Introducing a probabilistic definition of the target in a robust treatment planning framework

Gregory Buti[1], Kevin Souris[1], Ana Maria Barragán Montero[1], John Aldo Lee[1], Edmond Sterpin[1,2]

[1]Université Catholique de Louvain, Institut de Recherche Expérimentale et Clinique (IREC), Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Avenue Hippocrate 54 - box B1.54.07, 1200 Brussels, Belgium

[2]Katholieke Universiteit Leuven, Department of Oncology, Laboratory of Experimental Radiotherapy, UZ Herestraat 49 - box 7003, 3000 Leuven, Belgium

E-mail: gregory.buti@uclouvain.be

11 June 2021

Abstract. The "clinical target distribution" (CTD) has recently been introduced as a promising alternative to the binary clinical target volume (CTV). However, a comprehensive study that considers the CTD, together with geometric treatment uncertainties, was lacking. Because the CTD is inherently a probabilistic concept, this study proposes a fully probabilistic approach that integrates the CTD directly in a robust treatment planning framework. First, the CTD is derived from a reported microscopic tumor infiltration model such that it explicitly features the probability of tumor cell presence in its target definition. Second, two probabilistic robust optimization methods are proposed that evaluate CTD coverage under uncertainty. The first method minimizes the expected-value (EV) over the uncertainty scenarios and the second method minimizes the sum of the expected value and standard deviation (EV-SD), thereby penalizing the spread of the objectives from the mean. Both EV and EV-SD methods introduce the CTD in the objective function by using weighting factors that represent the probability of tumor presence. The probabilistic methods are compared to a conventional worst-case approach that uses the CTV in a worst-case optimization algorithm. To evaluate the treatment plans, a scenario-based evaluation strategy is implemented that combines the effects of microscopic tumor infiltrations with the other geometric uncertainties. The methods are tested for five lung tumor patients, treated with intensity-modulated proton therapy. The results indicate that for the studied patient cases, the probabilistic methods favour the reduction of the esophagus dose but compensate by increasing the high-dose region in a low conflicting organ such as the lung. These results show that a fully probabilistic approach has the potential to obtain clinical benefits when tumor infiltration uncertainties are taken into account directly in the treatment planning process.

Keywords: proton therapy, robust optimization, probabilistic, stochastic, worst-case, minimax, target definition

1. Introduction

In recent literature, the concept of a "clinical target distribution" (CTD) has been introduced in order to represent the probability of tumor cell presence away from the visible tumor mass (Shusharina et al. (2018), Unkelbach et al. (2020), Bortfeld et al. (2021)). This probabilistic definition of the target stands in clear contrast to current clinical practice where the binary *clinical target volume* (CTV) is used to encompass the microscopic spread of cancer cells. However, the question of how the CTD should be combined with the geometric treatment uncertainties remains open.

Current clinical practice typically treats tumor infiltration and geometric uncertainties separately by following a linear two step approach: first, the CTV is defined. According to the ICRU 83 report, the CTV contains the *gross tumor volume* (GTV) and/or regions where tumor cell presence is likely (*The International Commission on Radiation Units and Measurements* (2010)). Tumor cell presence may be the result of microscopic tumor infiltration at the boundary of the GTV or the possible infiltration into whole organs such as lymph nodes, among others. In this study, we focus on target volumes that are associated with the first type. In this case, the CTV is obtained as a geometric margin expansion of the GTV, followed by a correction for anatomical barriers. A second margin expansion then defines the *planning target volume* (PTV), to which the dose is prescribed. For particle therapy treatments, the PTV is considered inadequate due to its inability to deal with range uncertainties (Fredriksson and Bokrantz (2016), Unkelbach et al. (2018)). Therefore, state-of-the-art workflows replace the CTV-to-PTV margin expansion step by a robust optimization process where CTV coverage is evaluated in a set of geometric uncertainty scenarios. Research efforts have mainly focused on two types of robust optimization methods: (a) a worst-case formulation that uses the worst-case scenario to guide the optimization solution, such as *minimax* and voxel-wise worst-case (Pflugfelder et al. (2008), Fredriksson et al. (2011), Liu et al. (2012), Unkelbach et al. (2018)), and (b) a probabilistic formulation that minimizes the expected value of the objective function (Unkelbach et al. (2008), Fredriksson (2012)).

