Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER • **OPEN ACCESS**

Probabilistic proton treatment planning: a novel approach for optimizing underdosage and overdosage probabilities of target and organ structures

To cite this article: J R de Jong *et al* 2026 *Phys. Med. Biol.* **71** 025002

IOP Publishing Phys. Med. Biol. 71 (2026) 025002 https://doi.org/10.1088/1361-6560/ae351c

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
1 July 2025

REVISED
23 October 2025

ACCEPTED FOR PUBLICATION
7 January 2026

PUBLISHED
20 January 2026

PAPER

Probabilistic proton treatment planning: a novel approach for optimizing underdosage and overdosage probabilities of target and organ structures

J R de Jong[1,*], S Breedveld[3], S J M Habraken[2,4], M S Hoogeman[2,3], D Lathouwers[1] and Z Perkó[1]

[1] Department of Radiation, Science and Technology, Delft University of Technology, section Medical Physics and Technology, Mekelweg 15, Delft, The Netherlands
[2] Holland Proton Therapy Centre Delft, Delft, The Netherlands
[3] Department of Radiotherapy, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands
[4] Department of Radiation Oncology, Leiden University Medical Center, Leiden, The Netherlands
* Author to whom any correspondence should be addressed.

E-mail: j.r.dejong@tudelft.nl

Keywords: particle therapy, uncertainty, setup error, range error, polynomial Chaos, robust optimization, percentiles

Supplementary material for this article is available online

Abstract

Objective. Uncertainties in treatment planning are typically managed using either margin-based or robust optimization. Margin-based methods expand the clinical target volume (CTV) towards a planning target volume, which is generally unsuited for proton therapy. Robust optimization considers worst-case scenarios, but its quality depends on the chosen *uncertainty (scenario) set*: excluding extremes reduces robustness, while including too many make plans overly conservative. Probabilistic optimization overcomes these limitations by modeling a continuous scenario distribution, enabling the use of statistical measures. *Approach*. We propose a novel approach to probabilistic optimization that steers plans towards individualized probability levels, to control CTV and organs-at-risks (OARs) under- and overdosage. Voxel-wise dose percentiles (d) are estimated by expected value ($\mathbb{E}$) and standard deviation (SD) as $\mathbb{E}[d] \pm \delta \cdot \mathrm{SD}[d]$, where δ is iteratively tuned to match the target percentile of the underlying probability distribution (given setup and range uncertainties). The approach involves an inner optimization of $\mathbb{E}[d] \pm \delta \cdot \mathrm{SD}[d]$ for fixed δ, and an outer optimization loop that updates δ. Polynomial chaos expansion provides accurate and efficient dose estimates during optimization. We validated the method on a spherical CTV (prescribed 60 Gy) abutted by an OAR in different directions and a horseshoe-shaped CTV surrounding a cylindrical spine, under Gaussian-distributed setup (3 mm) and range (3%) uncertainties. *Main results*. For spherical cases with similar CTV coverage, $P(D_{2\%} > 30\,\mathrm{Gy})$ dropped by 10%–15%; for matched OAR dose, $P(D_{98\%} > 57\,\mathrm{Gy})$ increased by 67.5%–71%. In spinal plans, $P(D_{98\%} > 57\,\mathrm{Gy})$ increased by 10%–15% while $P(D_{2\%} > 30\,\mathrm{Gy})$ dropped by 24%–28% in the same plan. Probabilistic and robust optimization times were comparable for spherical (hours) but longer for spinal cases (7.5–11.5 h vs 9–20 min). *Significance*. Compared to discrete scenario-based optimization, the probabilistic approach offered better OAR sparing or target coverage, depending on individualized priorities.

1. Introduction

Intensity-modulated proton therapy (IMPT) has demonstrated improved sparing of organs at risk (OARs) and normal tissue compared to intensity-modulated radiation therapy (IMRT) for various treatment sites (Stuschke *et al* 2012, van de Sande *et al* 2016, Nguyen *et al* 2021). However, the precise dose delivery of IMPT makes it more sensitive to uncertainties that may occur during treatment. Uncertainties include patient setup misalignment, anatomical changes during the treatment, or range uncertainties due

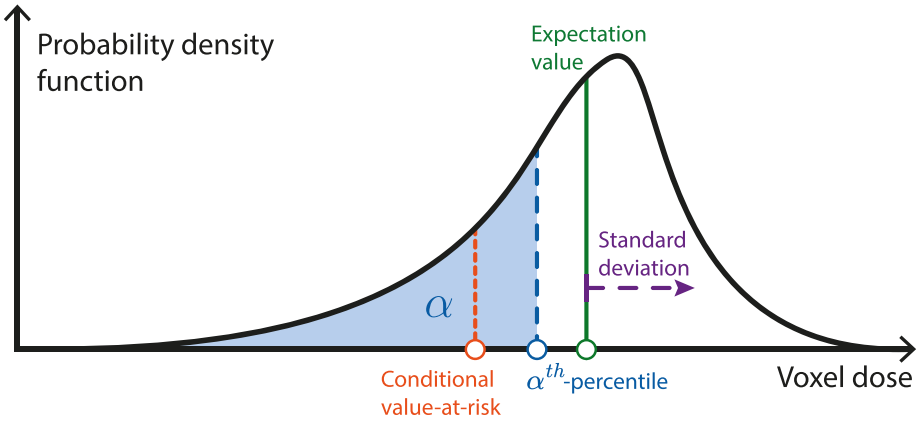

Figure 1. Illustration of the probability density function of the voxel dose, being a result of the statistical nature of the considered uncertainties. Shown are the expectation value (solid green), standard deviation (purple arrow) and the α^{th}-percentile of the voxel dose (dotted blue), which is the dose value for which $\alpha\%$ of the error scenarios (area under the curve) result in lower dose values. The conditional value-at-risk (dotted orange) corresponding to probability α is the mean of the lowest $\alpha\%$ of scenarios. This example is used to control target underdosage probability, while equivalent representations can be made for (target and OAR) overdosage probability.

to CT to stopping power conversions (Schaffner and Pedroni 1998, Lomax 2008). To ensure an effective treatment, it is crucial that treatment plans are robust against such uncertainties.

In photon therapy, uncertainties are typically managed using planning target volume (PTV) margins around the clinical target volume (CTV) (van Herk *et al* 2000). However, as the proton dose distribution is more sensitive to geometrical shifts, applying PTV margins is often ineffective in IMPT. Robust optimization offers an alternative by considering a predefined set of uncertainty scenarios, in practice corresponding to errors of a fixed magnitude. Multiple robust approaches exist (Unkelbach *et al* 2018), but most commonly used is mini–max robust optimization (Fredriksson *et al* 2011). It treats all scenarios in the uncertainty set as equally important and, for each iteration, optimizes for the worst-case among the scenario set. Depending on the desired conservativeness of the plan, one can choose to optimize using voxel-wise (Pflugfelder *et al* 2008, Liu *et al* 2012), objective-wise (Chen *et al* 2012) or composite-wise (van Dijk *et al* 2016, Janson *et al* 2024) worst-case objectives. While effective with a well-defined uncertainty set, robust optimization can lead to overly conservative plans if extreme scenarios dominate or to insufficient robustness if the uncertainty set is too narrow (van der Voort *et al* 2016, Zhang 2021). Unlike mini–max robust optimization, stochastic programming (Unkelbach *et al* 2007, 2008) assigns a probability weight to each scenario in the set. However, as the number of scenarios is still limited in this approach, it is not straightforward how the probability weights can be effectively assigned to a single scenario. Robust optimization methods that allow for intermediate robustness levels have been developed as well. Taasti *et al* (2020) proposed to optimize using the p-norm, for $1 \leqslant p \leqslant \infty$, where $p = 1$ and $p = \infty$ correspond to stochastic and composite-wise robust optimization. Alternatively, in scenario-based robust methods (Fredriksson and Bokrantz 2016, Bokrantz and Fredriksson 2017), each voxel is weighted according to how often it overlaps with the region of interest (ROI).

Probabilistic optimization (not to be confused with stochastic robust optimization) presents a promising alternative by defining uncertainty as a distribution, so that it simultaneously accounts for a full spectrum of possible errors. As quantities of interest (e.g. voxel dose) depend on the uncertainties, they become random variables as well, whose statistical effect can be presented by a probability density function (PDF), as shown in figure 1. The impact of uncertainty can be quantified by stochastic metrics (e.g. using expectation values and variances (Cristoforetti *et al* 2025), or using percentiles), which provide consistent and statistically interpretable results, thereby helping to reduce inter-patient variation (Rojo-Santiago *et al* 2023b, de Jong *et al* 2025). The α^{th}-percentile is the dose level such that $\alpha\%$ of the scenarios result in lower dose values. For small or large α, the percentile allows to quantify the most extreme scenarios, which in the context of radiotherapy can be used to respectively minimize for under- and overdosage probability. The conditional value-at-risk (CVaR) can be used alternatively, which quantifies the average of the worst $\alpha\%$ of scenarios. CVaR has been applied before in photon radiotherapy (Tilly *et al* 2019), IMRT (Chan *et al* 2014) and IMPT (An *et al* 2017), using a discrete number of error scenarios.

The calculation of percentiles and other probabilistic metrics require the evaluation of the PDF of the response of interest (e.g. voxel dose). This is challenging because these PDFs typically do not have an

analytical form. A common approximation is to assume that response PDFs have Gaussian-distributed tails, which is reasonable for e.g. the total dose in treatments involving a sufficient number of fractions (Sobotta *et al* 2010), but may be inadequate for other metrics. Chu *et al* (2005) approximated voxel-wise dose percentiles by expectation value ($\mathbb{E}$) and standard deviation (SD) as $\mathbb{E} \pm \delta \cdot \mathrm{SD}$, where δ is a constant that quantifies the displacement from the mean in units of SD. Fabiano *et al* (2022) optimized robustly for biological effective dose (BED) using $\delta = 2$, estimating the 2^{nd} and 98^{th} percentiles as $\mathbb{E} - 2 \cdot \mathrm{SD}$ and $\mathbb{E} + 2 \cdot \mathrm{SD}$, respectively. These correspond to the dose values below which 2% (or 98%) of the scenarios have lower values (if the response PDF would be truly Gaussian distributed). Although the Gaussian approximation potentially allows the PDF to be steered into the desired direction (Sobotta *et al* 2010), these methods do not allow to tune exactly for desired probability levels and dose thresholds. In fact, distribution shapes may be complex in practice and there is no guarantee that they are even near Gaussian distributed.

Probabilistic optimization studies have been done for photon therapy, optimizing for minimum dose percentiles in the objectives (Gordon *et al* 2010) and constraints (Mescher *et al* 2017), e.g. requiring that 90% of the scenarios lead to a minimum dose greater than 95% of the prescribed dose, as in the Van Herk margin recipe (VHMR) (van Herk *et al* 2000). An approach has been developed to optimize for target coverage (e.g. $D_{98\%}$) in general by using approximate DVH penalties (Wu and Mohan 2000), only including voxels within a rim around the CTV. This rim was defined using the VHMR, and dose-volume histograms (DVHs) were sampled under the static dose cloud approximation. Such margin recipes and the static dose cloud approximation do not hold in general for proton therapy, limiting the applicability of these approximations in proton plans.

These limitations motivate the need for a flexible probabilistic approach that can be applied to proton therapy as well (i.e. without relying on the static dose cloud), which allows us to optimize for specified dose thresholds for individualized probabilities of under- and overdosage, regardless of the PDF shape. Accurate probabilities are obtained by sampling thousands of error scenarios from the Gaussian distribution, resulting in the same number of dose distributions. As it is computationally expensive to perform these calculations by Monte-Carlo sampling, we use polynomial chaos expansion (PCE). It has proven to be an accurate meta-model of the dose-engine, also for probabilistic evaluation of treatment plans (Perkó *et al* 2016, Rojo-Santiago *et al* 2023a). Once constructed, PCE facilitates the efficient sampling of dose distributions in many error scenarios.

The probabilistic approach consists of an inner optimization and an outer optimization loop. The inner part optimizes the pencil-beam weights for a fixed set of δ-factors (as done by Fabiano *et al* (2022)). Qualitatively, the δ-factor is a bridge between the probability levels and dose thresholds, even when the underlying distribution is non-Gaussian. In other words, a probabilistic goal (e.g. at most 10% probability of underdosing) is translated into a specific dose threshold used in the optimization. The outer optimization loop updates the δ-factors to improve the $\mathbb{E} \pm \delta \cdot \mathrm{SD}$ approximation, effectively rescaling it based on the updated PDF.

The main goal of the probabilistic approach is to replace worst-case objectives by voxel-wise probabilistic objectives, keeping the spatial information of the ROIs and avoiding the under- and overconservatism that is inherent in mini–max robust optimization. This is achieved by efficient PCE sampling, which allows to optimize directly on statistical information (i.e. probability level and dose threshold). Regarding plan evaluation, often worst-case evaluation metrics (e.g. the $D_{98\%}$ of the voxel-wise minimum) are used (Korevaar *et al* 2019), which lack statistical insight into the plan quality (Park *et al* 2013, Sterpin *et al* 2021). Here, as we optimize for statistical objectives, the step towards plan evaluation with statistically meaningful metrics is a straightforward consequence.

This paper presents a proof-of-principle of a novel probabilistic approach to probabilistic treatment planning for systematic setup and range errors. In section 2 we discuss the PCE method that is used for the scenario sampling, the probabilistic approach and the (homogeneous) phantom geometries used: a simple spherical CTV with surrounding OARs, and a more complex horseshoe-shaped CTV around a cylindrical spine. In section 3, probabilistic VHMR equivalence with the probabilistic approach is validated for setup errors in a spherical CTV, after which comparisons to composite-wise mini–max robust plans are done, matching either CTV coverage or OAR sparing. In sections 4 and 5, a discussion and conclusion on the results are respectively presented.

2. Methods and materials

2.1. Polynomial Chaos Expansion

In this work, we consider Gaussian-distributed setup (in x and y) and range errors (denoted as $\boldsymbol{\xi}$) with respective SDs of $\sigma_{\mathrm{setup}} = 3$ mm and $\sigma_{\mathrm{range}} = 3\%$. The Gaussian distribution is truncated such that the

combined setup and range errors is cut at the 99% confidence level to include the majority of the uncertainty space used for sampling. Setup errors are modeled by shifting all pencil-beam spots with respect to the dose distribution in the shift-direction. Range errors are modeled by scaling all pencil-beam spots in the beam direction (i.e. the $\pm z$-direction) with the relative range uncertainty.

For the uncertainty quantification of the response, Polynomial Chaos (Wiener 1938) is used. The response is estimated as a PCE that is a function of the N uncertainty variables, in this work $N = 3$. As we assume these to be independent and Gaussian-distributed, their joint PDF $p(\boldsymbol{\xi})$ can be described as the product of the one-dimensional PDFs, such that $p(\boldsymbol{\xi}) = \prod_{j=1}^{N} p(\xi_j)$. We consider the assumption on uncertainty variable independence to be reasonable, as in robust optimization (the clinical practice) and plan evaluation (e.g. using voxel-wise minimum and maximum metrics (Rojo-Santiago *et al* 2021, de Jong *et al* 2025)), the discrete error scenarios are assumed independent as well. The PCE is an expansion using multi-dimensional basis vectors $\Psi_k(\boldsymbol{\xi})$, where the PCE of response $R(\boldsymbol{\xi})$ is defined as

$$R(\boldsymbol{\xi}) = \sum_{k=0}^{P} r_k \Psi_k(\boldsymbol{\xi}), \tag{1}$$

where r_k are the polynomial coefficients, such that $P+1$ basis vectors are used. The type of the basis vector is chosen based on the uncertainty distribution. For Gaussian input variables, the Wiener–Askey scheme (Xiu and Karniadakis 2002) proposes to use probabilists' Hermite polynomials $He_{\gamma_{k,j}}(\xi_j)$, where $\gamma_{k,j} = (\gamma_{k,1}, \ldots, \gamma_{k,N})$ denotes the polynomial order of the jth polynomial corresponding to basis vector k. The multi-dimensional basis vectors are thus given by $\Psi_k(\boldsymbol{\xi}) = \prod_{j=1}^{N} He_{\gamma_{k,j}}(\xi_j)$. The PCE corresponding to order O with a full basis set is defined by including the multi-dimensional polynomials for which $\sum_{j=1}^{N} \gamma_{k,j} \leqslant O$. As a result, the PCE in (1) has $P+1 = (N+O)!/(N!O!)$ basis vectors. In this work the full basis set is used.

Constructing the PCE comes down to determining the polynomial coefficients r_k. In this work we use spectral projection, such that

$$r_k = \frac{\langle R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})\rangle}{\langle \Psi_k(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})\rangle} = \frac{\int R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}}{\int \Psi_k(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}} = \frac{1}{h_k^2}\int R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}, \tag{2}$$

where $\langle\cdot\rangle$ denotes the inner product and $h_k^2 = \langle \Psi_k(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})\rangle$ is the norm of basis vector k. The integral in the nominator is determined by Gauss–Hermite cubature, for which a defined set of cubature points $\boldsymbol{\xi}_l$ is used with corresponding weights w_l to yield

$$\int R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \sum_l R(\boldsymbol{\xi}_l)\Psi_k(\boldsymbol{\xi}_l)p(\boldsymbol{\xi}_l)w_l. \tag{3}$$

For this computation, the exact response $R(\boldsymbol{\xi}_l)$ only has to be calculated on these cubature points, which is done using the dose engine. Instead of using full cubature grids we use Smolyak sparse grids (Smolyak 1963), in which higher-order cubature points that simultaneously occur in multiple dimensions are neglected, essentially reducing the number of terms in (3) and therefore the number of necessary dose computations without compromising accuracy. A more advanced form is to use extended Smolyak sparse grids, where the grid level along the single dimensions is increased by lev_{extra} levels, as often a significant increase in PCE accuracy can be obtained by only limited number of extra calculations. To reduce memory cost, we neglect voxels that have a dose lower than 0.01 Gy in the nominal and $\pm 3\sigma$ scenarios along the principal axes. This assumption has negligible effect on the plan quality (see section 1 of the supplementary material (SM)). Further details about PCE construction and numerical integration with sparse grids can be found in the SM of (Perkó *et al* 2016).

