CHARACTERIZING THE EFFECTS OF DELINEATION UNCERTAINTIES IN HEAD AND NECK CANCER RADIOTHERAPY

CHARACTERIZING THE EFFECTS OF DELINEATION UNCERTAINTIES IN HEAD AND NECK CANCER RADIOTHERAPY

by

André MAKKINJE

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday, March 08, 2023 at 10:00 AM.

Student number:	4162447	
Project duration:	May 19, 2022 - March 08, 2023	
Thesis Committee:	Dr. Z. Perkó	TU Delft supervisor
	Dr. ir. D. Lathouwers,	TU Delft
	Dr. ir. R. de Kruijff	TU Delft
	Dr. S.J.M. Habraken	Erasmus MC

ACKNOWLEDGEMENTS

I would like to express my gratitude to Zoltán for his time, help and guidance throughout this project. You really helped me reach the end. I also want to thank Steven for helping me out and making time whenever I had questions. Lastly I want to thank Michelle and Wens for their time and patience helping me with all the practical problems I encountered during this thesis.

ABSTRACT

The goal of radiotherapy is to maximize the dose to the target while minimizing the dose to normal tissue. Treatment plans are optimized for this goal and dose delivery is improved in accuracy and precision. The optimization is based on the delineations of the volumes of interest on the medical images of the patient. The problem is that these delineations contain uncertainties and the effect of these uncertainties become more pronounced as the accuracy and precision of dose delivery improves. The delineation uncertainties are caused by various factors such as knowledge and experience of the observers, guidelines, and image quality and modality.

The goal of this thesis is to research the effects of delineation uncertainties. Uncertainties are characterized using a rolling ball algorithm (RBA) to modify the delineation. The radius of the ball represents the uncertainty and ranges from -5 mm to 5 mm for the clinical target volume (CTV) and -2 mm to 2 mm for the brain stem. The effect of these uncertainties will be researched for both fixed and re-optimized dose distributions.

These are then used as input to create a model that will simulate the dose distributions for different uncertainties using Polynomial Chaos Expansion (PCE). PCE will be used to model dose volume histograms (DVH) of the CTV and the brain stem. The dosimetric effect will be determined by setting confidence intervals in $D_{98\%}$ of the CTV and $D_{2\%}$ of the brain stem. The widths of the confidence intervals in these two metrics represent the dose uncertainty.

This thesis uses three sets of patient data. Each patient had a CTV close to the brain stem. Dose uncertainty in the two metrics was found to increase as the uncertainty of CTV and brain stem delineation increases.

For a fixed dose distribution, a delineation uncertainty of 0.75 mm to 1.25 mm in the CTV would lead to a dose uncertainty of 2 Gy in the $D_{98\%}$ of the CTV. A delineation uncertainty of 0.25 mm to 0.5 mm in the brain stem would lead to the same dose uncertainty in $D_{2\%}$ of the brain stem.

For a re-optimized dose distribution, a combined delineation uncertainty of 1.25 mm for two patients and 4.5 mm for one patient would lead to a dose uncertainty of 2 Gy in $D_{98\%}$ of the CTV. For the same dose uncertainty in the $D_{2\%}$ of the brain stem, there would be a combined delineation uncertainty between 1.0 mm and 4.5 mm. For all three patients, the dose uncertainty in the $D_{2\%}$ was more sensitive to delineation uncertainties in the brain stem.

The effects of these uncertainties will depend on the nominal situation. For all patients a 2 Gy dose uncertainty corresponds to 2.85% of the prescribed dose to the CTV and 3.33% of the maximum dose constraint of the brain stem. One patient was on the boundary of underdosing the CTV in the nominal situation and a 2 Gy uncertainty would be sufficient to underdose the CTV. The nominal $D_{2\%}$ for all three patients was sufficiently low that it is unlikely it would exceed the maximum dose constraint.

Contents

1

INTRODUCTION

Head and neck cancer is a common cancer. It ranks seventh on a global level and accounts for more than 660,000 new cases and is the cause of 325,000 deaths a year [1, 2]. There also appears to be an increase in cases with a predicted 30% increase by 2030 [1, 2].

Radiotherapy is one of the methods that can be used to treat cancer. There are different variations of radiotherapy, such as brachytherapy and external beam radiotherapy. This thesis will focus on external beam radiotherapy and specifically proton beams.

1.1. RADIOTHERAPY

The focus of radiotherapy is to deliver a high dose to tumors while minimizing the dose to normal tissue [3, 4]. The dose can be delivered using different particles, such as photons and protons. Figure 1.1 shows the depth dose profiles of different particles in tissue. The figure shows how, for photons, there is a small build-up region until it reaches the peak, and then the dose decreases as the distance increases. Protons show a different behavior by depositing the dose mainly at a specific depth. This can be seen in the recognizable Bragg peak, after which the dose very quickly falls to zero. The blue line shows a spread-out Bragg peak by modifying the proton beam so that the Bragg peak can cover a larger range. The thesis will focus on radiotherapy using protons.

1.2. RADIOTHERAPY TREATMENT PLANNING

General approach after diagnosis is to scan the patient to generate medical images. This can be done with different modalities, such as magnetic resonance imaging, computed tomography (CT), and positron emission tomography. After the scans are performed, the volumes of interest should be delineated in these images. Figure 1.2 shows a visual example of relevant volumes. There is a gross tumor volume (GTV), which is the tumor that is actually visible on the images. Then there is the clinical target volume (CTV), which includes the GTV and minor microscopic extensions that are not visible on the images. If photons are used, then there is also a planning target volume (PTV). This is an expanded CTV to take into account multiple uncertainties, such as the setup error, organ motion, and delineation uncertainty. Lastly, there are organs at risk (OARs). These are

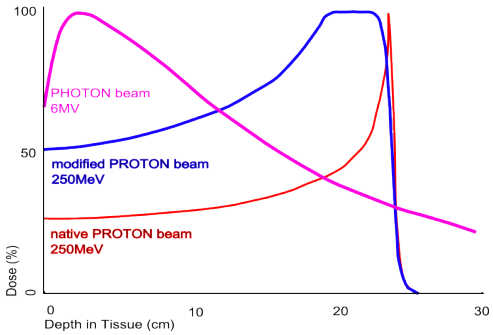


Figure 1.1.: Relative dose delivered by different particles in tissue. The red beam represents proton beams and shows a clear example of the Bragg Peak. The blue line shows a good example of a spread-out Bragg Peak. Figure reproduced from [5]

specific volumes that must be taken into account during planning, since there is usually a limit on how much dose these organs can tolerate before having adverse effects on the patient. Manual delineation of these volumes by an observer is a time consuming process [6–8]. However, there are options to do the delineations automatically using auto-contouring software and this can reduce the time [7, 8].

An example of auto-contouring software is based on the use of an atlas [9]. An atlas is a medical image of a different patient that has already been delineated. This medical image and its delineations will be warped to fit the current patient. Then, depending on the situation, some corrections might be necessary to finalize the delineations for the current patient.

A structure set contains all the delineated volumes of a patient. A treatment plan will be developed for this structure set. This plan will try to achieve the goal of delivering a high dose to tumors (GTV, CTV, PTV) while minimizing the dose to normal tissue such as the OARs. It starts by setting the treatment goals for each volume, and then the treatment planning software will use these goals and structure sets to determine the dose delivery and distribution within the patient. The dose distribution will be evaluated and, depending on the outcome, the plan will be accepted or the treatment goals will be adjusted and a new plan will be created for the new input.

The treatment plan can also be adjusted during treatment. With adaptive radiation therapy (ART), the plan can be modified based on possible variations in patient anatomy that might occur during the treatment period and can be performed between treatments, right before treatment or even during treatment [10]. This might require additional scans and new delineations. Depending on time constraints, the delineations can be made manually or automatically.

Based on this workflow, it becomes clear that delineations are one of the first steps taken in radiotherapy to cure a patient. Therefore, any errors or uncertainties made in this step will spread throughout patient treatment [4]. Furthermore, as dose delivery becomes more accurate and precise, the need for accurate volume delineations increases even more [11, 12]. As shown by the Bragg peak of protons in Figure 1.1, a small inac-

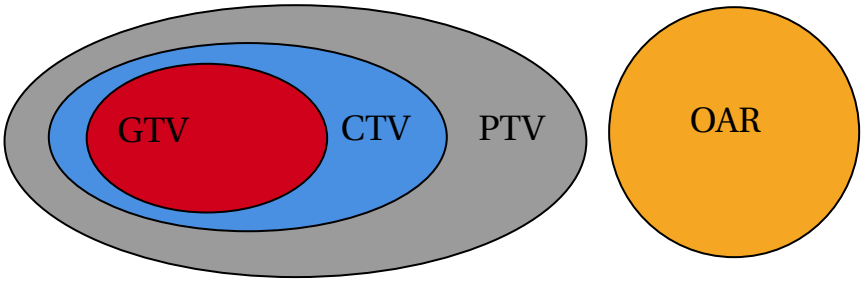


Figure 1.2.: Examples of the different volumes of interest that will be delineated on medical images. For this thesis the focus is on the clinical target volume (CTV) and the organs at risk (OAR).

curacy can have a noticeable impact on dose delivery. Even during ART, repeated delineations may still contain uncertainties depending on the situation.

1.3. UNCERTAINTIES IN RADIOTHERAPY

There are many different reasons for the uncertainty in the delineation. Most of them can be categorized as inter- or intraobserver variability (IOV). This refers to the difference in delineations between different observers or the difference between the delineations of the same observer at different times, respectively. These differences in turn have different causes. There can be a difference in experience, knowledge, and/or training. It can also be influenced by the type of imaging and the quality of these images; in addition, there can be different interpretations of these images. There is not always a guideline available and there are sometimes multiple guidelines that lead to different results. These guidelines can also be interpreted differently. The various factors will be explained in more detail in Section 2.1.2.

There are multiple different steps being taken to reduce IOV. Some of the steps focus on improving knowledge and experience through teaching to various degrees of success [13]. Other steps focus on creating guidelines that can be followed [14, 15].

There is also the possibility of using automatic contouring instead of manual contouring. This should lead to more consistent delineations and, therefore, limited IOV; however, these can have their own type of delineation uncertainties. The atlas is subject to the same type of uncertainty as mentioned above. Therefore, these uncertainties can be warped for the new patient.

1.4. RESEARCH QUESTION

The previous section has explained how uncertainties can be present and how they can affect the general workflow and treatment of patients. Now the next step is to determine what the actual effects are of these uncertainties. Without knowing the effects, it is difficult to estimate if the uncertainties give actual reason for concern. The effect will be investigated for two different situations.

In the first situation, the dose distribution of the nominal plan for the nominal structure set will be calculated and kept constant. The volumes of interest will then be modified to represent possible uncertainties and result in a new structure set. The dose delivered to this new volume will be calculated. This gives insight into possible consequences for the actual dose that is delivered to a volume. Therefore, the first research question is: "What is the dosimetric effect of delineation uncertainties for a fixed dose distribution?".

The second situation will use the same structure set of the first situation; however, the plan will be re-optimized for this new structure set. This gives a re-optimized dose distribution for this modified structure set. This gives insight into how uncertainties in volume A can affect the dose delivered to volume A, but also how uncertainties in volume B can affect the dose delivered to volume A. Therefore, the second research question is: "What is the dosimetric effect of delineation uncertainties for a reoptimized dose distribution?".

To find an answer to these questions, it is necessary to first characterize the uncertainty of the delineation. The volumes of interest can then be modified to simulate a possible uncertainty. This will result in the new structure sets that can be used in a fixed dose distribution or as the input for a new re-optimized dose distribution. This will result in a dose distribution that can be connected to a specific set of uncertainties in volumes of interest. A relation between the dose distribution and uncertainties in volumes of interest can then be found and used to answer the research questions.

1.5. THESIS STRUCTURE

After this general introduction, the theory will be discussed in Chapter 2. Chapter 3 will cover the methods and material used in this thesis. Results will be presented in Chapter 4 and those results will be discussed in Chapter 5. Finally Chapter 6 will contain the conclusion and recommendations for future research.

2

THEORY

2.1. DELINEATION UNCERTAINTY

This section will discuss the causes and factors that influence the uncertainty of the delineation and how these uncertainties can be characterized.

2.1.1. CAUSES OF DELINEATION UNCERTAINTIES

Delineation uncertainties are found in the form of inter- and intraobserver variability (IOV) [16]. Interobserver variability refers to the variability between the delineations of the same patient, made by different observers at the same time. Intraobserver variability refers to the variability between the delineation of the same patients, made by the same observer but at different times. There are several factors that can influence IOV [4]:

- Ground truth;
- Guidelines;
- Experience and knowledge;
- Image quality and modality.

GROUND TRUTH

A lack of ground truth makes it difficult to investigate IOV and is also the reason there is IOV. Therefore, a reference volume is necessary to estimate the IOV [17]. A delineated volume is a collection of voxels. A voxel is a volume element in, for example, a CT scan. The dimensions of the voxel depend on the machine settings. A reference volume can be created by using different approaches of either selecting delineations or more specifically selecting which voxels to include in a reference volume.

One approach relies on using a manual delineated contour; however, this contour can also contain an IOV. The influence of IOV can be minimized by creating a reference volume using multiple inputs from different observers. This can be a simple consensus between different experienced observers, or all of their delineated contours are taken

into account, and the reference volume is included based on probability. A voxel is included in a volume if it is included in at least x% of all delineations, where x must be greater than a predetermined threshold.

A different option uses the STAPLE algorithm [18]. The algorithm will take multiple different delineated contours as input and estimates an optimal combination to create this reference volume. Initially it is similar to the probability approach; it takes all delineated contours equally into account and creates a reference volume. Then each delineated contour is rated for accuracy compared to the reference volume; this accuracy score is used to give more weight to more accurate contours and less weight to less accurate contours. Now a new reference volume will be created by using the weighted delineated contours. This process is repeated until the reference volume is constant.

GUIDELINES

Guidelines can affect the magnitude of IOV. IOV can be reduced when the different observers use the same guidelines [19]. A variation that also decreases the IOV uses an automatically delineated contour as a base that observers can manually edit [20]. The effectiveness of the guidelines also depends on the interpretation of the observer and which guideline is used. Different guidelines can lead to different delineations [21]. Training can also affect IOV, but the degree of effectiveness varies depending on the type of training intervention [13].

EXPERIENCE AND KNOWLEDGE

The experience and knowledge of the observer also affect IOV [22]. The delineation of a target volume relies on the estimation of the observer on what tissue is a tumor. It is also possible that more experienced observers delineate smaller volumes than less experienced observers [23]. Furthermore, the agreements between contours can depend on the level of experience, where more experienced observers can have higher agreements with each other [24].

IMAGE QUALITY AND MODALITY

The last factor that can affect IOV is the quality of the image and the type of image [25]. IOV can differ depending on the imaging modality used [26]. This may be due to the quality of the image, but also depends on the interpretation of the image by the observer [27]. However, a lower IOV for a particular imaging modality does not necessarily mean it is superior, because there may be a lack of training that influences the IOV [17].

2.1.2. CHARACTERIZING DELINEATION UNCERTAINTIES

The metrics used to characterize delineation uncertainties can be divided into different groups [17].

- Comparisons of volume or center of mass;
- Measures of overlap;
- Distance-based.

Volume comparisons consist of measurements or ratios between volumes of different delineations. It is a relatively simple metric that can be used to characterize uncertainty; however, it does not provide information on the location of volumes. A high agreement in this metric can still have a high IOV. This is also true for comparisons of the center of mass.

The overlap of measures usually consists of a ratio between unions, intersections, or total volumes. An example is the Dice-Similarity Coefficient (DSC) defined as $\frac{2\cdot(A\cap B)}{A+B}$ which is the ratio between the intersection of volumes A and B and the total volume of A and B. A perfect overlap would give $DSC = 1$. As the value decreases to zero, the agreement between the two volumes decreases. These types of metric give more information regarding the location of the volumes and relative size of the volumes, but there are limitations, since it is insensitive near the edges, especially for larger volumes. Furthermore, these metrics may not be as useful in determining the difference in dose delivered [28, 29].

On the other hand, distance-based metrics are more sensitive to errors near the edges [30]. A common example is the Hausdorff distance. The Hausdorff distance is defined by finding the shortest distance between contour A and contour B for every point on contour A. From this group, the largest distance is the Hausdorff distance. Alternatives are to take a percentile such as 95% or the mean. A smaller Hausdorff distance indicates a better agreement. Distance-based metrics do not necessarily correlate with overlap measures and therefore might be complementary [28, 30].

2.2. MODELING DELINEATION UNCERTAINTY

This section will discuss several methods found in the literature to model uncertainty or incorporate uncertainties in the planning process. First, several different (an)isotropic margin recipes will be discussed, followed by models based on image contrast. Finally, the rolling ball algorithm will be discussed, including the traditional approach and the modifications that are used for the application within this thesis.

2.2.1. MARGIN RECIPES

A traditional approach to incorporate and model uncertainties involves the use of margin recipes to expand the CTV, to create the PTV. An example approach was to make this expansion depend on the different errors that can occur, such as motion, setup, and delineation errors. [31]. In this recipe, a combined standard deviation for each x-, y-, and z-component is calculated according to Equation 2.1. Σ_m^2 represents the standard deviation due to motion, Σ_s^2 represents the standard deviation due to setup errors, and lastly Σ_d^2 represents the standard deviation due to delineation errors. In addition, the recipe also includes a collective combined random error for the possible random errors due to motion and setup.

$$\Sigma^2 = \Sigma_m^2 + \Sigma_s^2 + \Sigma_d^2 \tag{2.1}$$

As stated above, this recipe is used to expand the CTV into a PTV. This is limited to tumors; however, there is an analogous expansion of the planning risk volume for some

organs at risk [32]. However, this has limited use for modeling uncertainty insofar only expansion is possible, where as delineation uncertainties can involve expansion and reduction at parts of the delineation. Furthermore, PTVs are not used in proton therapy.

In contrast to the previous margin recipe, which involves an expansion of the CTV based on Cartesian coordinates, Bell et al. researched a delineation margin for whole breast radiotherapy [33] using polar and spherical coordinate systems. The patients were divided into different groups according to the size of the CTV volume. These groups were labeled as small, medium, and large. A margin was created by calculating the standard deviation. The standard deviation was the difference between the radius of the contour of a specific patient for a specific angle and slice and the average value of the radius of the contour at the same angle and slice. The average value was taken over the entire data set, but also for each separate group labeled as small, medium, and large. Therefore, four different standard deviations were used to create four different margins. Their findings concluded that their approach was better than the conventional approach for larger target volumes.

2.2.2. ALTERNATIVE MARGIN STRATEGY

Bernstein et al. propose an alternative margin strategy [34]. In their approach, the clinician delineates two sets of GTV as shown in Figure 2.1.

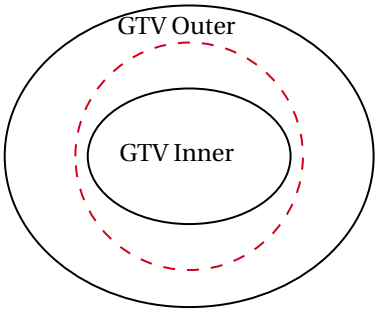


Figure 2.1.: An illustration of the alternative margin recipe [34]. Each full circle represents a GTV. The dashed red circle represents the delineated GTV based on a probability distribution.

These sets are called GTV Inner and GTV Outer. The GTV Inner is smaller than the GTV Outer and is fully within the GTV Outer. The GTV Inner contains all voxels of which the clinician is 100% certain that it is part of the target volume. Voxels outside of GTV Outer are voxels for which the clinician is 100% sure that they are not part of the target volume. The remaining voxels are between these two GTVs and include all voxels for which the clinician is not 100% certain if it is a part of the target volume. The clinician has to choose a probability density function (PDF) based on his expectations of where the real boundary of the GTV is located. The chosen PDF can bias the real boundary to be closer to either GTVs or could be uniform or Gaussian distributed.

Regardless of the chosen PDF, the PDF will depend on the voxel that is between the two GTVs and on the total path length. The total path length is the sum of the shortest

distance from that voxel to GTV Inner and the shortest distance from that voxel to GTV Outer. The PDF is normalized so that along the path starting from GTV Inner to the voxel to GTV Outer, there is exactly one voxel that is on the boundary of the real GTV. This is represented by the dashed red contour in Figure 2.1. This approach benefits from only needing a single clinician, but the downside is that it requires twice as many delineations made by a single person. Furthermore, as the authors state themselves, the delineation uncertainties are still influenced by the same general factors mentioned in Section 2.1.2. Also, like the other margin recipe, this is focused on target volumes and does not include OARs.

2.2.3. AVERAGE-SURFACE-OF-STANDARD-DEVIATION

Xu et al. model the delineation uncertainties using average-surface-of-standard-deviation (ASSD) [35]. Their model assumes that the source of the delineation uncertainties lies in the perception of image contrast. The uncertainty was modeled as a random perturbation scaled by a factor dependent on the intensity gradient of the CT image for that specific voxel. A larger intensity gradient results in a smaller magnitude of the uncertainty, and a lower intensity gradient results in a larger uncertainty in that voxel.

Aliotta et al. also used the ASSD model to simulate the uncertainty of delineation[36]. Their approach was similar insofar as the magnitude was dependent on the intensity gradient of the CT image in that voxel. The difference was in the magnitude of the random perturbation, which included three different values. The random perturbation was 2 mm, 5 mm, or 10 mm.

The delineation uncertainty was investigated in isolation and in combination with different setup errors. The generated delineation with an initial perturbation of 2 mm and 5 mm was considered realistic; however, the delineation that started with an initial perturbation of 10 mm was considered less realistic. Furthermore, their findings concluded that a small uncertainty of the delineation in an OAR close to a target volume led to clinically significant differences in the quality of the plan. However, if the OARs were not close to a target volume, only 5mm and 10mm led to significant differences in plan quality. Furthermore, these findings were for situations in which the delineation uncertainty was investigated in isolation. However, the impact decreased when the setup errors were taken into account.

2.2.4. ROLLING BALL ALGORITHM

Traditionally, the rolling ball algorithm (RBA) is used to expand the CTV to PTV [37], as shown in Figure 2.2. RBA would essentially be used as a margin recipe, where an isotropic expansion would be applied to the CTV. The expansion would depend on the radius of the ball. The method is shown in Figure 2.2. A ball of chosen radius R would roll around the CTV to apply an isotropic expansion of $2 \cdot R$.

A variation would involve setting up the confidence ellipsoid for translations around the center of the CTV and then moving the entire CTV while keeping the center within the range of the ellipsoid [38]. The confidence ellipsoid can contain the possible translations for the x, y, and z components, as well as the possible rotation uncertainties. This approach is equivalent to rolling the same ellipsoid around the CTV boundary.

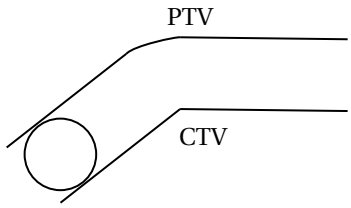


Figure 2.2.: Example of rolling ball algorithm applied to a CTV. The CTV is expanded by the diameter of the ball to a PTV.

