3000 spots resulted in adequate target coverage ($V_{95\%} \geq 95\%$ and $V_{107\%} \leq 2\%$) for 37/46 plans using the 1-plan-library and 41/46 plans for the 2-plan-library. Using two spot-addition iterations, the 2-plan-library approach could obtain acceptable coverage for all plans, while the 1-plan-library approach showed $V_{107\%} > 2\%$ for 3/46 plans. The approaches yielded similar results for the OARs. This shows that using a patient-specific plan-library (2-plan-library) instead of a single prior plan (1-plan-library) is beneficial for treatment sites showing large day-to-day variations.

Online plan adaptation can maintain high target coverage while increasing sparing of the healthy tissues when compared to robust treatment planning.

Having shown in the earlier chapters that dose restoration and full plan adaptation are feasible and can yield adequate treatment plans, the benefit compared to the current standard approach of robust treatment planning had to be quantified. In **Chapter 5** we therefore compared the three approaches for the prostate cancer database of 11 patients (88 repeat CT scans). For this dataset we first obtained the safety margins and robustness settings required for the robust treatment planning to achieve adequate target coverage for all repeat CT scans of at least 90% of the patients. For each repeat CT scan we adapted the corresponding prior plan using the simple dose restoration method and the full plan adaptation method. To account for uncertainties in the conversion from Hounsfield units to proton stopping power the full plan adaptation optimized the pencil-beam weights robustly using three scenarios (nominal, $\pm 3\%$ range). We found that adding margins of 4 and 8 mm around the high- and low-dose target regions and optimizing the targets robustly using a 6 mm setup error and a 3% range error yielded adequate target coverage ($V_{95\%} \geq 98\%$) for 10/11 patients. Both online-adaptive planning strategies yielded $V_{95\%} \geq 98\%$ and better OAR sparing in 11/11 patients. Largest improvements were seen for the rectum $V_{45 \text{ Gy(RBE)}}$, with improvements up to 11%-point and 16%-point when moving from robust planning to respectively restoration and full adaption. Excluding the initialization and final dose

calculations, plan restoration took on average 1.7 minutes and full plan adaptation took on average 6.6 minutes.

Dose restoration is mostly influenced by the choice of the prior plan, while the results of full plan adaptation depend more on the stopping criteria and optimization approach.

The effect of varying the parameter settings of the developed online-adaptive methods was investigated in **Chapter 6**. For each of the 88 repeat CT scans of the prostate cancer database of 11 patients we adapted a prior plan using both the simple dose restoration method and the full plan adaptation method. For all scans we evaluated the plans generated using variations on the default settings for the following parameters: 1) The amount of motion mitigation included in the prior plan, 2) the spot-addition sample size, 3) the spot-addition stopping criterion, and 4) the spot-intensity optimization approach. The first variation (prior plan) was evaluated for both simple dose restoration method and the full plan adaptation method, the last three were only evaluated for the full plan adaptation. For the dose restoration method, we found that changing the amount of motion mitigation accounted for in the prior plans with respect to the default can yield insufficient target coverage or increased OAR doses. For the full plan adaptation large effects were only observed when varying the stopping criterion and the optimization approach. Increasing the number of spot-addition iterations and changing the optimization approach from RPM to the 2-phase-epsilon-constraint (2pec) approach both resulted in similar target coverage but lower OAR doses compared to the default settings. As increasing the stopping criterion rapidly increases running times, and the 2pec approach is too time consuming for the online-adaptive workflow, efforts should be focused on improving the outcomes of the RPM optimization. As the RPM optimization is steered through its RPM-parameters, more elaborate tuning of these RPM-parameters is recommended.

Fully automated contour propagation for daily CT scans can be achieved through a combination of deep-learning and image registration.