Through PCE construction, we obtain a meta-model of the exact response that can be used for efficient sampling. Moreover, obtaining the first two moments from (1) is computationally simple. The mean of the response μ_R is equal to the zeroth polynomial coefficient r_0 and its variance is $\sigma_R^2 = \sum_{k=1}^{\infty} r_k^2 h_k^2 \approx \sum_{k=1}^{P} r_k^2 h_k^2$. These metrics become useful in the probabilistic evaluation of the treatment plans.

2.2. The probabilistic approach

Underdosage and overdosage probability can be quantified using percentiles, which define the worst voxel dose scenarios that may occur during a treatment. In the context of underdosing the CTV, we specifically aim to limit the probability (or in simpler terms, the fraction of scenarios) α where the dose falls below a voxel dose threshold γ_i, such that

$$P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \gamma_i) \leqslant \alpha \quad \forall i \in \text{CTV}. \tag{4}$$

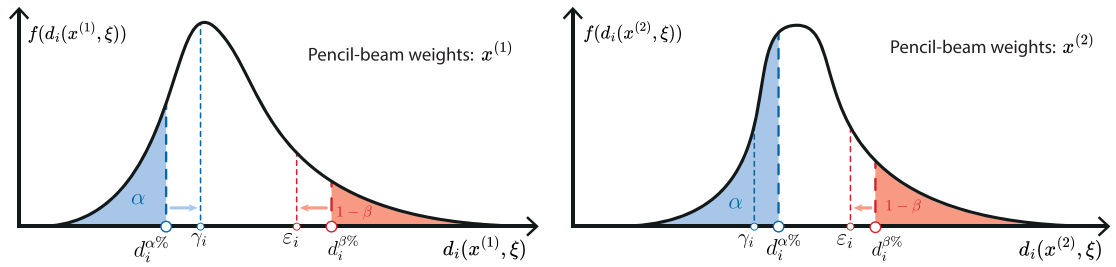

Figure 2. Probability density functions f of the voxel dose $d_i(\boldsymbol{x}^{(k)},\boldsymbol{\xi})$ for two different pencil-beam weights (e.g. at iterations k) during the optimization. Target underdosage probability is optimized for $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \gamma_i) \leqslant \alpha$, or equivalently $d_i^{\alpha\%}(\boldsymbol{x}) \geqslant \gamma_i$. Target and OAR overdosage can be optimized by $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \epsilon_i) \geqslant \beta$, i.e. $d_i^{\beta\%}(\boldsymbol{x}) \leqslant \epsilon_i$. The PDF shape has changed for iteration 2, resulting in sufficient target coverage in this voxel ($d_i^{\alpha\%}(\boldsymbol{x}^{(2)}) > \gamma_i$).

For example, one may aim for the voxel dose to fall below $\gamma_i = 0.95 \cdot d_i^p$ (i.e. underdosing) in at most $\alpha = 10\%$ of the error scenarios. In this work, we aim to reformulate (4) into an objective, such that dose threshold γ_i and probability α are quantities that can be controlled as part of the objective in our optimization approach. For this purpose, we define the α^{th}-percentile of the voxel dose PDF as $d_i^{\alpha\%}(\boldsymbol{x})$. Its definition is such that $\alpha\%$ of the error scenarios lead to voxel doses smaller than $d_i^{\alpha\%}(\boldsymbol{x})$, such that

$$\alpha = P\left(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant d_i^{\alpha\%}(\boldsymbol{x})\right) = \int_{-\infty}^{d_i^{\alpha\%}(\boldsymbol{x})} f(d_i(\boldsymbol{x},\boldsymbol{\xi}))\,\mathrm{d}(d_i(\boldsymbol{x},\boldsymbol{\xi})), \tag{5}$$

where the voxel dose PDF $f(d_i(\boldsymbol{x},\boldsymbol{\xi}))$ is analytically unknown in general. Substituting (5) into (4) for voxel i gives

$$P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \gamma_i) \leqslant \alpha = P\left(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant d_i^{\alpha\%}(\boldsymbol{x})\right), \tag{6}$$

which holds if α is such that $d_i^{\alpha\%}(\boldsymbol{x}) \geqslant \gamma_i$. This implies that reducing $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \gamma_i)$ below α can be achieved by increasing $d_i^{\alpha\%}(\boldsymbol{x})$ above γ_i. Practically, this means that obtaining the α-th percentile of the voxel dose and setting threshold γ_i is sufficient to optimize for such a percentile description. This is visualized in figure 2, where the PDFs corresponding to two example pencil-beam weight vectors ($\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$) during the optimization are illustrated. For pencil-beam weights $\boldsymbol{x}^{(2)}$, the target underdosage is lower than the desired level, so $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \gamma_i) < \alpha$.

Similarly, for overdosing structure Σ (e.g. for CTV or OAR), we aim to limit the probability $1-\beta$ where the dose exceeds voxel dose threshold ϵ_i, such that $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \epsilon_i) \geqslant \beta \quad \forall i \in \Sigma$. Equivalently, we use $d_i^{\beta\%}(\boldsymbol{x}) \leqslant \epsilon_i$. In this context, one may aim for the voxel dose to exceed $\epsilon_i = 1.07 \cdot d_i^p$ (i.e. overdosing) in at most $1-\beta = 10\%$ of the error scenarios. Or to put it differently, we aim for the voxel dose to fall below $\epsilon_i = 1.07 \cdot d_i^p$ in at least $\beta = 90\%$ of the error scenarios.

In order to determine the voxel dose percentiles, $d_i(\boldsymbol{x},\boldsymbol{\xi})$ must be computed in a large number of uncertainty scenarios. The voxel dose $d_i(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{j\in\mathbb{B}} D_{ij}(\boldsymbol{\xi})x_j$ is determined by summing over all physical proton pencil-beams ($j \in \mathbb{B}$) with corresponding intensity x_j (i.e. pencil-beam weights, also referred to as beam weights). Each pencil-beam's Bragg peak is positioned at a predefined spot position in the pencil-beam grid. Its contribution to each voxel $i \in \mathbb{V}$ is defined by the dose-influence matrix $D_{ij}(\boldsymbol{\xi})$ and volume $\mathbb{V}$. A PCE of the dose-influence matrix is approximated as

$$D_{ij}(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} R_{ij}^{(k)} \Psi_k(\boldsymbol{\xi}), \tag{7}$$

with $P+1$ number of basis vectors, k^{th} coefficient $R_{ij}^{(k)}$ and multi-dimensional Hermite basis vector $\Psi_k(\boldsymbol{\xi})$. Sampling directly from (7) is possible, but requires the construction of maximally $N_v \cdot N_b$ (the number of $D_{ij}(\boldsymbol{\xi})$ elements) PCEs. The number of needed PCEs can be reduced to only N_v (the number of $d_i(\boldsymbol{x},\boldsymbol{\xi})$ elements) by converting (7) to a PCE of the voxel dose $d_i(\boldsymbol{x},\boldsymbol{\xi})$ for all voxels $i = 1,\ldots,N_v$ as

$$d_i(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{j\in\mathbb{B}} D_{ij}(\boldsymbol{\xi})x_j \approx \sum_{j\in\mathbb{B}} \left(\sum_{k=0}^{P} R_{ij}^{(k)} \Psi_k(\boldsymbol{\xi})\right) x_j \tag{8}$$

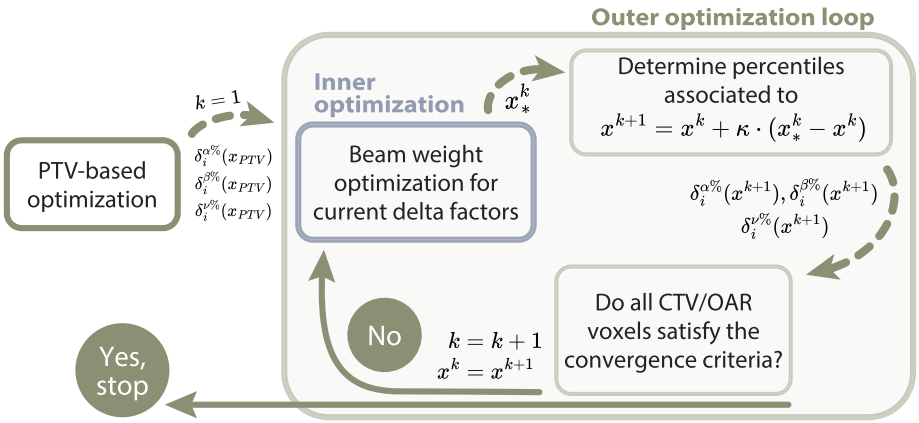

Figure 3. The proposed probabilistic optimization approach has a nested structure: the inner optimization optimizes pencil-beam weights, while the outer optimization loop ensures convergence to desired probability levels.

$$= \sum_{k=0}^{P} \left(\sum_{j \in \mathbb{B}} R_{ij}^{(k)} x_j \right) \Psi_k(\boldsymbol{\xi}) = \sum_{k=0}^{P} q_i^{(k)} \Psi_k(\boldsymbol{\xi}), \tag{9}$$

where the PCE coefficients of the voxel dose $q_i^{(k)} = \sum_{j \in \mathbb{B}} R_{ij}^{(k)} x_j$ can be obtained from the PCE coefficients of D_{ij} (i.e. from $R_{ij}^{(k)}$). Since (7) is independent of the beam weights, constructing it once before the optimization is sufficient (as opposed to constructing the voxel dose PCE for each iteration). To summarize, voxel dose percentiles are determined from voxel dose samples (we use $N_s = 100.000$ number of samples), by constructing the PCE coefficients of D_{ij} (i.e. (7)), then converting (7)–(9), so that the approximated voxel dose can be efficiently sampled.

The remainder of section 2.2 introduces the probabilistic approach. The corresponding optimization scheme is a nested structure consisting of an inner optimization and outer optimization loop, as is illustrated in figure 3. The inner optimization serves as an optimizer for the beam weights by optimizing using a given percentile estimate. The outer optimization loop makes sure that the percentile estimate remains accurate during the optimization.

2.2.1. The inner optimization

The inner optimization focuses on optimizing the beam weights for a given percentile estimate $d_i^{\alpha\%}(\boldsymbol{x})$, i.e. with fixed δ-factors, as was done by Fabiano *et al* (2022). We choose to optimize the CTV underdosage probability by defining a quadratic underdose penalty as

$$f_{\text{CTV}}^{\alpha,\gamma}(\boldsymbol{x}) = \frac{1}{N_{\text{CTV}}} \sum_{i \in \text{CTV}} w_i^{\text{CTV}} \left[\gamma_i - d_i^{\alpha\%}(\boldsymbol{x}) \right]_+^2, \tag{10}$$

where w_i^{CTV} is the CTV voxel weight and N_{CTV} is the number of CTV voxels. We use $[h]_+ = \max\{0, h\}$ to only penalize voxels if $h \geqslant 0$, meaning that only underdosed CTV voxels are penalized ($d_i^{\alpha\%}(\boldsymbol{x}) < \gamma_i$).

In order to use an analytical gradient and Hessian of the objective, we rewrite $d_i^{\alpha\%}(\boldsymbol{x})$ using the expectation value $\mathbb{E}[d_i(\boldsymbol{x}, \boldsymbol{\xi})]$ and standard deviation $\text{SD}[d_i(\boldsymbol{x}, \boldsymbol{\xi})]$ of the voxel dose (Fabiano *et al* 2022), as

$$\begin{aligned} d_i^{\alpha\%}(\boldsymbol{x}) &= \mathbb{E}[d_i(\boldsymbol{x}, \boldsymbol{\xi})] - \delta_i^{\alpha\%}(\boldsymbol{x}) \, \text{SD}[d_i(\boldsymbol{x}, \boldsymbol{\xi})] \\ &= \mathbb{E}[d_i(\boldsymbol{x}, \boldsymbol{\xi})] - \delta_i^{\alpha\%}(\boldsymbol{x}) \left[\mathbb{E}\left[d_i^2(\boldsymbol{x}, \boldsymbol{\xi}) \right] - \mathbb{E}^2[d_i(\boldsymbol{x}, \boldsymbol{\xi})] \right]^{1/2}, \end{aligned} \tag{11}$$

where $\delta_i^{\alpha\%}(\boldsymbol{x}) \in \mathbb{R}$ (also referred to as δ-factor) defines the number of SDs the percentile $d_i^{\alpha\%}(\boldsymbol{x})$ is displaced from the expectation value $\mathbb{E}[d_i(\boldsymbol{x}, \boldsymbol{\xi})]$. The δ-factor is determined by using the accurate percentile $d_i^{\alpha\%}(\boldsymbol{x})$ obtained from the PCE as

$$\delta_i^{\alpha\%}(\boldsymbol{x}) = \frac{\mathbb{E}[d_i(\boldsymbol{x}, \boldsymbol{\xi})] - d_i^{\alpha\%}(\boldsymbol{x})}{\text{SD}[d_i(\boldsymbol{x}, \boldsymbol{\xi})]}. \tag{12}$$

Similarly, quadratic overdose penalties for CTV ($d_i^{\beta} > \epsilon_i$) and OAR ($d_i^{\nu} > \mu_i$) voxels are defined as

$$f_{\text{CTV}}^{\beta,\epsilon}(\boldsymbol{x}) = \frac{1}{N_{\text{CTV}}} \sum_{i \in \text{CTV}} w_i^{\text{CTV}} \left[d_i^{\beta\%}(\boldsymbol{x}) - \epsilon_i \right]_+^2, \tag{13}$$

$$f_{\text{OAR}}^{\nu,\mu}(\boldsymbol{x}) = \frac{1}{N_{\text{OAR}}} \sum_{i \in \text{OAR}} w_i^{\text{OAR}} \left[d_i^{\nu\%}(\boldsymbol{x}) - \mu_i \right]_+^2, \tag{14}$$

where w_i^{OAR} is the OAR voxel weight and N_{OAR} is the number of OAR voxels. The δ-factors are similarly calculated assuming that the β^{th}-percentile and ν^{th}-percentiles are written as

$$d_i^{\beta\%}(\boldsymbol{x}) = \mathbb{E}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] + \delta_i^{\beta\%}(\boldsymbol{x}) \,\text{SD}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right], \tag{15}$$

$$d_i^{\nu\%}(\boldsymbol{x}) = \mathbb{E}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] + \delta_i^{\nu\%}(\boldsymbol{x}) \,\text{SD}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right], \tag{16}$$

leading to the corresponding multiplicative factors $\delta_i^{\beta\%}(\boldsymbol{x}), \delta_i^{\nu\%}(\boldsymbol{x}) \in \mathbb{R}$. The probabilistic objectives in (11), (15) and (16) are convex if the δ-factors in each term are non-negative for all voxels in the corresponding structure (see section 4 of the SM). In the experiment for setup errors in X and Y (CTV-only, see figure 8(b)), a few δ-factors were slightly negative ($\delta > -4 \cdot 10^{-3}$). In all other experiments the δ-factors were non-negative, ensuring convexity.

The objectives in (10), (13) and (14) vanish for $d_i^{\alpha\%}(\boldsymbol{x}) > \gamma_i$, $d_i^{\beta\%}(\boldsymbol{x}) < \epsilon_i$ and $d_i^{\nu\%}(\boldsymbol{x}) < \mu_i$, respectively. To ensure that CTV and OAR voxels are always included in the optimization, we further optimize for the expected quadratic dose difference (we have a similar term for normal tissue) with a low weight, respectively as

$$f_{\text{CTV}}(\boldsymbol{x}) = \frac{1}{N_{\text{CTV}}} \sum_{i \in \text{CTV}} w_i^{\text{CTV}} \mathbb{E}\left[\left(d_i(\boldsymbol{x},\boldsymbol{\xi}) - d_i^{p} \right)^2 \right], \tag{17}$$

$$f_{\text{OAR}}(\boldsymbol{x}) = \frac{1}{N_{\text{OAR}}} \sum_{i \in \text{OAR}} w_i^{\text{OAR}} \mathbb{E}\left[\left(d_i(\boldsymbol{x},\boldsymbol{\xi}) \right)^2 \right], \tag{18}$$

$$f_{\text{Tissue}}(\boldsymbol{x}) = \frac{1}{N_{\text{Tissue}}} \sum_{i \in \text{Tissue}} w_i^{\text{Tissue}} \mathbb{E}\left[\left(d_i(\boldsymbol{x},\boldsymbol{\xi}) \right)^2 \right], \tag{19}$$

where d_i^{p} is the prescribed voxel dose, N_{Tissue} is the number of tissue voxels and w_i^{Tissue} is the tissue weight. The tissue objective in (19) aims to achieve dose conformity to the CTV. Voxel weights of CTV, OAR and Tissue are chosen based on clinical priority (see appendix).