This thesis will also use RBA. The traditional approach creates a PTV large enough to be x% sure that the CTV is contained within that volume. The modified approach of this thesis uses the RBA to model uncertainties for a specific delineation. It features both expansion and reduction of any delineation by allowing for positive and negative values of the radius R. The approach can be used for both target volumes and OARs. The magnitude of the radius will always be in millimetres. The radius of the ball can have different values for different volumes for the same structure set. For each structure set the dose distribution will be calculated. Then confidence intervals can be set to determine the dosimetric effect of a specific uncertainty.

2.3. GENERAL TREATMENT PLANNING

This section will first discuss treatment planning in general. It is followed by an introduction to the treatment planning software iCycle.

2.3.1. TREATMENT PLANNING

The goal of radiation therapy is to maximize the dose to the target volume while minimizing the dose to normal tissue [39]. The process starts by making the necessary images, usually CT images, to characterize the anatomy of the patient. The target volume and the OARs will be delineated in these images. Then specific treatment goals will be set for different volumes. Target volumes tend to have a minimal prescribed dose, while OARs will have a maximum allowable dose.

Treatment planning software will use these delineations and treatment aims to determine the correct dose delivery and dose distribution. The software will determine the direction, number, intensity, and angle of the beams. This will also give a calculated dose distribution. Depending on the outcome, the plan is accepted or adjustments must be made to the treatment goals until an acceptable plan is formed.

2.3.2. ICYCLE

The previous section described a more manual approach to treatment planning based on trial-and-error; however, iCycle generates treatment plans automatically based on a user-chosen wish-list [40]. Each wish-list contains constraints and objectives for different volumes. Every constraint has to be met, while the objectives are optimized accord-

ing to their priority. For proton therapy, iCycle will choose the optimal locations and intensities of the pencil beam [41].

Originally an iterative loop would be used to select optimal pencil beams [40, 41], however, this thesis used a more recently developed method of spot selection. 20,000 spot candidates are uniformly distributed and the optimized spot weights are chosen based on a weighted-sum cost function that is derived from the wish-list. 20% of the spots with the highest monitor units are selected for optimization. Spots that fall below a minimum monitor unit threshold during optimization are removed and followed by a re-optimization. The end result will be a Pareto optimal plan, which is a plan that cannot improve performance in a criteria without sacrificing performance in a different criteria.

The optimization method consists of two phases [42]. At first, it starts to create a plan that achieves the first objective within the limitations of the original constraints. After each optimization of an objective, it will become a constraint that will be used to optimize the next objective. Therefore, higher priority objectives are subject to less constraints than lower priority objectives. In the second phase, the objectives that have met the goal, but could be further optimized, will be optimized again. This process also occurs on the basis of the priority order.

Furthermore, it is possible to use robust treatment planning with iCycle. As stated in section 2.2.4 the concept of PTV is not used in proton therapy [43]. PTV is based on the static dose cloud approximation, which is true for photons [44]. Small shifts have a limited impact on the dose. This is not true for protons. Range or setup errors can have a noticeable impact on the dose distribution [43]. This can be handled by using robust treatment planning. Robust treatment planning creates plans for different error scenarios. Each error scenario will usually contain a setup error in one direction and may contain a range error.

An example of robust treatment planning could calculate the dose distribution for 19 scenarios. One scenario for the nominal situation without errors. 6 scenarios for the maximum setup error in the positive/negative x, y, z directions with zero range error. Those last 6 scenarios will then be repeated for a different range error. In this example a maximum positive range error and a maximum negative range error. Then a minimax optimization is used to minimize the penalty in the worst performing scenario [45].

2.4. THEORY OF POLYNOMIAL CHAOS EXPANSION

In this section, the theoretical basis of Polynomial Chaos Expansion (PCE) is introduced. This will be used to quantify the dosimetric effects of the delineation uncertainties mentioned in Section 2.1.2. PCE is a spectral modeling approach to quantifying uncertainty. Spectral methods (re)construct the dependence of the response on a set of random variables. Equation 2.2 shows the general expression of the dependence [46].

$$R(\vec{\xi}) = \sum_{k=0}^{\infty} r_k \Psi_k(\vec{\xi}) \tag{2.2}$$

In equation 2.2 R is the solution dependent on the random set of variables $\vec{\xi}$. Ψ_k are basis vectors and r_k are the corresponding coefficients.

Although Equation 2.2 shows the general expansion of an infinite sum, the sum has to be limited to a finite number for computational reasons [46]. Equation 2.3 shows the expansion that will be used for this thesis. This limits the infinite sum to $P+1$. R refers more specifically to the responses of interest, Ψ_k are the polynomial chaos(PC) basis vectors, and r_k are the polynomial chaos coefficients(PCC).

$$R(\vec{\xi}) = \sum_{k=0}^{P} r_k \Psi_k(\vec{\xi}) \tag{2.3}$$

The variability of R as a function of the random variables $\vec{\xi}$ will be investigated. $\vec{\xi} = (\xi_1, \cdots, \xi_N) \in \mathbb{R}^N$ are the input variables and the total N determines the number of dimensions. This vector has a joint probability density function $p_{\vec{\xi}}$. The random variables are assumed to be independent and therefore the joint probability density function will be the product of each individual distribution. This leads to $p_{\vec{\xi}}(\vec{\xi}) = \prod_{j=1}^{N} p_{\xi_j}(\xi_j)$.

2.4.1. PC BASIS VECTORS

The $P+1$ PC basis vectors Ψ_k need to be defined and the corresponding PCC r_k needs to be calculated. A polynomial family will be chosen for each random variable ξ_j. The polynomial family will be chosen using the Wiener-Askey scheme [47]. This scheme states that the polynomial family orthogonal to the probability distribution p_{ξ_j} of ξ_j is the optimal choice. Legendre polynomials are the optimal choice for random variables with a uniform distribution.

Continuing with the example of Legendre polynomials, the PC basis vectors can be written as $\Psi_k\left(\vec{\xi}\right) = \prod_{j=1}^{N} P_{\gamma_{k,j}}(\xi_j)$. Here P is the Legendre polynomial for the random variable ξ_j and the order is determined by $\gamma_{k,j}$, where $\vec{\gamma}_k = (\gamma_{k,1}, \gamma_{k,2}, \cdots, \gamma_{k,N})$. Equation 2.4 shows the complete base set for a specific order O. A complete set contains $P+1 = \frac{(N+O)!}{(N!O!)}$ basis vectors.

$$\Gamma(O) = \left\{ \prod_{j=1}^{N} P_{\gamma_{k,j}}(\xi_j) \;:\; \sum_{j=1}^{N} \gamma_{k,j}(\xi_j) \leq O \right\} = \left\{ \Psi_k(\vec{\xi}) \;:\; \sum_{j=1}^{N} \gamma_{k,j}(\xi_j) \leq O \right\} \tag{2.4}$$

It is possible to reduce the full set to a sparse set by eliminating higher-order multidimensional PC vectors. This is possible on the basis of the principle of sparsity of effects [48]. The principle states that most responses are caused by the main parameters or low-order interactions; therefore, it is possible to remove the higher-order interactions with limited effect on the response.

$$\left\| \vec{\gamma}_k \right\|_q \equiv \left(\sum_{j=1}^{N} \gamma_{k,j}^q \right)^{1/q} \tag{2.5}$$

Equation 2.5 shows the quasi-norm, where only the orders for which $\left\| \vec{\gamma}_k \right\|_q \leq O$ is valid are kept. With $q \in (0,1]$ dictates which orders are eliminated. For $q = 1$, none of the orders are eliminated and the whole set is used. As q goes to zero, more higher orders are eliminated.

Whether the full or reduced set of PC basis vectors is used, the set will be orthogonal to each other with respect to the joint probability distribution $p_{\vec{\xi}}(\vec{\xi})$. Equation 2.6 shows

the value of the inner product of two PC basis vectors. h_k is the norm of $\Psi_k(\vec{\xi})$ and $\delta_{k,l}$ is the Kronecker delta.

$$\left\langle \Psi_k(\vec{\xi}), \Psi_l(\vec{\xi}) \right\rangle = h_k^2 \delta_{k,l} \tag{2.6}$$

2.4.2. EXPANSION COEFFICIENTS

The PCCs r_k will be calculated using non-intrusive spectral projection. The method is non-intrusive since it depends on the actual output R of specific realizations of $\vec{\xi}$ [46]. Therefore, the actual model used to calculate the output R can be treated as a black box, and no adjustments to the model are necessary [49].

Equation 2.7 shows how the PCCs will be calculated. The denominator is simplified using Equation 2.6. The integral will be approximated using a cubature formula. The cubature formula will be built using quadratures.

$$r_k = \frac{\left\langle R(\vec{\xi}), \Psi_k(\vec{\xi}) \right\rangle}{\left\langle \Psi_k(\vec{\xi}), \Psi_k(\vec{\xi}) \right\rangle} = \frac{\int_D R(\vec{\xi}) \Psi_k(\vec{\xi}) p_{\vec{\xi}}(\vec{\xi}) d(\vec{\xi})}{h_k^2} \tag{2.7}$$

Equation 2.8 shows how the integral for a function $f(\xi_j)$ dependent on the random variable ξ_j can be approximated by a quadrature. The quadrature points are $\xi_{j,lev}^{(i)} \in [a,b]$. $\omega_{lev}^{(i)} \in \mathbb{R}$ are the weights. These depend on the probability density function p_ξ and the quadrature rule. To distinguish between different levels of accuracy between quadratures, the level index lev is used. The accuracy increases as lev increases. The total number of function evaluations n_{lev} also increases as lev increases.

$$I^{(1)} f = \int_a^b f(\xi_j) p_{(\xi_j)}(\xi_j) d\xi_j \approx Q_{lev}^{(1)} f = \sum_{i=1}^{n_{lev}} f\left(\xi_{j,lev}^{(i)}\right) \omega_{lev}^{(i)} \tag{2.8}$$

For this thesis, the chosen quadrature rule is Gauss quadratures. Gauss quadratures are accurate for polynomial orders up to $2 \cdot n_{lev} - 1$ [50]. The relation between the level index and the function evaluations is $n_{lev} = 2 \cdot lev - 1$. Therefore, the Gauss quadratures will be accurate up to a polynomial order of $4 \cdot lev - 3$.

As stated before, the cubature formula is built from the quadratures. This is done by tensorization. Equation 2.9 shows the different form of equation 2.8 but now for N different random variables. In this thesis, the same level index will be used for all variables, therefore $lev_1 = lev_2 = lev_N = lev$. Then Equation 2.9 also shows a problem of how the total number of points and the number of function evaluations scale. Function evaluation scales with n_{lev}^N where N is the number of input variables. As the number of variables increases, the number of function evaluations increases exponentially. This is the curse of dimensionality [46]. This effect can be limited by using sparse grids, which reduces the function evaluations by decreasing the number of quadrature points.

$$
\begin{aligned}
I^{(N)} f \approx Q_{lev}^{(N)} f &= \left(Q_{lev_1}^{(1)} \otimes \cdots \otimes Q_{lev_N}^{(1)}\right) f \\
&= \sum_{i_1=1}^{n_{lev_1}} \cdots \sum_{i_N=1}^{n_{lev_N}} f\left(\xi_{1,lev_1}^{(i_1)}, \cdots, \xi_{N,lev_N}^{(i_N)}\right) \omega_{lev_1}^{i_1} \cdots \omega_{lev_N}^{i_N} \\
&= \sum_{i}^{n} f\left(\vec{\xi^{(i)}}\right) \omega^{(i)}
\end{aligned}
\tag{2.9}
$$

2.4.3. SPARSE GRIDS

Sparse grids follow a logic similar to that of the set of reduced PC basis vectors. The principle of sparsity of effects also applies in this situation [48]. Therefore, higher-order grids can be left out with a similar method to that used in Equation 2.5.

First, the difference formula needs to be defined, and then the quadrature rules need to be rewritten as the sum of these difference formulas. Equation 2.10 shows the definition of the difference formula and Equation 2.11 shows the rewritten form of the quadrature formulas for 1 variable and the complete tensorization for N variables.

$$
\begin{aligned}
\Delta_{lev}^{(1)} f &= Q_{lev}^{(1)} f - Q_{lev-1}^{(1)} f \\
Q_0^{(1)} f &= 0
\end{aligned}
\tag{2.10}
$$

$$
\begin{aligned}
Q_{lev}^{(1)} f &= \sum_{l=1}^{lev} \Delta_l^{(1)} f \\
Q_{\vec{lev}}^{(N)} f &= \sum_{l_1=1}^{lev_1} \cdots \sum_{l_N=1}^{lev_N} \left(\Delta_{l_1}^{(1)} \otimes \cdots \otimes \Delta_{l_N}^{(1)}\right) f
\end{aligned}
\tag{2.11}
$$

$\vec{lev} = (lev_1, \cdots, lev_N)$ is the vector containing the different levels for each input variable. Combining with the new multi-index $\vec{l} = (l_1, \cdots, l_N)$, equation 2.11 can be rewritten in a more general form, as shown by equation 2.12. $\mathscr{I}(lev)$ refers to the collection of grid points based on the level index lev.

$$
\begin{aligned}
Q_{lev}^{(N)} f &= \sum_{\vec{l} \in \mathscr{I}(lev)} \left(\Delta_{l_1}^{(1)} \otimes \cdots \otimes \Delta_{l_N}^{(1)}\right) f \\
&= \sum_{\vec{l} \in \mathscr{I}(lev)} \Delta_{\vec{l}}^{(N)} f
\end{aligned}
\tag{2.12}
$$

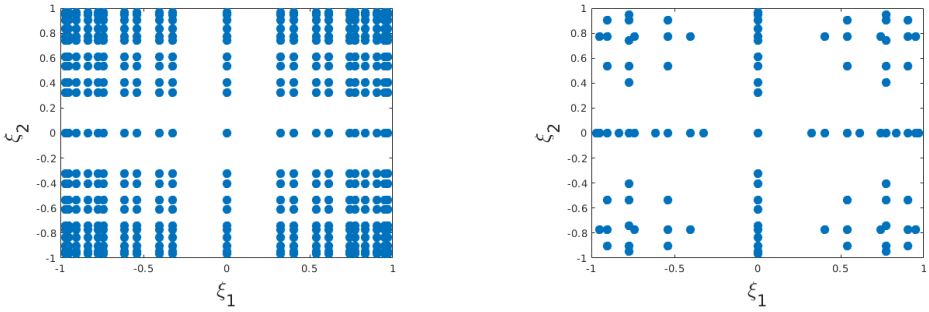
Then the Smolyak sparse grid can be defined according to Equation 2.13. The intention of using Smolyak sparse grids is to reduce the number of points and as a result decrease the computation time. In addition to the Smolyak sparse grid, there is also the extended Smolyak sparse grid.

$$
\mathscr{I}^{Smolyak}(lev) = \left\{ \vec{l} : \sum_{j=1}^{N} l_j \leq lev + N - 1 \right\}
\tag{2.13}
$$

The extended Smolyak sparse grid slightly increases the number of points to increase the accuracy of the response. This is done by replacing the highest 1D grid order with an extended grid, as shown by Equation 2.14.

$$\begin{aligned} l_j &= lev + N - 1 \\ l_j &= lev + N - 1 + lev_{extra} \end{aligned} \tag{2.14}$$

Figure 2.3 shows a comparison between a full grid for $lev = 5$ and a sparse grid for the same lev. The full grid contains 441 points, while the sparse grid contains 97 points. Extension has not been applied to the sparse grid; however, if the grid was extended by 1 level, the number of points would have increased to 101.



(a) Full grid.

(b) Sparse grid.

Figure 2.3.: A comparison of a full grid and sparse grid for $lev = 5$ and two input variables. The sparse grid is not extended. Full grid contains 441 points and the sparse grid contains 97 points.

2.4.4. PROJECTION & REGRESSION

Now Equation 2.7 can be solved. The integral can be solved using equation 2.12 and substituting the general function f with $R\Psi_k$. This leads to equation 2.15. Choosing the level index, also called grid order (GO), will determine the accuracy. The weights and grid points depend on the choice of (extended) Smolyak sparse grid. The response R is the real response for the random variable determined by the points on the grid. All of this can be combined with the basis vectors chosen to calculate the PCC.

$$\begin{aligned} r_k &= \frac{1}{h_k^2} \sum_{\vec{l} \in \mathscr{G}^{Smolyak}(lev)} \Delta_{\vec{l}}^{(N)} (R\Psi_k) \\ &= \frac{1}{h_k^2} \sum_{i}^{n} R\left(\vec{\xi}^{(i)}\right) \Psi_k\left(\vec{\xi}^{(i)}\right) \omega^{(i)} \end{aligned} \tag{2.15}$$

In addition to projection, the coefficients can also be calculated using a regression method [46]. This thesis will also calculate the PCCs using a least-squares minimization problem.

The real model is replaced by an alternative model with an expansion similar to that of Equation 2.3. The grid points used for projection will also be used for regression.

$$\hat{R}(\vec{\xi}) = \sum_{k=0}^{P} \hat{r}_k \Psi_k(\vec{\xi}) \tag{2.16}$$

The solution is found by minimizing Equation 2.17 where the sum is taken over all points of the grid M.

$$\sum_{i}^{M} \left(R(\vec{\xi}^{(i)}) - \hat{R}(\vec{\xi}^{(i)}) \right)^2 \tag{2.17}$$

The equation to be solved can be rewritten as shown by equation 2.18. Here Z is a matrix and each element $Z_{ij} = \Psi_j\left(\xi^{(i)}\right)$.

$$\begin{aligned} \left(Z^t Z\right) \hat{R} &= Z^t R \\ \hat{R} &= \left(Z^t Z\right)^{-1} Z^t R(\vec{\xi}) \end{aligned} \tag{2.18}$$

3

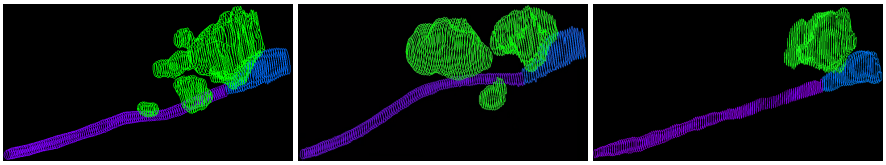
METHOD

This chapter will cover the methods and data used. First, the patient data will be discussed. This contains information on the data, volumes of interest, and the wish-list and robustness parameters. The next section will explain how the rolling ball algorithm is applied in modeling the delineation uncertainty. This is followed by an explanation of the parameters that will be used to construct the PCE models, what the responses of interest are, and how the models will be validated. The chapter ends with sections covering the characterization of the dosimetric effects of the uncertainty of the delineation.

3.1. iCYCLE & PATIENTS

3.1.1. PATIENT DATA

There are three sets of patient files that are being used and are differentiated by referring to their patient ID number. Figure 3.1 contains the three volumes of interest. For each patient, that is, the high-dose clinical target volume (CTV), the brain stem, and the spinal cord. The target volume is close to the brain stem and spinal cord. Other target volumes and organ at risk (OAR) are delineated; however, these remain constant. CT scans have a resolution of 0.9766 x 0.9766 x 2.0 mm. There are 192 slices, and each slice is 511 by 192 voxels.



(a) Patient 2 (b) Patient 69 (c) Patient 155

Figure 3.1.: The three structures of interest for all three patients. Green is the high-dose clinical target volume, purple is the spinal cord and dark blue is the brain stem.

3.1.2. WISH-LIST

Table 3.1 shows the constraints and objectives for the CTV, brain stem, and spinal cord. As stated in Section 2.3.2 constraints must be met, and the objectives are optimized in priority order [40]. The same wish-list is used for all patients. There is a goal for the objectives, however, with the new method only objectives for the target volume will be optimized towards the specific goal. Other objectives will focus only on minimizing the dose. This affects the objectives of the brain stem and spinal cord. The prescribed dose for the high-dose CTV is 70 Gy and 54.25 Gy for the low-dose CTV. Serial_OARs refers to the OARs and they are expanded by 15 mm. Therefore, there are two different CTV_7000 volumes, one that does not overlap with expanded OARs, and one that does. The CTV that does not overlap with the expanded OARs has a minimum dose constraint, while the CTV that does overlap with the expanded OARs has an objective.

iCycle will calculate the dose distribution for a total of 19 scenarios per structure set. One nominal scenario and 18 error scenarios. Each error scenario contains the largest setup error in either the x, y, or z direction and may contain a range error. The setup error is 3 mm and the range error is 3%.

Priority	Volume	Min/Max	Type	Goal [Gy]
Constraint	Brainstem_Surf	Minimize (maximum)	linear	60
Constraint	Brainstem_Core	Minimize (maximum)	linear	54
Constraint	SpinalCord_Surf	Minimize (maximum)	linear	60
Constraint	SpinalCord_Core	Minimize (maximum)	linear	50
Constraint	CTV_7000_substract_Serial_OARs_expanded	Maximize (minimum)	linear	67.9
Constraint	CTV_7000	Minimize (maximum)	linear	95
Objective	CTV_7000_AND_Serial_OARs_expanded	Maximize (minimum)	linear	67.9
Objective	CTV_7000	Minimize (maximum)	linear	53.2
Objective	Brainstem	Minimize (maximum)	linear	30
Objective	SpinalCord	Minimize (maximum)	linear	30

Table 3.1.: Constraints and objectives on the high-dose clinical target volume, brain stem volumes, and spinal cord volumes. Other constraints and objectives are not shown.

3.2. ROLLING BALL ALGORITHM

The starting point for the algorithm is the structure set contained in the DICOM files. A volume of interest is represented by a collection of N connected points. The coordinates of every point are stored in the DICOM files as a vector of length 3N. The points are connected on the basis of their order in the vector. The points will form a closed contour. The goal of the algorithm is to expand or shrink the contour, which involves displacing the points in the desired direction. A positive radius R expands the contour by R, and a negative radius shrinks the contour by R.

DIRECTION OF THE POINTS

First, the desired direction must be determined. This is done by finding the vectors that points from a point P_n of the contour to a point P_{n+1} for every point. Then this vector will be rotated counterclockwise with a rotation matrix to get the normal to this vector and then normalized to form the unit vector. The correct direction of the unit vector depends on the sign of the radius R and if the points P are ordered in a (counter)clockwise direction.

The direction of the points can be (counter)clockwise and, within the patient data sets, there is no consistent direction between slices of the same volume. The direction of the points is determined using a geometric solution by calculating the light blue area enclosed by the contour shown in Figure 3.2. If the sign of the area is positive, the direction of the points is clockwise. Otherwise, it is counterclockwise. For each point P on the contour, the vector toward the point $P+1$ is calculated. If the vector points in the positive x-direction, then that line segment will have a positive area under the curve, else it has a negative area under the curve. With this approach, a positive enclosed area indicates clockwise direction and a negative enclosed area indicates counterclockwise direction.