Stroom et al. (2014) proposed a GTV-to-PTV margin recipe that treated tumor infiltration and geometric uncertainties together, leading to smaller overall margins. However, in the context of robust optimization, tumor infiltration uncertainties remain largely unaddressed. In this regard, an approach that combines a fixed target volume (the CTV) with worst-case robust optimization holds some limitations. For instance, we can define the CTV with a tumor infiltration uncertainty model by following ICRU recommendations, which state that "a probability of occult disease higher than from 5% to 10% is assumed to require treatment" (*The International Commission on Radiation Units and Measurements* (2010)). However, by evaluating such a resulting CTV in a set of extreme geometric scenarios (e.g. a 5 mm setup error and 3% range error), some of these scenarios become overly conservative, when taking into account the probability of the combined tumor infiltration and geometric uncertainty. The overly

conservative nature of these scenarios can be considered analogous to the overestimation of conventional margins, observed by Stroom et al. (2014). In practice, this implies that for clinical cases where the organs-at-risk are in close proximity to the CTV, a worst-case optimizer will need to balance excessive conflicts among the planning objectives. Given that worst-case optimization algorithms already show a tendency to overemphasize a limited number of scenarios (Fredriksson and Bokrantz (2014), Unkelbach et al. (2018)), this strategy can lead to overdose in critical organs.

An alternative approach could consist of defining a CTD according to a tumor infiltration uncertainty model, followed by a robust optimization process where CTD coverage is evaluated in a set of geometric error scenarios. As the CTD is inherently a probabilistic concept, we elect for an approach where the CTD is integrated in a – *fully probabilistic* – robust optimization setting. This will allow us to extend the definition of the objective function, using weighting factors that represent the probability of tumor presence, as defined by the CTD.

The aim of this study is therefore threefold. First, we propose a procedure to construct a probabilistic target, i.e. the CTD. In Shusharina et al. (2018), the CTD was composed of several shells, for instance delineated by a physician, with each shell defining the probability of tumor presence. Similar to Shusharina et al. (2018), we adopt the notion that the CTD represents the probability of tumor presence. However, this study proposes a voxel-wise approach where the probability of the target voxels will be derived from a reported probability distribution of tumor infiltrations. Deriving the target from an uncertainty model will allow us to compare different treatment planning strategies in a statistically consistent way. Second, we extend the use of the CTD in the treatment planning process by developing a fully probabilistic optimization framework that includes the CTD, in conjunction with other geometric uncertainties. Moreover, a scenario-based evaluation strategy is developed that evaluates the effect of tumor infiltration uncertainties, together with the other considered uncertainties. Third, we illustrate the features of fully probabilistic optimization for five lung tumor cases, treated with intensity-modulated proton therapy (IMPT) using the pencil beam scanning (PBS) technique.

2. Methods

This section is organized as follows: Section 2.1 reviews the main features of the treatment planning system (TPS) in which the methods have been implemented. In Section 2.2, the uncertainty models and their assumed magnitudes are described. Section 2.3 details the procedures to obtain the target, i.e. the CTV for worst-case optimization and CTD for fully probabilistic optimization. Section 2.4 introduces the robust optimization methods, that is, the reference method (worst-case optimization) and the proposed method (fully probabilistic optimization). The section focuses specifically on the integration of tumor infiltration uncertainties into the optimization process. Section 2.5 features the evaluation procedure used to make comprehensive assessments of the

performance of each method. Finally, Section 2.6 gives an overview of the patient data and treatment plan characteristics.

2.1. Treatment planning system

Treatment plans are created with the open-source TPS MIROpt, coded in Matlab, MathWorks - Natick, United States (Barragán-Montero (n.d.)). MIROpt uses the Monte Carlo proton dose engine MCsquare, available open-source (Souris (n.d.)), for its dose calculations with 10^4 ions per spot and a $2{\times}2{\times}2$ mm^3 dose calculation grid. Plan optimization is performed with the large-scale nonlinear solver Ipopt (Wächter and Biegler (2005)), through its Matlab interface. The Ipopt solver was employed for the optimization of both the worst-case and probabilistic methods. Treatment plans were generated on a 256GB RAM system with a 2x8 Core Intel Xeon processor (E5-2667 v3) @3.20 GHz.

In MIROpt, the objective function consists of several quadratic dose-fidelity terms which penalize deviations from the pre-defined planning objectives. During the optimization, the dose d in voxel i is evaluated by summing the contribution of all m beamlets:

$$d_i = \sum_{j=1}^{m} P_{ij} \cdot w_j, \tag{1}$$

where $\boldsymbol{P} \in \mathbb{R}^{\mathrm{n}\times\mathrm{m}}$ represents the dose-influence matrix (with n the total number of voxels of the dose grid) and $\boldsymbol{w} \in \mathbb{R}^{\mathrm{m}}$ the spot weight vector.