For clinical use of the full plan adaptation method to be feasible, daily contours of the repeat CT scans have to be acquired. **Chapter 7** presents an automated method to obtain these. The approach first uses a deep-learning network to automatically delineate the bladder of a daily CT scan. After this, gas pockets in the rectum and intestines are detected and inpainted (filled) using a generative adversarial network (GAN). Finally, using the inpainted image and the obtained bladder contour, the manual contours of the planning CT scan are propagated onto the daily CT scan using deformable image registration (DIR). Three different prostate cancer databases from three

different institutes were used. The first database was used to train and test the 3D convolutional neural network for automatic bladder delineation. The remaining two datasets were used for evaluation of the complete workflow. The obtained contours were evaluated by comparing them to the manual gold-standard contours using the dice similarity coefficient (DSC), the mean surface distance (MSD) and the 95% Hausdorff distance (95%HD). A clinical validation of the contours was performed for 88 repeat CT scans of 11 patients. The contours were considered clinically acceptable if their treatment plans achieved clinically acceptable target coverage on the manual contours. This evaluation showed that for 80% of these scans the automatically obtained contours did not require user alterations, as adequate target coverage was obtained for all targets using these contours.

Measured prompt gamma ray emission profiles can be used to detect changes in PTV coverage due to day-to-day anatomical variations.

In **Chapter 8** we investigated the final step of the proposed online-adaptive workflow, which is to monitor treatment delivery using prompt gamma (PG) ray emission profiles as part of treatment delivery quality assurance (QA). PG rays result from nuclear interactions between the incoming protons and the patients' tissue, and emission profiles can be measured outside the patient. We investigated whether PG ray emission profiles can be used to detect changes in target coverage due to inter-fraction anatomical variations with respect to the planning CT scan. A treatment plan was generated for each planning CT scan of the prostate cancer database (11 patients, 88 repeat CT scans) using Erasmus-iCycle. Treatment plan delivery was then simulated on the planning CT scan as well as all repeat CT scans using Monte Carlo. For each dose distribution dosimetric parameter values ($D_{x\%}$) describing target coverage were determined, as well as the halfway-point of the falloff region of the PG emission profile (X_{50}). Differences between planned values and repeat CT values were then evaluated and the Pearson correlation coefficients between functions of the ΔX_{50} and $\Delta D_{x\%}$ were determined. Strongest correlations were found between changes in the coverage of the high- and low-dose PTV and the median and mean absolute values of ΔX_{50} when including only the 5% most intense pencil-beams in the PG emission profile. These correlations show as a proof of principle that it is feasible to use PG ray emission profiles to detect dosimetric changes. In order to determine the exact relation between the emission profiles and the dosimetric parameters more research is needed.

Samenvatting

Intensiteitsgemoduleerde protonetherapie (IMPT) is een vorm van radiotherapie waarin de patiënt wordt bestraald met groepen protonen met een hoge energie, zogenoemde protonenbundels of spots. Protonen zijn positief geladen deeltjes met een beperkte dracht in het weefsel. De meeste dosis wordt afgegeven aan het eind van de dracht in een zogenoemde Braggpiek, waarna geen dosis meer wordt afgegeven. Met behulp van de karakteristieke Braggpiek kan een gelokaliseerde dosis worden afgegeven, met een hoge dosis in het doelvolumen en sparing van de gezonde weefsels (kritieke organen) eromheen. De Braggpiek maakt de bestralingsvorm echter ook gevoelig voor dagelijkse variaties in de anatomie, zoals variaties in de weefseldichtheden op het pad van de protonenbundels en variaties in de vorm en locatie van de organen. Deze variaties kunnen leiden tot een vervormde dosisverdeling, met mogelijk lokale over- en onderdosering tot gevolg.

Om te kunnen omgaan met deze dagelijkse anatomische veranderingen hebben we online-adaptieve strategieën ontwikkeld om voldoende dosis per fractie te garanderen en een lage dosis in de kritieke organen te behouden. Met de ontwikkeling van deze strategieën hebben we de haalbaarheid van online-adaptieve IMPT bewezen en tegelijkertijd zeven onderzoeksvragen beantwoord.

Een bestralingsplan kan worden teruggezet naar zijn originele staat door achtereenvolgend een energie-adaptatie en een heroptimalisatie van de spot-gewichten toe te passen.