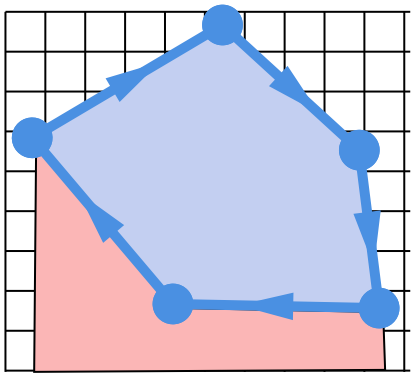
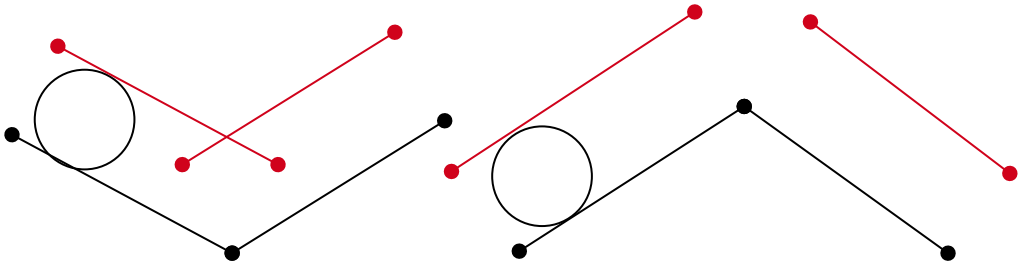


Figure 3.2.: Geometric method to determine the clockwise direction of the contour. Thick dark blue represents the contour with points in a clockwise ordering. The area under each line segment is calculated and the metric of interest is the sign of the light blue area. If the sign of the blue area is positive, the direction is clockwise. If the sign is negative, the direction is counterclockwise.

CREATING AND INSERTING NEW POINTS

Now, every point will have two unit vectors pointing in the desired direction, one unit vector per connected line segment. This is used to create two new collections of points. A collection of points where the nominal points are translated by unit vectors that belong to the line segments of the point $P_n -> P_{n+1}$. The other collection where the nominal points are translated by the unit vectors belonging to the line segments from point $P_{n-1} -> P_n$. This translation then results in two different types of situation, as shown in Figure 3.3. Each situation has a different problem that needs to be solved. The differen-

tiation between the two situations depends on the angle between two connecting line segments. Therefore, all angles are found to filter out the two different situations.



(a) (b)

Figure 3.3.: Two different situations after the initial expansion. The black line segments represents the nominal contour. The red line segments represent the initial translation by the radius R. In both cases new points need to be generated by the two translated points. For angles smaller than 180°, the new points have to be created by intersection. For the remaining angles, these points need to be interpolated.

The problem in Figure 3.3a is that there are new points for which the distance between the point and a nominal line segment is smaller than the radius $|R|$. In both cases, the points need to be removed and replaced to maintain the correct shape. For every nominal point that belongs to this situation, the intersection between the two red line segments is calculated and inserted between the two points to maintain the correct order of the points. The removal of points will be done in a later step.

The problem in Figure 3.3b is that there is a gap between the red line segments. DICOM will connect these directly, but then the distance between that line segment and a nominal point is less than the radius $|R|$. The solution is to insert a collection of points along an arc of radius R with the nominal point at its center. The computation time and accuracy of the contour scale with an increase in points. For this thesis, an additional 15 evenly distributed points will be inserted.

REMOVING INVALID POINTS

The current approach will create invalid points that need to be removed. A new point is invalid if the distance between the new point and the nominal contour is smaller than $|R|$. The points are invalid since every point on the new contour need to be a distance R away. These points will be removed in a two step process. First the distance between every new point and every nominal point is calculated. If there is a distance smaller than $|R|$, the corresponding new point is removed. Then for the remaining new points, the distance between that point and a nominal line segment is calculated. If there is a distance smaller than $|R|$, then that point is removed.

3.3. POLYNOMIAL CHAOS EXPANSION

This section will cover how polynomial chaos expansion (PCE) will be used in this thesis. PCE is used to model the Dose Volume Histogram (DVH) of the CTV and brain stem. Additionally, it is also used to model the dose distribution when the dose distribution is re-optimized. The subsections have been divided into the construction of the PCE model in 3.3.1 and model validation in 3.3.2.

Instead of creating separate sections for fixed and reoptimized dose distributions, each subsection will cover both due to the significant overlap in the methods.

3.3.1. BUILDING PCE MODEL

The PCE models are built with the OpenGPC package in Matlab [49, 51]. Each model requires a specific grid order (GO), extra level (EL), polynomial order (PO), trimming, and random distributions for every input variable. EL was fixed at 1 for all models and trimming was not used. GO was in the interval of [3,6] for a fixed dose distribution and [3,5] for re-optimized dose distributions. PO was always 1 larger and therefore ranged from [4,7] and [4,6], respectively.

There were a total of two random distributions that were used for fixed and re-optimized dose distributions. Both input variables had a uniform distribution. The first input variable represented the uncertainty in the CTV delineation and was in the interval of [-5 mm, 5 mm]. The second input variable represented the uncertainty in the brain stem and was in the interval of [-2 mm, 2 mm].

As stated above, each GO requires a different number of grid points, and the points themselves might also be different. A sparse grid has been chosen for all models. For each point on the grid, there are combinations of radii that are chosen from the two random distributions. These radii are then used in the rolling ball algorithm to modify the delineations, and this results in a new structure set. The remaining steps will depend on whether the dose distribution is kept constant.

If the dose distribution is fixed, then the new DVH of the CTV and the three volumes of the brain stem will be calculated and used to build the PCE model. Spinal cord volumes are not included because these volumes are constant. Therefore, DVH cannot change because the dose distribution is also kept constant. The dose distribution used to calculate the new DVH is always the dose distribution that belongs to the nominal structure set.

If the dose distribution is re-optimized for the new structure set, then the new DVH of the CTV, the three volumes of the brain stem and the three volumes of the spinal cord are included. The three volumes of the OARs refer to the whole OAR, the core, and the surface of that OAR. The reason for adjusting all three volumes for the brain stem is that there is a constraint on the core and surface of the brain stem but not on the entire brain stem.

Additionally, the entire dose distribution is also included to build the PCE model.

The specific combination of radii, DVH, and dose distribution will be called a scenario. Therefore, each grid point has a scenario, and building a PCE model for different GO will require a different number of scenarios. Table 3.2 contains the total number of scenarios needed to build the PCE model for a specific GO. As shown, the number of scenarios

increases as the GO increases. This is also the reason why the GO has been limited to 5 for the re-optimized dose distributions. Each re-optimization takes 6-8 hours, and it was not feasible to perform all re-optimizations for GO6 for each patient within the available time frame.

	# of Scenarios
GO3	21
GO4	49
GO5	101
GO6	125

Table 3.2.: The number of scenarios needed to build the PCE models for a specific GO

3.3.2. VALIDATING PCE MODEL

The PCE models will be validated using evaluation scenarios. As stated above, a scenario contains a combination of radii, DVH of volumes of interest, and a dose distribution if it has been re-optimized. To save time, the same scenarios that have been used to build the PCE models will also be used as evaluation scenarios. Table 3.3 show the total number of scenarios needed to build the PCE models for the selected GOs. The table also shows the number of unique scenarios that were used since some scenarios were used to build multiple GOs. All unique scenarios are used as evaluation scenarios.

	Fixed	Re-optimized
Total	296	171
Unique	225	121

Table 3.3.: The total sum of scenarios needed to build the PCE model for the selected GOs. The unique row shows the number of unique scenarios needed to build all the GOs for either a fixed or re-optimized dose distributions. Every unique scenario is used as an evaluation scenario.

For each evaluation scenario, the response of a specific model will be created and compared. The metrics of interest for each DVH are the mean error in the DVH and the mean and the 90th percentile error in specific $D_{x\%}$. This metric represents the minimal dose that is delivered to x% of the volume. A $D_{98\%}$ therefore means that 98% of the volume receives at least this dose. The errors of interest are $\Delta D_{0\%,mean}$, $\Delta D_{0\%,90}$, $\Delta D_{100\%,mean}$, $\Delta D_{100\%,90}$, $\Delta D_{2\%,mean}$, $\Delta D_{2\%,90}$, $\Delta D_{98\%,mean}$ and $\Delta D_{98\%,90}$. For each unique scenario, the error is the absolute difference of $D_{x\%}$ between the PCE response and the real value is calculated as shown in Equation 3.1. Then from this group, the mean and the 90th percentile is calculated.

$$\Delta D_{x\%} = |D_{x\%,PCE} - D_{x\%,real}| \tag{3.1}$$

If the dose distribution has been re-optimized, then the mean error in the dose distribution will also be included. The error is calculated in a similar way to Equation 3.1 where the absolute difference between the PCE response and the real value is found.

Although all of these metrics are being calculated, they do not carry the same weight to determine the validity of a model. For CTV, the mean error in DVH and the mean and the 90th percentile error in $D_{98\%}$ are the most important. The errors in $D_{98\%}$ are given more weight, since it is important to know if there is a possibility of underdosing the CTV.

For all volumes of the brain stem and spinal cord, the mean error in the DVH and the mean and the 90th percentile error in $D_{2\%}$ are the most important. The errors in $D_{2\%}$ were chosen to keep track of the maximum dose delivered to an organ at risk (OAR). Although there is no possibility of this value exceeding the planning constraint, it is in the best interest of the patient to minimize the maximum dose to these OAR's.

The extreme values of $D_{0\%}$ and $D_{100\%}$ do not carry the same weight, as PCE tends to have higher errors for these metrics that depend on a singular voxel. They have been included to confirm whether that is also the case for these patients.

In addition to these metrics, different DVH's will be plotted and compared. Both the fixed and re-optimized dose distributions will compare the PCE response to the real DVH. A third DVH is also built for the re-optimized dose distribution. The PCE response to the dose distribution will be used as a basis to reconstruct a DVH for a given volume. This type of DVH is called PCE-based DVH. For each patient and each volume of interest, the largest error scenarios are identified. These scenarios include the scenario with the largest error between the real DVH and the PCE response, and, if applicable, the scenario with the largest error between the PCE response and the PCE-based DVH. Largest error is defined as the largest sum of all errors, where error is defined as the absolute difference between the PCE response and either the real value or the PCE-based value.

Lastly there is also the voxel acceptance that will be used for the PCE responses to the dose distribution. For each evaluation scenario, all the PCE responses are compared against their real counterpart. If the absolute dose difference is below a specific threshold, then the voxel is accepted. The threshold is 0.1 Gy.

3.4. DOSIMETRIC EFFECTS

After building and evaluating, the patient's PCE model can be used to investigate the dosimetric effects of delineation uncertainties. The general method that will be used is to create data sets after simulating a large number of scenarios for limited random distributions of the input. This will be done for each PCE model that has been built for each patient. From this data set, the confidence intervals of the DVH will be calculated. Additionally, the dependence of the PCE response on the delineation uncertainties will be characterized. The PCE responses will correspond to the metrics of interest mentioned in section 3.3.2. Lastly, the histogram in those metrics of interest will be created and from those histograms the 95% and 98% confidence intervals will be found. This information will then be used to determine the maximum delineation uncertainty that will not exceed a specific width of the aforementioned confidence intervals.

3.4.1. CREATING THE DATA SET

First, the general approach must be explained. Since the PCE models have already been validated according to the steps of Section 3.3.2, it will not be necessary to use iCycle for re-optimizations of new scenarios or to use the rolling ball algorithm to create new structure sets. This gives the option of creating a large number of new scenarios with random combinations of the radii used for the adjustments to the CTV and the volumes of the brain stem.

As stated in Section 3.3.1 the uncertainty of delineation in the CTV and the brain stem is modeled by drawing the radius of the rolling ball from a uniform distribution. Each uniform distribution is of the form $[-R, R]$, where R represents the largest possible radius in millimetres of the rolling ball used to expand or reduce the delineation of a specific volume. This R also represents the uncertainty in that particular volume. To investigate the effects of different uncertainties, it is necessary to investigate by drawing from different uniform distributions by altering R.

In addition, the combination of uniform distributions used will also influence the results. For the PCE models built using re-optimized dose distributions, the effect of an uncertainty in volume A, can affect the DVH in volume B. This relation is not true for a fixed dose distribution; in that situation, the uncertainty in volume A cannot affect DVH in volume B.

Therefore, different cases have to be set up to investigate the dosimetric effects of different magnitudes of delineation uncertainty. These cases will be set up using different (combinations of) uniform distributions for the volumes of interest. For each case, 100,000 random radii will be pulled from the available uniform distributions.

Table 3.4 shows all possible uniform distributions for CTV and brain stem volumes. It shows the general form of the uniform distribution, the smallest value of R, the largest value of R, and the increments R_{step} used. R_{max} is limited by the value of R used to build the initial PCE model. This results in 21 different uniform distributions for the CTV and 9 different uniform distributions for the volumes of the brain stems.

VOI	Uniform Distribution	R_{min}	R_{max}	R_{step}
CTV	$[-R, R]$	0	5	0.25
Brain stem	$[-R, R]$	0	2	0.25

Table 3.4.: The possible uniform distribution for each volume. Radius R represents the uncertainty in mm.

For a fixed dose distribution, the dosimetric effects of the two uniform distributions will be investigated separately. This results in a total of 30 cases that will be investigated, 21 CTV cases, and 9 brain stem cases.

For a re-optimized dose distribution, the dosimetric effects of the two distributions needs to be investigated simultaneously. Every combination of the uniform distribution will be investigated, which results in a total of 189 cases.

3.4.2. DVH, DEPENDENCIES, AND CONFIDENCE INTERVALS

From this data set, the confidence intervals for the DVH will be calculated. All PCE responses to a specific metric will be sorted, independently of which scenario it belongs to. From this sorted array, the confidence intervals will be found. These confidence intervals will be limited to 98%, 95%, 75%, 50% and 25%.

For dependencies, there is a slight variation depending on whether the data set is created for a fixed or a re-optimized dose distribution. For a fixed dose distribution, only the dependency of a volume of interest on its own uncertainty will be investigated. In the case of a re-optimized dose distribution, the dependence on both uncertainties will be investigated to see the general behavior.

The histogram and confidence intervals will be set only for the metrics of interest in Section 3.3.2 that passed the validation. The confidence intervals are set with the same method as the confidence intervals for the DVH. The confidence intervals are limited to 98% and 95%. The quantity of interest is the dependence of the width of the confidence intervals on the delineation uncertainty.

Then, specific thresholds for the width of the confidence intervals are chosen. For this thesis, these thresholds are set at 0.5 Gy, 1.0 Gy, 1.5 Gy, and 2.0 Gy. The next steps depend on the dose distribution being kept constant. If the dose distribution is kept constant, then for each threshold, the largest uncertainty is found for which the width of the confidence interval does not exceed the threshold.

The situation for a re-optimized dose distribution will find three different sets of maximum uncertainties for each threshold. First all possible combinations of uncertainties are found for which the width of the confidence interval does not exceed the threshold. From this group, the largest sum of uncertainties will be calculated. That is the first set of maximum uncertainty of interest. For the second set, a subgroup is formed by finding the scenarios that contain the largest uncertainty in the CTV. Then from this subgroup, the scenario with the largest uncertainty in the brain stem is selected. This gives the second set of maximum uncertainties. For the third set, the order is reversed. Form the subgroup by finding the largest uncertainty in the brain stem, then from that subgroup find the scenario with the largest uncertainty in the CTV.

4

RESULTS

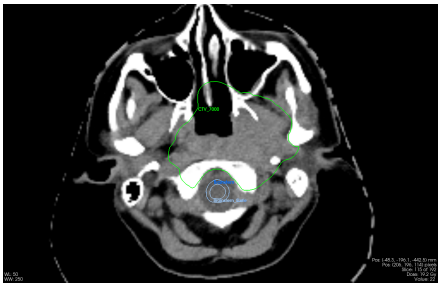
4.1. CHARACTERIZATION OF DELINEATION UNCERTAINTIES

This section is intended to show some final results of applying the rolling ball algorithm to the existing delineations. The rolling ball algorithm was applied only to the delineations of the high-dose clinical target volume (CTV) and to the three different delineations of the brain stem, which consist of the core, surface, and entire brain stem. The radius of the rolling ball was in the interval of [-5 mm, 5 mm] for the CTV and in the interval of [-2 mm, 2 mm] for the brain stem related delineations. For scenarios that adjust the brain stem, the same radius of the rolling ball is used for the delineation of the core, surface, and entire brain stem. This ensures that the original proportions between the core and the surface of the brain stem are maintained.

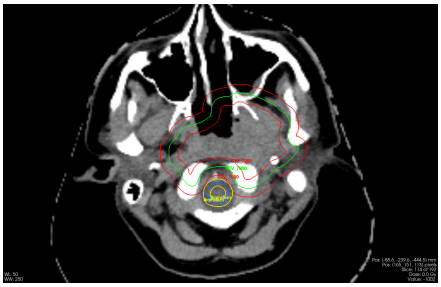
Figure 4.1a shows the original delineation of the CTV and the core, surface and entire brain stem. Figure 4.1b shows the maximum expansion and shrinkage of the CTV and the brain stem. Adjustments to the surface and core of the brain stem are not shown. The delineation of the surface of the brain stem would be similar to the delineation of the entire brain stem and would therefore be poorly visible. The delineation of the core of the brain stem is 2 mm smaller than that of the surface and the entire brain stem. Therefore, a maximum expansion of 2 mm of the brain stem core would overlap and be poorly visible.

Figure 4.2 shows the optimized dose distribution for two different structure sets. Both images contain the same slice and include the brain stem and two different CTVs belonging to two different structure sets. Although both CTVs are shown in figure 4.2, the dose is optimized for the structure set that contains only one of the two CTVs.

Figure 4.2a shows the optimized dose distribution for the original delineations and figure 4.2b shows the optimized dose distribution after a 5 mm expansion of the CTV delineation, shown in red. The CTV expansion is the only difference between the two structure sets. As expected, there is a different dose distribution. The total area within the CTV that receives at least 70 Gy has increased as the CTV increases in size, and in the brain stem you can see an increase in the minimal dose. In figure 4.2a the minimal dose of the brain stem could be as low as 10 Gy, but in figure 4.2b the minimal dose has increased to at least 20 Gy.

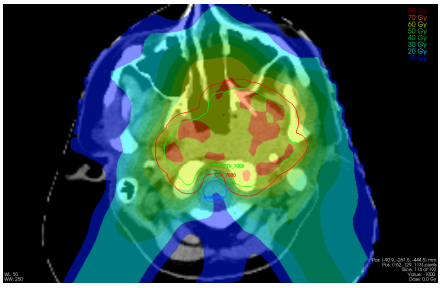


(a) Original delineations

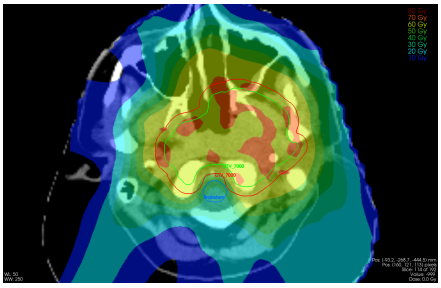


(b) Maximum expansion and reduction of CTV and brain stem delineations.

Figure 4.1.: The original CTV delineation is shown in green. The 5 mm expansion/reduction of the CTV delineation is shown in red. The original delineation of the core, surface and the entire brain stem are shown in light and dark blue. The 2 mm expansion/reduction of the delineation of the brain stem is shown in yellow.



(a) The optimized dose distribution for the structure set containing the green CTV.



(b) The optimized dose distribution for the structure set containing the red CTV.

Figure 4.2.: A comparison of the difference in dose distribution as a result of changing the CTV delineation. The green CTV is the original delineation. The red CTV is the result of a 5 mm expansion.

4.2. PCE VALIDATION

In this section, the results of the validation of the PCE model are shown for a fixed dose distribution in section 4.2.1 and for a reoptimized dose distribution in section 4.2.2.

4.2.1. PCE VALIDATION FOR A FIXED DOSE DISTRIBUTION

As mentioned in Section 3.3.1, the different PCE models are built by varying the GO at the interval $[3,6]$. The corresponding PO is always 1 greater and therefore on the interval $[4,7]$. For evaluation, the mean error will be taken into account on the DVH of the CTV and the brain stem. Furthermore, the mean and the 90th percentile error in $\Delta D_{98\%}$ for the CTV and $\Delta D_{2\%}$ for the brain stem will be taken into account.

Errors are calculated by comparing the PCE response of the different models with the evaluation scenarios mentioned in Section 3.3.2. Table 3.3 have shown that there are 225 evaluation scenarios.

The error values in the PCE responses for the CTV DVH are shown in table 4.1. The metrics of interest can be found in Table 4.1 under the columns of E_{mean}, $\Delta D_{98\%,mean}$, and $\Delta D_{98\%,90}$.

For each patient, the general error in the CTV DVH is low, regardless of the GO and PO used. The errors in $\Delta D_{98\%}$ are low except for the case of GO3PO4 for patient 2. In the case of patient 2, the mean and 90th percentile error in $\Delta D_{98\%}$ is relatively large compared to the other patients.

Given that the errors are so close in magnitude for most combinations of GO and PO, a combination cannot be clearly chosen. GO3PO4 will be excluded from consideration due to the relatively large error for patient 2 compared to the other combinations of GO and PO.

	CTV_7000								
Patient 2	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.01	0.36	0.81	0.03	0.06	0.08	0.19	0.01	0.01
GO4PO5	0.01	0.72	1.77	0.03	0.07	0.03	0.08	0.00	0.00
GO5PO6	0.01	0.41	0.78	0.03	0.05	0.04	0.07	0.00	0.00
GO6PO7	0.01	0.38	1.10	0.03	0.06	0.04	0.08	0.00	0.00
Patient 69	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.01	0.34	0.84	0.08	0.21	0.03	0.08	0.00	0.01
GO4PO5	0.01	0.53	1.10	0.08	0.19	0.03	0.07	0.00	0.00
GO5PO6	0.01	0.55	1.33	0.12	0.22	0.03	0.07	0.00	0.01
GO6PO7	0.01	0.48	0.93	0.11	0.26	0.04	0.10	0.00	0.00
Patient 155	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.01	0.30	0.62	0.01	0.03	0.03	0.07	0.00	0.00
GO4PO5	0.01	0.25	0.57	0.01	0.03	0.03	0.07	0.00	0.01
GO5PO6	0.01	0.26	0.62	0.01	0.01	0.02	0.04	0.00	0.01
GO6PO7	0.01	0.31	0.66	0.01	0.03	0.02	0.05	0.00	0.01

Table 4.1.: Errors for different PCE responses of the CTV DVH for a fixed dose distribution. The errors were calculated using all evaluation scenarios.

The error values in the PCE responses for the DVH of the brain stem are shown in Table 4.2. As stated previously, the PCE responses of interest for the brain stem are the mean error of the entire DVH, and the mean and the 90th percentile error in $\Delta D_{2\%}$. These values can be found in Table 4.2 under the columns of E_{mean}, $\Delta D_{2\%,mean}$, and $\Delta D_{2\%,90}$, respectively.

For each patient, the general error in the DVH of the brain stem is very similar, regardless of the combination of GO and PO used. For all patients and all combinations of GO and PO, the general error is greater for the brain stem than for the CTV. In all patients, the errors in $\Delta D_{2\%}$ appear to be the smallest for GO3PO4 and GO6PO7.