The complete probabilistic (inner) optimization for a given set of δ-factors and objective weights $\Pi = \{\pi_{\text{CTV}}^{\alpha}, \pi_{\text{CTV}}^{\beta}, \pi_{\text{OAR}}^{\nu}, \pi_{\text{CTV}}^{\text{low}}, \pi_{\text{OAR}}^{\text{low}}, \pi_{\text{Tissue}}\}$ is given by

$$\begin{aligned} \min_{\boldsymbol{x}} \Big[& \pi_{\text{CTV}}^{\alpha} f_{\text{CTV}}^{\alpha,\gamma}(\boldsymbol{x}) + \pi_{\text{CTV}}^{\beta} f_{\text{CTV}}^{\beta,\epsilon}(\boldsymbol{x}) + \pi_{\text{OAR}}^{\nu} f_{\text{OAR}}^{\nu,\mu}(\boldsymbol{x}) \\ & + \pi_{\text{CTV}}^{\text{low}} f_{\text{CTV}}(\boldsymbol{x}) + \pi_{\text{OAR}}^{\text{low}} f_{\text{OAR}}(\boldsymbol{x}) + \pi_{\text{Tissue}} f_{\text{Tissue}}(\boldsymbol{x}) \Big] \end{aligned} \tag{20}$$

$$\text{s.t.} \quad d_i^{\alpha\%}(\boldsymbol{x}) = \mathbb{E}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] - \delta_i^{\alpha\%}(\boldsymbol{x}) \,\text{SD}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] \quad \forall i \in \text{CTV} \tag{21}$$

$$d_i^{\beta\%}(\boldsymbol{x}) = \mathbb{E}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] + \delta_i^{\beta\%}(\boldsymbol{x}) \,\text{SD}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] \quad \forall i \in \text{CTV} \tag{22}$$

$$d_i^{\nu\%}(\boldsymbol{x}) = \mathbb{E}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] + \delta_i^{\nu\%}(\boldsymbol{x}) \,\text{SD}\left[d_i(\boldsymbol{x},\boldsymbol{\xi})\right] \quad \forall i \in \text{OAR} \tag{23}$$

$$d_i(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{j \in \mathbb{B}} D_{ij}(\boldsymbol{\xi}) x_j, \; x_j \geqslant 0, \quad \forall j \in \mathbb{B}, \tag{24}$$

where the probabilistic objective weights Π were determined empirically, depending on the phantom experiment, aiming to balance target coverage and tissue sparing.

We solve the optimization using the interior-point method provided by `fmincon` in Matlab (The MathWorks, Inc. 2024), with an optimality tolerance of 10^{-8}. To make sure the optimality tolerance is reached before the step- and function tolerance, we define the latter two to be 10^{-25}. The analytical gradient and Hessian of the objective in (20) are derived in section 4 of the SM.

2.2.2. The outer optimization loop

The outer loop is responsible for updating the percentile estimates (by the δ-factors), so that $\mathbb{E}[d_i(\boldsymbol{x},\boldsymbol{\xi})] \pm \delta \cdot \mathrm{SD}[d_i(\boldsymbol{x},\boldsymbol{\xi})]$ remains an accurate estimation of the percentile (obtained from PCE sampling). For current iteration k, the inner optimization starts with initial guess $\boldsymbol{x}_{\mathrm{init}}$ and results in beam weight $\boldsymbol{x}_*^k$, which in general is significantly different from $\boldsymbol{x}_{\mathrm{init}}$. As a result, the PDF associated to each voxel dose (and thus the voxel dose percentiles) may have changed. We warm-start the next (inner) optimization (iteration $k+1$) using the previous initial ($\boldsymbol{x}^k$) and final ($\boldsymbol{x}_*^k$) beam weights as

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \kappa \cdot \left(\boldsymbol{x}_*^k - \boldsymbol{x}^k\right), \tag{25}$$

where the damping factor $\kappa = 0.2$. Once the beam weights are updated, in the outer optimization loop the new δ-factors are determined by (21), (22) and (23), making use of the fast PCE sampling. The dampening of the beam weights implicitly dampens the δ-factors as well (i.e. the percentiles that are optimized for), because the δ-factors depend on the beam weights.

The outer optimization loop is terminated when the voxel dose percentiles corresponding to the damped pencil-beam weights converge for all CTV and OAR voxels. Since the voxel dose percentiles are determined by PCE sampling, a sampling noise is involved (which is propagated to the δ-factors). As a result, the percentiles can only converge within a tolerance that is larger than the sampling noise. We consider the percentiles to be converged if their trend does not change within a certain tolerance. To quantify this trend, we smooth the percentiles at iteration k associated with the damped beam weight $\boldsymbol{x}^k$, using a moving average (MA) of window ΔW (denoted as MAΔW). For the lower percentile of the CTV (as in (21)) this is done as

$$d_i^{\alpha\%,\mathrm{MA}\Delta W}\left(\boldsymbol{x}^k\right) = \frac{1}{\Delta W} \sum_{t=k-\Delta W+1}^{k} d_i^{\alpha\%}\left(\boldsymbol{x}^t\right), \qquad \forall i \in \mathrm{CTV}, \tag{26}$$

for $k \geqslant \Delta W$. We define the convergence criteria such that the relative change of $d_i^{\alpha\%,\mathrm{MA}\Delta W}(k)$ within Δk iterations is smaller than $\tau_{\mathrm{CTV},\alpha}$ for all CTV voxels, i.e. the convergence criterion for probability level α is given by

$$\left| \frac{d_i^{\alpha\%,\mathrm{MA}\Delta W}(k) - d_i^{\alpha\%,\mathrm{MA}\Delta W}(k-\Delta k)}{d_i^{\alpha\%,\mathrm{MA}\Delta W}(k)} \right| < \tau_{\mathrm{CTV},\alpha}, \qquad \forall i \in \mathrm{CTV}, \tag{27}$$

for $k \geqslant \Delta W + \Delta k$. The same convergence criteria are applied to $d_i^{\beta\%,\mathrm{MA}\Delta W}$ and $d_i^{\nu\%,\mathrm{MA}\Delta W}$ for all voxels in the structure, with corresponding convergence tolerances of $\tau_{\mathrm{CTV},\beta}$ and $\tau_{\mathrm{OAR},\nu}$.

2.2.3. Initialization of the probabilistic optimization

As will be shown in section 3.1, the probabilistic optimization is closely related to a PTV-based optimization with a spherically symmetric dose distribution and static dose cloud approximation. Therefore, it serves as a good initial estimate to warm-start the probabilistic optimization. The PTV is defined by extending the CTV by a PTV-margin M_{PTV} isotropically. The PTV-optimization is initialized by using uniform beam weights (all 0.01) and minimizes for the quadratic difference of the nominal voxel dose $d_i^{\mathrm{nom}}(\boldsymbol{x})$ and the prescribed dose d_i^p as $\sum_{i\in\mathbb{V}} w_i \left(d_i^{\mathrm{nom}}(\boldsymbol{x}) - d_i^p\right)^2$. The resulting output beam weights $\boldsymbol{x}_{\mathrm{PTV}}$ are used in the first $k=1$ iteration of the probabilistic optimization, i.e. $\boldsymbol{x}_{\mathrm{init}} = \boldsymbol{x}_{\mathrm{PTV}}$.

Before starting the inner optimization, we specify the voxel dose thresholds γ_i, ϵ_i and μ_i with the corresponding desired probability levels α, β and ν. Then, we determine the voxel dose percentiles by PCE, for the CTV and OAR voxels that correspond to the current beam weights (so for $k=1$ that is $\boldsymbol{x}_{\mathrm{PTV}}$). After the percentiles are converted to the δ-factors by (21)–(23), the probabilistic optimization is started.

2.3. Composite-wise robust optimization as comparison

Various types of robust treatment planning exist (Unkelbach and Paganetti 2018), where one can choose for voxel-wise, objective-wise or composite-wise mini–max robust optimization (for decreasing level of conservativeness). Therefore, we restrict ourselves to compare to the composite-wise approach. We robustly optimize for the worst-case scenario within scenario set S for CTV and OAR with a nominal tissue objective, as

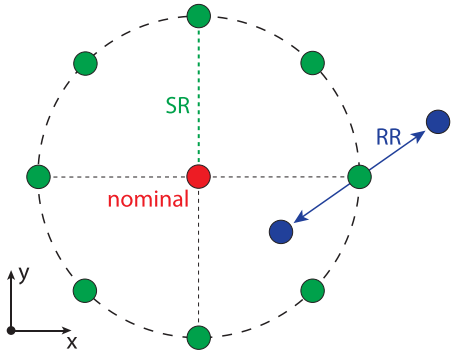

Figure 4. The robust error scenario set used during optimization. For every setup error with given setup robustness (SR) a range error with given range robustness (RR) is included. The nominal scenario (red) is always included.

$$\min_{\boldsymbol{x}} \Big[\max_{s\in S} \{\omega_{\mathrm{CTV}} f_{\mathrm{CTV}}(d(\boldsymbol{x},s)) + \omega_{\mathrm{OAR}} f_{\mathrm{OAR}}(d(\boldsymbol{x},s)) + \omega^{\max}_{\mathrm{OAR}} f_{\mathrm{maxOAR}}(d(\boldsymbol{x},s))\} \\ + \omega^{\mathrm{nom}}_{\mathrm{CTV}} f^{\mathrm{nom}}_{\mathrm{CTV}}(d(\boldsymbol{x})) + \omega_{\mathrm{Tissue}} f^{\mathrm{nom}}_{\mathrm{Tissue}}(d(\boldsymbol{x})) \Big] \tag{28}$$

$$\text{s.t.} \quad \boldsymbol{x} \geqslant 0,$$

$$f_{\mathrm{CTV}}(d(\boldsymbol{x},s)) = \frac{1}{N_{\mathrm{CTV}}} \sum_{i\in\mathrm{CTV}} w_i^{\mathrm{CTV}} \cdot \left(d_i(\boldsymbol{x},s) - d_i^p\right)^2,$$

$$f_{\mathrm{OAR}}(d(\boldsymbol{x},s)) = \frac{1}{N_{\mathrm{OAR}}} \sum_{i\in\mathrm{OAR}} w_i^{\mathrm{OAR}} \cdot \left(d_i(\boldsymbol{x},s) - d_i^p\right)^2,$$

$$f_{\mathrm{maxOAR}}(d(\boldsymbol{x},s)) = \frac{1}{N_{\mathrm{OAR}}} \sum_{i\in\mathrm{OAR}} w_i^{\mathrm{OAR}} \cdot \left(d_i(\boldsymbol{x},s) - d_i^{\mathrm{maxOAR}}\right)_+^2,$$

$$f^{\mathrm{nom}}_{\mathrm{Tissue}}(d(\boldsymbol{x})) = \frac{1}{N_{\mathrm{Tissue}}} \sum_{i\in\mathrm{Tissue}} w_i^{\mathrm{Tissue}} \cdot \left(d_i^{\mathrm{nom}}(\boldsymbol{x}) - d_i^p\right)^2,$$

$$f^{\mathrm{nom}}_{\mathrm{CTV}}(d(\boldsymbol{x})) = \frac{1}{N_{\mathrm{CTV}}} \sum_{i\in\mathrm{CTV}} w_i^{\mathrm{CTV}} \cdot \left(d_i^{\mathrm{nom}}(\boldsymbol{x}) - d_i^p\right)^2,$$

where the set of objective weights is given by $\Omega = \{\omega_{\mathrm{CTV}}, \omega_{\mathrm{OAR}}, \omega^{\max}_{\mathrm{OAR}}, \omega^{\mathrm{nom}}_{\mathrm{CTV}}, \omega_{\mathrm{Tissue}}\}$ and determined as explained in section 2.5.3. Large OAR voxel doses are penalized by a piecewise quadratic dose difference between the scenario voxel dose $d_i(\boldsymbol{x},s)$ and the voxel dose threshold d_i^{maxOAR}. As composite-wise robust optimization risks to neglect scenarios involving the smaller displacements from the nominal scenario (Fredriksson and Bokrantz 2016, Bokrantz and Fredriksson 2017), the importance of the nominal scenario (tuned by objective weight $\omega^{\mathrm{nom}}_{\mathrm{CTV}}$) is increased for some phantom experiments.

Scenario set S is commonly defined as illustrated in figure 4, consistent with approaches implemented in commercial treatment planning systems, such as by RayStation (RaySearch Laboratories 2023). For setup errors in X and Y, scenarios are included at primary axes corresponding to the used setup robustness (SR). Scenarios in the $(\pm1, \pm1)$ direction of the XY-plane are included and lie on the circle defined by the SR. For every setup scenario, range error scenarios are included with a fixed range robustness (RR). The nominal scenario is always included.

2.4. Phantom geometries

Figure 5 shows the three-dimensional homogeneous (water) phantom geometry that is used in this work for the spherical (left) and spinal (right) case in the XZ-plane. The spherical geometry has dimensions $(L_x, L_y, L_z) = (45\text{ mm}, 45\text{ mm}, 130\text{ mm})$ and consists of $1 \times 1 \times 1$ mm^3 voxels. As the ROI in the z-direction is limited from $L_z = 85$ mm to 130 mm, the number of considered voxels is $N_v^{\mathrm{spheres}} = 91.125$. The spherical CTV has radius $r_{\mathrm{CTV}} = 9$ mm. As a first case, only the spherical CTV is considered (referred to as CTV-only), i.e. no OARs are included so that the remaining volume is normal tissue. For the CTV+OAR case, we include a single OAR to the geometry. We distinguish between the XZ-displaced OAR (radius $r^{XZ}_{\mathrm{OAR}} = 9$ mm) and X-displaced OAR (radius $r^{X}_{\mathrm{OAR}} = 5$ mm), which centers are respectively located at $(x,y,z)_{XZ} = (44.5\text{ mm}, 22.5\text{ mm}, 129.5\text{ mm})$ and $(x,y,z)_{X} = (44.5\text{ mm}, 22.5\text{ mm}, 107.5\text{ mm})$.

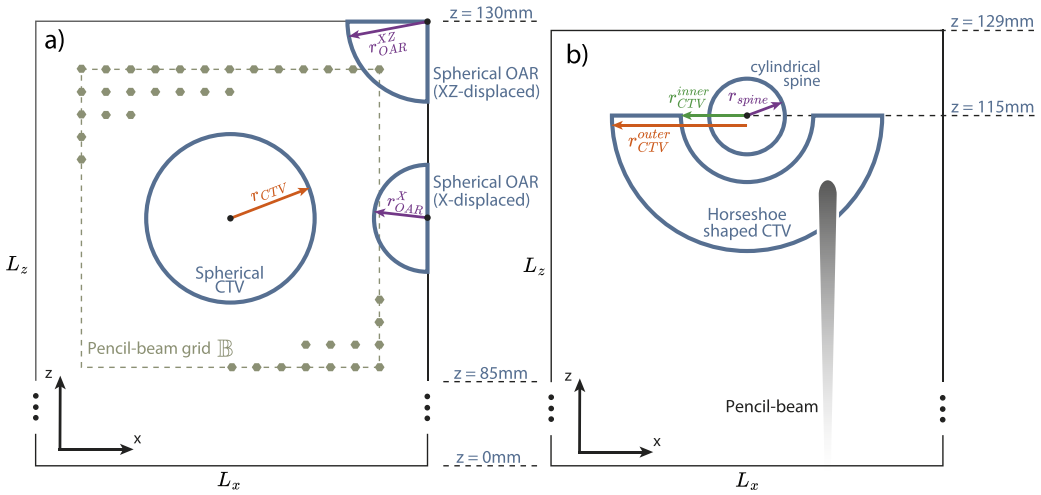

Figure 5. Illustration of the three-dimensional homogeneous (water) phantom geometries used in this work, for the (a) spherical and (b) spinal case in the XZ-plane. Dimensions are $(L_x, L_y, L_z) = (45\text{ mm}, 45\text{ mm}, 130\text{ mm})$ and $(70\text{ mm}, 30\text{ mm}, 129\text{ mm})$ for the respective cases. Voxel grids are defined for $z > 85$ mm, with respectively $1 \times 1 \times 1$ mm^3 and $2 \times 2 \times 2$ mm^3 voxels. Pencil-beam $j \in \mathbb{B}$ is directed at the pencil-beam spot in the grid and travels in the positive z-direction.

As only part of the OARs fall within the ROI, only a quarter of the XZ-displaced and half of the X-displaced OAR is included. The remaining part of the geometry is normal tissue. The CTV, OAR and normal tissue voxels form the ROI $\mathbb{V}$, i.e. $\mathbb{V} = \text{CTV} \cup \text{OAR} \cup \text{Tissue}$.

The spinal geometry has dimensions $(L_x, L_y, L_z) = (70\text{ mm}, 30\text{ mm}, 129\text{ mm})$ consisting of $2 \times 2 \times 2\,\text{mm}^3$ voxels, thus $N_v^{\text{spinal}} = 11.550$. It consists of a cylindrical spine that is surrounded by a horseshoe shaped CTV. The spine has radius $r_{\text{Spine}} = 6$ mm and is positioned parallel to the y-direction along the entire geometry at $(x, z) = (35\text{ mm}, 115\text{ mm})$. Parallel to the spine lies the CTV (for 9 mm $< y <$ 21 mm), extending from $r_{\text{CTV}}^{\text{inner}} = 12$ mm to $r_{\text{CTV}}^{\text{outer}} = 24$ mm in the XZ-plane for $z < 115$ mm. The remaining part of the geometry is normal tissue.