In **Hoofdstuk 2** hebben we een 2-staps dosisrestauratie methode ontwikkeld, die een vervormde dosisverdeling kan terugbrengen naar de originele staat door te compenseren voor de dagelijkse variaties in weefseldichtheden. De methode gebruikt de dagelijkse (herhaal) CT-scan in combinatie met de contouren van de planning CT-scan geprojecteerd op de herhaal CT-scan. Voor een startplan gemaakt op de planning CT-scan wordt (1) een restauratie van de spot posities (Braggpieken) uitgevoerd door de energieën van de protonbundels aan te passen naar de nieuwe water equivalente padlengte; en (2) een heroptimalisatie van de spotgewichten uitgevoerd door de dosimetrische verschillen met het startplan te minimaliseren met behulp van een snelle en exacte kwadratische optimalisatie. Voor 10 prostaatcancer patiënten met in totaal 80 herhaal CT-scans hebben we verschillende manieren geëvalueerd van het toewijzen van optimalisatiegewichten aan de heroptimalisatie van de spotgewichten.

Resultaten toonden aan dat een hoog optimalisatiegewicht voor het planning doelvolume (planning target volume, PTV) resulteert in adequate restauraties. Voor alle scans werd een klinisch acceptabele dekking ($V_{95\%} \geq 98\%$ en $V_{107\%} \leq 2\%$) behaald voor het geprojecteerde PTV. Dosimetrische parameters voor het rectum en de blaas bleven veelal binnen 2 Gy en 2%-punt van de originele waarden. De eerste stap van de methode kostte gemiddeld 5,4 seconden, spotgewicht heroptimalisatie kostte gemiddeld 3,8 seconden.

Volledige planadaptatie kan zowel de dekking van het doelvolume als sparing van de kritieke organen verhogen in vergelijking met dosisrestauratie.

In **Hoofdstuk 3** hebben we de dosisrestauratie methode uitgebreid naar een volledige planadaptatie methode die ook kan corrigeren voor veranderingen in de vorm en locatie van het doelvolume en de kritieke organen (organs at risk, OARs). De methode gebruikt de dagelijkse herhaal CT-scan in combinatie met de contouren van de herhaal CT-scan. Beginnend met een startplan voert de methode (1) een restauratie van de spot posities uit zoals beschreven in Hoofdstuk 2; en (2) een toevoeging van nieuwe spots om het doelvolume van die dag volledig te kunnen dekken, gevolgd door een snelle optimalisatie met de referentie punt methode (RPM) van de spotgewichten. Dit resulteert in een Pareto optimaal plan voor de dagelijkse anatomie. De methode is ontwikkeld en geëvalueerd op 88 herhaal CT-scans van 11 prostaatkanker patiënten. Voor alle scans is de dosisrestauratie methode en de nieuwe online-adaptieve methode toegepast. Met de dosisrestauratie methode werd een klinisch acceptabele dekking gehaald voor het doelvolume van de dag ($V_{95\%} \geq 98\%$ en $V_{107\%} \leq 2\%$) voor 51,1% van de scans. Met de automatische volledige planadaptatie methode werd een klinisch acceptabele dekking gehaald voor 96,6% van de scans na één iteratie van het toevoegen van 2500 spots, en voor 100% van de scans na twee iteraties. Volledige planadaptatie resulteerde in lagere dosis in de OARs dan dosisrestauratie alleen. De grootste verbeteringen van dosisrestauratie naar volledige planadaptatie met twee iteraties werden behaald voor de $V_{45 \text{ Gy(RBE)}}$ van het rectum en de blaas, met verschillen in de mediaan van 10,3%-punt en 10,8%-punt respectievelijk (maximaal 22%-punt). De twee stappen van de volledige planadaptatie namen gemiddeld 7,3 seconden en 1,7 minuten respectievelijk in beslag.

Voor doelgebieden met grote dag-tot-dag variaties resulteert volledige planadaptatie in betere bestralingsplannen wanneer het gecombineerd is met een patiënt-specifieke planbibliotheek.