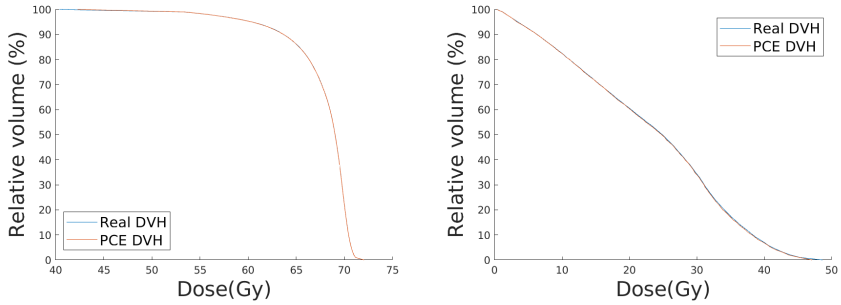
Given the CTV errors in GO3PO4, which have been determined to be too inaccurate, GO6PO7 appears to be the most accurate combination to build PCE models for the three patients. The PCE responses of the model built with GO6PO7 will be used in Section 4.3.

For each patient, the scenario with the largest overall error between the real DVH and the PCE response to the real DVH has been identified and plotted. Figure 4.3 shows the most inaccurate graphs for one of the three patients for CTV and brain stem. No plots have been made for the spinal cord related volumes. This is because those delineations

Brain stem									
Patient 2	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.03	0.03	0.07	0.24	0.45	0.02	0.06	0.04	0.11
GO4PO5	0.05	0.01	0.03	0.25	0.49	0.03	0.07	0.07	0.16
GO5PO6	0.05	0.02	0.06	0.25	0.53	0.03	0.07	0.09	0.17
GO6PO7	0.04	0.02	0.05	0.32	0.64	0.02	0.05	0.05	0.11
Patient 69	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.03	0.01	0.01	0.24	0.56	0.02	0.03	0.05	0.15
GO4PO5	0.04	0.01	0.02	0.47	1.01	0.02	0.06	0.11	0.26
GO5PO6	0.04	0.01	0.02	0.35	0.78	0.02	0.04	0.10	0.22
GO6PO7	0.04	0.01	0.01	0.46	0.98	0.02	0.04	0.07	0.15
Patient 155	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.03	0.01	0.02	0.28	0.69	0.01	0.02	0.08	0.28
GO4PO5	0.06	0.01	0.02	0.40	0.95	0.01	0.02	0.21	0.39
GO5PO6	0.05	0.01	0.02	0.40	0.99	0.01	0.02	0.20	0.39
GO6PO7	0.04	0.01	0.03	0.36	0.80	0.01	0.03	0.12	0.27

Table 4.2.: Errors for different PCE responses of the brain stem DVH for a fixed dose distribution. Errors were calculated using all evaluation scenarios.

have not been altered, and therefore the DVH would be constant for a fixed dose distribution.



(a) Patient 2 CTV

(b) Patient 155 brain stem

Figure 4.3.: Each plot shows the DVH for a volume for the scenario with the largest error between the real DVH and the PCE response.

The plots show that the PCE responses are close to the real DVH, even for the most inaccurate scenarios. The remaining plots of the most inaccurate scenarios are found in Appendix A.

4.2.2. PCE VALIDATION FOR A RE-OPTIMIZED DOSE DISTRIBUTION

In a similar manner, the PCE model for a re-optimized dose distribution will be constructed as described in Section 3.3.1. The adjusted parameters are GO and PO at the intervals $[3,5]$ and $[4,6]$, respectively. For evaluation, the mean error in the DVH of the CTV and the brain stem will be taken into account. Furthermore, the mean and the 90th percentile error in $\Delta D_{98\%}$ for the CTV and $\Delta D_{2\%}$ of the brain stem will be taken into account.

Errors are also calculated by comparing the PCE response of the different models with the evaluation scenarios. Table 3.3 have shown that there are 121 evaluation scenarios.

Table 4.3 shows the mean error for the dose distribution of a patient for various GO and PO. According to these results, the mean error for the entire dose distribution is fairly high. However, the error decreases as the GO and PO increases.

Mean dose error	Patient 2 (Gy)	Patient 69 (Gy)	Patient 155 (Gy)
GO3PO4	1.82	2.17	1.85
GO4PO5	0.85	1.07	0.69
GO5PO6	0.76	0.90	0.60

Table 4.3.: Mean error of the dose distribution of each patient for various combinations of GO and PO.

Table 4.4 shows the error values in the PCE responses of the CTV DVH. As indicated above, the mean error E_{mean} and the mean and the 90th percentile error in $D_{98\%}$ are the most important values. For each patient, the three errors decrease with higher GO and PO.

	CTV_7000								
Patient 2	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.10	0.33	0.89	0.25	0.57	0.20	0.63	0.11	0.24
GO4PO5	0.08	0.26	0.59	0.17	0.35	0.17	0.31	0.08	0.18
GO5PO6	0.07	0.23	0.50	0.15	0.33	0.10	0.26	0.07	0.14
Patient 69	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.14	0.52	1.18	0.40	0.85	0.19	0.42	0.22	0.44
GO4PO5	0.13	0.42	0.99	0.39	0.80	0.14	0.29	0.20	0.44
GO5PO6	0.11	0.34	0.71	0.33	0.63	0.12	0.27	0.17	0.36
Patient 155	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	0.11	0.58	1.43	0.29	0.72	0.18	0.43	0.14	0.34
GO4PO5	0.10	0.50	1.13	0.24	0.49	0.12	0.24	0.11	0.21
GO5PO6	0.08	0.42	0.88	0.18	0.40	0.11	0.22	0.10	0.20

Table 4.4.: Errors for different PCE responses of the CTV DVH for a re-optimized dose distribution. Errors were calculated using all evaluation scenarios.

Table 4.5 shows the error values in the PCE responses of the brain stem DVH. As indicated above, the mean error E_{mean} and the mean and the 90th percentile error in $D_{2\%}$ are the most important values. For each patient, the three errors decrease with higher GO and PO. However, the mean error is too large and, therefore, the PCE response is too inaccurate for all models. The mean error and the 90th percentile error in $D_{2\%}$ are of a more acceptable magnitude.

Based on these results, GO5PO6 has been selected to build the PCE model for the re-optimized dose distributions. The PCE responses of this model will be used in Section 4.4.

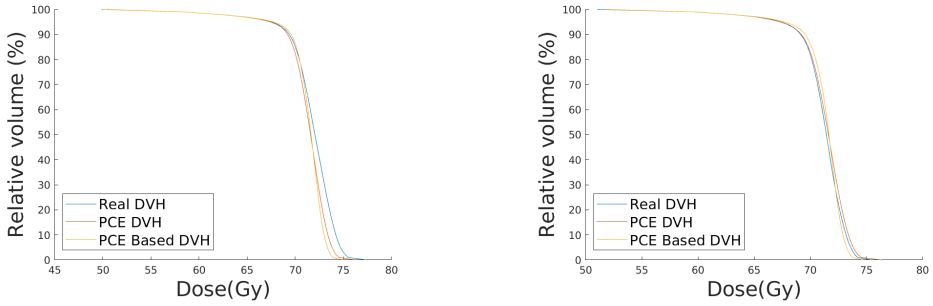
Similar to Section 4.2.1, for each patient, the scenario with the largest overall error between the real DVH and the PCE response to the real DVH has been identified and plotted. In addition to the scenario with the highest error between the real DVH and the PCE response to the real DVH, this section will also include the scenario with the largest error between the PCE response to the real DVH and the PCE-based DVH as explained in Section 3.3.2. These two scenarios have been found for CTV, brain stem, and spinal cord. For each volume, only the patient with the largest error is shown.

Brain stem									
Patient 2	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	1.07	0.37	0.82	0.48	0.98	0.81	1.87	0.20	0.48
GO4PO5	0.95	0.30	0.75	0.40	0.81	0.68	1.44	0.18	0.36
GO5PO6	0.77	0.28	0.62	0.34	0.70	0.60	1.49	0.16	0.34
Patient 69	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	1.18	0.27	0.54	0.65	1.62	0.48	1.00	0.21	0.49
GO4PO5	0.80	0.19	0.40	0.49	1.02	0.30	0.72	0.17	0.31
GO5PO6	0.72	0.16	0.37	0.36	0.79	0.29	0.70	0.15	0.30
Patient 155	E_{mean}(Gy)	$\Delta D_{100\%,mean}$(Gy)	$\Delta D_{100\%,90}$(Gy)	$\Delta D_{0\%,mean}$(Gy)	$\Delta D_{0\%,90}$(Gy)	$\Delta D_{98\%,mean}$(Gy)	$\Delta D_{98\%,90}$(Gy)	$\Delta D_{2\%,mean}$(Gy)	$\Delta D_{2\%,90}$(Gy)
GO3PO4	1.45	0.58	1.86	0.59	1.33	1.31	3.63	0.53	1.28
GO4PO5	1.31	0.58	1.80	0.51	1.13	1.29	3.30	0.43	1.06
GO5PO6	1.18	0.51	1.54	0.43	0.99	1.15	2.94	0.39	0.80

Table 4.5.: Errors for different PCE responses of the brain stem DVH for a re-optimized dose distribution. Errors were calculated using all evaluation scenarios.

These two largest error scenarios are shown in figure 4.4 for the CTV of patient 69. Figure 4.4a shows the largest error between the real DVH and the PCE response to the real DVH. The errors start to become noticeable from $D_{80\%}$ and worsen with decreasing volume fraction. However, the error in $D_{98\%}$ is very small and that is an error of interest for the CTV.

Figure 4.4b shows the largest error between the PCE response to the real DVH and the PCE-based DVH. The graph shows a large overlap between the two different DVH's, and this is an indication that there is no significant difference between the two methods that were used to build the DVH.



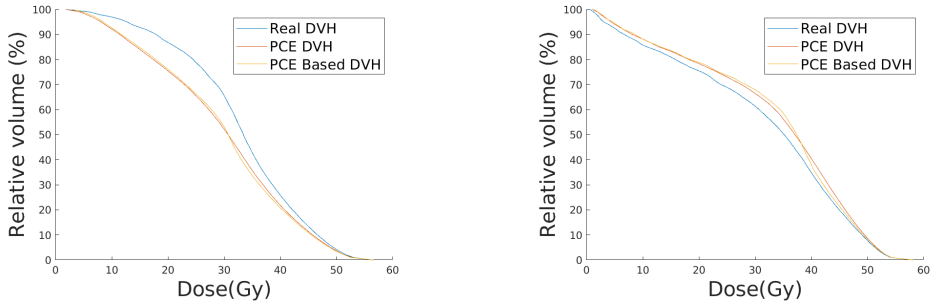
(a) (b)

Figure 4.4.: Comparison of the scenarios with the largest error between two DVHs. Left image shows the scenario with the largest error between the real DVH and the PCE response to the real DVH. Right image shows the scenario with the largest error between the PCE response to the real DVH and the PCE-based DVH. Blue line is the real DVH, red line is the PCE response to the real DVH and the yellow line is the PCE-based DVH.

Figure 4.5 shows the two largest error scenarios for the brain stem of patient 155. The scenario with the highest error between the real DVH and the PCE response to the real DVH is shown in Figure 4.5a. Errors begin to appear from $D_{5\%}$ and worsen with increasing volume fraction. However, the error in $D_{2\%}$ is very small, which is the error of interest

for the brain stem.

The scenario with the largest error between the PCE response to the real DVH and the PCE-based DVH is shown in Figure 4.5b. This graph does not show a large overlap between the two different DVH's over the entire range of volume fractions. This is an indication that there is a difference between the two methods that were used. The PCE-based DVH is built on the basis of the PCE response to the real dose distribution. As shown in Table 4.6, there is a large overall error for the response to the entire dose distribution. This error could explain the difference between the two DVH's.



(a) (b)

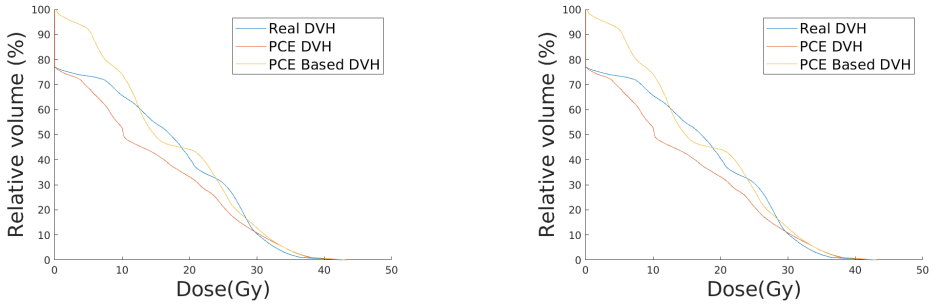
Figure 4.5.: Comparison of the scenarios with the largest error between two DVHs. Left image shows the scenario with the largest error between the real DVH and the PCE response to the real DVH. Right image shows the scenario with the largest error between the PCE response to the real DVH and the PCE-based DVH. Blue line is the real DVH, red line is the PCE response to the real DVH and the yellow line is the PCE-based DVH.

Lastly, the two largest error scenarios for the spinal cord of patient 2 are shown in figure 4.6. Figure 4.6a shows the scenario with the largest error between the real DVH and the PCE response to the real DVH. Noticeable errors are present throughout the range of the DVH. The model is not sufficiently accurate for the spinal cord.

Figure 4.6b shows the scenario with the largest error between the PCE response to the real DVH and the PCE-based DVH. As with the brain stem of patient 155, there are noticeable errors throughout the DVH range. However, in contrast to that situation, it is possible that both DVHs are inaccurate. PCE-based DVH could be similarly affected by the general error in the PCE response to the real distribution. However, the PCE response to the real DVH was also shown to have a possible inaccuracy. Therefore, it is possible that both DVH's are wrong.

The remaining figures of the largest error scenarios for these volumes for each patient can be found in Appendix A.

As can be seen in Figure 4.7 the voxel acceptance for the different combinations of GO and PO for each patient is low. In the best case scenario, it still has around a voxel acceptance of 40%, which is insufficient accuracy to use the PCE response to the real dose distribution as a basis for conclusions.

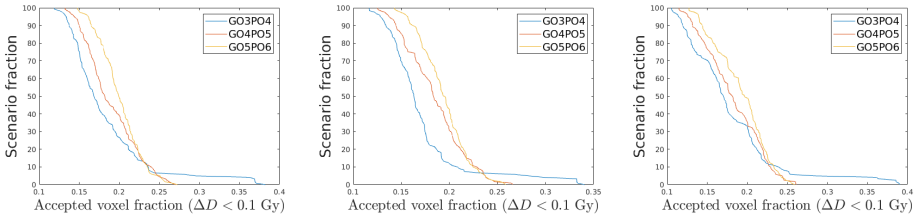


(a) (b)

Figure 4.6.: Comparison of the scenarios with the largest error between two DVHs. Left image shows the scenario with the largest error between the real DVH and the PCE response to the real DVH. Right image shows the scenario with the largest error between the PCE response to the real DVH and the PCE-based DVH. Blue line is the real DVH, red line is the PCE response to the real DVH and the yellow line is the PCE-based DVH.

A possible cause might be due to re-optimization in iCycle. Re-optimization in iCycle happens according to the wish-list. In the wish-list, there are only constraints and objectives on the minimum / maximum dose delivered to a voxel or the mean dose for an entire volume. There is no constraint or objective on the actual dose distribution within a volume.

The same wish-list, patient, and nominal structure set have been used for multiple re-optimizations with iCycle. Originally, the random seed is reset between the plans, resulting in the exact same dose distribution after each run. To investigate possible variability, this reset has been removed.



(a) Patient 2 (b) Patient 69 (c) Patient 155

Figure 4.7.: The voxel acceptance for the entire dose distribution for each patient for different combinations of GO and PO in each evaluation scenario.

Removing the reset resulted in four different distributions for the same patient, for the same structure set and the same wish-list. The different dose distributions will be referenced using the letters A,B,C, and D. Table 4.6 shows the results of comparing the different dose distributions. $\Delta D_{all,100}$ shows the 100th percentile, or maximum error,

between the entire dose distribution. These errors are large, except for the comparison between dose distribution C and dose distribution D.

$\Delta D_{>0,90}$ contains the 90th percentile error of a smaller subset of voxels. The subset contains the voxels that have an error larger than 0. This filter was used to eliminate the influence of the large number of voxels that receive no dose in both dose distributions.

$\Delta D_{all,mean}$ shows the mean error of all voxels. Except for the comparison between distributions C and D, the mean error is already larger than the 0.1 Gy threshold that has been set for voxel acceptance. The mean error becomes even greater when the subset of voxels is restricted to the same voxels that are included in the construction of the PCE model, as shown in column $\Delta D_{mask,mean}$. $\Delta D_{>0,mean}$ is the mean error of the subset of voxels that contain an error greater than zero.

$Voxels_{Mask}$ and $Voxels_{>0}$ show the total number of voxels that are included in the masked subset and the subset with an error greater than zero.

Excluding voxels that contain no error means that it excludes voxels that are zero in both distributions, but it also excludes voxels that happen to receive an identical dose. Therefore, the number of voxels that have been wrongly excluded has been found and is shown as $Voxels_{Equal}$. A voxel is wrongly excluded if the dose difference between the two distributions in that voxel is zero, while the dose in that voxel is not zero.

As can be seen from the column, the maximum number of voxels that are wrongly excluded is 172. This is the case in the comparison between distributions C and D. This comparison already has the smallest dose difference, and even for this comparison, the number of voxels is too small to have an impact on the dose differences found.

Dose Comparisons	$\Delta D_{all,100}(Gy)$	$\Delta D_{>0,90}(Gy)$	$\Delta D_{all,mean}(Gy)$	$\Delta D_{mask,mean}(Gy)$	$\Delta D_{>0,mean}(Gy)$	$Voxels_{Mask}$	$Voxels_{>0}$	$Voxels_{Equal}$
Dose Comparisons A-B	11.98	2.02	0.12	0.76	0.72	3.77E+06	4.00E+06	1
Dose Comparisons A-C	21.60	2.02	0.13	0.83	0.78	3.77E+06	4.00E+06	5
Dose Comparisons A-D	22.09	2.02	0.13	0.83	0.78	3.77E+06	4.00E+06	7
Dose Comparisons B-C	15.77	1.69	0.11	0.69	0.65	3.77E+06	4.00E+06	3
Dose Comparisons B-D	16.25	1.70	0.11	0.69	0.65	3.77E+06	4.00E+06	2
Dose Comparisons C-D	0.56	0.02	0.00	0.01	0.01	3.77E+06	3.97E+06	172

Table 4.6.: The results of comparing the four different dose distributions A, B, C, and D of the exact same situation but for a different random seed.

4.3. DOSIMETRIC EFFECT OF DELINEATION UNCERTAINTIES FOR A FIXED DOSE DISTRIBUTION

This section describes the dosimetric effect of the uncertainty of the delineation in fixed dose distributions. The dosimetric effects on CTV will be covered in Section 4.3.1, while the dosimetric effects on the brain stem will be covered in Section 4.3.2. Other regions are not discussed because the delineations are unchanged; therefore, there is no dosimetric effect due to the delineation uncertainties for a fixed dose distribution. As mentioned in section 3.4.1 a total of 30 cases, 21 CTV cases and 9 brain stem cases, will be investigated and each case will contain 100,000 random scenarios. Only the figures of one patient will be shown. The figures of the other patients can be found in Appendix B.

4.3.1. CTV

100,000 different CTVs have been made by drawing the radius from the uniform distribution [-5 mm, 5 mm] and resulted in the same number of different DVH's. Figure 4.8 shows different confidence intervals for the DVH's for patient 2. Additionally, the 98% volume fraction and the planning constraint/objective have been highlighted. It is noticeable that approximately 10% to 30% of the volume do not receive the planned dose. This is not surprising since it is a planning constraint for parts of the CTV that are far away from any OAR and it is an objective for parts that are close to the OAR. It seems that for patient 2 the objective is not met due to the close proximity of the brain stem to the CTV.

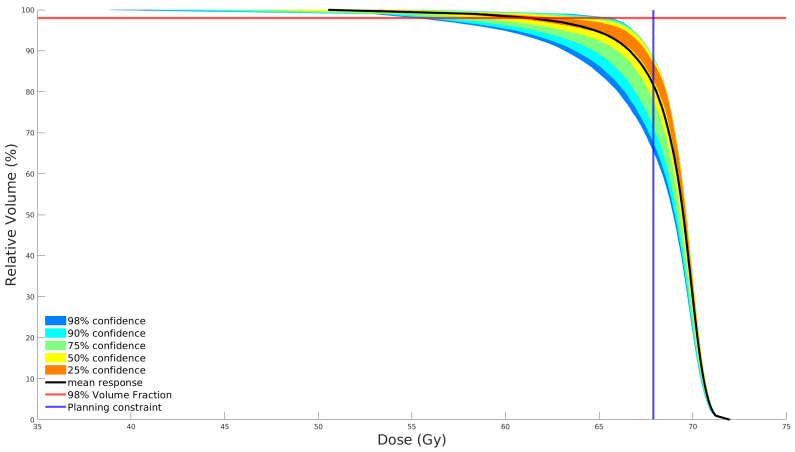


Figure 4.8.: Overview of the confidence intervals for the 100,000 DVH's that have been created for patient 2 for a fixed dose distribution.

Figure 4.9 shows the histogram of the dose delivered to 98% of the CTV volume. The figure also shows the planning constraint/objective. Unsurprisingly, there are no values at the planning constraint, since at least 10% of the volume in all cases does not receive that much dose. The width of the confidence intervals is quite wide, but this is not unexpected. The dose distribution was fixed for the original delineations. Generally, the larger the distance between a voxel and the CTV, the lower the dose. Therefore, an expansion of the CTV should generally lead to a decrease in the dose delivered to the same volume fraction. As shown in Figure 4.1 a 5 mm expansion of the CTV can significantly decrease the distance between the CTV and the brain stem. Therefore, a drastic reduction in $D_{98\%}$ is not surprising.

Figure 4.10 shows the behavior of the PCE response of $D_{98\%}$ as the CTV delineation changes in size. The correlation between $D_{98\%}$ and the adjustment of the CTV is as expected. As the CTV expands, it will expand into lower dose regions and therefore $D_{98\%}$ should decrease. As the CTV decreases in volume, it will move its boundaries into higher-dose regions and therefore $D_{98\%}$ should increase. The figure also highlights the signific-

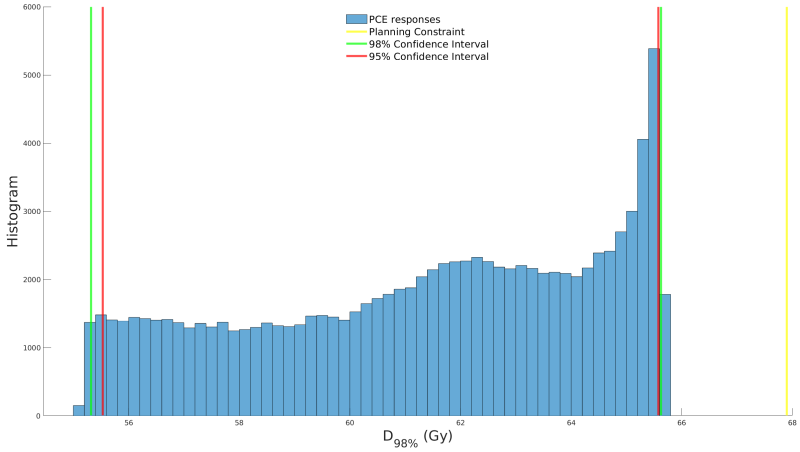


Figure 4.9.: Histogram for $D_{98\%}$ of patient 2 for a fixed dose distribution.

ant difference in dose that is possible due to the uncertainty of the delineation.