In both cases, the CTV is irradiated by pencil-beams traveling in the positive z-direction. In the spherical geometry we use 13 pencil-beam spots in each direction (spaced 3 mm apart), so that the number of pencil-beam spots $N_b^{\text{spherical}} = 13^3 = 2197$. The center pencil-beam spot is located at the CTV center, such that the grid extends 18 mm from the CTV center in all directions. In the spinal geometry, the number of pencil-beam spots in the respective directions are $(N_b^x, N_b^y, N_b^z) = (21, 9, 13)$, spaced 3 mm apart ($N_b^{\text{spinal}} = 2457$). Dose dependencies in this work are obtained by analytical approximations of the Bragg curve (Bortfeld 1997), where dose values below 0.01% of the maximum in D_{ij} are neglected. Each beam has a Gaussian-distributed lateral profile that is assumed to have an energy-independent initial width of $\sigma_b = 3$ mm, increasing in depth.

2.5. Probabilistic evaluation of treatment plans

2.5.1. Cost accuracy analysis of the PCEs

To ensure that the dose approximation is sufficiently accurate in the relevant uncertainty domain during optimization, a cost-accuracy analysis of the PCE is done in section 1 of the SM. For this purpose, a Γ-evaluation (Biggs *et al* 2022) is done (with distance-to-agreement 0.1 Gy and 1% dose difference criteria) for two different pencil-beams (and voxel doses $\geqslant 0.1$ Gy), in 123 different error scenarios that lie within the 99% confidence ellipsoid of the input phase space (taking into account all uncertain variables simultaneously). For every scenario in the 99%-ellipsoid we check the accepted voxel fraction.

Moreover, the PCE accuracy is quantified by determining the dose difference between the PCE and the dose engine for both test pencil-beams. For all test scenarios within the 99% confidence ellipsoid, we determine the minimum voxel dose difference among the 2% of the voxels having the largest dose difference, which we denote by $\Delta D_{2\%}$. Then, we calculate the scenario fraction for which the $\Delta D_{2\%}$ is larger than a certain dose value. Moreover, we determine the voxel dose difference averaged over all test scenarios (denoted by ΔD), and check what voxel fraction has ΔD larger than a certain dose value.

After the optimization is done, treatment plan quality is checked by constructing an independent PCE of the voxel dose, in line with previous work (Rojo-Santiago *et al* 2021, 2023a, 2023b, 2024, Oud *et al* 2024). As the voxel dose distribution for a single pencil-beam is different than for the final

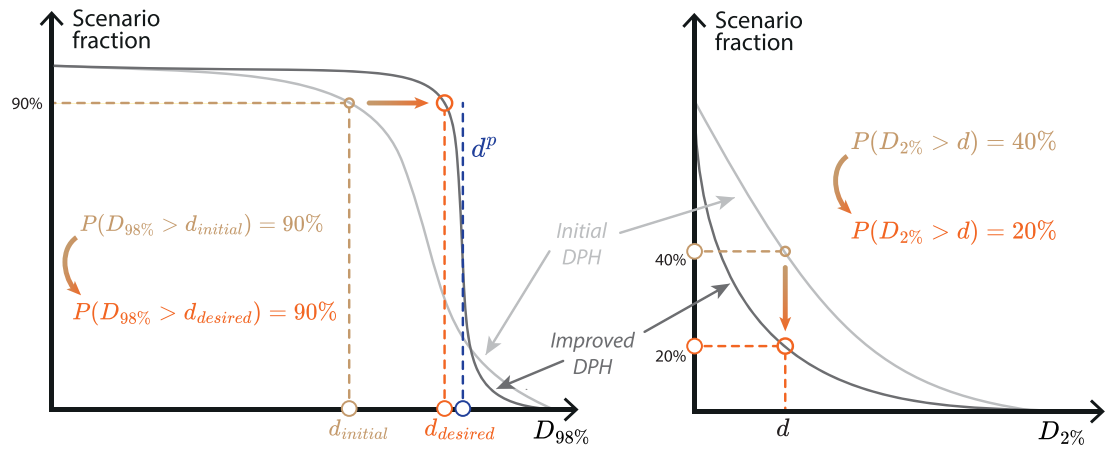

Figure 6. Illustration of dose population histograms for the CTV metric $D_{98\%}$ (left) and OAR metric $D_{2\%}$ (right), showing the fraction of error scenarios that exceed a given dose level (e.g. 90% of scenarios yield $D_{98\%} \geqslant d_{\text{desired}}$, where d_{desired} may for example be 95% of prescription dose d^p).

treatment plan, the necessary PCE accuracy is determined by a separate Γ-analysis (with distance-to-agreement 0.1 Gy and 1% dose difference criteria for voxel doses $\geqslant$ 0.1 Gy).

2.5.2. Probabilistic evaluation metrics

In the following we discuss the probabilistic evaluation metrics that are used to get insights in the probabilistic outcomes of the treatment plans (Perkó *et al* 2016). The well-known DVH can be probabilistically extended towards the DVH-distribution (Trofimov *et al* 2012), where the plan robustness is captured by the width of the DVH bands, representing confidence intervals of the DVHs. For example, the 95% confidence band is defined such that in 95% of the error scenarios—or in other words with 95% probability—the DVH-curves lie within the 2.5^{th} and 97.5^{th} percentile of the dose value.

Probabilities of voxel-wise under- and overdosage can be shown to understand to which extent the treatment plan reaches the probabilistic objectives. The probabilities are obtained by counting the fraction of error scenarios for which a voxel is below or above the desired threshold.

A more complete understanding can be obtained by the dose population histogram (DPH) (van Herk *et al* 2000), which is strongly related to the cumulative distribution function (CDF). It shows the (error) scenario fraction (i.e. probability) in which a dose metric of interest exceeds a given dose level. An illustrative example of the DPH is shown in figure 6 for the $D_{98\%}$ (left) and $D_{2\%}$ (right), but other dose metrics may be used. For DVH-metrics of the CTV (e.g. $D_{98\%}$) we aim to have steep DPH curves, ideally being a step function that starts at a scenario fraction of 100% and drops to 0% at prescription dose d^p. This ideal shape corresponds to a perfectly homogeneous dose distribution, because in all error scenarios the $D_{98\%}$ is exactly equal to d^p. In practice, some scenarios would result in the $D_{98\%} > d^p$ (as a result of partially overdosing the CTV), or would result in $D_{98\%} < d^p$ (as a result of partially underdosing the CTV). The corresponding DPH curve would be less steep and deviates from the ideal curve. In figure 6, the initial DPH curve shows that 90% of the error scenarios has a $D_{98\%}$ of at least d_{initial}. For the same fraction of error scenarios, the minimum $D_{98\%}$ can be increased (i.e. improved) towards $d_{\text{desired}} > d_{\text{initial}}$ (in practice, d_{desired} is for example 95% of the prescription dose), such that 90% of the error scenarios has a $D_{98\%}$ of at least d_{desired}.

For OARs we aim to have DVH-metrics (e.g. $D_{2\%}$) that have low dose values in most scenarios, which correspond to DPH curves in the bottom left of the figure. By going from the initial to the improved DPH we improve the probability in which $D_{2\%} > d$.

2.5.3. Probabilistic scaling to compare probabilistic and robust plans

The probabilistic plans (section 2.2) are compared to robust optimizations (section 2.3). The comparison is done by obtaining a robust plan that has either 1) a *similar CTV coverage*, or 2) a *similar OAR dose*, as the probabilistic plan. The used robust objective weights are found by manual tuning, and by checking what combination results in similar CTV or OAR dose, depending on the comparison method. The CTV coverage is defined as the 10^{th} percentile of the $D_{98\%}$ (as used by Tilly *et al* (2019), denoted as $D_{98\%}^{10\text{th}}$), which is the maximum $D_{98\%}$ value among the 10% of scenarios with the lowest $D_{98\%}$ values. Alternatively, when the OAR dose is matched, the 90^{th} percentile of the $D_{2\%}$ is used (denoted as $D_{2\%}^{90\text{th}}$),

which is the minimum $D_{2\%}$ value among the 10% of scenarios with the highest $D_{2\%}$ values. Thus, once the probabilistic plan is done, either its $D_{98\%}^{10\text{th}}$ or its $D_{2\%}^{90\text{th}}$ is determined, and robust plans are made and similarly evaluated using PCE.

As a next step, the probabilistic plan is scaled such that the 50^{th} percentile of $D_{50\%}$ (the median dose, an ICRU-recommended metric (The International Commission on Radiation Units and Measurements 2010)), matches the prescribed dose ($D_{50\%}^{50\text{th}} = 100\%\, d^p$). For the robust and probabilistic plans that are matched based on CTV coverage, the robust plan is scaled identically as the probabilistic plan, i.e. so that the 50^{th} percentile of the $D_{50\%}$ equals $100\%\, d^p$. For similar OAR dose, the robust plan is scaled to the probabilistic plan by matching the $D_{2\%}^{90\text{th}}$.

3. Results

In the following, results are first presented on the spherical CTV-only case (section 3.2) and compared with the VHMR. The spherical CTV+OAR case (section 3.3) and the spinal case (section 3.4) are shown as well. The optimization parameters for each geometry case are discussed in appendix and are summarized in tables A1 and A2 for the probabilistic and robust optimizations.

Expectation values in (11), (15) and (16) ($\mathbb{E}[D_{ij}(\boldsymbol{\xi})]$ and $\mathbb{E}[D_{ij}(\boldsymbol{\xi})D_{ij'}(\boldsymbol{\xi})]$) are calculated for CTV and OAR by Gauss–Hermite cubature (so without PCE) using 105 dose calculations. The PCE of the dose-influence matrix is constructed using 1637 dose calculations. Details on their cost-accuracy analyses are shown in section 1 of the SM.

3.1. Comparison with the VHMR

The PTV margin recipes (van Herk *et al* 2000) are based on a spherical CTV with ideal dose conformation (i.e. spherical symmetry in the dose) and static dose cloud approximation (i.e. invariance of the dose distribution shape under uncertainties). As only systematic errors are considered in this work, we compare to the VHMR without random errors and maintain spherical symmetry by assuming setup errors in the X and Y directions (i.e. *setupXY*). The phantom itself is 3D.

The VHMR ensures that a certain patient population (e.g. 90%) receives at least a minimum CTV dose, i.e. $P(d_{\min} \geqslant 57\,\text{Gy}) \geqslant 90\%$. To compare with our probabilistic approach, we optimize to avoid voxel-wise CTV underdosage probability by prioritizing $P(d_i \geqslant 57\,\text{Gy}) \geqslant 98\%$. The resulting probabilistic plan is shown in figure 7 with the cross section along the x-axis, together with the CTV edge (dashed black), the Van Herk margin (M_{VH}, dashed orange), and the probabilistic margin ($M_{\text{probabilistic}}$, dotted brown) corresponding to the start of the dose fall-off at $x = 36$ mm. Plan evaluation by PCE sampling yields $P(d_{\min} \geqslant 57\,\text{Gy}) \approx 90.1\%$. The corresponding (2D) Van Herk margin is $M_{\text{VH}} = \Sigma_{\text{setup}}\sqrt{-2\cdot\ln(1-P)} \approx 2.15\Sigma_{\text{setup}}$, where $\Sigma_{\text{setup}} = 3$ mm and $P = 0.901$.

The Van Herk margin for systematic errors depends only on the Gaussian uncertainty, not on the dose-falloff, therefore assuming a hard cutoff at the CTV edge (i.e. the 57 Gy dose level was not considered). In an extreme scenario (e.g. a $2.15\Sigma_{\text{setup}}$ shift), the entire CTV still receives the full prescription dose. When the dose fall-off is included, the dose decreases gradually, so that after a $2.15\Sigma_{\text{setup}}$ shift, parts of the CTV edge still receive at least $\geqslant 95\%d^p$. The dose fall-off in the probabilistic plan is a 2D Gaussian with $\mu_{\text{lateral}} = 36$ mm and $\sigma_{\text{lateral}} = 5.6$ mm (dashed gray, see figure 7).

Accounting for this dose fall-off allows reducing the Van Herk margin to $M_{\text{VH}}^{\text{reduced}} = 4.6$ mm (see section 5 of the SM), showing that accounting for dose fall-off makes the Van Herk margin less conservative (but more realistic). In the probabilistic plan, this corresponds to the distance between the CTV edge (at $x = 31.5$ mm) and the start of the dose fall-off (at $x = 36$ mm), i.e. 4.5 mm. The 95% dose level in this reduced-margin model lies 6.4 mm beyond the CTV—nearly identical to the probabilistically obtained margin of 6.2 mm. This shows that full probabilistic planning and evaluation yield margins comparable to the Van Herk approach when dose fall-off is included, while automatically generalizing to dose fall-offs of any shape (as this is captured by the PCE model).

3.2. Probabilistic and robust plans for the spherical CTV-only case

For the spherical CTV-only case we consider two combinations of systematic uncertainties: (1) setup errors in X and Y (i.e. *setupXY*) and (2) setup errors in X and Y together with range errors (i.e. *setupXYrange*).

3.2.1. Probabilistic and robust optimization

The probabilistic optimization minimizes for CTV underdosage ($P(d_i \leqslant 0.95\cdot d_i^p) \leqslant 10\%$) and overdosage ($P(d_i \geqslant 1.07\cdot d_i^p) \leqslant 10\%$) probabilities, at the same time pushing the expected dose in the CTV to $d_i^p = 60$ Gy. As a comparison to the probabilistic plan, we perform a robust optimization for the

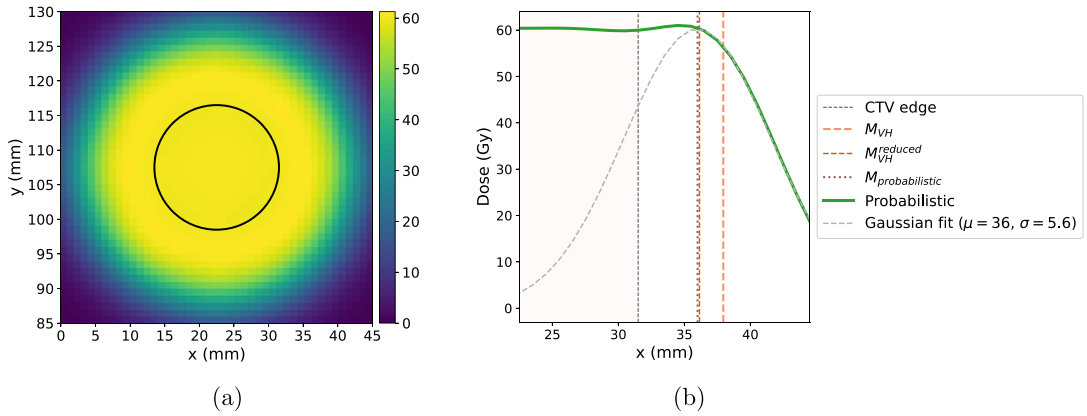

Figure 7. The probabilistic plan (*setupXY*) as a comparison with the Van Herk margin recipe, where avoiding CTV underdosage probability was prioritized. The (a) nominal dose distribution in the *XY*-plane is shown with the (b) cross section through the CTV center along the *x*-axis.

Table 1. Comparison of the CTV coverage in the robust plans for a setup robustness (SR) ranging from 4 mm till 7 mm versus the probabilistic plan in the CTV-only *setupXY* case.

Treatment Plan	$D_{98\%}^{2nd}$ (Gy)	$D_{98\%}^{5th}$ (Gy)	$D_{98\%}^{10th}$ (Gy)	$P(D_{98\%} \geqslant 0.95 \cdot d^p)$
SR of 4 mm	35.95	41.36	46.30	55%
SR of 5 mm	43.98	48.70	52.55	75%
SR of 6 mm	49.36	53.18	55.96	87%
SR of 7 mm	53.70	56.42	57.98	94%
Probabilistic plan	50.19	53.59	55.98	85%

Table 2. Comparison of the CTV coverage of the robust plan for a setup robustness (SR) of 6 mm and range robustness (RR) of 4% and 5%, versus the probabilistic plan in the CTV only case.

Treatment plan	$D_{98\%}^{2nd}$ (Gy)	$D_{98\%}^{5th}$ (Gy)	$D_{98\%}^{10th}$ (Gy)	$P(D_{98\%} \geqslant 0.95 \cdot d^p)$
SR/RR: 6 mm/4%	42.94	48.93	53.01	77%
SR/RR: 6 mm/5%	45.86	50.84	54.31	81%
Probabilistic plan	46.58	50.92	54.03	79%

CTV and use a nominal objective for tissue, with the same objective weights as in the probabilistic optimization.

Robust plans with setup errors ranging from 4 till 7 mm were made and compared to the probabilistic plan by matching their CTV coverage probability ($D_{98\%}^{10th}$). Additionally, we checked the $D_{98\%}^{2nd}$ and $D_{98\%}^{5th}$, and determined the probability of $D_{98\%}$ exceeding 95% of the prescribed dose d^p, i.e. $P(D_{98\%} \geqslant 0.95 \cdot d^p)$. As table 1 shows, the required SR to achieve similar CTV coverage ($D_{98\%}^{10th}$) as in the probabilistic *setupXY* plan, is SR = 6 mm. Figure 8(top) compares the dose distributions for the robust (SR = 6 mm) and probabilistic *setupXY* plans for the XY-plane ($z = 117.5$ mm). Additionally, cross sections through the CTV center for both plans along the *x*-axis are shown.

As the SR was tuned to the $D_{98\%}^{10th}$, the dose extension beyond the CTV is very similar for the probabilistic and robust plans. Moreover, both plans are very conformal to the CTV, because of the way the uncertainty set is defined. According to RaySearch Laboratories (2023), the uncertainty set for SR = 6 mm includes the nominal scenario and eight others located on a circle of radius SR: $(x, y) = (0, 0)$, $(x, y) = (\pm SR, 0)$, $(x, y) = (0, \pm SR)$ and $(x, y) = \left(\pm SR/\sqrt{2}, \pm SR/\sqrt{2}\right)$. No intermediate scenarios are included.