De volledige planadaptatie methode uit Hoofdstuk 3 gebruikt een startplan als beginpunt voor de adaptatie naar de dagelijkse anatomie. In **Hoofdstuk 4** hebben we onderzocht of doelgebieden met grote dag-tot-dag variaties voordeel kunnen hebben van de toevoeging van een patiënt-specifieke planbibliotheek aan de volledige planadaptatie methode. Dit is geëvalueerd op CT data van zes patiënten met een cervixcarcinoom (baarmoederhalskanker) met in totaal 23 herhaal CT-scans, en dosis voorgeschreven voor de primaire tumor en het nodale klinische doelvolumen (clinical target volume, CTV). In plaats van een enkel startplan per patiënt te genereren werd een patiëntspecifieke bibliotheek van startplannen gegenereerd. De bibliotheekplannen werden gegenereerd op variaties van het doelvolumen die met behulp van een van te voren vastgesteld bewegingsmodel van het doelvolumen werden bepaald. Voor elke fractie werd het best passende startplan geselecteerd. Het geselecteerde startplan werd vervolgens aangepast door middel van de adaptatie methode uit Hoofdstuk 3. De RPM-parameters werden automatisch getuned voor deze studie. Drievoudige kruisvalidatie, i.e. driemaalig tunen op één derde van de data en de gevonden parameters gebruiken op het resterende deel van de data, werd gebruikt om mogelijke invloeden van de parameters op de uitkomsten van adaptatie te minimaliseren. Het toevoegen van spots en optimalisatie van de spot-gewichten werd één of tweemaal gedaan. Het gebruik van een 1-plan-bibliotheek (één startplan dat gebaseerd is op alle beweging in het bewegingsmodel) is vergeleken met de selectie van een plan uit een 2-plan-bibliotheek (twee startplannen gebaseerd op een deel van de beweging). Het uitvoeren van een volledige planadaptatie met één iteratie van 3000 spots toevoegen resulteerde in een adequate dekking ($V_{95\%} \geq 95\%$ en $V_{107\%} \leq 2\%$) voor 37/46 plannen bij de 1-plan-bibliotheek en 41/46 plannen bij de 2-plan-bibliotheek. Met twee iteraties van spots toevoegen kon de 2-plan-bibliotheek een acceptabele dekking halen voor alle plannen, terwijl de 1-plan-bibliotheek een $V_{107\%} > 2\%$ behaalde voor 3/46 plannen. De 1-plan-bibliotheek en 2-plan-bibliotheek behaalden vergelijkbare resultaten voor de kritieke organen. Dit laat zien dat het gebruik van een patiëntspecifieke plan-bibliotheek (2-plan-bibliotheek) in plaats van een enkel startplan (1-plan-bibliotheek) gunstig is voor doelgebieden met grote dag-tot-dag variaties.

Online planadaptatie kan de dekking van het doelvolumen behouden en tegelijkertijd dosissen in de kritieke organen verlagen in vergelijking met robuust plannen.

Nu in eerdere hoofdstukken is aangetoond dat dosisrestauratie en volledige planadaptatie mogelijk zijn en kunnen leiden tot adequate bestralingsplannen, hebben we het voordeel ten opzichte van de huidige standaardmethode van robuust plannen in kaart gebracht. In **Hoofdstuk 5** hebben we daarom de drie methoden met elkaar vergeleken