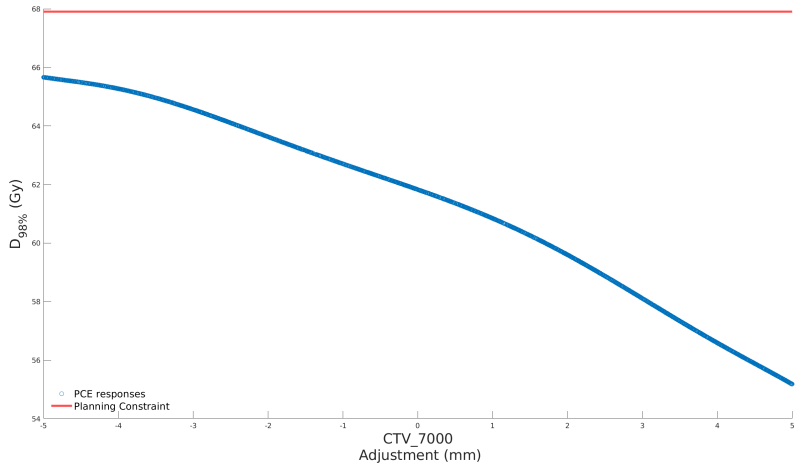


Figure 4.10.: PCE response of the $D_{98\%}$ as a function of CTV expansion/reduction.

Figure 4.11 shows the width of the 95% and 98% confidence intervals of $D_{98\%}$ as a function of the uncertainty in the CTV for all patients. There is an expected correlation between the increasing uncertainty in the CTV and the increasing width of the confidence intervals. As the uncertainty in the CTV increases, as modeled by the larger uniform distribution from which the radius of the rolling ball can be drawn, the width of any confidence interval increases.

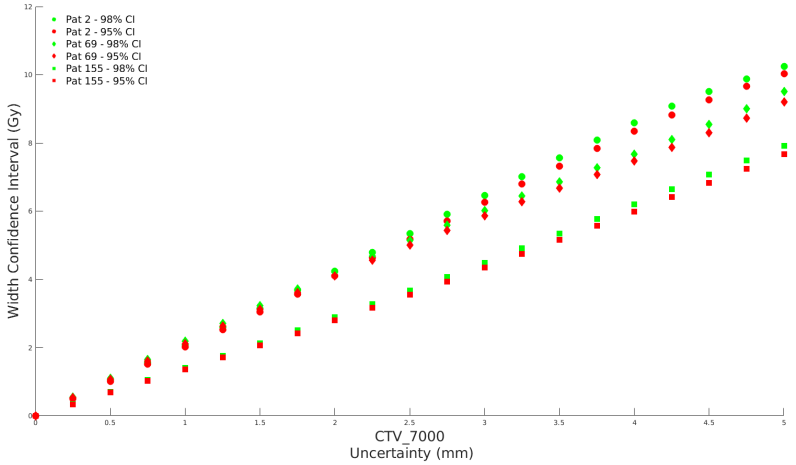


Figure 4.11.: Bandwidths of the 95% and 98% confidence intervals of the $D_{98\%}$ for different combinations of uniform distributions for all three patients

Table 4.7 shows how large the uncertainty in the CTV delineation can be without exceeding a specified width of the 95% or 98% confidence intervals.

95% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.5	0.25	0	0.25
1.0	0.25	0.25	0.5
1.5	0.50	0.5	1
2.0	0.75	0.75	1.25
98% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.5	0	0	0.25
1.0	0.25	0.25	0.5
1.5	0.5	0.5	1
2.0	0.75	0.75	1.25

Table 4.7.: Maximum delineation uncertainty of CTV that does not exceed a specific width of the confidence interval in the $D_{98\%}$ for each patient.

4.3.2. BRAINSTEM

Similarly to the previous section, 100,000 different brain stems have been made by drawing a radius from the uniform distribution [-2 mm, 2 mm] and resulted in an equivalent amount of different DVH's. Figure 4.12 shows different confidence intervals for the DVH's for patient 2. In addition, the 2% volume fraction and the planning constraint have been highlighted. The figure shows that there are some situations in which a part of the brain stem receives a higher dose than the planning constraint. Since this dose distribution has been optimized for the original delineation, it is possible that a 2 mm expansion of the brain stem expands the volume to a region where the dose exceeds the planning constraint as shown in Figure 4.12.

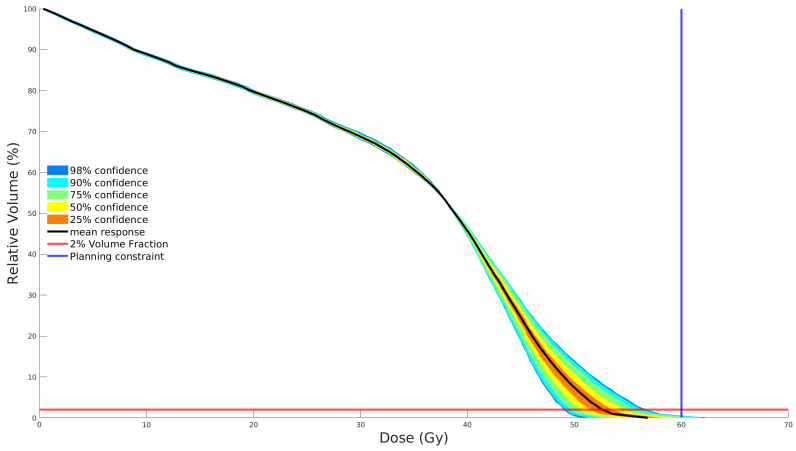


Figure 4.12.: Overview of the different confidence intervals for the 100,000 DVH's that have been created for patient 2 for a fixed dose distribution.

Figure 4.13 shows the histogram of the dose delivered to 2% of the volume of the brain stem. The figure also shows the planning constraint and there are no values at the planning constraint. Figure 4.12 already showed that there are no situations in which $D_{2\%}$ is equal to or greater than the planning constraint. Just as with the CTV the 95% and 98% confidence interval covers a wide range.

Figure 4.14 shows the behavior of the PCE response of $D_{2\%}$ as the size of the brain stem delineation changes. The correlation between $D_{2\%}$ and the adjustment of the brain stem is as expected. As the brain stem expands, it will expand to higher-dose regions, increasing $D_{2\%}$. The opposite is also true; as the brain stem shrinks in volume, it shrinks into lower dose regions, resulting in a decrease in $D_{2\%}$.

Figure 4.15 shows the width of the 95% and 98% confidence intervals of $D_{2\%}$ as a function of the uncertainty in the brain stem for all patients. There is an expected correlation between the increasing uncertainty in the brain stem and the increasing width of the confidence intervals. As the uncertainty in the brain stem increases, the width of any confidence interval increases.

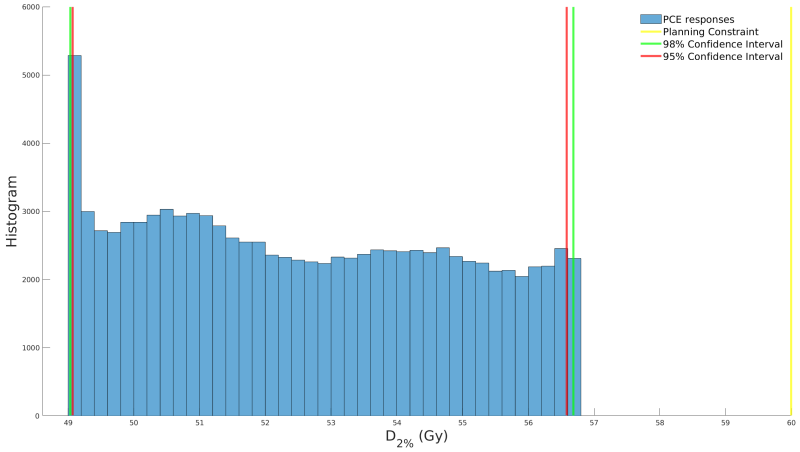


Figure 4.13.: Histogram for $D_{2\%}$ of patient 2 for a fixed dose distribution.

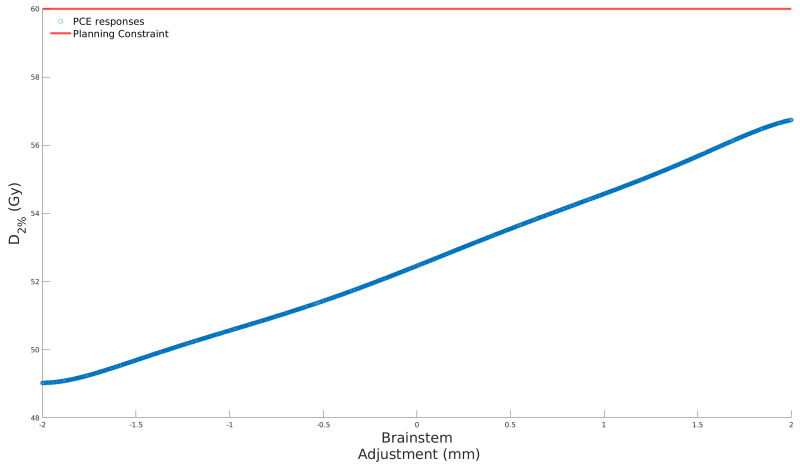


Figure 4.14.: PCE response of the $D_{2\%}$ as a function of brain stem expansion/reduction.

Table 4.8 shows how large the uncertainty in the delineation of the brain stem can be without exceeding a specified width of the 95% or 98% confidence intervals.

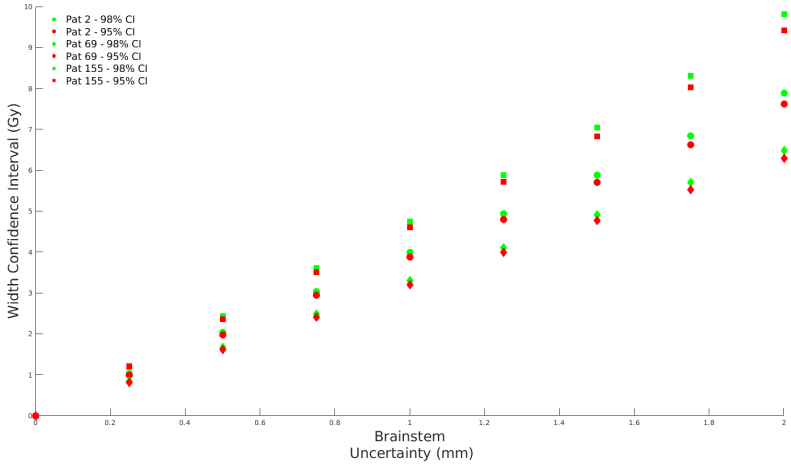


Figure 4.15.: Bandwidths of the 95% and 98% confidence intervals of the $D_{2\%}$ for different uniform distributions for all patients

95% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.5	0	0	0
1.0	0.25	0.25	0
1.5	0.25	0.25	0.25
2.0	0.50	0.50	0.25
98% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.5	0	0	0
1.0	0	0.25	0
1.5	0.25	0.25	0.25
2.0	0.25	0.50	0.25

Table 4.8.: Maximum delineation uncertainty of the brain stem to obtain a specific width of the confidence interval in the $D_{2\%}$ for each patient.

4.4. DOSIMETRIC EFFECT OF DELINEATION UNCERTAINTIES FOR A RE-OPTIMIZED DOSE DISTRIBUTION

This section describes the dosimetric effect of the uncertainty of the delineation for re-optimized dose distributions. The dosimetric effects on CTV will be covered in Section 4.4.1, while dosimetric effects on the brain stem will be covered in Section 4.4.2. Other regions are not discussed due to the inaccuracy of the PCE response, as shown in Section

4.2.2.

As mentioned in section 3.4.1 a total of 189 cases will be investigated. This covers all combinations of the 21 CTV cases and 9 brain stem cases. Each case will contain 100,000 random scenarios. The figures of the other patients can be found in Appendix B.

4.4.1. CTV

Figure 4.16 shows the confidence intervals for the PCE responses of the 100,000 random scenarios. For each random scenario, the radius of the rolling ball was drawn from the largest possible uniform distribution of the CTV and the brain stem, which are [-5 mm, 5 mm] and [-2 mm, 2 mm], respectively. When comparing the DVH's for a re-optimized dose distribution of figure 4.16 with the DVHs for a fixed dose distribution shown in figure 4.8, it becomes noticeable that after re-optimization a larger volume fraction receives a dose equal to or greater than the planning constraint/objective. Furthermore, the confidence intervals are slimmer after re-optimization compared to the fixed dose distribution.

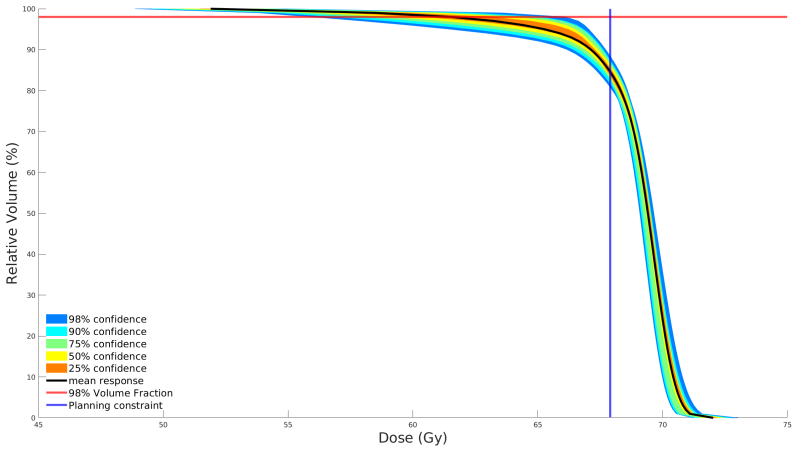


Figure 4.16.: The confidence intervals for the DVH's that have been generated for a uniform distribution of [-5 mm, 5 mm] for the CTV and [-2 mm, 2 mm] for the brain stem.

Figure 4.17 shows the histogram for $D_{98\%}$ for a re-opimized dose distribution. This is for the 100,000 scenarios generated using the largest possible uniform distribution of the CTV and the brain stem. It also shows that the planning constraint/objective falls outside the 95% and 98% confidence intervals but that is not surprising since Figure 4.16 has already shown that there is no scenario in which the $D_{98\%}$ is equal to or greater than the planning constraint/objective. This can be explained by the proximity of the brain stem to the CTV, which increases the volume fraction of the CTV that is close to an OAR. For this volume fraction, there is only a planning objective instead of a constraint, and

an objective does not have to be met.

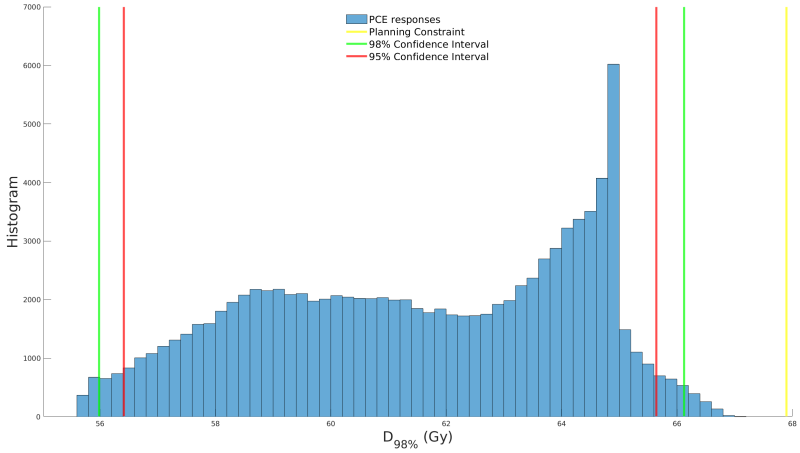


Figure 4.17.: Histogram for $D_{98\%}$ of patient 2. Also shows the planning constraint/objective and the 95% and 98% confidence intervals.

Figure 4.18 shows the behavior of the PCE response of $D_{98\%}$ as a function of CTV and expansion or shrinkage of the brain stem. The general correlation between $D_{98\%}$ and the adjustment of the CTV is similar to the situation for a fixed dose distribution. As the CTV expands, $D_{98\%}$ decreases and as the CTV shrinks, $D_{98\%}$ increases. There is a similar correlation between $D_{98\%}$ and brain stem adjustment.

Both correlations can be explained by the proximity of the brain stem to the CTV. The planning constraint on the brain stem is 60 Gy, which is less than the planning constraint/objective on the CTV, which is set at 67.9 Gy. This puts pressure on the dose delivered to voxels that are near the CTV boundary and close to the brain stem. Any expansion of the CTV and the brain stem decreases the distance of the voxels near the boundary of the CTV and the brain stem. Furthermore, as the distance between the two regions decreases, the volume fraction of the CTV near an OAR increases. The planning constraint becomes an objective for this volume fraction, and the objectives can fail, unlike constraints.

Figure 4.19 shows the width of the 95% and 98% confidence intervals of $D_{98\%}$ as a function of uncertainty in the CTV and the brain stem. There is an expected correlation between the increasing uncertainty in either of the two volumes and the increasing width of the confidence intervals.

Table 4.9 shows the largest combined uncertainty in CTV and in the brain stem without exceeding a specific width of the confidence interval for $D_{98\%}$. There is a clear difference between the largest combined uncertainties in patient 155 compared to the other two patients. Although the CTV of patient 155 is smaller than the CTVs of the other two patients as shown in Figure 3.1, the additional volume of CTV should have a major effect on the spinal cord and vice versa. It is unexpected that the sensitivity to uncertainty is so

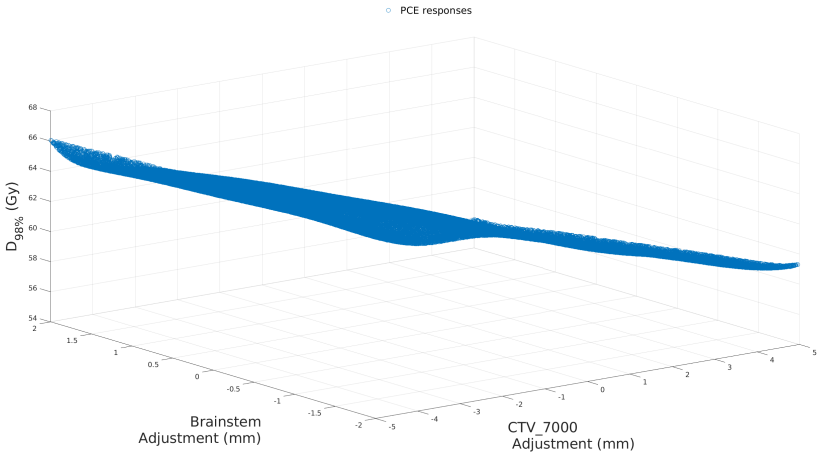


Figure 4.18.: PCE response of the $D_{98\%}$ as a function of both CTV and brain stem expansion/reduction

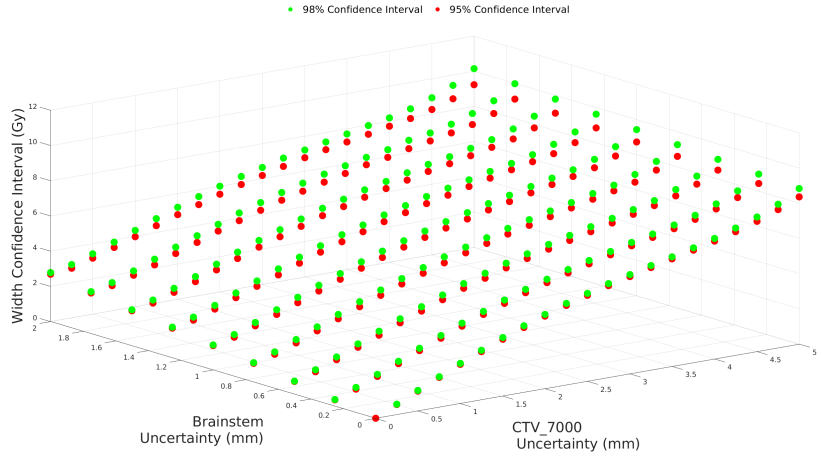


Figure 4.19.: Bandwidths of the 95% and 98% confidence intervals of the $D_{98\%}$ for different combinations of uniform distributions

low for the CTV of patient 155.

Although Table 4.9 shows the maximum combined uncertainty, that does not mean that the uncertainty for a single volume is allowed to be that great. Table 4.10 shows the highest uncertainty for a specific volume that does not exceed the specified width of the confidence intervals. Then it shows how large the uncertainty in the other volume

95% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.5	0.25	0.25	2.25
1.0	0.5	0.5	3.25
1.5	1	1	4.25
2.0	1.5	1.25	4.75
98% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.5	0.25	0.25	2
1.0	0.5	0.5	3
1.5	1	1	3.75
2.0	1.25	1.25	4.5

Table 4.9.: Maximum combined delineation uncertainty of CTV and brain stem that does not exceed a specific width of the confidence interval in the $D_{98\%}$ for each patient.

could be without exceeding the specified width of the confidence interval. The gray rows show the greatest uncertainty allowed in the CTV. Then, given that uncertainty, what the maximum allowed uncertainty in the brain stem could be. The white rows show the largest uncertainty allowed in the brain stem. Then, given that uncertainty, what the greatest allowed uncertainty in the CTV could be.

4.4.2. BRAINSTEM

Figure 4.20 shows the confidence intervals for the PCE responses of the 100,000 random scenarios. For each random scenario, the radius of the rolling ball was drawn from the largest possible uniform distribution of the CTV and the brain stem, which were [-5 mm, 5 mm] and [-2 mm, 2 mm], respectively. The figure also contains the planning constraint of a maximum of 60 Gy. Due to re-optimization, every scenario complies with the dose constraint, unlike the scenarios with the fixed dose distribution.

Figure 4.21 shows the histogram for $D_{2\%}$ for a re-optimized dose distribution. This is for the 100,000 scenarios generated using the largest possible uniform distribution of the CTV and the brain stem. It also shows that the planning constraint falls outside the 95% and 98% confidence intervals but that is not surprising since figure 4.20 has already shown that there is no scenario in which the $D_{2\%}$ is equal to or greater than the planning constraint.

Figure 4.22 shows the width of the 95% and 98% confidence intervals of $D_{2\%}$ as a function of uncertainty in the CTV and brain stem. There is an expected correlation between the increasing uncertainty in either of the two volumes and the increasing width of the confidence intervals.