For SR = 6 mm, we proceed to determine the necessary RR to achieve comparable CTV coverage (where we use range errors additional to setup errors in the *X* and *Y*-directions). The results are listed in table 2, showing that an RR of 5% gives the closest match of CTV coverage to the probabilistic plan. The robust (SR/RR: 6 mm/5%) and probabilistic *setupXYrange* plans are compared in figure 8(bottom) for the *XZ*-plane ($y = 22.5$ mm). Cross sections through the CTV center along the diagonal ($z = x + 85$ mm) for $y = 22.5$ mm are shown as well.

Compared to the probabilistic plan, the robust plan shows a larger (more conservative) dose expansion along the diagonals of the *XZ*-plane. This is a consequence of the way the discrete error scenarios

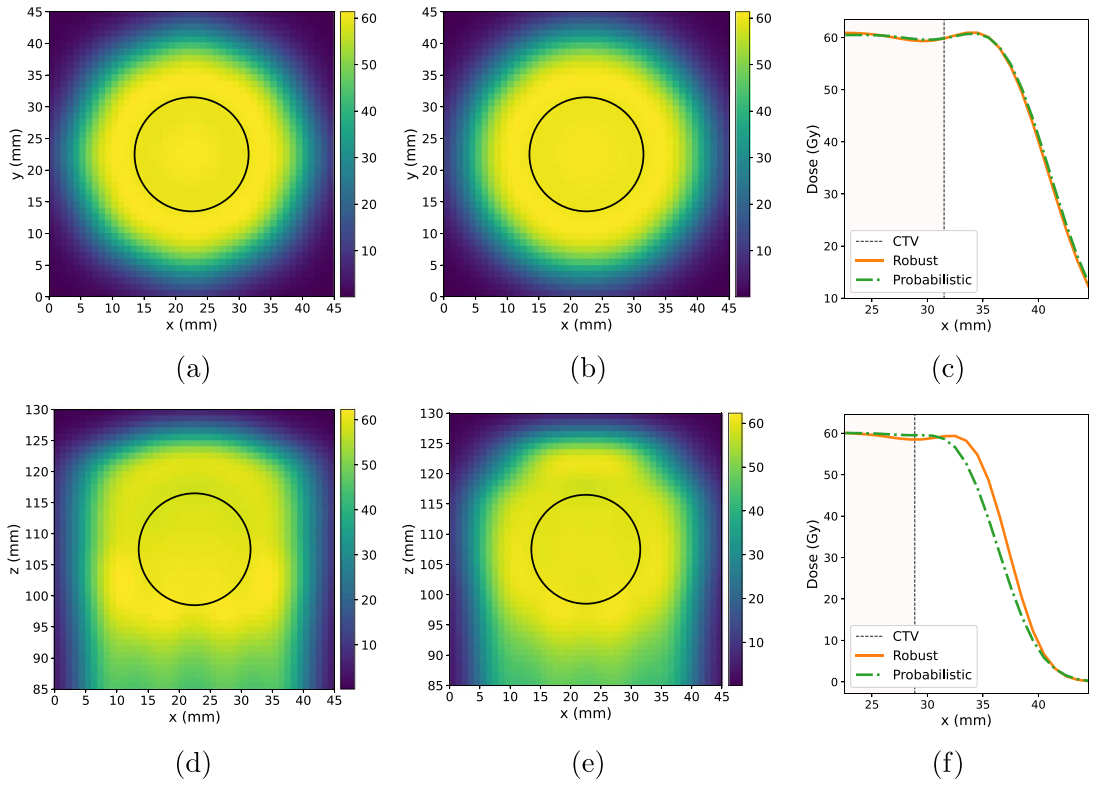

Figure 8. Nominal dose distributions for the (top) *setupXY* and (bottom) *setupXYrange* cases, shown in the *XY*-plane and *XZ*-plane through the CTV center, respectively. For each case: (left) the robust plan, (middle) the probabilistic plan, and (right) a cross-section through the CTV center (along the *x*-axis for *setupXY* and along the line $z = x + 85$ mm for *setupXYrange*) are shown.

Table 3. Statistical DVH-metrics to compare the CTV coverage ($D^{10\text{th}}_{98\%}$) and OAR overdosage ($D^{90\text{th}}_{2\%}$) of the robust and probabilistic plans for the *XZ*-displaced spherical CTV+OAR case. Metrics corresponding to the robust plans after scaling the beam weights are shown in brackets. The objective weights used in the robust plans are shown in the corresponding rows as $\{\omega_{\text{CTV}}, \omega_{\text{OAR}}, \omega^{\text{max}}_{\text{OAR}}, \omega_{\text{Tissue}}\}$.

	$D^{10\text{th}}_{98\%}$ (Gy)	$D^{50\text{th}}_{50\%}$ (Gy)	$D^{90\text{th}}_{2\%}$ (Gy)
Robust $\{120, 1, 1, 160\}$	52.4 (52.3)	60.1 (60.0)	40.8 (40.8)
Robust ($\{100, 10, 10, 100\}$)	45.4 (44.6)	60.2 (59.1)	27.6 (27.1)
Probabilistic	53.3 (53.3)	60.0 (60.0)	27.0 (27.1)

are constructed (see figure 4): scenarios that include a range-shift error are positioned farther from the nominal case than those without a shift. Probabilistic optimization, however, takes into account that large-shift scenarios are less probable to occur (compared to other scenarios in the set), resulting in a more conformal margin. In the *XY*-plane, where all scenarios in the set are equidistant from the nominal scenario (with distance SR), the dose distribution ends up as conformal as in the probabilistic plan.

3.3. The spherical CTV and OAR case

We perform similar probabilistic optimizations as in section 3.2, but here we additionally optimize probabilistically for a spherical OAR, either for the *XZ*-displaced case or for the *X*-displaced case. Accordingly, besides optimizing for CTV under- and overdosage probability, the OAR overdosage probability is limited as $P(d_i \geqslant 30\,\text{Gy}) \leqslant 10\%$. In the following, results for the *XZ*-displaced case are shown. Results of the *X*-displaced case are shown in section 2 of the SM. Robust plans are obtained by tuning its objective weights, to achieve 1) a similar CTV coverage ($D^{10\text{th}}_{98\%}$) and 2) a similar OAR dose ($D^{90\text{th}}_{2\%}$) as in the probabilistic plan. The robust plan with objective weights of $\{\omega_{\text{CTV}}, \omega_{\text{OAR}}, \omega^{\text{max}}_{\text{OAR}}, \omega_{\text{Tissue}}\} = \{120, 1, 1, 160\}$ and $\{100, 10, 10, 100\}$ give similar CTV coverage and OAR dose, respectively. The resulting probabilistic DVH-metrics before and after scaling (within brackets) are shown in table 3.

The dose distributions for the *XZ*-displaced case are shown in figure 9, where the probabilistic plan (left) is shown together with the robust plans, scaled by matching CTV coverage (middle) and OAR dose (right). We show the *XZ*-slice (top) and *XY*-slice (bottom) through the CTV center. The dose distribution of the probabilistic plan shows a slight reduction in the CTV margin on the OAR-side, at the same

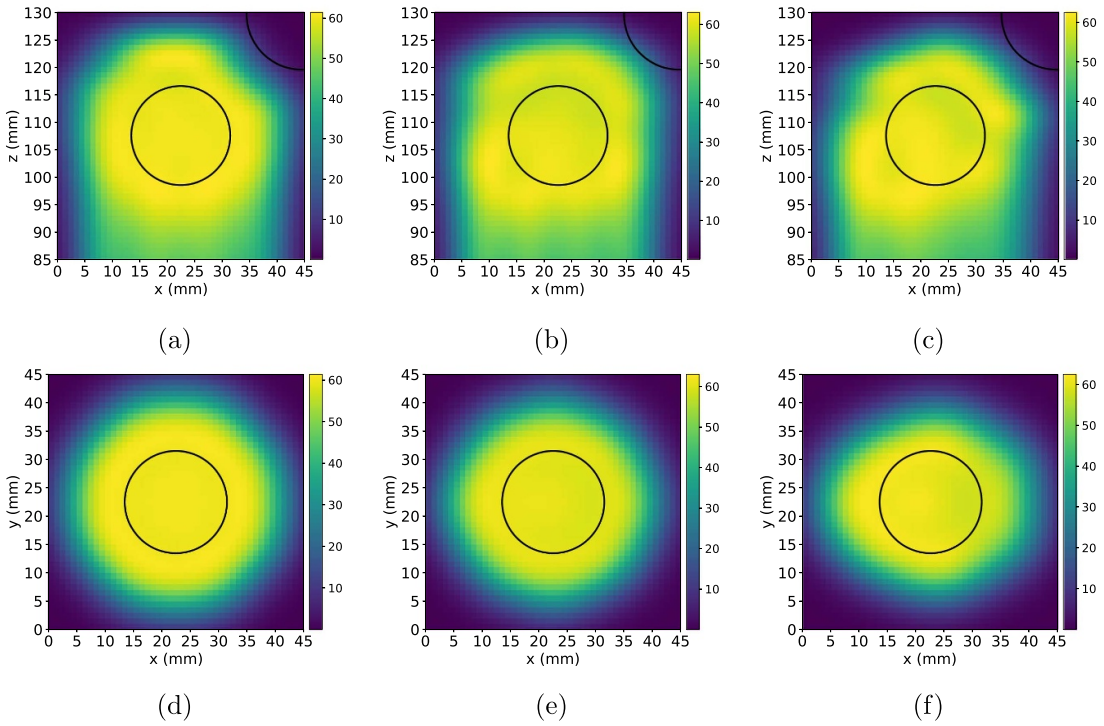

Figure 9. Comparison of nominal XZ-displaced dose distributions for the (left) probabilistic plan, with the robust plans that are matched based on the (middle) $D_{50\%}^{50\text{th}}$ and (right) $D_{2\%}^{90\text{th}}$ metrics. Both the (bottom) XY-plane and (top) XZ-plane through the CTV center are shown.

time being conformal to the other parts of the CTV. Compared to the probabilistic plan, the robust plan with similar CTV coverage (middle row) is less conformal in general and has larger margins on the OAR-side, leading to higher OAR doses.

By increasing the relative OAR objective weight (i.e. its importance in the optimization) with respect to the $D_{50\%}^{50\text{th}}$ comparison, the margin at the OAR-side is reduced, leading to a similar OAR dose ($D_{2\%}^{90\text{th}}$). At the same time, this leads to reduced CTV conformity in both the XY- and XZ-plane. Since the OAR has an increased importance, shifts into the OAR (i.e. diagonal shifts) become worst-case scenarios more frequently during the optimization. This has two main effects. Firstly, the CTV margin in the y-direction is reduced because y-shifts are less likely to be worst-case scenarios (diagonal shifts lead to more OAR overdosage than y-shifts). In turn, this leads to a significant drop in CTV coverage. Secondly, a dose extension appears on the opposite side of the OAR (around $(x, z) = (12.5\text{ mm}, 97.5\text{ mm})$), ensuring that CTV coverage remains in cases where high-dose regions are shifted into the OAR.

The CTV coverage of the plans is compared in figure 10, where the DPH of various DVH-metrics is compared between the probabilistic plan and either the (top) $D_{50\%}^{50\text{th}}$ or the (bottom) $D_{2\%}^{90\text{th}}$ scaled robust plan. For the former, CTV coverage (especially the high scenario fraction region) is similar and significant differences in the near-maximum OAR DVH-metrics can be seen. Specifically, in the probabilistic plan the $D_{2\%}$ is larger than 30 Gy in about 7.5% of the scenarios whereas this probability is increased to about 22.5% in the scaled robust case.

The $D_{2\%}^{90\text{th}}$ scaled robust plan has similar near-maximum OAR DVH-metrics, whereas significant difference in the CTV coverage is visible. In the scaled robust plan, $P(D_{98\%} > 57\,\text{Gy}) \approx 6\%$, whereas this probability increases to approximately 77% in the robust case (i.e. about 70% increase).

In figure 11, we compare the DVH bands of CTV (top) and OAR (bottom) between the probabilistic (left) and robust plans that are matched based on the (middle) $D_{50\%}^{50\text{th}}$ and (right) $D_{2\%}^{90\text{th}}$ metrics. The probabilistic plan shows more CTV homogeneity in the presence of uncertainty (as well as in the nominal plan) and shows a smaller spread in the DVH distributions of the CTV, compared to both robust plans. The scaled $D_{2\%}^{90\text{th}}$ robust plan was matched to the probabilistic plan by OAR dose, resulting in similar DVH-distributions for the OAR. Compared to the probabilistic plan, the DVH-distributions of the $D_{50\%}^{50\text{th}}$ scaled robust plan are wider (i.e. less robust) and shifted towards larger dose (i.e. increased OAR overdosage probability).

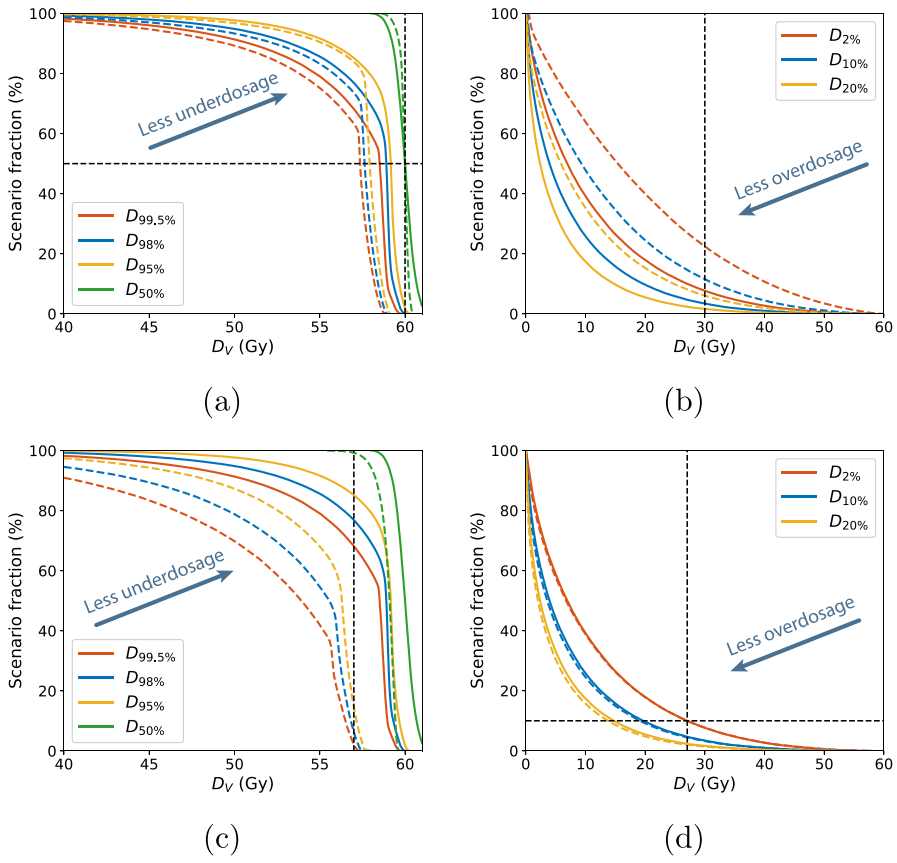

Figure 10. Dose population histograms of various DVH metrics (D_V, e.g. $V = 98\%$) comparing XZ-displaced probabilistic (solid) and robust (dashed) plans in the CTV (left) and OAR (right). Plans matched on CTV coverage (top) have similar CTV coverage (a) and reduced OAR overdosage probability (b), whereas matching on (bottom) $D_{2\%}^{90\text{th}}$ reduces CTV underdosage probability (c) with similar OAR dose (d).

3.4. The spinal case

We perform 3 optimizations for the spinal geometry as depicted as in figure 5, where the OAR overdosage (> 54 Gy) is prioritized over CTV coverage, with the 3 cases corresponding to optimizing for different probability levels for the OAR ($\nu = 90\%$, $\nu = 95\%$ and $\nu = 98\%$). Robust plans are optimized by matching their CTV coverage (and scaling the $D_{50\%}^{50\text{th}}$) to the probabilistic plan. The statistical DVH-metrics that are used for scaling are shown in table 4. Dose distributions of the $\nu = 90\%$ case are shown in this section, while the $\nu = 95\%$ and $\nu = 98\%$ cases are similar and are presented in section 3 of the SM.

The resulting dose distributions for $\nu = 90\%$ are shown in figure 12, where the XZ-slice (top) and XY-slice (bottom) through the spine center are shown, respectively. The probabilistic plans show improved CTV conformity compared to the robust plans, the difference being especially significant in the XY-plane. A dose build-up occurs at the inner CTV edges (on the spinal side) in order to reduce the spinal dose (at the same time preserving CTV coverage). This effect is seen more clearly in figure 13, where cross sections (for $\nu = 90\%$) along the X- and Z-axis passing through the spine center are shown, with the cross sections for the $\nu = 95\%$ and $\nu = 98\%$ plan comparisons for completeness. Compared to the robust plans, all probabilistic plans lead to reduction of spinal dose, at the expense of having more inhomogeneous CTV dose. As expected, the dose margin in the probabilistic plan is reduced if spinal overdosage is allowed in less error scenarios (i.e. increased ν). This is especially visible in the range direction (along the z-axis).