voor de dataset van 11 prostaatkanker patiënten (88 herhaal CT-scans). Voor deze dataset hebben we eerst de veiligheidsmarges en robuustheidswaarden bepaald die nodig zijn om voor alle herhaal CT-scans van ten minste 90% van de patiënten een adequate dosis in het doelvolumen te behalen. Voor elke herhaal CT-scan hebben we daarnaast een bijbehorend startplan aangepast met de dosisrestauratie methode en met de volledige planadaptatie methode. Om rekening te houden met de onzekerheden in de dracht van de protonenbundel die ontstaan door de omzetting van Hounsfield-eenheden naar proton stopping powers heeft de volledige planadaptatie de gewichten van de protonenbundels robuust geoptimaliseerd met drie scenario's (nominaal, $\pm 3\%$ dracht). Door 4 en 8 mm rond het hoge- en lage-doelvolumen te plaatsen en de doelvolumen robuust te optimaliseren met een 6 mm positioneringsonzekerheid en een 3% drachtonzekerheid werd een adequate dekking ($V_{95\%} \geq 98\%$) behaald voor 10/11 patiënten. Beide online-adaptieve methoden behaalden $V_{95\%} \geq 98\%$ en betere OAR sparing voor alle 11/11 patiënten. De grootste verbeteringen werden gezien voor de $V_{45 \text{ Gy(RBE)}}$ van het rectum, met verbeteringen tot 11%-punt en 16%-punt tussen robuust en dosisrestauratie en tussen robuust en volledige adaptatie respectievelijk. Exclusief de initialisatie en de uiteindelijke dosisberekening duurde de dosisrestauratie gemiddeld 1,7 minuten en volledige planadaptatie gemiddeld 6,6 minuten.

Dosisrestauratie wordt voornamelijk beïnvloed door de keuze van het startplan, terwijl de resultaten van volledige planadaptatie meer afhankelijk zijn van het stopcriterium en de optimalisatiemethode.

Het effect van het variëren van de parameters van de ontwikkelde online-adaptieve methoden is onderzocht in **Hoofdstuk 6**. Voor elk van de 88 herhaal CT-scans van de prostaatkanker database bestaande uit 11 patiënten is een startplan aangepast met behulp van de dosisrestauratie methode en met de volledige planadaptatie methode. Voor alle scans hebben we plannen gegenereerd met variaties op de standaardinstellingen van de volgende parameters: 1) de hoeveelheid bewegingsmitigatie die is meegenomen in het startplan, 2) de sample grootte van de spot toevoeging, 3) het stopcriterium van de spot toevoeging, en 4) de optimalisatiemethode voor de spotgewichten. De eerste variatie (startplan) is geëvalueerd voor zowel de dosisrestauratie als voor de volledige planadaptatie. De laatste drie variaties zijn enkel geëvalueerd voor de volledige planadaptatie. Voor de dosisrestauratie hebben we gevonden dat het veranderen van de hoeveelheid bewegingsmitigatie in het startplan ten opzichte van de standaardrange kan leiden tot onvoldoende dosis in het doelvolumen of hogere dosissen in de kritieke organen. Voor de volledige planadaptatie zijn grote effecten alleen gevonden door het stopcriterium of de optimalisatiemethode.

ode te variëren. Het verhogen van het aantal iteraties waarin spots worden toegevoegd en de optimalisatiemethode van RPM te vervangen door de '2-phase-epsilon-constraint' (2pec) methode resulteerde bij beiden in vergelijkbare dekking van de doelvolumes maar lagere dosissen in de kritieke organen ten opzichte van de standaardinstellingen. Gezien het verhogen van het stopcriterium de rekestijden snel verhoogt en de 2pec methode te veel tijd kost voor een online-adaptieve werkwijze moet de aandacht worden gericht op het verbeteren van de uitkomsten van de RPM-optimalisaties. Aangezien de RPM-optimalisatie gestuurd wordt door de RPM-parameters is een uitgebreidere tuning van deze parameters aanbevolen.

Volledig automatische contourpropagatie voor herhaal CT-scans is haalbaar door een combinatie van deep-learning en beeldregistratie.