Table 4.11 shows the largest combined uncertainty in CTV and in the brain stem that does not exceed a specific width of the confidence interval for $D_{2\%}$. There is a lot of

Confidence Interval 95%	Patient 2		Patient 69		Patient 155	
Bandwidth (Gy)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)
0.5	0.00	0.25	0.25	0.00	1.50	0.75
0.5	0.00	0.25	0.00	0.25	0.25	1.50
1.0	0.25	0.25	0.50	0.00	2.75	0.25
1.0	0.00	0.50	0.00	0.50	1.00	2.00
1.5	0.50	0.25	1.00	0.00	3.50	0.50
1.5	0.00	1.00	0.00	0.75	2.00	2.00
2.0	0.75	0.25	1.25	0.00	4.25	0.25
2.0	0.00	1.50	0.25	1.00	2.75	2.00
Confidence Interval 98%	Patient 2		Patient 69		Patient 155	
Bandwidth (Gy)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)
0.5	0.00	0.25	0.25	0.00	1.50	0.50
0.5	0.00	0.25	0.25	0.00	0.50	1.25
1.0	0.25	0.25	0.50	0.00	2.50	0.50
1.0	0.00	0.50	0.25	0.25	0.75	2.00
1.5	0.50	0.25	0.75	0.25	3.25	0.50
1.5	0.00	1.00	0.00	0.75	1.75	2.00
2.0	0.75	0.25	1.25	0.00	4.00	0.25
2.0	0.00	1.25	0.00	1.00	2.25	2.00

Table 4.10.: Combinations of maximum uncertainties that will not exceed a specified bandwidth of a confidence interval for $D_{98\%}$. The gray rows maximized the uncertainty in the CTV and then in the brain stem. The white rows maximized the uncertainty in the brain stem and then in the CTV.

variability between the different patients.

Although Table 4.11 shows the maximum combined uncertainty, that does not mean that the uncertainty for a single volume is allowed to be that great. Table 4.12 shows the largest uncertainty for a specific volume that does not exceed the specified width of the confidence intervals. Then it shows how large the uncertainty in the other volume could be without exceeding the specified width of the confidence interval. The gray rows show the greatest uncertainty allowed in the CTV. Then, given that uncertainty, what the maximum allowed uncertainty in the brain stem could be. The white rows show the largest uncertainty allowed in the brain stem. Then, given that uncertainty, what the greatest allowed uncertainty in the CTV could be.

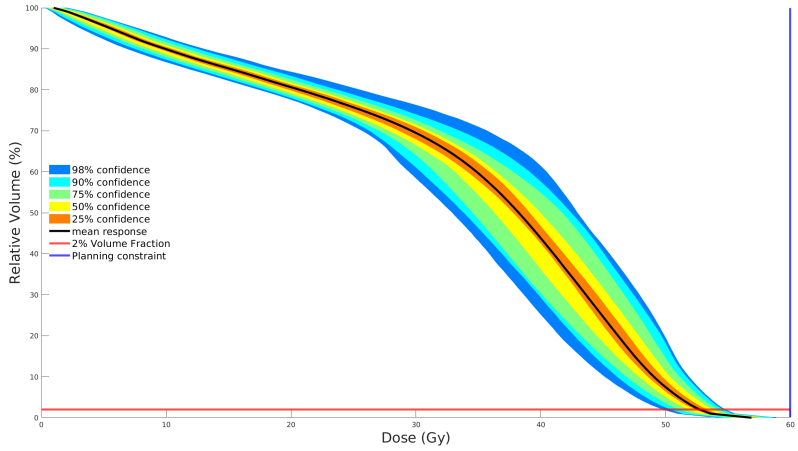


Figure 4.20.: The confidence intervals for the DVH's that have been generated for a uniform distribution of [-5 mm, 5 mm] for the CTV and [-2 mm, 2 mm] for the brain stem.

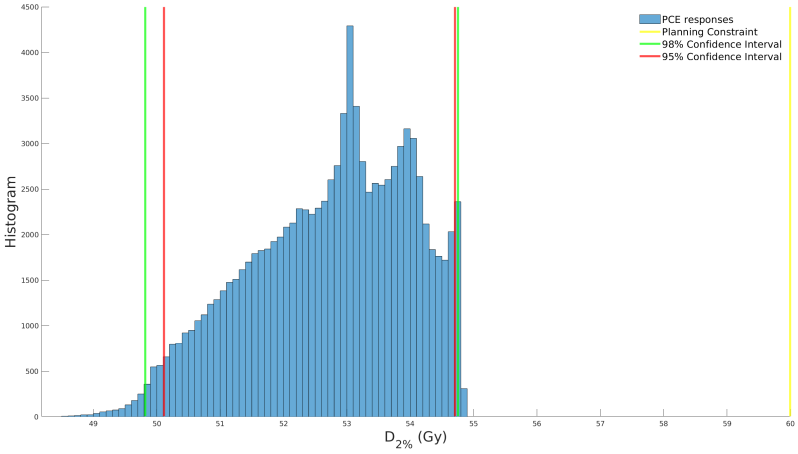


Figure 4.21.: Histogram for $D_{2\%}$ of patient 2. Also shows the planning constraint and the 95% and 98% confidence intervals.

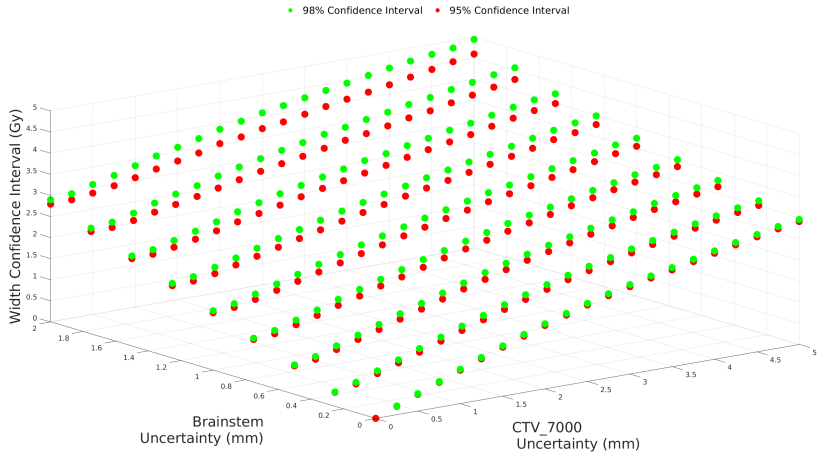


Figure 4.22.: Bandwidths of the 95% and 98% confidence intervals of the $D_{2\%}$ for different combinations of uniform distributions

95% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.50	0.50	1.00	0.25
1.00	1.25	2.00	0.50
1.50	2.00	3.00	0.75
2.00	2.75	4.75	1.25
98% Confidence Interval			
Bandwidth (Gy)	Patient 2 (mm)	Patient 69 (mm)	Patient 155 (mm)
0.50	0.50	0.75	0.25
1.00	1.00	1.75	0.50
1.50	1.75	2.75	0.75
2.00	2.75	4.50	1.00

Table 4.11.: Maximum combined delineation uncertainty of CTV and brain stem that does not exceed a specific width of the confidence interval in the $D_{2\%}$ for each patient.

Confidence Interval 95%	Patient 2		Patient 69		Patient 155	
Bandwidth (Gy)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)
0.5	0.50	0.00	0.75	0.25	0.25	0.00
0.5	0.25	0.25	0.00	0.75	0.25	0.00
1.0	1.25	0.00	1.75	0.00	0.50	0.00
1.0	0.50	0.50	0.75	1.00	0.25	0.25
1.5	1.75	0.25	2.75	0.25	0.75	0.00
1.5	0.25	1.00	0.50	1.50	0.25	0.50
2.0	2.75	0.00	4.75	0.00	1.00	0.25
2.0	0.75	1.25	0.75	1.75	0.25	0.75
Confidence Interval 98%	Patient 2		Patient 69		Patient 155	
Bandwidth (Gy)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)	R_{CTV} (mm)	$R_{brainstem}$ (mm)
0.5	0.50	0.00	0.75	0.00	0.25	0.00
0.5	0.25	0.25	0.25	0.50	0.25	0.00
1.0	1.00	0.00	1.75	0.00	0.50	0.00
1.0	0.50	0.50	0.50	1.00	0.25	0.25
1.5	1.75	0.00	2.75	0.00	0.75	0.00
1.5	0.25	1.00	0.00	1.50	0.25	0.50
2.0	2.75	0.00	4.50	0.00	1.00	0.00
2.0	0.50	1.25	0.25	1.75	0.00	0.75

Table 4.12.: The combinations of maximum uncertainties that will not exceed a specified bandwidth of a confidence interval for $D_{2\%}$. The gray rows maximized the uncertainty in the CTV and then in the brain stem. The white rows maximized the uncertainty in the brain stem and then in the CTV.

5

DISCUSSION

5.1. PCE VALIDATION

The discussion for the PCE validation will focus on the results of the validation for the re-optimized dose distribution. As stated above, the model was found to contain too large errors in selected ranges of the DVH and the dose distribution. The results will be compared with the results of Perko et al. [49] and Rojo-Santiago et al. [52].

Both studies have used PCE to model dose distributions in head and neck patients. There were 4 input variables, 3 for setup errors and 1 for range error. All input variables had a Gaussian distribution with a mean of zero. The largest standard deviation used was approximately 2.8 mm for setup errors and 2% for range error. The model used in this thesis had two input variables and each input variable represented the uncertainty in either the CTV delineation or the brain stem delineation. The variables were uniformly distributed over a range of [-5 mm, 5 mm] for the CTV and [-2 mm, 2 mm] for the brain stem.

One of the studies reported a voxel acceptance of at least 80% for all scenarios [49]. The voxel acceptance in this thesis was at most 40% for a single scenario.

DVHs for the brain stem [52] and for the spinal cord [49] showed a large overlap between the PCE DVH and the real DVH or at the very least a similar behavior. For this thesis, this was only true in a subset of the scenarios. Scenarios with the largest error were shown, and possible inaccuracies were large.

The $D_{98\%}$ for the CTV has been reported and showed that all robust plans delivered at least 95% of the prescribed dose. For the non-robust plan, there was a small probability that $D_{98\%}$ was smaller than 95%. For two patients in this thesis, the probability was almost 100% that $D_{98\%}$ would be less than 95% of the prescribed dose and approximately 75% chance for the remaining patient to receive less than 95% of the prescribed dose for uncertainties up to 5 mm. The caveat is that a mistake was found that was made during the re-optimizations. In the last step, spots that are below a MU threshold are thrown away, and then the remaining spots are re-optimized for the wish-list. The spots were thrown away, but re-optimization did not happen. This could be a reasonable explanation for the almost guaranteed under dosing of the CTV.

Overall, the PCE model showed poor accuracy compared to the other two studies. The main difference between this thesis and those two studies is the type of error sources

that are being investigated. Setup errors and range errors do not change the delineation, whereas the input variables used in this thesis actively change the delineations. Depending on the magnitude of the change, the resulting structure set could be considered a completely different patient, which can result in a different set of beams. It could be compared to building a single PCE model using the responses of multiple different patients.

The last factor to consider is of course the mistake of missing the last re-optimization step. It is currently unclear how much of an impact this has on the final plan and, therefore, how optimal the plans are. It is possible that the last re-optimization step could make a noticeable difference in the large error scenarios.

5.2. DOSIMETRIC EFFECTS

Aliotta et al. [36] have researched the dosimetric impact on organs at risk, which was reported as the probability of achieving a plan metric such as D_{max}. Their paper was focused on both delineation uncertainty and setup uncertainty, but they have included results for delineation uncertainty only. The plan quality metric for the brain stem was a D_{max} of 54 Gy, which is less than 60 Gy for this thesis. The delineation uncertainty was modeled for three different initial magnitudes of random anisotropic adjustments of the delineation, however, the largest magnitude was considered unrealistic by the authors and therefore the results will be excluded. The remaining initial magnitudes were 2 mm and 5 mm. Although these were the initial magnitudes, the actual median Hausdorff distances were approximately 5 mm and 7.5 mm. For the smallest magnitudes, there was no impact on achieving their plan metric, and for the medium magnitude there was a 14% chance of not achieving their plan metric. Furthermore, the initial D_{max} was around 50 Gy for the nominal situation, increased to at most 52 Gy for the small magnitude and up to 62 Gy for the medium magnitude.

Although this thesis evaluated $D_{2\%}$ and used an isotropic adjustment to the contour, there is still a noticeable difference. A 2 mm expansion of the brain stem for a fixed dose distribution increased the $D_{2\%}$ by approximately 3-5 Gy depending on the patient. In a re-optimized dose distribution the difference was of a smaller magnitude depending on the distance between the brain stem and the CTV. If the confidence interval is extended to 2.0 Gy, the largest possible uncertainty on the brain stem would be 1.75 mm.

There can be several reasons as to why the sensitivty of the brain stem in their research is lower than in this thesis. Main reason is that although the reported Hausdorff distance is at least 5 mm, the corresponding DSC is 0.88 which indicates a large degree of similarity between the nominal delineation and the distorted delineation. In addition their tumor was not located near the brain stem as it is for the patients in this thesis. Their results also indicated that OARs next to tumors were more sensitive to uncertainties than those that are not [36].

6

CONCLUSION

This chapter will cover the conclusions of the research. After the conclusions are reached, there will be a section that covers the recommendations for future research.

6.1. PCE VALIDATION

Different PCE models have been constructed and validated with evaluation scenarios. Comparisons were made for DVH of the CTV, brain stem, and spinal cord. Additionally, the dose distribution was included for re-optimized dose distributions.

For a fixed dose distribution, the model was constructed with GO6PO7 and for a re-optimized dose distribution with GO5PO6. The model for a fixed dose distribution had small errors for the CTV and the brain stem. The largest error was 0.27 Gy in $D_{2\%}$ of the brain stem. The DVHs of the CTV and brain stem showed great agreement even for the scenarios with the largest error between the PCE DVH and the real DVH.

The model for a re-optimized dose distribution had limited validity due to the magnitude of errors. In the end, only $D_{98\%}$ from the CTV and $D_{2\%}$ from the brain stem were used to determine the dosimetric effects. The DVHs of the CTV and brain stem showed limited agreement between PCE DVH and real DVH. The comparison of DVHs of the spinal cord was inaccurate. The dose distribution had at most 40% voxel acceptance for a threshold of 0.1 Gy for a small number of scenarios.

To check a possible reason for this inaccuracy, the same patient, structure set, and wish-list has been re-optimized several times to see what potential noise the re-optimization could have. This resulted in a comparison between four different dose distributions. Two of those distributions were very similar and had a small mean dose difference of 0.01 Gy. However, any other comparison resulted in large mean dose differences between 0.69 Gy and 0.83 Gy. These mean dose differences are greater than most of the errors found in the PCE responses of interest.

In addition, a mistake has been found in the re-optimizations during the last step. In the last step, spots are thrown away if they are below a certain monitor unit threshold, and then the remaining spots are re-optimized. The last re-optimization did not take place and this could be the reason why for two of the three patients the CTV is underdosed regardless of the uncertainty in the delineation.

6.2. DOSIMETRIC EFFECTS

Dosimetric effects have been investigated by calculating the width of the 98% and the 95% confidence intervals for $D_{98\%}$ of the CTV and $D_{2\%}$ of the brain stem for different values of the radius R that have been drawn from uniform distributions ranging from [0 mm, 0 mm] to [-5 mm, 5mm] and [-2 mm, 2mm] for the CTV and the brain stem, respectively. This resulted in 21 different distributions for the CTV and 9 different distributions for the brain stem.

For fixed dose distributions, these distributions have been used in isolation for a total of 30 different cases. For re-optimized dose distribution, all the different combinations have been used for a total of 189 different cases. For each case 100,000 samples were drawn. DVHs, histograms, and confidence intervals have been plotted.

For both situations, an increase in uncertainty resulted in a larger confidence interval and therefore a larger uncertainty in the $D_{98\%}$ in the CTV and the $D_{2\%}$ in the brain stem.

For a fixed dose distribution, it was found that the width of the 98% confidence interval in the $D_{98\%}$ in the CTV would increase to approximately 2 Gy for a uncertainty in the delineation of the CTV between 0.75 mm and 1.25 mm depending on the patient.

The width of the 98% confidence interval in the $D_{2\%}$ in the brain stem would increase to approximately 2 Gy for a delineation uncertainty in the brain stem between 0.25 mm and 0.5 mm.

For a re-optimized distribution, it was found that the width of the 98% confidence interval in the $D_{98\%}$ in the CTV would increase to approximately 2 Gy for a combined uncertainty of 1.25 mm for two patients and 4.5 mm for one patient. For two patients, the dose uncertainty was more sensitive to delineation uncertainties in the brain stem, and for one patient, it was more sensitive to delineation uncertainties in the CTV.

The width of the 98% confidence interval in the $D_{2\%}$ in the brain stem would increase to approximately 2 Gy for a combined delineation uncertainty between 1.0 mm and 4.5 mm depending on the patient. For all three patients, the dose uncertainty was more sensitive to delineation uncertainty in the brain stem than the delineation uncertainty in the CTV.

The effect of these uncertainties will depend on the nominal situation. For all patients, a dose uncertainty of 2 Gy corresponds to approximately 2.85% of the prescribed dose for the CTV and approximately 3.33% of the maximum dose constraint for the brain stem. Two of the patients were already under dosed in the nominal situation, the remaining patient was near the boundary of under dosing. A 2 Gy dose uncertainty would most likely result in under dosing the CTV. The $D_{2\%}$ in the brain stem for the three patients was around 49 Gy to 52 Gy. A 2 Gy dose uncertainty is unlikely to result in exceeding the maximum dose constraint on the brain stem in these patients.

6.3. LIMITATIONS & RECOMMENDATIONS

A clear limitation of these findings is that the range of uniform distributions increased in steps of 0.25 mm. Smaller steps would increase the total number of cases that have to be investigated but would allow for more accurate results. It is likely that the widths of the confidence intervals could not be limited to 0.5 Gy due to the step size of 0.25 mm in the uniform distributions.

There are also several limitations to the method of modeling uncertainty. The rolling ball algorithm that is used in this is only applied in 2D. Therefore, it is more of a rolling circle. The results might be different for an actual rolling ball applied in 3D. However, even with this change, there is an inherent limitation in modeling uncertainty with a single radius applied to an entire volume. There can be different amounts of uncertainty for different parts of a volume, as shown when comparing the delineation of different observers [36] and the improved results of anisotropic margins [33]. A more realistic variation of the rolling ball would require a variable radius that depends on the current location of the ball. An obstacle that needs to be overcome would be for the algorithm to recognize parts of the contour and the radius that needs to be applied at that location. An additional obstacle would be that the adjustments to each point of the contour are likely not independent of each other. The adjustment of a point will most likely be influenced and dependent on the adjustment of neighboring points.

Another modification that might increase the range of uncertainties that could be modeled would be to split the radius of the ball into components. An obvious split would be into x-, y-, and z-components.

In addition to using the rolling ball, it might also be interesting to incorporate shifts and rotations. Theoretically, this can be done with a variable rolling ball, but implementation is probably easier with shifts and rotations.

Another problem found in this thesis was modeling the DVH of the spinal cord. Although there were some problems in modeling the entire DVH of the CTV and brain stem, there was still sufficient accuracy in two metrics of interest. However, there was not enough accuracy for the spinal cord. one of the differences between the spinal cord and the brain stem was that the spinal cord was not adjusted. Although the dose distribution within the spinal cord is not independent of changes in brain stem volume and CTV, it is likely that the dependence on direct changes in the spinal cord is stronger. Finding out whether the inclusion of a volume in the construction of a PCE model matters for the accuracy of the PCE responses for metrics in that volume might be interesting.

An additional problem was the modeling of PCE for the entire dose distribution. Models were considered invalid due to poor accuracy. A possible cause might be that the ranges of the uniform distributions were too large. Perhaps, smaller deviations might lead to a more accurate model, since the structure sets will not deviate from each other by such a large margin.

A different aspect to consider would be the acceptable dose range for a voxel. The CTV was more accurate than the other volumes and had a relatively narrow dose range for its voxels, as it had both a maximum and partial minimum dose constraint. This could be investigated by lowering the maximum dose constraint on OARs to narrow the acceptable dose range for a voxel in those volumes.

Last recommendation is related to auto-contouring. This study was limited to looking at the dosimetric effects of contouring uncertainties. However an interesting recommendation would be to look at the uncertainty propagation and the dosimetric effect of that propagation. These uncertainties can also be present in the images that are used as an atlas for auto-contouring. How does that uncertainty propagate, and what is the dosimetric effect of the initial uncertainty. The software itself might also have an uncertainty, what is the interaction between the two uncertainties and what is the magnitude of the

final uncertainty. What would happen if multiple different atlases are used, would that lessen or worsen the final uncertainty. There are a lot of interactions and consequences of uncertainties and auto-contouring that should be investigated.

BIBLIOGRAPHY

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray. 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries'. In: *CA: A Cancer Journal for Clinicians* 71.3 (May 2021), pp. 209–249. ISSN: 0007-9235, 1542-4863. DOI: 10.3322/caac.21660.

[2] D. E. Johnson, B. Burtness, C. R. Leemans, V. W. Y. Lui, J. E. Bauman and J. R. Grandis. 'Head and Neck Squamous Cell Carcinoma'. In: *Nature Reviews Disease Primers* 6.1 (Nov. 2020), p. 92. ISSN: 2056-676X. DOI: 10.1038/s41572-020-00224-3.

[3] M. Mukesh, R. Benson, R. Jena, A. Hoole, T. Roques, C. Scrase, C. Martin, G. A. Whitfield, J. Gemmill and S. Jefferies. 'Interobserver Variation in Clinical Target Volume and Organs at Risk Segmentation in Post-Parotidectomy Radiotherapy: Can Segmentation Protocols Help?' In: *The British Journal of Radiology* 85.1016 (Aug. 2012), e530–e536. ISSN: 0007-1285, 1748-880X. DOI: 10.1259/bjr/66693547.

[4] S. Sadeghi, Z. Siavashpour, A. Vafaei Sadr, M. Farzin, R. Sharp and S. Gholami. 'A Rapid Review of Influential Factors and Appraised Solutions on Organ Delineation Uncertainties Reduction in Radiotherapy'. In: *Biomedical Physics & Engineering Express* 7.5 (Sept. 2021), p. 052001. ISSN: 2057-1976. DOI: 10.1088/2057-1976/ac14d0.

[5] *BraggPeak-en.Svg*. https://commons.wikimedia.org/wiki/File:BraggPeak-en.svg. Aug. 2022.

[6] V. Andrianarison, M. Laouiti, O. Fargier-Bochaton, G. Dipasquale, X. Wang, N. Nguyen, R. Miralbell and V. Vinh-Hung. 'Contouring Workload in Adjuvant Breast Cancer Radiotherapy'. In: *Cancer/Radiothérapie* 22.8 (Dec. 2018), pp. 747–753. ISSN: 12783218. DOI: 10.1016/j.canrad.2018.01.008.

[7] D. N. Teguh, P. C. Levendag, P. W. Voet, A. Al-Mamgani, X. Han, T. K. Wolf, L. S. Hibbard, P. Nowak, H. Akhiat, M. L. Dirkx, B. J. Heijmen and M. S. Hoogeman. 'Clinical Validation of Atlas-Based Auto-Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck'. In: *International Journal of Radiation Oncology*Biology*Physics* 81.4 (Nov. 2011), pp. 950–957. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2010.07.009.