To understand to which extent the probabilistic objectives are reached in the $\nu = 90\%$, $\nu = 95\%$ and $\nu = 98\%$ case, we show the probability of underdosing and overdosing CTV and spinal voxels in figure 14(in the XZ-plane). The probabilities of CTV underdosage, CTV overdosage and spinal overdosage are shown from top to bottom, where the $\nu = 90\%$, $\nu = 95\%$ and $\nu = 98\%$ cases are shown from left to right. CTV voxels that do not reach the probabilistic objectives are red (i.e. are under- or overdosed in more than 10% of the scenarios). For the $\nu = 90\%$ plan, spinal overdosage probabilities are below 10% (i.e. the probability that was optimized for) for all voxels. Probability levels of overdosage in

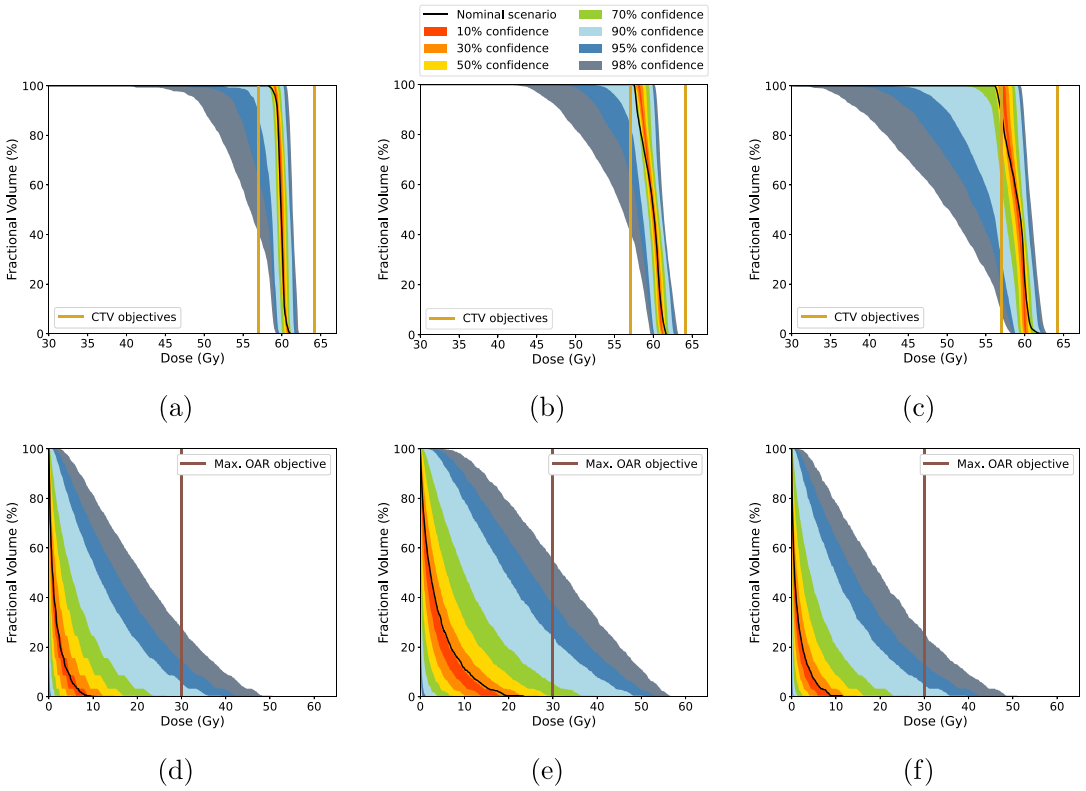

Figure 11. Dose volume histogram distributions for the CTV (top) and OAR (bottom), comparing the XZ-displaced probabilistic (left) and robust plans that are matched based on the (middle) $D_{50\%}^{50\text{th}}$ and (right) $D_{2\%}^{90\text{th}}$ metrics. The nominal scenario (black) is shown together with various confidence bands.

Table 4. Statistical DVH-metrics to compare the CTV coverage and spinal overdosage of the robust and probabilistic spinal plans. Metrics corresponding to the robust plans after scaling the beam weights are shown in brackets.

	$D_{98\%}^{10\text{th}}$ (Gy)	$D_{50\%}^{50\text{th}}$ (Gy)	$D_{2\%}^{90\text{th}}$ (Gy)
Robust (equivalent of $\nu = 90\%$)	33.1 (33.1)	60.0 (60.0)	54.2 (54.3)
Probabilistic ($\nu = 90\%$)	32.0 (31.6)	60.7 (60.0)	54.8 (54.1)
Robust (equivalent of $\nu = 95\%$)	30.4 (30.4)	59.9 (60.0)	52.9 (52.9)
Probabilistic ($\nu = 95\%$)	31.0 (30.6)	60.8 (60.0)	52.7 (52.1)
Robust (equivalent of $\nu = 98\%$)	27.5 (27.6)	59.9 (60.0)	51.2 (51.3)
Probabilistic ($\nu = 98\%$)	27.9 (27.5)	60.9 (60.0)	48.7 (48.0)

the spinal edge voxels reach approximately 5% and 2% for the $\nu = 95\%$ and $\nu = 98\%$ plans, respectively. This shows that the probabilistic approach allows for tuning desired probability levels.

A comparison of the DVH-metrics is done in figure 15, where the DPH of various DVH-metrics is shown for the probabilistic and robust ($\nu = 90\%$) plans. As both plans are matched by the CTV coverage $D_{98\%}^{10\text{th}}$, both plans are similar the region of large scenario fraction (i.e. in the lower tails of the DVH-metrics). This means that in the worst error scenarios (where the CTV receives the lowest dose), the plan quality is similar. Also, the region of low scenario fraction of the spinal DVH-metrics (e.g. $D_{2\%}^{90\text{th}}$) is similar between both plans. All DPHs deviate significantly in the central region of the distributions. For the majority of the dose values, the probabilistic plan shows smaller probability of CTV underdosage (e.g. $P(D_{98\%} > 57\,\text{Gy})$ increased by about 14%) and spinal overdosage (e.g. the $P(D_{2\%} > 30\,\text{Gy})$ is reduced by about 28.5%) in the same plan.

DVH bands of the CTV and spine are compared between the probabilistic (top) and robust (bottom) plans in figure 16 for the $\nu = 90\%$ case. The probabilistic CTV objectives at $0.95 \cdot d_p = 57\,\text{Gy}$ and $1.07 \cdot d_p = 64.2\,\text{Gy}$ are shown, together with the maximum OAR dose threshold at 54 Gy. As the probabilistic plan focused on preventing spinal overdosage probability, the spinal DVH has lower dose values associated to the same fractional volume, but similar DVH spread. Only for approximately 5% of the scenarios (corresponding to the upper bound of the 90% confidence level), the robust plan reaches lower

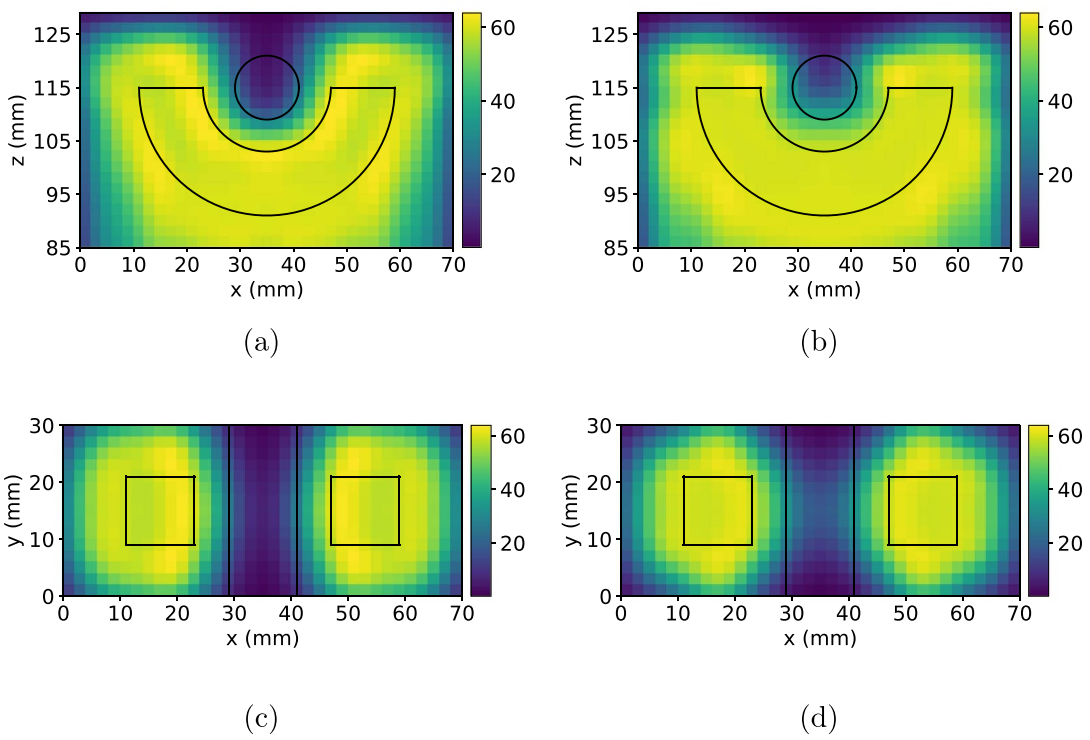

Figure 12. Comparison of nominal dose distributions for the (left) probabilistic and (right) robust plan in the spinal case ($\nu = 90\%$). Both the (bottom) XY-plane and (top) XZ-plane through the CTV center are shown.

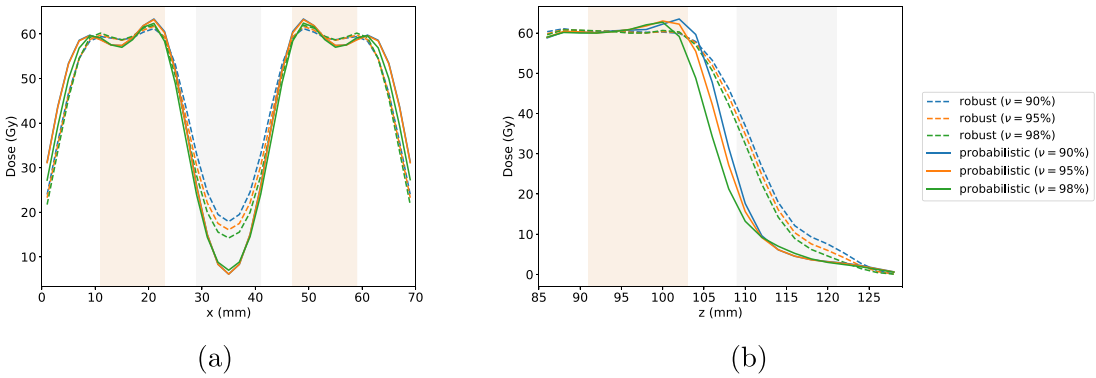

Figure 13. Cross sections through the spine center along the X and Z axes for the probabilistic ($\nu = 90\%$, $\nu = 95\%$ and $\nu = 98\%$) and robustly matched plans.

doses than in the probabilistic plan. Improvement in the spine is possible at the expense of a less homogeneous CTV, with a slight overdosage (as allowed for). Additionally, the DVH bands are wider, though they remain within the desired thresholds.

4. Discussion

In this work we present a proof-of-concept of a novel approach to probabilistic treatment planning, that allows for the precise tuning of voxel-wise under- and overdosage probabilities.

Probabilistic planning—based on re-optimization after probabilistic evaluation—has shown potential to reduce inter-patient variation and improve trade-offs between target coverage and OAR sparing (de Jong *et al* 2025). Earlier developed probabilistic approaches for IMRT optimize for minimum dose objectives (Gordon *et al* 2010), which was later extended to minimum dose constraints (Mescher *et al* 2017). Approaches for probabilistic target coverage have been proposed as well (Tilly *et al* 2019), where the dose-volume metric (e.g. $D_{98\%}$) is approximated by a surrogate function. Our approach adds to these previous developments by optimizing probabilistic voxel dose directly in a convex formulation when

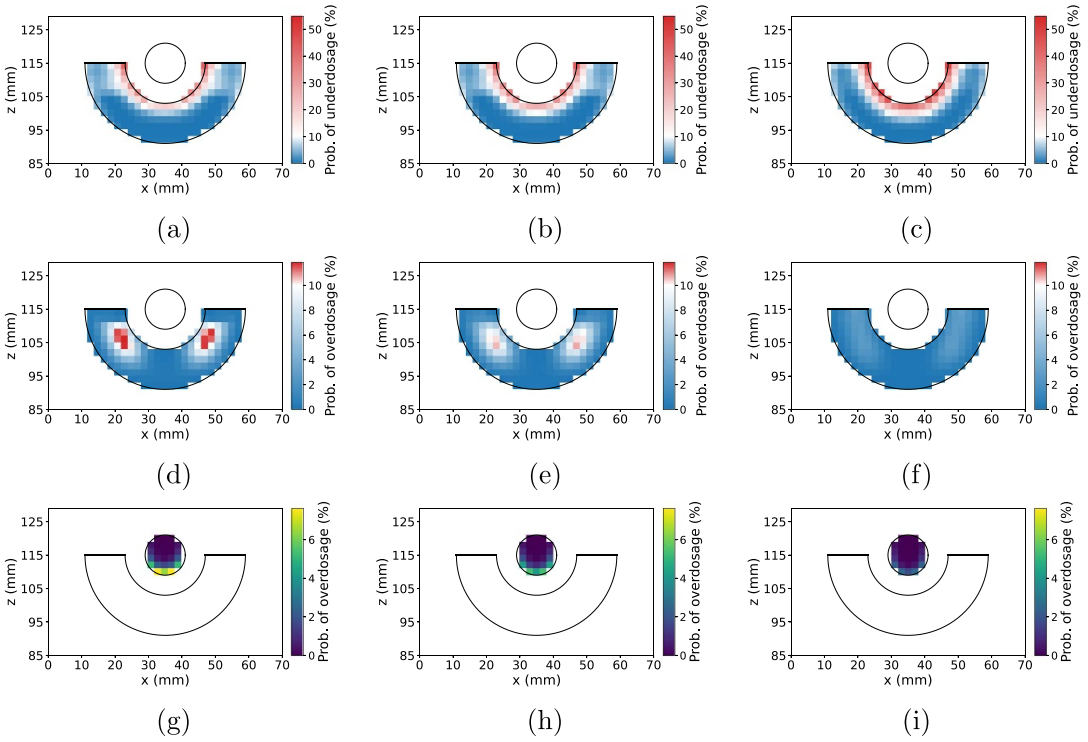

Figure 14. The probability of CTV underdosage (top), CTV overdosage (middle) and spine overdosage (bottom) probability, for the (left to right) $\nu = 90\%$, $\nu = 95\%$ and $\nu = 98\%$ cases. Probabilistic CTV objectives are reached if the probability of under- and overdosage is below 10% (blue).

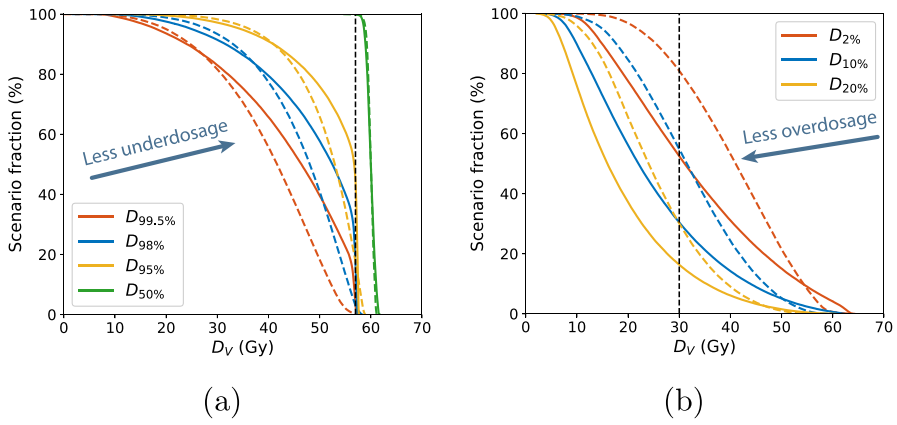

Figure 15. Dose population histograms of various DVH metrics (D_V, e.g. $V = 98\%$) comparing spinal probabilistic (solid) and robust (dashed) plans for $\nu = 90\%$ in the (a) CTV and (b) spine, scaled by $D_{50\%}^{50\text{th}}$. The probabilistic plan shows reduced spinal overdosage probability and reduced CTV underdosage probability in the same plan.

$\delta > 0$ for all active voxels. Since the coverage probability and dose threshold can be selected without proxies, the approach is personalized and not tied to patient populations and tumor sites. The resulting formulation is conceptually similar to the VHMR but is based on fewer assumptions, as it does not rely on spherical target geometry, the static dose cloud, or uniform population-based uncertainty (i.e. the same SD of setup or range uncertainties for all patients).

The approach is sufficiently general to be used for different purposes, e.g. to optimize for probabilistic minimum, maximum or mean dose. By associating the δ-factors to CVaR estimates rather than percentile estimates, our approach can likewise optimize the CVaR of the voxel dose (as was used in Tilly *et al* (2019)). In future work, we can explore a broader range of probabilistic objectives, including commonly used radio-biological metrics, such as generalized equivalent uniform dose, BED and tumor control probability (van Haveren and Breedveld 2019).