Om klinisch gebruik van de volledige planadaptatie mogelijk te maken moeten contouren op de dagelijkse herhaal CT-scans beschikbaar zijn. **Hoofdstuk 7** presenteert een automatische methode om deze te verkrijgen. De methode gebruikt eerst een 'deep-learning network' om automatisch de blaas in te tekenen op een dagelijkse herhaal CT-scan. Hierna worden luchtruimtes in het rectum en de darmen gelokaliseerd en opgevuld met behulp van een 'generative adversarial network' (GAN). Als laatst worden met behulp van het opgevulde beeld en de verkregen blaascontour de handmatige intekeningen van de planning CT-scan gepropageerd naar de dagelijkse herhaal CT-scan met 'deformable image registration' (DIR). Drie verschillende databases van prostaatkankerpatiënten van drie verschillende instituten zijn gebruikt. De eerste database is gebruikt voor training en het testen van het '3D convolutional neural network' die wordt gebruikt voor het automatisch intekenen van de blaas. De overige twee datasets zijn gebruikt voor evaluatie van de volledige workflow. De verkregen contouren zijn geëvalueerd door ze te vergelijken met de gouden standaard handmatige intekeningen door middel van de 'dice similarity coefficient' (DSC), de 'mean surface distance' (MSD) en de 95% 'Hausdorff distance' (95%HD). Een klinische validatie van de contouren is uitgevoerd voor de 88 herhaal CT-scans van 11 patiënten. De contouren werden klinisch acceptabel geacht wanneer het bestralingsplan verkregen met deze automatische contouren een acceptabele dekking van het doelvolumen behaalden op de handmatige intekeningen. Deze evaluatie toonde aan dat voor 80% van de automatisch verkregen contouren direct een acceptabele dekking werd behaald en er dus geen handmatige toepassingen noodzakelijk waren.

Gemeten prompt gamma straling emissieprofielen kunnen worden gebruikt om veranderingen in PTV dosis als gevolg van dag-tot-dag anatomische veranderingen te detecteren.

In **Hoofdstuk 8** hebben we de laatste stap van de voorgestelde online-adaptieve workflow onderzocht, namelijk het monitoren van de afgifte van de behandeling door middel van prompt gamma (PG) straling emissieprofielen als onderdeel van behandeling kwaliteitscontrole. PG stralen zijn het resultaat van nucleaire interacties tussen inkomende protonen en de weefsels van de patiënt, en emissieprofielen kunnen buiten de patiënt worden gemeten. We hebben onderzocht of PG straling emissieprofielen gebruikt kunnen worden om verschillen in PTV dekking door inter-fractie anatomische variaties ten opzichte van de planning CT-scan te detecteren. Een bestralingsplan is gegenereerd voor elke planning CT-scan van de prostaat kanker database (11 patiënten, 88 herhaal CT-scans) met behulp van Erasmus-iCycle. Afstraling van het plan werd met Monte-Carlo gesimuleerd op de planning CT-scan en de herhaal CT-scans. Voor elke dosisverdeling werden dosimetrische parameters ($D_{x\%}$) die de tumor dekking beschrijven gescoord, alsmede het punt halverwege de afname regio van het PG emissieprofiel (X_{50}). Verschillen tussen de geplande waarden en de waarden van de herhaal CT-scans werden geëvalueerd, en de Pearson correlatie coëfficiënten tussen functies van de ΔX_{50} en $\Delta D_{x\%}$ werden bepaald. De sterkste correlaties werden gevonden tussen veranderingen in de dekking van het hoge- en lage-dosis PTV en de mediaan en het gemiddelde van de absolute waarden van ΔX_{50} , bij het gebruiken van enkel de 5% meest intense protonbundels in het PG emissieprofiel. De correlaties laten zien dat het in principe mogelijk is om PG emissieprofielen te gebruiken om dosimetrische veranderingen te detecteren. Om een exacte relatie tussen de emissieprofielen en de dosimetrische parameters te bepalen is meer onderzoek noodzakelijk.

PhD portfolio



PhD student	Ir. T.Z. Jagt
PhD period	2015 – 2020
Erasmus MC Department	Radiation Oncology
Research School	Molecular Medicine
Promotor	Prof. dr. M.S. Hoogeman
Copromotor	Dr. ir. S. Breedveld

General courses	Year
Programming with Python	2015
Giving and receiving feedback	2015
Presenting skills	2016
Scientific writing	2016
ESTRO course Particle Therapy	2017
1st ESTRO Physics Workshop (Glasgow, oral presentation)	2017
ESTRO Pre-meeting Radiobiology course	2018
Scientific Integrity	2018
ESTRO Pre-meeting Machine Learning course	2019