[8] A. Isambert, F. Dhermain, F. Bidault, O. Commowick, P.-Y. Bondiau, G. Malandain and D. Lefkopoulos. 'Evaluation of an Atlas-Based Automatic Segmentation Software for the Delineation of Brain Organs at Risk in a Radiation Therapy Clinical Context'. In: *Radiotherapy and Oncology* 87.1 (Apr. 2008), pp. 93–99. ISSN: 01678140. DOI: 10.1016/j.radonc.2007.11.030.

[9] B. Schipaanboord, D. Boukerroui, D. Peressutti, J. van Soest, T. Lustberg, T. Kadir, A. Dekker, W. van Elmpt and M. Gooding. 'Can Atlas-Based Auto-Segmentation Ever Be Perfect? Insights From Extreme Value Theory'. In: *IEEE Transactions on Medical Imaging* 38.1 (Jan. 2019), pp. 99–106. ISSN: 0278-0062, 1558-254X. DOI: 10.1109/TMI.2018.2856464.

[10] C. K. Glide-Hurst, P. Lee, A. D. Yock, J. R. Olsen, M. Cao, F. Siddiqui, W. Parker, A. Doemer, Y. Rong, A. U. Kishan, S. H. Benedict, X. A. Li, B. A. Erickson, J. W. Sohn, Y. Xiao and E. Wuthrick. 'Adaptive Radiation Therapy (ART) Strategies and Technical Considerations: A State of the ART Review From NRG Oncology'. In: *International Journal of Radiation Oncology*Biology*Physics* 109.4 (Mar. 2021), pp. 1054–1075. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2020.10.021.

[11] C. Njeh. 'Tumor Delineation: The Weakest Link in the Search for Accuracy in Radiotherapy'. In: *Journal of Medical Physics* 33.4 (2008), p. 136. ISSN: 0971-6203. DOI: 10.4103/0971-6203.44472.

[12] B. Segedin and P. Petric. 'Uncertainties in Target Volume Delineation in Radiotherapy Are They Relevant and What Can We Do about Them?' In: *Radiology and Oncology* 50.3 (Sept. 2016), pp. 254–262. ISSN: 1581-3207. DOI: 10.1515/raon-2016-0023.

[13] S. K. Vinod, M. Min, M. G. Jameson and L. C. Holloway. 'A Review of Interventions to Reduce Inter-Observer Variability in Volume Delineation in Radiation Oncology'. In: *Journal of Medical Imaging and Radiation Oncology* 60.3 (June 2016), pp. 393–406. ISSN: 17549477. DOI: 10.1111/1754-9485.12462.

[14] D. B. Eekers, L. in 't Ven, E. Roelofs, A. Postma, C. Alapetite, N. G. Burnet, V. Calugaru, I. Compter, I. E. Coremans, M. Høyer, M. Lambrecht, P. W. Nyström, A. Méndez Romero, F. Paulsen, A. Perpar, D. de Ruysscher, L. Renard, B. Timmermann, P. Vitek, D. C. Weber, H. L. van der Weide, G. A. Whitfield, R. Wiggenraad and E. G. Troost. 'The EPTN Consensus-Based Atlas for CT- and MR-based Contouring in Neuro-Oncology'. In: *Radiotherapy and Oncology* 128.1 (July 2018), pp. 37–43. ISSN: 01678140. DOI: 10.1016/j.radonc.2017.12.013.

[15] D. B. Eekers, D. Di Perri, E. Roelofs, A. Postma, J. Dijkstra, T. Ajithkumar, C. Alapetite, M. Blomstrand, N. G. Burnet, V. Calugaru, I. Compter, I. E. Coremans, S. Harrabi, A. Iannalfi, Y. L. Klaver, M. Lambrecht, A. M. Romero, F. Paulsen, B. Timmermann, P. Vitek, H. L. van der Weide, G. A. Whitfield, P. W. Nyström, J. Zindler, D. de Ruysscher, J. Langendijk, D. C. Weber and E. G. Troost. 'Update of the EPTN Atlas for CT- and MR-based Contouring in Neuro-Oncology'. In: *Radiotherapy and Oncology* 160 (July 2021), pp. 259–265. ISSN: 01678140. DOI: 10.1016/j.radonc.2021.05.013.

[16] E. Weiss and C. F. Hess. 'The Impact of Gross Tumor Volume (GTV) and Clinical Target Volume (CTV) Definition on the Total Accuracy in Radiotherapy: Theoretical Aspects and Practical Experiences'. In: *Strahlentherapie und Onkologie* 179.1 (Jan. 2003), pp. 21–30. ISSN: 0179-7158. DOI: 10.1007/s00066-003-0976-5.

[17] S. K. Vinod, M. G. Jameson, M. Min and L. C. Holloway. 'Uncertainties in Volume Delineation in Radiation Oncology: A Systematic Review and Recommendations for Future Studies'. In: *Radiotherapy and Oncology* 121.2 (Nov. 2016), pp. 169–179. ISSN: 01678140. DOI: 10.1016/j.radonc.2016.09.009.

[18] S. Warfield, K. Zou and W. Wells. 'Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation'. In: *IEEE Transactions on Medical Imaging* 23.7 (July 2004), pp. 903–921. ISSN: 0278-0062. DOI: 10.1109/TMI.2004.828354.

[19] J. Yang, W. A. Woodward, V. K. Reed, E. A. Strom, G. H. Perkins, W. Tereffe, T. A. Buchholz, L. Zhang, P. Balter, L. E. Court, X. A. Li and L. Dong. 'Statistical Modeling Approach to Quantitative Analysis of Interobserver Variability in Breast Contouring'. In: *International Journal of Radiation Oncology*Biology*Physics* 89.1 (May 2014), pp. 214–221. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2014.01.010.

[20] C.-J. Tao, J.-L. Yi, N.-Y. Chen, W. Ren, J. Cheng, S. Tung, L. Kong, S.-J. Lin, J.-J. Pan, G.-S. Zhang, J. Hu, Z.-Y. Qi, J. Ma, J.-D. Lu, D. Yan and Y. Sun. 'Multi-Subject Atlas-Based Auto-Segmentation Reduces Interobserver Variation and Improves Dosimetric Parameter Consistency for Organs at Risk in Nasopharyngeal Carcinoma: A Multi-Institution Clinical Study'. In: *Radiotherapy and Oncology* 115.3 (June 2015), pp. 407–411. ISSN: 01678140. DOI: 10.1016/j.radonc.2015.05.012.

[21] E. Lee, W. Park, S. H. Ahn, J. H. Cho, J. H. Kim, K. H. Cho, Y. M. Choi, J.-S. Kim, J. H. Kim, H.-S. Jang, Y.-S. Kim and T.-K. Nam. 'Interobserver Variation in Target Volume for Salvage Radiotherapy in Recurrent Prostate Cancer Patients after Radical Prostatectomy Using CT versus Combined CT and MRI: A Multicenter Study (KROG 13-11)'. In: *Radiation Oncology Journal* 36.1 (Mar. 2018), pp. 11–16. ISSN: 2234-1900, 2234-3156. DOI: 10.3857/roj.2017.00080.

[22] C. Weltens, J. Menten, M. Feron, E. Bellon, P. Demaerel, F. Maes, W. Van den Bogaert and E. van der Schueren. 'Interobserver Variations in Gross Tumor Volume Delineation of Brain Tumors on Computed Tomography and Impact of Magnetic Resonance Imaging'. In: *Radiotherapy and Oncology* 60.1 (July 2001), pp. 49–59. ISSN: 01678140. DOI: 10.1016/S0167-8140(01)00371-1.

[23] C. W. Hurkmans, J. H. Borger, B. R. Pieters, N. S. Russell, E. P. Jansen and B. J. Mijnheer. 'Variability in Target Volume Delineation on CT Scans of the Breast'. In: *International Journal of Radiation Oncology*Biology*Physics* 50.5 (Aug. 2001), pp. 1366–1372. ISSN: 03603016. DOI: 10.1016/S0360-3016(01)01635-2.

[24] M. Visser, D. Müller, R. van Duijn, M. Smits, N. Verburg, E. Hendriks, R. Nabuurs, J. Bot, R. Eijgelaar, M. Witte, M. van Herk, F. Barkhof, P. de Witt Hamer and J. de Munck. 'Inter-Rater Agreement in Glioma Segmentations on Longitudinal MRI'. In: *NeuroImage: Clinical* 22 (2019), p. 101727. ISSN: 22131582. DOI: 10.1016/j.nicl.2019.101727.

[25] I. Fotina, C. Lütgendorf-Caucig, M. Stock, R. Pötter and D. Georg. 'Critical Discussion of Evaluation Parameters for Inter-Observer Variability in Target Definition for Radiation Therapy'. In: *Strahlentherapie und Onkologie* 188.2 (Feb. 2012), pp. 160–167. ISSN: 0179-7158, 1439-099X. DOI: 10.1007/s00066-011-0027-6.

[26] G. Altorjai, I. Fotina, C. Lütgendorf-Caucig, M. Stock, R. Pötter, D. Georg and K. Dieckmann. 'Cone-Beam CT-Based Delineation of Stereotactic Lung Targets: The Influence of Image Modality and Target Size on Interobserver Variability'. In: *International Journal of Radiation Oncology*Biology*Physics* 82.2 (Feb. 2012), e265–e272. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2011.03.042.

[27] R. J. Steenbakkers, J. C. Duppen, I. Fitton, K. E. Deurloo, L. J. Zijp, E. F. Comans, A. L. Uitterhoeve, P. T. Rodrigus, G. W. Kramer, J. Bussink, K. De Jaeger, J. S. Belderbos, P. J. Nowak, M. van Herk and C. R. Rasch. 'Reduction of Observer Variation Using Matched CT-PET for Lung Cancer Delineation: A Three-Dimensional Analysis'. In: *International Journal of Radiation Oncology*Biology*Physics* 64.2 (Feb. 2006), pp. 435–448. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2005.06.034.

[28] M. V. Sherer, D. Lin, S. Elguindi, S. Duke, L.-T. Tan, J. Cacicedo, M. Dahele and E. F. Gillespie. 'Metrics to Evaluate the Performance of Auto-Segmentation for Radiation Treatment Planning: A Critical Review'. In: *Radiotherapy and Oncology* 160 (July 2021), pp. 185–191. ISSN: 01678140. DOI: 10.1016/j.radonc.2021.05.003.

[29] E. Cha, S. Elguindi, I. Onochie, D. Gorovets, J. O. Deasy, M. Zelefsky and E. F. Gillespie. 'Clinical Implementation of Deep Learning Contour Autosegmentation for Prostate Radiotherapy'. In: *Radiotherapy and Oncology* 159 (June 2021), pp. 1–7. ISSN: 01678140. DOI: 10.1016/j.radonc.2021.02.040.

[30] G. Sharp, K. D. Fritscher, V. Pekar, M. Peroni, N. Shusharina, H. Veeraraghavan and J. Yang. 'Vision 20/20: Perspectives on Automated Image Segmentation for Radiotherapy: Perspectives on Automated Image Segmentation for Radiotherapy'. In: *Medical Physics* 41.5 (Apr. 2014), p. 050902. ISSN: 00942405. DOI: 10.1118/1.4871620.

[31] M. van Herk, P. Remeijer, C. Rasch and J. V. Lebesque. 'The Probability of Correct Target Dosage: Dose-Population Histograms for Deriving Treatment Margins in Radiotherapy'. In: *International Journal of Radiation Oncology*Biology*Physics* 47.4 (July 2000), pp. 1121–1135. ISSN: 03603016. DOI: 10.1016/S0360-3016(00)00518-6.

[32] N. G. Burnet. 'Defining the Tumour and Target Volumes for Radiotherapy'. In: *Cancer Imaging* 4.2 (2004), pp. 153–161. ISSN: 1470-7330. DOI: 10.1102/1470-7330.2004.0054.

[33] L. R. Bell, E. M. Pogson, P. E. Metcalfe and L. C. Holloway. 'Defining and Assessing an Anisotropic Delineation Margin for Modern Radiotherapy: Defining an Anisotropic Delineation Margin for Modern Radiotherapy'. In: *Medical Physics* 43.12 (Nov. 2016), pp. 6644–6653. ISSN: 00942405. DOI: 10.1118/1.4967942.

[34] D. Bernstein, A. Taylor, S. Nill and U. Oelfke. 'New Target Volume Delineation and PTV Strategies to Further Personalise Radiotherapy'. In: *Physics in Medicine & Biology* 66.5 (Mar. 2021), p. 055024. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/1361-6560/abe029.

[35] H. Xu, J. J. Gordon and J. V. Siebers. 'Coverage-Based Treatment Planning to Accommodate Delineation Uncertainties in Prostate Cancer Treatment: Coverage-based Treatment Planning to Accommodate Delineation Uncertainties'. In: *Medical Physics* 42.9 (Aug. 2015), pp. 5435–5443. ISSN: 00942405. DOI: 10.1118/1.4928490.

[36] E. Aliotta, H. Nourzadeh and J. Siebers. 'Quantifying the Dosimetric Impact of Organ-at-Risk Delineation Variability in Head and Neck Radiation Therapy in the Context of Patient Setup Uncertainty'. In: *Physics in Medicine & Biology* 64.13 (July 2019), p. 135020. ISSN: 1361-6560. DOI: 10.1088/1361-6560/ab205c.

[37] M. P. Langer, L. Papiez, S. Spirydovich and V. Thai. 'The Need for Rotational Margins in Intensity-Modulated Radiotherapy and a New Method for Planning Target Volume Design'. In: *International Journal of Radiation Oncology*Biology*Physics* 63.5 (Dec. 2005), pp. 1592–1603. ISSN: 03603016. DOI: 10.1016/j.ijrobp.2005.08.021.

[38] P. Remeijer, C. Rasch, J. V. Lebesque and M. van Herk. 'Margins for Translational and Rotational Uncertainties: A Probability-Based Approach'. In: *International Journal of Radiation Oncology*Biology*Physics* 53.2 (June 2002), pp. 464–474. ISSN: 03603016. DOI: 10.1016/S0360-3016(02)02749-9.

[39] S. J. Gardner, J. Kim and I. J. Chetty. 'Modern Radiation Therapy Planning and Delivery'. In: *Hematology/Oncology Clinics of North America* 33.6 (Dec. 2019), pp. 947–962. ISSN: 08898588. DOI: 10.1016/j.hoc.2019.08.005.

[40] S. Breedveld, P. R. M. Storchi, P. W. J. Voet and B. J. M. Heijmen. 'iCycle: Integrated, Multicriterial Beam Angle, and Profile Optimization for Generation of Coplanar and Noncoplanar IMRT Plans: iCycle: Multicriteria Beam Angle Optimization'. In: *Medical Physics* 39.2 (Jan. 2012), pp. 951–963. ISSN: 00942405. DOI: 10.1118/1.3676689.

[41] S. van de Water, A. C. Kraan, S. Breedveld, W. Schillemans, D. N. Teguh, H. M. Kooy, T. M. Madden, B. J. M. Heijmen and M. S. Hoogeman. 'Improved Efficiency of Multi-Criteria IMPT Treatment Planning Using Iterative Resampling of Randomly Placed Pencil Beams'. In: *Physics in Medicine and Biology* 58.19 (Oct. 2013), pp. 6969–6983. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/58/19/6969.

[42] S. Breedveld, P. R. M. Storchi and B. J. M. Heijmen. 'The Equivalence of Multi-Criteria Methods for Radiotherapy Plan Optimization'. In: *Physics in Medicine and Biology* 54.23 (Dec. 2009), pp. 7199–7209. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/54/23/011.

[43] J. Unkelbach, M. Alber, M. Bangert, R. Bokrantz, T. C. Y. Chan, J. O. Deasy, A. Fredriksson, B. L. Gorissen, M. van Herk, W. Liu, H. Mahmoudzadeh, O. Nohadani, J. V. Siebers, M. Witte and H. Xu. 'Robust Radiotherapy Planning'. In: *Physics in Medicine & Biology* 63.22 (Nov. 2018), 22TR02. ISSN: 1361-6560. DOI: 10.1088/1361-6560/aae659.

[44] J. Unkelbach, T. Bortfeld, B. C. Martin and M. Soukup. 'Reducing the Sensitivity of IMPT Treatment Plans to Setup Errors and Range Uncertainties via Probabilistic Treatment Planning'. In: *Medical Physics* 36.1 (Jan. 2009), pp. 149–163. ISSN: 0094-2405, 2473-4209. DOI: 10.1118/1.3021139.

[45] A. Fredriksson, A. Forsgren and B. Hårdemark. 'Minimax Optimization for Handling Range and Setup Uncertainties in Proton Therapy: Minimax Optimization for Handling Uncertainties in Proton Therapy'. In: *Medical Physics* 38.3 (Mar. 2011), pp. 1672–1684. ISSN: 00942405. DOI: 10.1118/1.3556559.

[46] O. P. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Scientific Computation. Dordrecht: Springer Netherlands, 2010. ISBN: 978-90-481-3519-6 978-90-481-3520-2. DOI: 10.1007/978-90-481-3520-2.

[47] D. Xiu and G. Karniadakis. 'The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations'. In: *SIAM Journal on Scientific Computing* 24.2 (2002), pp. 619–644. DOI: 10.1137/S1064827501387826.

[48] D. C. Montgomery. *Design and Analysis of Experiments*. Eighth edition. Hoboken, NJ: John Wiley & Sons, Inc, 2013. ISBN: 978-1-118-14692-7.

[49] Z. Perkó, S. R. van der Voort, S. van de Water, C. M. H. Hartman, M. Hoogeman and D. Lathouwers. 'Fast and Accurate Sensitivity Analysis of IMPT Treatment Plans Using Polynomial Chaos Expansion'. In: *Physics in Medicine and Biology* 61.12 (June 2016), pp. 4646–4664. ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/61/12/4646.

[50] Z. Perkó. 'Supplemental Material for Fast and Accurate Sensitivity Analysis of IMPT Treatment Plans Using Polynomial Chaos Expansion'. In: *Physics in Medicine and Biology* 61.12 (June 2016).

[51] Z. Perkó, L. Gilli, D. Lathouwers and J. L. Kloosterman. 'Grid and Basis Adaptive Polynomial Chaos Techniques for Sensitivity and Uncertainty Analysis'. In: *Journal of Computational Physics* 260 (Mar. 2014), pp. 54–84. ISSN: 00219991. DOI: 10.1016/j.jcp.2013.12.025.

[52] J. Rojo-Santiago, S. J. Habraken, D. Lathouwers, A. Méndez Romero, Z. Perkó and M. S. Hoogeman. 'Accurate Assessment of a Dutch Practical Robustness Evaluation Protocol in Clinical PT with Pencil Beam Scanning for Neurological Tumors'. In: *Radiotherapy and Oncology* 163 (Oct. 2021), pp. 121–127. ISSN: 01678140. DOI: 10.1016/j.radonc.2021.07.028.

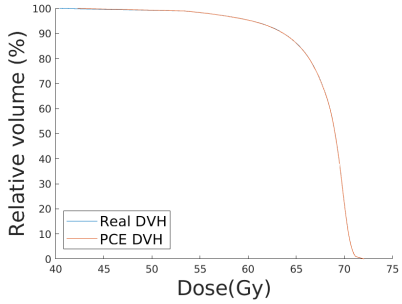
A

LARGEST ERROR SCENARIOS

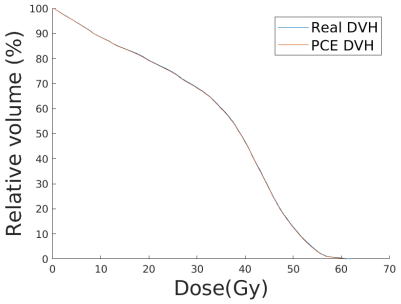
A.1. PATIENT 2

This subsection contains all the figures that show the largest error scenarios for patient 2 for either a fixed dose distribution or a reoptimized dose distribution. Each figure will specify if it shows the largest error scenario between the PCE response to the real DVH and the real DVH or if it shows the largest error scenario between the PCE response to the real DVH and the PCE-based DVH.

A.1.1. FIXED DOSE DISTRIBUTION



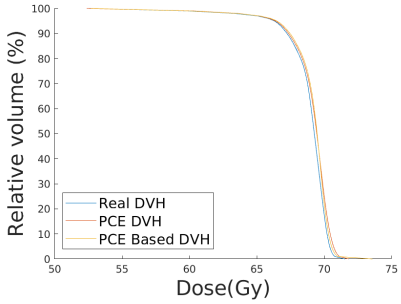
(a) Largest error between the real DVH and the PCE DVH of the CTV



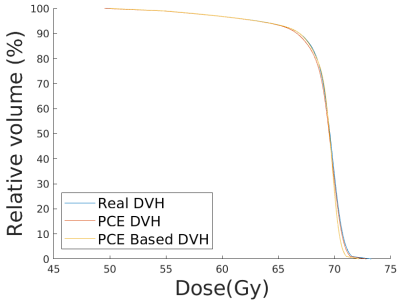
(b) Largest error between the real DVH and the PCE DVH of the brain stem

A

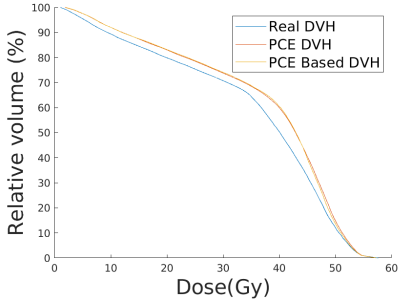
A.1.2. RE-OPTIMIZED DOSE DISTRIBUTION



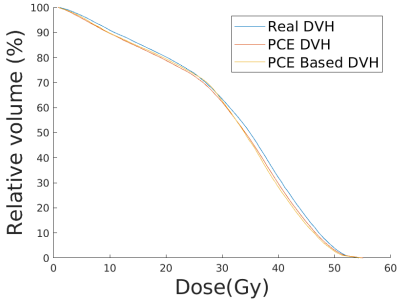
(a) Largest error between the real DVH and the PCE DVH of the CTV



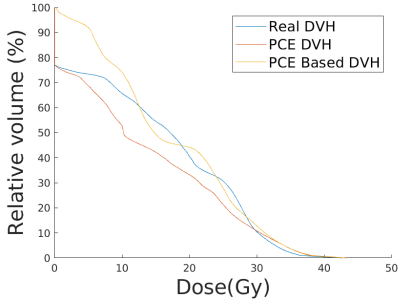
(b) Largest error between the PCE DVH and the PCE-based DVH of the CTV



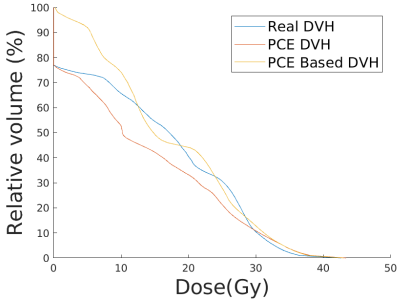
(a) Largest error between the real DVH and the PCE DVH of the brain stem



(b) Largest error between the PCE DVH and the PCE-based DVH of the brain stem



(a) Largest error between the real DVH and the PCE DVH of the spinal cord



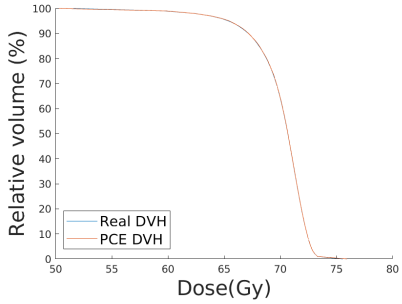
(b) Largest error between the PCE DVH and the PCE-based DVH of the spinal cord

A

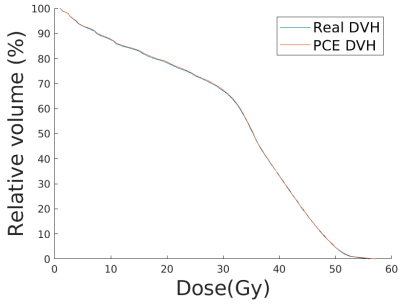
A.2. PATIENT 69

This subsection contains all the figures that show the largest error scenarios for patient 69 for either a fixed dose distribution or a reoptimized dose distribution. Each figure will specify if it shows the largest error scenario between the PCE response to the real DVH and the real DVH or if it shows the largest error scenario between the PCE response to the real DVH and the PCE-based DVH.