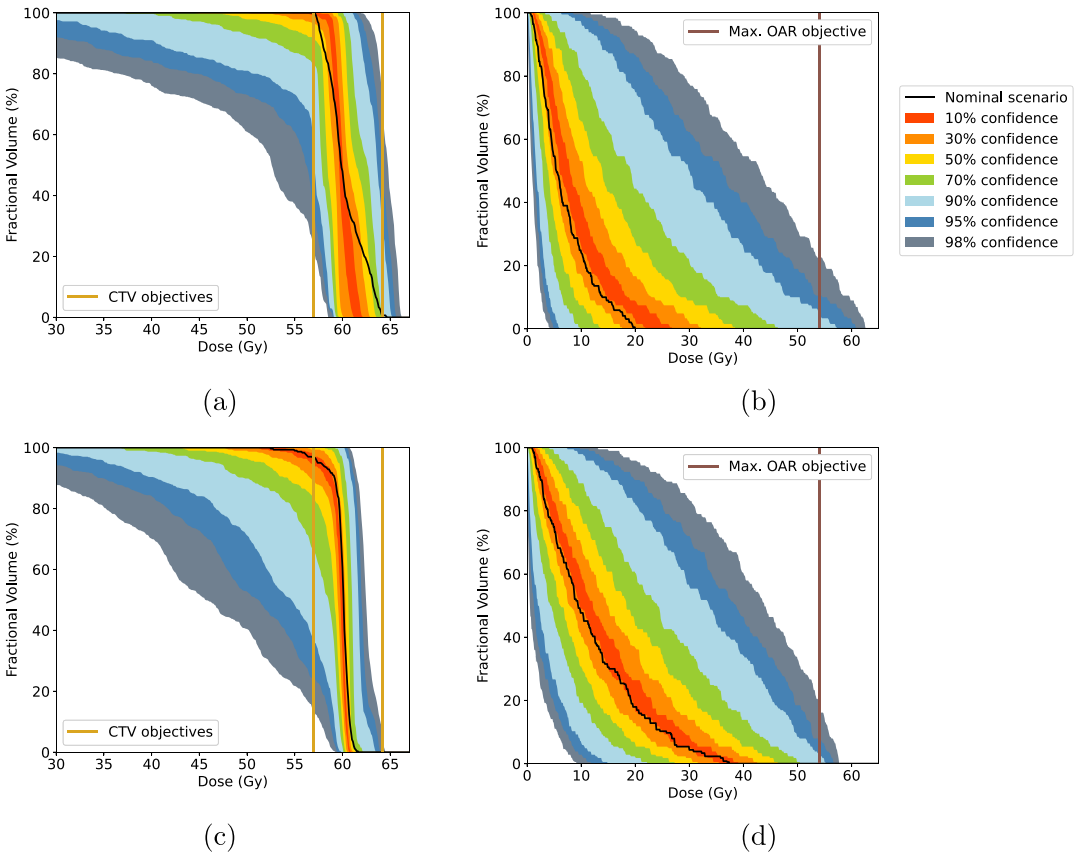

Figure 16. Dose volume histogram distributions for the CTV (left) and spine (right) for the $\nu = 90\%$ plans. Probabilistic (top) and robust (bottom) plans are compared, showing the nominal scenario (black) with various confidence bands.

Percentile optimization has been approximated in previous studies (Chu *et al* 2005, Sobotta *et al* 2010, Fabiano *et al* 2022) by assuming Gaussian-distributed responses, which does not allow for precise tuning towards desired probabilities or percentiles. Our approach directly applies to non-Gaussian response distributions, because PCE allows to model non-Gaussian responses and its efficient sampling allows to yield sufficient statistics for accurate percentile prediction. This was demonstrated in the spinal case, where the probability of spinal overdosage in the treatment plan matched the desired values or reached values below the threshold.

PCE is not intrinsic to the current approach, but other fast methods can be used, such as Monte-Carlo based methods (Holmes *et al* 2024), potentially using artificial intelligence (Pastor-Serrano and Perkó 2022). As sampling thousands of scenarios is preferable for obtaining accurate statistical estimations, the efficiency of the alternatives should be comparable to PCE.

Although improved robust plans (e.g. in terms of conformity) can possibly be achieved by addition of other objectives or other types of robust optimization (e.g. objective-wise), the spherical and spinal probabilistic plans consistently show that OAR/spinal overdosage is reduced for identical CTV coverage. The reason for this is that, as opposed to probabilistic optimization, mini–max robust optimization relies on a chosen uncertainty set and treats every scenario within it as equally probable.

The effect of using a discrete uncertainty set in robust optimization can be clearly seen in the CTV-only *setupXYrange* case (figure 8, bottom). The dose margin is extended into all directions where an error scenario is defined. As a result, the dose margins along the diagonals of the *XZ*-plane are more conservative (and thus less conformal) compared to the probabilistic plan. Similar results are seen in figure 9 for the $D_{2\%}^{90\text{th}}$ comparison, where the increase of the OAR objective weight leads to overcompensation of single shifts, at the same time giving less importance to other (potentially more important, higher probability) shifts. Besides a reduction in conformity, this leads to a smaller CTV margin in the *y*-direction, in turn resulting in lower CTV coverage. In the spherical *XZ*-displaced case, the probabilistic plan automatically leads to a conformal plan, because (1) there are no competing objectives and (2) it does not rely on single worst-case scenario that could overcompensate other scenarios.

For the spinal phantoms, we prioritized to limit spinal overdosage probabilities, resulting in large CTV underdosage probabilities. Still, CTV underdosage probabilities were larger in the robust case (for

Table 5. Computation times of expectation values ($\mathbb{E}[D_{ij}(\boldsymbol{\xi})]$, $\mathbb{E}[D_{ij}(\boldsymbol{\xi})D_{ij'}(\boldsymbol{\xi})]$ and $\sum_{i\in\text{nonCTV}} w_i\,\mathbb{E}[D_{ij}(\boldsymbol{\xi})D_{ij'}(\boldsymbol{\xi})]$) and PCE construction of the dose-influence matrix.

	spherical	spinal	Number dose calculations
Expectation values of D_{ij}	8435 s	1412 s	105
PCE construction of D_{ij} (CTV)	6416 s	1738 s	1637
PCE construction of D_{ij} (OAR)	2435 s	1143 s	1637

Table 6. Comparison of total probabilistic and robust optimization times.

	Probabilistic	Robust
Spherical *XZ*-displaced	3.7 h	4.0 h
Spherical *X*-displaced	6.8 h	4.4 h
Spinal ($\nu = 90\%$)	7.6 h	9 min
Spinal ($\nu = 95\%$)	7.3 h	11 min
Spinal ($\nu = 98\%$)	11.4 h	20 min

similar spine overdosage). The CTV coverage in these plans did not meet the clinical criterion that is often aimed for ($D^{10\text{th}}_{98\%} = 95\% d_i^p$), meaning that these plans are non-robust. This can be seen from figure 16, where the lower tails of the near-maximum DVH metrics of the CTV extend to dose values much smaller than was prescribed. Sufficient CTV coverage can be reached by using a stricter CTV underdosage objective, or by increasing the importance of the particular objective. A probabilistic optimization of the horseshoe-shaped CTV shows that the clinical criterion can be reached if there would be no spine (see section 3 of the SM).

To enable a fair comparison, plans were scaled by the $D_{50\%}$. This ICRU-based metric describes the median CTV dose and is numerically robust. We want to emphasize that the exact choices and priorities chosen in treatment planning do not invalidate the approach. In fact, we observed that in some probabilistic optimizations, even one or two outer loop iterations were sufficient to achieve the desired under- and overdosage probabilities. Further research on clinical datasets is needed to explore how the probabilistic approach performs under realistic clinical trade-offs and priorities.

The current approach is memory intensive, because of the combination of quadratic underdosage and overdosage objectives with the $\text{SD}[d_i(\boldsymbol{x},\boldsymbol{\xi})]$ term. Summation over the voxels must be done during the optimization, because for every iteration the active voxel subset (that is under- and overdosed) must be determined. Therefore, $\mathbb{E}[D_{ij}(\boldsymbol{\xi})D_{ij'}(\boldsymbol{\xi})]$ (with maximally N_b^2 non-zero elements for every active voxel) is computed and stored (see section 4 of the SM). The probabilistic approach should especially upscale by number of voxels. In the current approach, voxels were excluded from the dose-influence matrices using conservative bounds, no sparsity was used in the optimization, and the approach was not optimized for computational efficiency. Besides improving on these aspects, upscaling to clinical grid sizes becomes more feasible if probabilistic terms only optimize for substructures of the ROIs (e.g. for rims or checkerboard-like structures). Automatic methods can select voxel subsets using adaptive (Martin *et al* 2007) or deep-learning based (Quarz *et al* 2024) sampling.

The expectation value calculation in (11), (15) and (16) and PCE construction of the dose-influence matrix have been parallelized by 16 CPU-cores (2x Intel XEON E5-6248R 24 C 3.0 GHz) (Delft High Performance Computing Centre DHPC). Computation times are reported in table 5. Calculations for the spinal plans were significantly lower, because of the courser voxel grid. Long computation times for the expectation values are especially due to $\sum_{i\in\text{nonCTV}} w_i\,\mathbb{E}[D_{ij}(\boldsymbol{\xi})D_{ij'}(\boldsymbol{\xi})]$, which is computed for non-CTV voxels. This computational expense can be partly reduced by optimizing over tissue voxels without using expectation values. Computation speeds are expected to be improved by using GPU-cores.

PCE construction times in this work can be treated as a conservative upper bound to what is clinically often regarded as sufficient, because stricter Γ-evaluation criteria are used here. Namely, a Γ-evaluation using 3 mm/3% instead of 1 mm/0.1 Gy settings lead to 98% of the voxels being accepted in all test scenarios for both pencil-beams.

Cumulative optimization times are listed in table 6. Spherical *X*-displaced plans took longer to optimize than *XZ*-displaced plans, due to more conflicting probabilistic objectives (CTV and OAR are closer). For the same reason, even though the spinal geometry contains less voxels than the spherical geometry, the probabilistic spinal plans took longer to optimize than the spherical plans. Robust optimization times for spherical plans were similar to probabilistic ones, but spinal robust optimizations were much faster, likely because the choice of objective weights made the plans more challenging

to optimize. Probabilistic optimization times could be improved by parallelization and by using different warm-start strategies. Large computation times are primarily due to the inner optimization, which is further analyzed in appendix A.1. The computational expense of the current approach may benefit from operator-splitting methods, as was shown to be the case in previous studies on mini–max robust (Fan *et al* 2024, Liu *et al* 2025) and robust-CVaR optimization (Luxenberg *et al* 2025).
Other suggestions to be included into future work are as follows:

(i) Although the current work only focuses on probabilistic objectives, the approach can be extended to handle probabilistic constraints.
(ii) The presented approach is not limited to proton therapy, but can likewise be applied to photon therapy, or radiation therapy in general. Although photons may be less sensitive to uncertainties than protons, photons—particularly volumetric modulated arc therapy—have shown inter-patient variation in PTV coverage (Rojo-Santiago *et al* 2023b). In cases involving complex anatomies, challenging trade-offs between target and OARs, or hypo-fractionated treatments, the probabilistic approach could potentially have great added value for photon therapy as well. Even when the expected improvements are less significant, the probabilistic approach can help in the interpretation of dosimetric outcomes.
(iii) To extend the work to hypo-fractionated treatments, random errors should be included in the probabilistic approach. Possibly, fractionation schemes can be explicitly included into the optimization process by accounting for the number of fractions, similar to previous approaches for photon (Unkelbach and Oelfke 2004) and proton therapy (Wahl *et al* 2018).

5. Conclusions

This work presents a new approach to probabilistic treatment planning, that is able to optimize for exact underdosage and overdosage probabilities of multiple structures (which are not site-specific), for personalized probability levels and dose thresholds. For systematic setup errors, including dose fall-off—which reflects the actual dose distribution—makes the Van Herk margin less conservative, yielding values that closely correspond to those from the probabilistic approach. Compared to composite-wise robust plans, the probabilistic plans achieve more OAR sparing with similar target coverage (or improved target coverage with similar OAR sparing) for all spherical and spinal comparisons. Probabilistic plans were found to be more conformal to the CTV, as probabilistic optimization accounts for the probability of different error scenarios rather than relying only on a predefined uncertainty set and optimizing with a single worst-case scenario. As the proposed method is sufficiently general to be extended to dose-coverage or CVaR optimization, this is an obvious follow-up. Besides that, following work should focus on improving computational efficiency through time and memory optimization techniques, so that clinical feasibility (i.e. the application to real clinical cases) can be demonstrated.

Data availability statement

The Open Source Generalized Polynomial Chaos Expansion Toolbox (Perkó *et al* 2014) (https://gitlab.com/zperko/opengpc/-/tree/4a2e11cc77617c705f8a555996aee3c7cd4a61db/) is used to construct and evaluate Polynomial Chaos Expansions. The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Supplementary data available at https://doi.org/10.1088/1361-6560/ae351c/data1.

Acknowledgment

The authors would like to acknowledge that this work is funded by RaySearch Laboratories. Zoltán Perkó is an associate professor at TU Delft and is employed as a Senior Applied Scientist at Radformation Inc.. His industry employment is unrelated to the submitted work. The remaining authors have no conflicts of interest to declare.

CRediT author and contributor statement

Jelte Rinus de Jong: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—Original Draft, Writing—Review and Editing

Zoltán Perkó: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Resources, Writing—Review and Editing
Danny Lathouwers: Conceptualization, Methodology, Project Administration, Supervision, Resources, Writing—Review and Editing
Mischa Hoogeman: Conceptualization, Funding Acquisition, Writing—Review and Editing
Steven Habraken: Conceptualization, Writing—Review and Editing
Sebastiaan Breedveld: Conceptualization, Writing—Review and Editing

Appendix. Probabilistic and robust optimizations details

This section presents a summary of the probabilistic (appendix A.1) and robust (appendix A.2) optimizations that are applied to all phantom geometries, as presented in section 2.4.

A.1. Probabilistic optimizations

Table A1 summarizes the probabilistic optimization parameters that are used for each geometry. All variables are in accordance with section 2.2.1, where the general (inner) probabilistic optimization is shown in (20). As an example, for the spherical CTV-only case, we aim to limit the CTV underdosage probability as $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant \gamma_i) \leqslant \alpha\% = P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \leqslant 57\,\text{Gy}) \leqslant 10\%$ and CTV overdosage probability as $P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \geqslant \epsilon) \leqslant (100-\beta)\% = P(d_i(\boldsymbol{x},\boldsymbol{\xi}) \geqslant 64.2\,\text{Gy}) \leqslant 10\%$.

Convergence criteria are defined in accordance with section 2.2.2. A representative example of the percentile convergence (corresponding to the damped beam weights) is shown on top in figure A1 for the spherical *setupXYrange* plan, where the 10^{th}- and 90^{th}-percentiles of the CTV voxels and the 90^{th}-percentiles of the OAR voxels are optimized for. The bottom of figure A1 shows the respective convergence criteria with the convergence thresholds (in dashed black).

Probabilistic optimization times. Large computational times are primarily due to the inner optimization. Figure A2 shows the inner optimization times for the spherical (CTV + OAR) and spinal plans, which decreases over the course of the optimization for all plans. This occurs because the probabilistic objectives are only evaluated for voxels that have not yet met the target probability level. Voxels that meet the target do not necessarily remain passing throughout the optimization; they can fall below the target again and be re-included in the probabilistic objective. However, as optimization continues, more voxels consistently satisfy the objectives. This reduces the number of voxels needing probabilistic optimization and thus lowers the overall computational load.

Table A1. Probabilistic optimization parameters.

	Spherical			Spinal		
	van Herk comparison	CTV-only	CTV and OAR	$\nu_i = 90\%$	$\nu_i = 95\%$	$\nu_i = 98\%$
w_i^{CTV}	100	100	100	100	100	100
w_i^{OAR}	20	20	20	20	20	20
w_i^{Tissue}	1	1	1	1	1	1
d_i^p [Gy]	60	60	60	60	60	60
α_i [%]	2	10	10	10	10	10
β_i [%]	90	90	90	90	90	90
ν_i [%]	—	—	90	90	95	98
γ_i [Gy]	57	57	57	57	57	57
ϵ_i [Gy]	64.2	64.2	64.2	64.2	64.2	64.2
μ_i [Gy]	—	—	30	54	54	54
$\pi_{\text{CTV}}^{\alpha}$	750	15	15	15	15	15
π_{CTV}^{β}	15	15	15	15	15	15
π_{OAR}^{ν}	—	—	15	750	750	750
$\pi_{\text{CTV}}^{\text{low}}$	1	1	1	5	5	5
$\pi_{\text{OAR}}^{\text{low}}$	—	—	1	15	15	15
π_{Tissue}	1	1	1	1	1	1
ΔW	15	20	15	15	15	15
Δk	5	10	5	5	5	5
$\tau_{\text{CTV},\alpha}$	$5 \cdot 10^{-3}$	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$5 \cdot 10^{-3}$
$\tau_{\text{CTV},\beta}$	$1 \cdot 10^{-3}$	$5 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$4 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$1 \cdot 10^{-3}$
$\tau_{\text{OAR},\nu}$	—	—	$7 \cdot 10^{-3}$	0.1	0.1	$5 \cdot 10^{-2}$

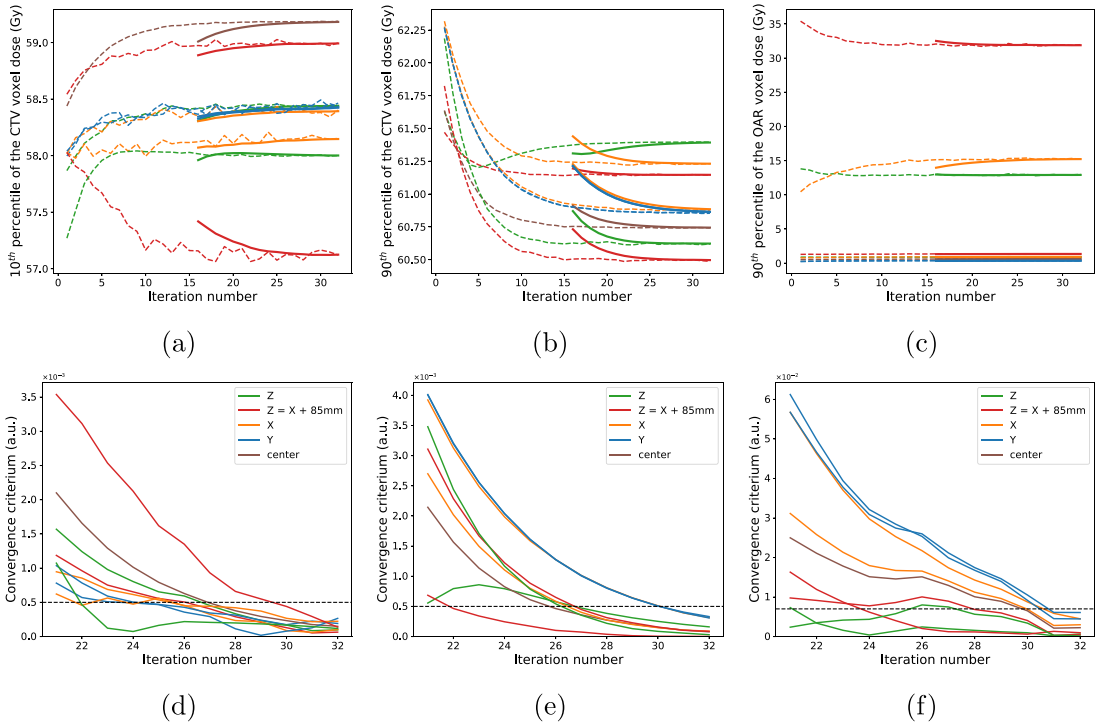

Figure A1. The convergence behavior of the damped (10^{th} and 90^{th}) CTV and (90^{th}) OAR voxel dose percentiles (dashed) for the *XZ*-displaced *setupXYrange* probabilistic optimization, with their moving average (MA15, solid). Only representative outer voxels of both structures are shown, along the CTV and OAR centered *X*, *Y*, *Z* and $Z = X + 85$ mm axes. The relative change in MA15 as in (27) with $\Delta k = 5$ is shown (bottom) with the convergence tolerance (dashed black).