International presentations	Year
ICCR 2016, London (oral presentation)	2016
Enlight 2016, the Netherlands (poster (awarded) + oral presentation)	2016
ESTRO 36, Vienna (oral presentation)	2017
ESTRO 37, Barcelona (oral presentation)	2018
AAPM 2018, Nashville (oral presentation)	2018
ESTRO 38, Milan (oral presentation)	2019
PTCOG, Manchester (oral presentation)	2019
ESTRO 39, Vienna (accepted for poster discussion)	2020

National presentations	Year
LPRGT, Utrecht (oral presentation)	2018
Kringdag, AMC Amsterdam (oral presentation)	2016, 2019
Implementation and collaboration hub, Rotterdam	2019

In-house presentations	Year
Journal Club (5x)	2015 - 2018
R&D (4x)	2016 - 2019
Referee (3x)	2016 - 2019
Research day (awarded for best oral presentation in 2018)	2016, 2018

Teaching activities	Year
Introduction to proton therapy for students of the Haagse Hogeschool	2016
Guest lecturer for proton course ZRTI Vlissingen	2018
Guest lecturer for RTT proton course 'Planning for proton therapy' (4x)	2015 - 2019
Supervising master-students	2016 - 2019

Other	Year
Educational sessions, organization and attendance (12x)	2018 - 2019

List of Publications

PEER-REVIEWED JOURNAL PAPERS

1. T.Z. Jagt, S. Breedveld, S. van de Water, B.J.M. Heijmen, and M.S. Hoogeman. "Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations in prostate cancer". In: *Physics in Medicine and Biology* 62.11 (2017), pp. 4254–4272. DOI: 10.1088/1361-6560/aa5c12
Chapter 2
2. T.Z. Jagt, S. Breedveld, R. van Haveren, B.J.M. Heijmen, and M.S. Hoogeman. "An automated planning strategy for near real-time adaptive proton therapy in prostate cancer". In: *Physics in Medicine and Biology* 63.13 (2018), p. 135017. DOI: 10.1088/1361-6560/aacaa7
Chapter 3
3. T.Z. Jagt, S. Breedveld, R. van Haveren, R.A. Nout, E. Astreinidou, B.J.M. Heijmen, and M.S. Hoogeman. "Plan-library supported automated replanning for online-adaptive intensity-modulated proton therapy of cervical cancer". In: *Acta Oncologica* 58.10 (2019), pp. 1440–1445. DOI: 10.1080/0284186X.2019.1627414
Chapter 4
4. T.Z. Jagt, S. Breedveld, R. van Haveren, B.J.M. Heijmen, and M.S. Hoogeman. "Online-adaptive versus robust IMPT for prostate cancer: How much can we gain?" In: *Radiotherapy and Oncology* 151 (2020), pp. 228–233. DOI: <https://doi.org/10.1016/j.radonc.2020.07.054>
Chapter 5
5. M.S. Elmahdy, T.Z. Jagt, R.T. Zinkstok, Y. Qiao, R. Shahzad, H. Sokooti, S. Yousefi, L. Incrocci, C.A.M. Marijnen, M.S. Hoogeman, and M. Staring. "Robust contour propagation using deep learning and image registration for online adaptive proton therapy of prostate cancer". In: *Medical Physics* 46.8 (2019), pp. 3329–3343. DOI: 10.1002/mp.13620
Chapter 7
6. E. Lens, T.Z. Jagt, M.S. Hoogeman, and D.R. Schaart. "Correlations between the shifts in prompt gamma emission profiles and the changes in daily target cover-

age during simulated pencil beam scanning proton therapy". In: *Physics in Medicine and Biology* 64.8 (2019), p. 085009. DOI: 10.1088/1361-6560/ab145e
Chapter 8

7. Y. Qiao, T.Z. Jagt, M.S. Hoogeman, B.P.F. Lelieveldt, and M. Staring. "Evaluation of an Open Source Registration Package for Automatic Contour Propagation in Online Adaptive Intensity-Modulated Proton Therapy of Prostate Cancer". In: *Frontiers in Oncology* 9 (2019), p. 1297. DOI: 10.3389/fonc.2019.01297