A.2.1. FIXED DOSE DISTRIBUTION

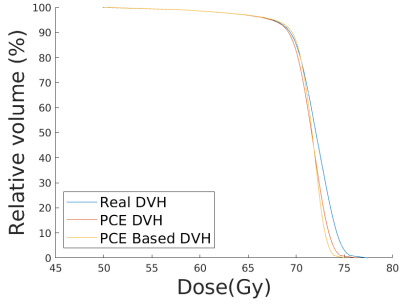


(a) Largest error between the real DVH and the PCE DVH of the CTV

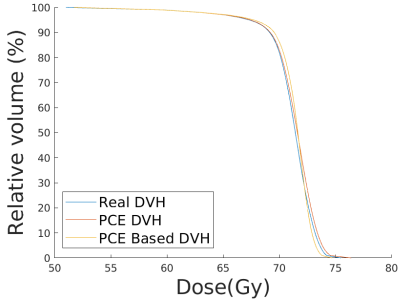


(b) Largest error between the real DVH and the PCE DVH of the brain stem

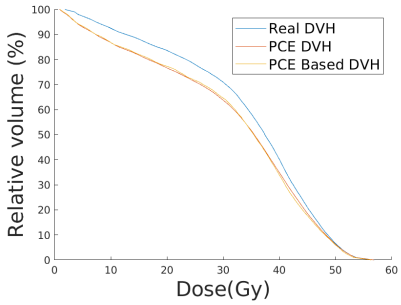
A.2.2. RE-OPTIMIZED DOSE DISTRIBUTION



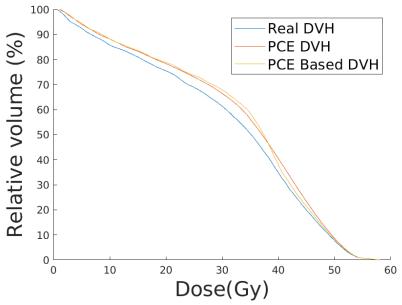
(a) Largest error between the real DVH and the PCE DVH of the CTV



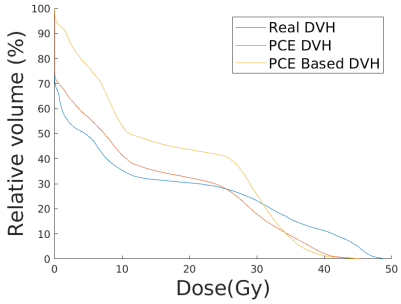
(b) Largest error between the PCE DVH and the PCE-based DVH of the CTV



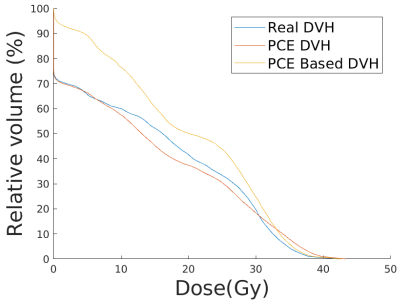
(a) Largest error between the real DVH and the PCE DVH of the brain stem



(b) Largest error between the PCE DVH and the PCE-based DVH of the brain stem



(a) Largest error between the real DVH and the PCE DVH of the spinal cord

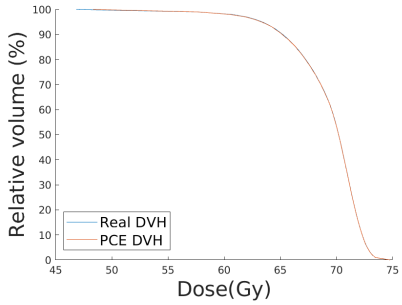


(b) Largest error between the PCE DVH and the PCE-based DVH of the spinal cord

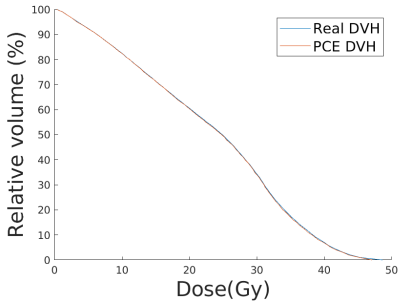
A.3. PATIENT 155

This subsection contains all the figures that show the largest error scenarios for patient 155 for either a fixed dose distribution or a reoptimized dose distribution. Each figure will specify if it shows the largest error scenario between the PCE response to the real DVH and the real DVH or if it shows the largest error scenario between the PCE response to the real DVH and the PCE-based DVH.

A.3.1. FIXED DOSE DISTRIBUTION

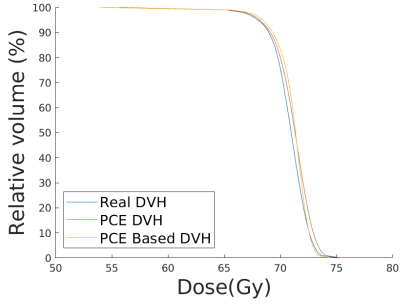


(a) Largest error between the real DVH and the PCE DVH of the CTV

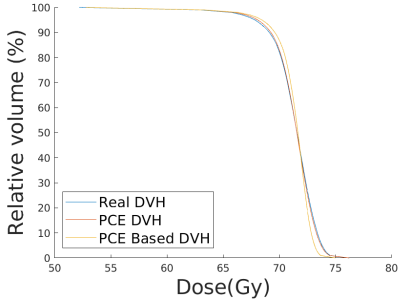


(b) Largest error between the real DVH and the PCE DVH of the brain stem

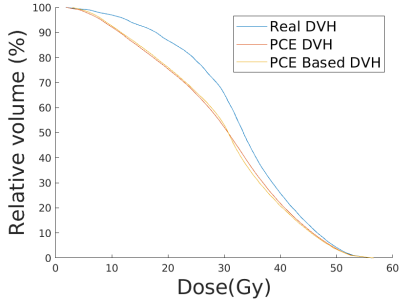
A.3.2. RE-OPTIMIZED DOSE DISTRIBUTION



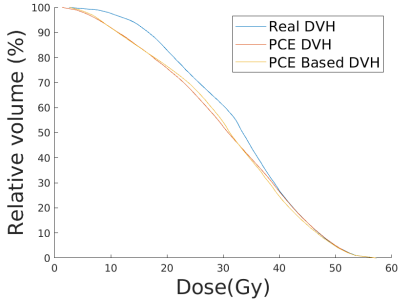
(a) Largest error between the real DVH and the PCE DVH of the CTV



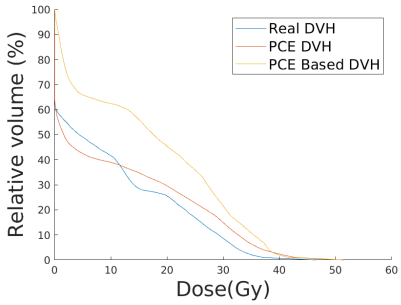
(b) Largest error between the PCE DVH and the PCE-based DVH of the CTV



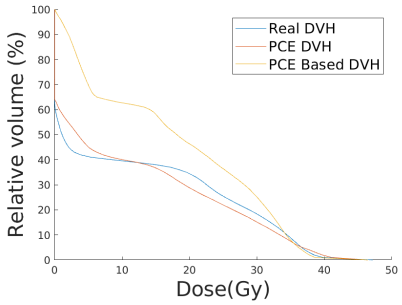
(a) Largest error between the real DVH and the PCE DVH of the brain stem



(b) Largest error between the PCE DVH and the PCE-based DVH of the brain stem



(a) Largest error between the real DVH and the PCE DVH of the spinal cord



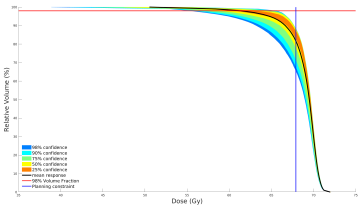
(b) Largest error between the PCE DVH and the PCE-based DVH of the spinal cord

B

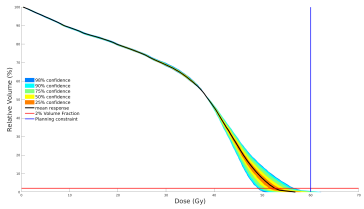
DOSIMETRIC EFFECTS

B.1. PATIENT 2

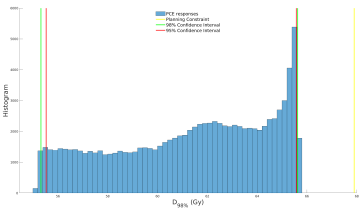
B.1.1. FIXED DOSE DISTRIBUTION



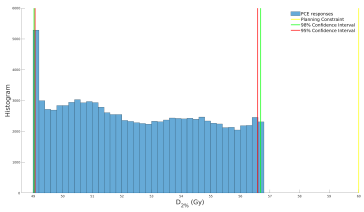
(a) Confidence intervals for the DVH of the CTV



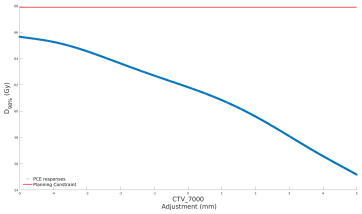
(b) Confidence intervals for the DVH of the Brain stem



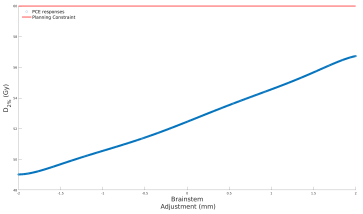
(a) Histogram for the $D_{98\%}$ of the CTV



(b) Histogram for the $D_{2\%}$ of the Brain stem

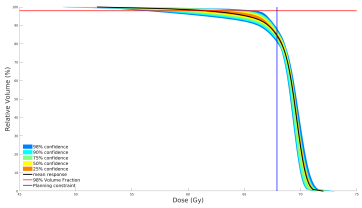


(a) Dependence of $D_{98\%}$ of the CTV on the delineation uncertainty

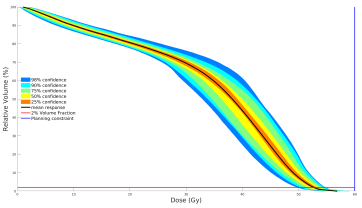


(b) Dependence of $D_{2\%}$ of the Brain stem on the delineation uncertainty

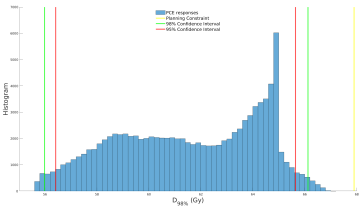
B.1.2. RE-OPTIMIZED DOSE DISTRIBUTION



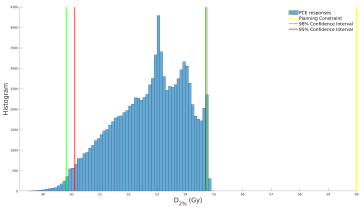
(a) Confidence intervals for the DVH of the CTV



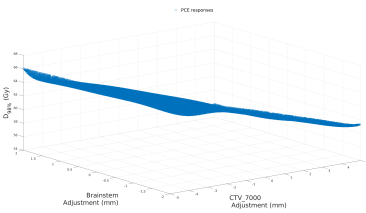
(b) Confidence intervals for the DVH of the Brain stem



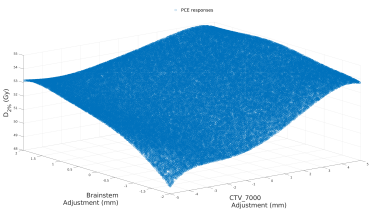
(a) Histogram for the $D_{98\%}$ of the CTV



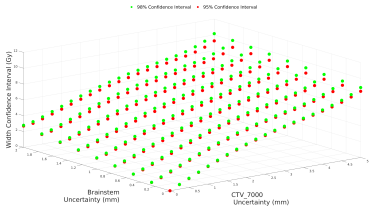
(b) Histogram for the $D_{2\%}$ of the Brain stem



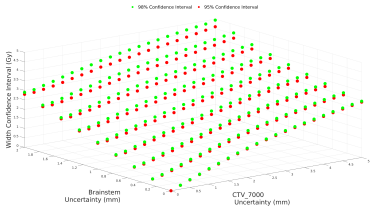
(a) Dependence of $D_{98\%}$ of the CTV on the delineation uncertainty



(b) Dependence of $D_{2\%}$ of the Brain stem on the delineation uncertainty



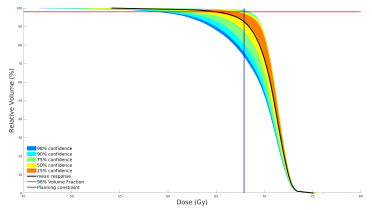
(a) Dependence of the width of the confidence intervals on the delineation uncertainty for the CTV



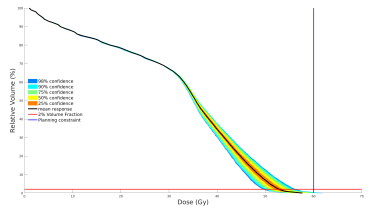
(b) Dependence of the width of the confidence intervals on the delineation uncertainty for the brain stem

B.2. PATIENT 69

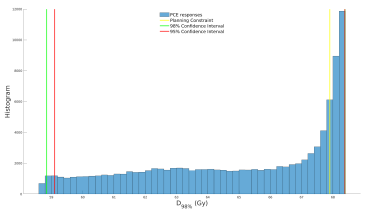
B.2.1. FIXED DOSE DISTRIBUTION



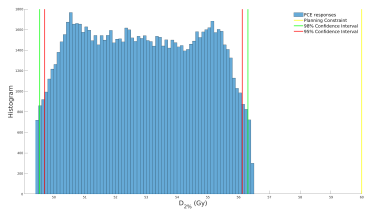
(a) Confidence intervals for the DVH of the CTV



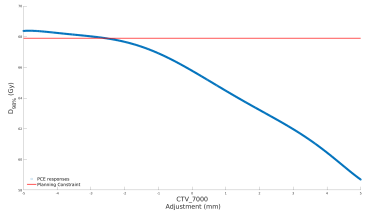
(b) Confidence intervals for the DVH of the Brain stem



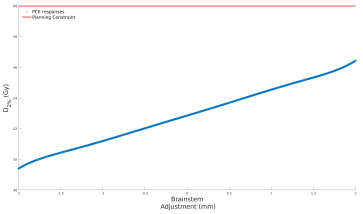
(a) Histogram for the $D_{98\%}$ of the CTV



(b) Histogram for the $D_{2\%}$ of the Brain stem

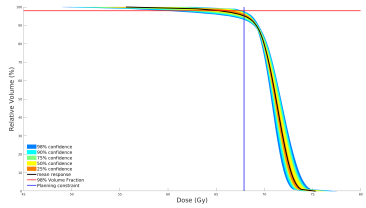


(a) Dependence of $D_{98\%}$ of the CTV on the delineation uncertainty

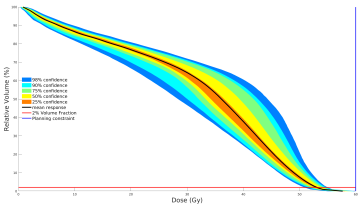


(b) Dependence of $D_{2\%}$ of the Brain stem on the delineation uncertainty

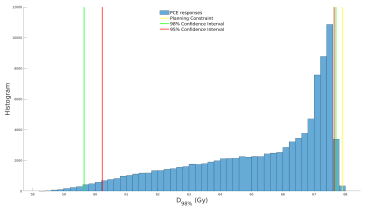
B.2.2. RE-OPTIMIZED DOSE DISTRIBUTION



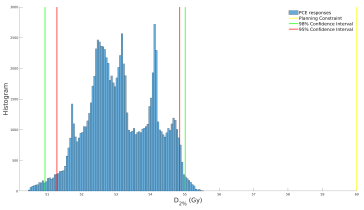
(a) Confidence intervals for the DVH of the CTV



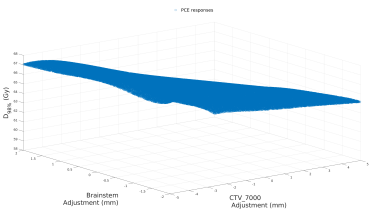
(b) Confidence intervals for the DVH of the Brain stem



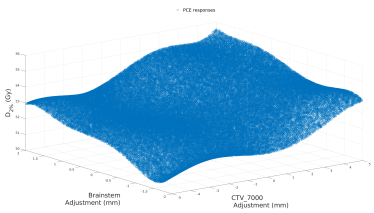
(a) Histogram for the $D_{98\%}$ of the CTV



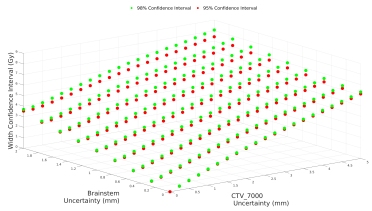
(b) Histogram for the $D_{2\%}$ of the Brain stem



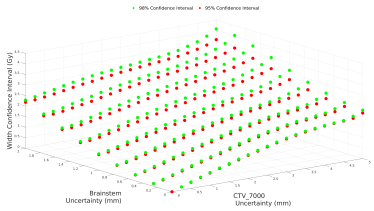
(a) Dependence of $D_{98\%}$ of the CTV on the delineation uncertainty



(b) Dependence of $D_{2\%}$ of the Brain stem on the delineation uncertainty



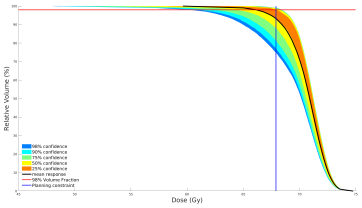
(a) Dependence of the width of the confidence intervals on the delineation uncertainty for the CTV



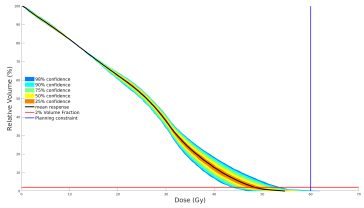
(b) Dependence of the width of the confidence intervals on the delineation uncertainty for the brain stem

B.3. PATIENT 155

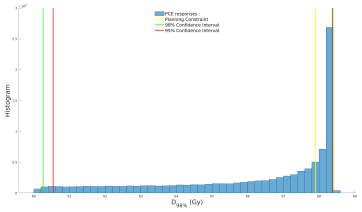
B.3.1. FIXED DOSE DISTRIBUTION



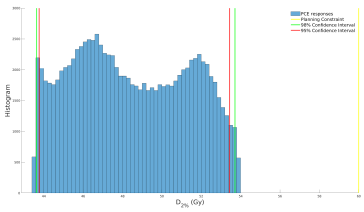
(a) Confidence intervals for the DVH of the CTV



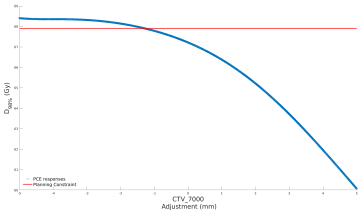
(b) Confidence intervals for the DVH of the Brain stem



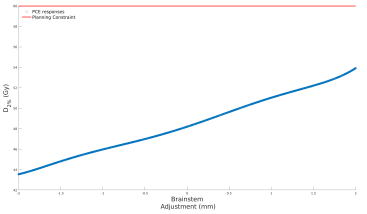
(a) Histogram for the $D_{98\%}$ of the CTV



(b) Histogram for the $D_{2\%}$ of the Brain stem

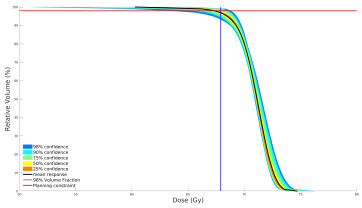


(a) Dependence of $D_{98\%}$ of the CTV on the delineation uncertainty

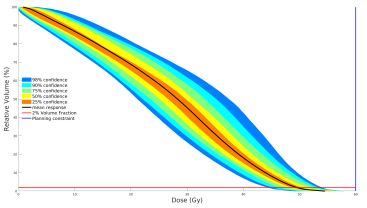


(b) Dependence of $D_{2\%}$ of the Brain stem on the delineation uncertainty

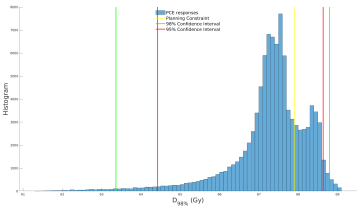
B.3.2. RE-OPTIMIZED DOSE DISTRIBUTION



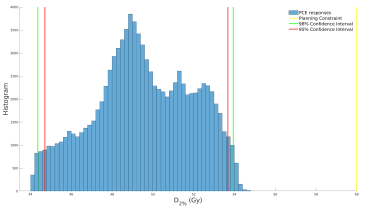
(a) Confidence intervals for the DVH of the CTV



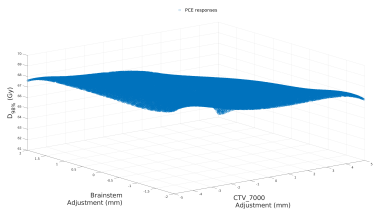
(b) Confidence intervals for the DVH of the Brain stem



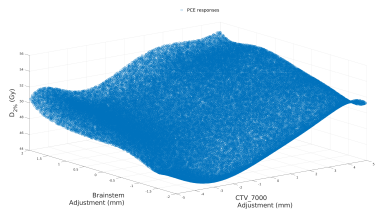
(a) Histogram for the $D_{98\%}$ of the CTV



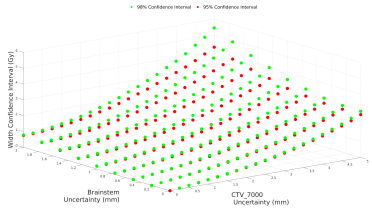
(b) Histogram for the $D_{2\%}$ of the Brain stem



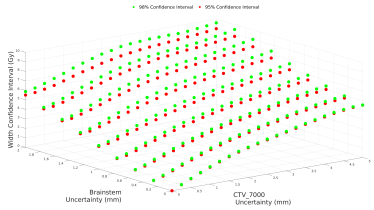
(a) Dependence of $D_{98\%}$ of the CTV on the delineation uncertainty



(b) Dependence of $D_{2\%}$ of the Brain stem on the delineation uncertainty



(a) Dependence of the width of the confidence intervals on the delineation uncertainty for the CTV



(b) Dependence of the width of the confidence intervals on the delineation uncertainty for the brain stem