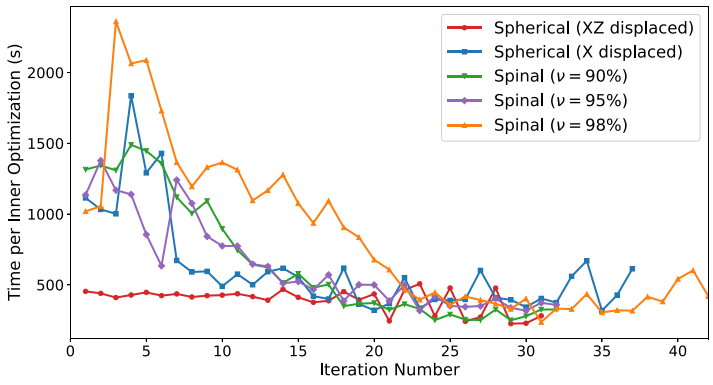

Figure A2. Inner optimization times of the probabilistic spherical and spinal plans.

The $\nu = 98\%$ case took particularly longer than the other cases, mainly because of increased inner optimization times during the early iterations (before iteration 21). In addition to having slightly stricter convergence criteria compared to $\nu = 90\%$ and $\nu = 95\%$, stricter spinal overdose probabilities caused more voxels to not reach the target probability level, making them active in the optimization more often.

The fact that iterations take longer if more voxels have not reached the desired thresholds yet, means at the same time that optimization times are sensitive to the choice of initial beam weights. Inner optimization times may be improved by terminating the first iterations before fully converged beam weights are obtained. This can be done since the main purpose of these initial optimizations is not to obtain converged beam weights, but rather to direct the voxel dose percentiles (i.e. δ-factors) to the desired levels. This may reduce the optimization time significantly, while the number of (outer loop) iterations is barely affected.

Table A2. Robust optimization parameters for the *XZ*-displaced (*XZ*) and *X*-displaced (X) geometries, that are matched by either CTV coverage ($D_{98\%}^{10\text{th}}$) or OAR dose ($D_{2\%}^{90\text{th}}$). Prescribed dose $d_i^p = 60$ Gy and voxel weights for CTV, spine/OAR and tissue are $w_i^{\text{CTV}} = 100$, $w_i^{\text{OAR}} = 20$, $w_i^{\text{Tissue}} = 1$, respectively.

	ω_{CTV}	ω_{OAR}	$\omega_{\text{OAR}}^{\max}$	$\omega_{\text{CTV}}^{\text{nom}}$	ω_{Tissue}	d_i^{maxOAR}(Gy)
CTV-only	15	—	—	—	1	30
XZ ($D_{98\%}^{10\text{th}}$)	120	1	1	—	160	30
XZ ($D_{2\%}^{90\text{th}}$)	100	10	10	—	100	30
X ($D_{98\%}^{10\text{th}}$)	120	1	1	—	160	30
X ($D_{2\%}^{90\text{th}}$)	100	15	1	—	100	30
Spinal ($\nu = 90\%$)	2	1	1	4	1	54
Spinal ($\nu = 95\%$)	3	2	2	6	2	54
Spinal ($\nu = 98\%$)	11	10	10	22	10	54

A.2. Robust optimizations

All probabilistic plans in this work are compared to a composite-wise mini–max robust plan as in (28). Table A2 shows the resulting robust objective weights $\omega \in \Omega$, obtained from tuning the CTV coverage or OAR/spinal dose towards the probabilistic plan outcome. For the spherical CTV + OAR case, combinations of *XZ*-displaced or *X*-displaced and types of scaling (either by CTV coverage or OAR dose) are done. Three different spinal plans are compared, where the spinal overdosage probability is limited by different probability levels ($\nu_i = 90\%$, $\nu_i = 95\%$ and $\nu_i = 98\%, \forall i \in \mathit{Spine}$).

ORCID iDs

J R de Jong 0009-0007-1963-7377
S Breedveld 0000-0001-8954-4554
S J M Habraken 0009-0002-1168-3479
D Lathouwers 0000-0003-3810-1926
Z Perkó 0000-0002-0975-4226

References

An Y, Liang J, Schild S E, Bues M and Liu W 2017 Robust treatment planning with conditional value at risk chance constraints in intensity-modulated proton therapy *Med. Phys.* **44** 28–36

Biggs S *et al* 2022 Pymedphys: a community effort to develop an open, python-based standard library for medical physics applications *J. Open Source Softw.* **7** 4555

Bokrantz R and Fredriksson A 2017 Scenario-based radiation therapy margins for patient setup, organ motion and particle range uncertainty *Phys. Med. Biol.* **62** 1342–57

Bortfeld T 1997 An analytical approximation of the Bragg curve for therapeutic proton beams *Med. Phys.* **24** 2024–33

Chan T C, Mahmoudzadeh H and Purdie T G 2014 A robust-CVaR optimization approach with application to breast cancer therapy *Eur. J. Oper. Res.* **238** 876–85

Chen W, Unkelbach J, Trofimov A, Madden T, Kooy H, Bortfeld T and Craft D 2012 Including robustness in multi-criteria optimization for intensity-modulated proton therapy *Phys. Med. Biol.* **57** 591–608

Chu M, Zinchenko Y, Henderson S G and Sharpe M B 2005 Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty *Phys. Med. Biol.* **50** 5463–77

Cristoforetti R, Hardt J J and Wahl N 2025 Scenario-free robust optimization algorithm for IMRT and IMPT treatment planning (arXiv:2501.05818)

de Jong J I, Habraken S J, Rojo-Santiago J, Lathouwers D, Perkó Z, Breedveld S and Hoogeman M S 2025 Probabilistic evaluation guided IMPT planning with realistic setup and range uncertainties improves the trade-off between OAR sparing and target coverage in neuro-oncological patients *Radiother. Oncol.* **213** 111171

Delft High Performance Computing Centre (DHPC) 2024 DelftBlue supercomputer (Phase 2) (available at: www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2)

Fabiano S, Torelli N, Papp D and Unkelbach J 2022 A novel stochastic optimization method for handling misalignments of proton and photon doses in combined treatments *Phys. Med. Biol.* **67** 185006

Fan Q *et al* 2024 A novel fast robust optimization algorithm for intensity-modulated proton therapy with minimum monitor unit constraint *Med. Phys.* **51** 6220–30

Fredriksson A and Bokrantz R 2016 The scenario-based generalization of radiation therapy margins *Phys. Med. Biol.* **61** 2067–82

Fredriksson A, Forsgren A and Hardemark B 2011 Minimax optimization for handling range and setup uncertainties in proton therapy *Med. Phys.* **38** 1672–84

Gordon J J, Sayah N, Weiss E and Siebers J V 2010 Coverage optimized planning: probabilistic treatment planning based on dose coverage histogram criteria *Med. Phys.* **37** 550–63

Holmes J, Feng H, Zhang L, Fix M K, Jiang S B and Liu W 2024 Fast Monte Carlo dose calculation in proton therapy *Phys. Med. Biol.* **69** 17TR01

Janson M, Glimelius L, Fredriksson A, Traneus E and Engwall E 2024 Treatment planning of scanned proton beams in raystation *Med. Dosim.* **49** 2–12

Korevaar E W et al 2019 Practical robustness evaluation in radiotherapy - a photon and proton-proof alternative to PTV-based plan evaluation *Radiother. Oncol.* **141** 267–74
Liu J, Zhu Y, Zhang X and Gao H 2025 Efficient operator-splitting minimax algorithm for robust optimization *Med. Phys.* **52** e17929
Liu W, Zhang X, Li Y and Mohan R 2012 Robust optimization of intensity modulated proton therapy: robust optimization of IMPT *Med. Phys.* **39** 1079–91
Lomax A J 2008 Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions *Phys. Med. Biol.* **53** 1043–56
Luxenberg E, Pérez-Piñeiro D, Diamond S and Boyd S 2025 An operator splitting method for large-scale CVaR-constrained quadratic programs (arXiv:2504.10814)
Martin B C, Bortfeld T R and Castañon D A 2007 Accelerating IMRT optimization by voxel sampling *Phys. Med. Biol.* **52** 7211–28
Mescher H, Ulrich S and Bangert M 2017 Coverage-based constraints for IMRT optimization *Phys. Med. Biol.* **62** N460–73
Nguyen M-L, Cantrell J N, Ahmad S and Henson C 2021 Intensity-modulated proton therapy (IMPT) versus intensity-modulated radiation therapy (IMRT) for the treatment of head and neck cancer: A dosimetric comparison *Med. Dosim.* **46** 259–63
Oud M, Breedveld S, Rojo-Santiago J, Gizyńska M K, Kroesen M, Habraken S, Perkó Z, Heijmen B and Hoogeman M 2024 A fast and robust constraint-based online re-optimization approach for automated online adaptive intensity modulated proton therapy in head and neck cancer *Phys. Med. Biol.* **69** 075007
Park P C et al 2013 Statistical assessment of proton treatment plans under setup and range uncertainties *Int. J. Radiat. Oncol. Biol. Phys.* **86** 1007–13
Pastor-Serrano O and Perkó Z 2022 Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy *Phys. Med. Biol.* **67** 105006
Perkó Z, Gilli L, Lathouwers D and Kloosterman J L 2014 Grid and basis adaptive polynomial chaos techniques for sensitivity and uncertainty analysis *J. Comput. Phys.* **260** 54–84
Perkó Z, van der Voort S R, van de Water S, Hartman C M H, Hoogeman M and Lathouwers D 2016 Fast and accurate sensitivity analysis of IMPT treatment plans using polynomial chaos expansion *Phys. Med. Biol.* **61** 4646–64
Pflugfelder D, Wilkens J J and Oelfke U 2008 Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy *Phys. Med. Biol.* **53** 1689–700
Quarz A, Volz L, Antink C H, Durante M and Graeff C 2024 Deep learning-based voxel sampling for particle therapy treatment planning *Phys. Med. Biol.* **69** 155014
RaySearch Laboratories 2023 Raystation 12a instructions for use *User Manual RSL-D-RS-12A-IFU-EN-1.1* (RaySearch Laboratories AB)
Rojo-Santiago J, Habraken S J M, Romero A M, Lathouwers D, Wang Y, Perkó Z and Hoogeman M S 2023a Robustness analysis of CTV and OAR dose in clinical PBS-PT of neuro-oncological tumors: prescription-dose calibration and inter-patient variation with the dutch proton robustness evaluation protocol *Phys. Med. Biol.* **68** 175029
Rojo-Santiago J, Habraken S J, Lathouwers D, Méndez Romero A, Perkó Z and Hoogeman M S 2021 Accurate assessment of a Dutch practical robustness evaluation protocol in clinical PT with pencil beam scanning for neurological tumors *Radiother. Oncol.* **163** 121–7
Rojo-Santiago J, Habraken S J, Unipan M, Both S, Bosmans G, Perkó Z, Korevaar E and Hoogeman M S 2024 A probabilistic evaluation of the dutch robustness and model-based selection protocols for head-and-neck IMPT: a multi-institutional study *Radiother. Oncol.* **199** 110441
Rojo-Santiago J, Korevaar E, Perkó Z, Both S, Habraken S J and Hoogeman M S 2023b PTV-based VMAT vs. robust IMPT for head-and-neck cancer: a probabilistic uncertainty analysis of clinical plan evaluation with the dutch model-based selection *Radiother. Oncol.* **186** 109729
Schaffner B and Pedroni E 1998 The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power *Phys. Med. Biol.* **43** 1579–92
Smolyak S A 1963 Quadrature and interpolation formulas for tensor products of certain classes of functions *Soviet Math. Dokl.* **4** 240–3
Sobotta B, Söhn M and Alber M 2010 Robust optimization based upon statistical theory *Med. Phys.* **37** 4019–28
Sterpin E, Rivas S T, Van den Heuvel F, George B, Lee J A and Souris K 2021 Development of robustness evaluation strategies for enabling statistically consistent reporting *Phys. Med. Biol.* **66** 045002
Stuschke M, Kaiser A, Pöttgen C, Lübcke W and Farr J 2012 Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to intensity modulated photon plans *Radiother. Oncol.* **104** 45–51
Taasti V T, Hong L, Deasy J O and Zarepisheh M 2020 Automated proton treatment planning with robust optimization using constrained hierarchical optimization *Med. Phys.* **47** 2779–90
The International Commission on Radiation Units and Measurements 2010 *J. ICRU* **10** N2–N
The MathWorks, Inc. 2024 MATLAB version R2024a The MathWorks, Inc.
Tilly D, Holm A s, Grusell E and Ahnesjö A 2019 Probabilistic optimization of dose coverage in radiotherapy *Phys. Imaging Radiat. Oncol.* **10** 1–6
Trofimov A, Unkelbach J, DeLaney T F and Bortfeld T 2012 Visualization of a variety of possible dosimetric outcomes in radiation therapy using dose-volume histogram bands *Pract. Radiat. Oncol.* **2** 164–71
Unkelbach J et al 2018 Robust radiotherapy planning *Phys. Med. Biol.* **63** 22TR02
Unkelbach J, Bortfeld T, Martin B C and Soukup M 2008 Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning *Med. Phys.* **36** 149–63
Unkelbach J, Chan T C Y and Bortfeld T 2007 Accounting for range uncertainties in the optimization of intensity modulated proton therapy *Phys. Med. Biol.* **52** 2755–73
Unkelbach J and Oelfke U 2004 Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions *Phys. Med. Biol.* **49** 4005–29
Unkelbach J and Paganetti H 2018 Robust proton treatment planning: physical and biological optimization *Sem. Radiat. Oncol.* **28** 88–96
van de Sande M A, Creutzberg C L, van de Water S, Sharfo A W and Hoogeman M S 2016 Which cervical and endometrial cancer patients will benefit most from intensity-modulated proton therapy? *Radiother. Oncol.* **120** 397–403
van der Voort S, van de Water S, Perkó Z, Heijmen B, Lathouwers D and Hoogeman M 2016 Robustness recipes for minimax robust optimization in intensity modulated proton therapy for oropharyngeal cancer patients *Int. J. Radiat. Oncol. Biol. Phys.* **95** 163–70
van Dijk L V, Steenbakkers R J H M, ten Haken B, van der Laan H P, van't Veld A A, Langendijk J A and Korevaar E W 2016 Robust intensity modulated proton therapy (IMPT) increases estimated clinical benefit in head and neck cancer patients *PLoS One* **11** e0152477

van Haveren R and Breedveld S 2019 Fast and exact Hessian computation for a class of nonlinear functions used in radiation therapy treatment planning *Phys. Med. Biol.* **64** 16NT01

van Herk M, Remeijer P, Rasch C and Lebesque J V 2000 The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **47** 1121–35

Wahl N, Hennig P, Wieser H and Bangert M 2018 Analytical incorporation of fractionation effects in probabilistic treatment planning for intensity-modulated proton therapy *Med. Phys.* **45** 1317–28

Wiener N 1938 The homogeneous chaos *Am. J. Math.* **60** 897

Wu Q and Mohan R 2000 Algorithms and functionality of an intensity modulated radiotherapy optimization system *Med. Phys.* **27** 701–11

Xiu D and Karniadakis G E 2002 The Wiener–Askey polynomial chaos for stochastic differential equations *SIAM J. Sci. Comput.* **24** 619–44

Zhang X 2021 A review of the robust optimization process and advances with Monte Carlo in the proton therapy management of head and neck tumors *Int. J. Part. Therapy* **8** 14–24