CONFERENCE PROCEEDINGS

1. T.Z. Jagt, S. Breedveld, S. van de Water, B.J.M. Heijmen, and M.S. Hoogeman. "OC-0448: Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations". In: *Radiotherapy and Oncology* 123 (2017), S238. DOI: 10.1016/S0167-8140(17)30890-3
Oral Presentation ESTRO 2017
2. T.Z. Jagt, S. Breedveld, R. van Haveren, B.J.M. Heijmen, and M.S. Hoogeman. "Scientific Abstracts and Sessions - An Automated Planning Strategy for Near Real-Time Adaptive Proton Therapy". In: *Medical Physics* 45.6 (2018), e120–e706. DOI: 10.1002/mp.12938
Oral Presentation AAPM 2018
3. T.Z. Jagt, S. Breedveld, B.J.M. Heijmen, and M.S. Hoogeman. "OC-0090: A Planning Strategy for Near Real-Time Adaptive Proton Therapy". In: *Radiotherapy and Oncology* 127 (2018), S47–S48. DOI: 10.1016/S0167-8140(18)30400-6
Oral Presentation ESTRO 2018
4. E. Lens, T.Z. Jagt, M.S. Hoogeman, M. Staring, and D.R. Schaart. "OC-0082: Using prompt gamma emission profiles to monitor day-to-day dosimetric changes in proton therapy". In: *Radiotherapy and Oncology* 127 (2018), S40–S41. DOI: 10.1016/S0167-8140(18)30392-X
Oral Presentation ESTRO 2018
5. M.S. Elmahdy, T.Z. Jagt, S. Yousefi, H. Sokooti, R.T. Zinkstok, M.S. Hoogeman, and M. Staring. "Evaluation of Multi-metric Registration for Online Adaptive Proton Therapy of Prostate Cancer". In: *Biomedical Image Registration*. 2018, pp. 94–104. DOI: 10.1007/978-3-319-92258-4_9
6. T.Z. Jagt, S. Breedveld, R. Van Haveren, R.A. Nout, E. Astreinidou, M. Staring, B.J.M. Heijmen, and M.S. Hoogeman. "OC-0081 Plan-library supported auto-

mated replanning for online-adaptive IMPT of cervical cancer". In: *Radiotherapy and Oncology* 133 (2019), S38–S39. doi: 10.1016/S0167-8140(19)30501-8

Oral Presentation ESTRO 2019

7. M.S. Elmahdy, T.Z. Jagt, R. Zinkstok, C. Marijnen, M.S. Hoogeman, and M. Starling. "PO-0989 Deep learning improves robustness of contour propagation for online adaptive IMPT of prostate cancer". In: *Radiotherapy and Oncology* 133 (2019), S543–S544. doi: 10.1016/S0167-8140(19)31409-4

Poster ESTRO 2019

Curriculum Vitae

Thyrza Jagt was born on April 22, 1991, in Lelystad, the Netherlands. After graduating high school in 2008, she started the bachelor program Applied Mathematics at the Delft University of Technology. She obtained the degree of Bachelor of Science after completing a minor on Science Education; an internship on treatment planning for hyperthermia at the Erasmus Medical Center Cancer Institute in Rotterdam, the Netherlands; and a graduation project on optimization in scheduling at ABB in Mannheim, Germany. After this she continued with the master Applied Mathematics, for which she received a Master of Science degree in Numerical Mathematics. During her graduation project she focused on the modeling of the development of pressure ulcers. She also completed the master Science Education for which she received a Teaching degree in Mathematics.

In April 2015 she started her PhD at the department of Radiation Oncology at the Erasmus Medical Center Cancer Institute in Rotterdam. In this period, she developed automated methods for online-adaptive intensity-modulated proton therapy, of which the publications resulted in this thesis. Since January 2020, she has been working as a postdoctoral researcher at the department of Radiation Oncology at the Dutch Cancer Institute (NKI) in Amsterdam. She is currently focusing on the clinical aspects of online-adaptive planning for treatments using the MR-Linac.