2.2. Uncertainty models for optimization

In this study, the optimization methods are applied to lung tumor patients, treated with IMPT-PBS. We consider tumor infiltration uncertainties as well as geometric uncertainties with a systematic component (systematic setup errors and range errors) and a random component (random setup errors and respiratory motion). This section presents a brief overview of the way in which MIROpt models each uncertainty. More details can be found in Barragán-Montero et al. (2017) and Barragán-Montero (2017). Except for tumor motion, all uncertainties are assumed to follow Gaussian probability density functions (PDF).

2.2.1. Setup errors

Similar to other IMPT studies (for example Unkelbach et al. (2008) and Fredriksson et al. (2011)), systematic setup errors are modeled by rigid shifts of the spot weight grid. The systematic setup errors are assumed to be normally distributed with a magnitude of zero mean and 2.4 mm standard deviation Σ_s, identical in each direction x, y and z.‡

‡ Following our institution guidelines, the magnitude of the setup errors includes the effect of baseline shifts.

Random setup errors are modeled by randomly shifting the incident position of each proton during the Monte Carlo simulation of each beamlet, according to the assumed probability distribution (Barragán-Montero et al. (2016)). This method resembles the way random errors are sometimes treated in conventional radiotherapy, i.e. by blurring the dose distribution (see for example Bohoslavsky et al. (2013)). However, shifting protons is more appropriate for proton-based dose calculation as the range variation associated with each shift is simulated. This method implicitly assumes that the treatment is delivered in an infinite number of fractions. Fredriksson (2012) reports that for IMPT, the infinite fraction assumption can be considered a valid approximation provided firstly, treatment delivery in at the minimum 30 fractions and secondly, the incorporation of an uncertainty on the standard deviation of the random errors. Both conditions are satisfied as the methods are tested for IMPT treatments of 30 fractions (see Section 2.6), and an uncertainty on the standard deviation is considered by simulating random setup errors with a standard deviation of both 0 mm and 3 mm (see Section 2.4).

2.2.2. Range errors

Range errors are modeled by uniformly scaling the CT densities during the computation of the dose-influence matrices. Following the study of Paganetti (2012), range errors are assumed to follow a 1-D Gaussian distribution with zero mean and standard deviation of $\Sigma_r = 1.6\%$.§

2.2.3. Respiratory motion

Respiratory motion is represented by 10 respiratory phases, equally spaced in time. The dosimetric effect is simulated by accumulating the dose along all 10 respiratory phases on the mid-position CT (MidP-CT) (Wanet et al. (2014)). The deformable registration algorithm, from the open-source platform OpenReggui (Janssens (n.d.)), is used in order to register all respiratory phases to the MidP-CT.

2.2.4. Tumor infiltration uncertainties

The tumor infiltration uncertainty is defined according to the histological study of Meng et al. (2012). Meng et al. (2012) report an uncertainty distribution for maximum tumor infiltrations of mean 3.4 mm $\pm$ 2.8 mm, for a population of non-small-cell lung cancer patients. Assuming this data corresponds to radial infiltration of tumor cells, a 1-D truncated Gaussian PDF $\rho(x)$ is then used in order to approximate the probability distribution, i.e. negative values ($x < 0$) are removed, followed by a normalization of the distribution (see Fig 1a). As will be detailed in Section 2.3, tumor infiltration uncertainties will be incorporated in the robust optimization process through the definition of the target.

§ Paganetti (2012) reports a 2.4% range uncertainty, evaluated at $1.5\Sigma_r$ which translates to a $\Sigma_r = 1.6\%$.

2.3. Target definitions

This section presents the procedures used to define the CTD and CTV, i.e. the targets considered in fully probabilistic optimization and worst-case optimization, respectively. Both targets are derived from the tumor infiltration uncertainty model defined in Section 2.2.4.

2.3.1. CTD

The *clinical target distribution* (CTD) represents a 3D-distribution where the value of each voxel defines the probability of tumor presence p_t, based on a population of tumor infiltrations x. The CTD is constructed by first computing p_t as a function of radial distance r from the GTV edge. $p_t(r)$ can be obtained from $\rho(x)$, the assumed tumor infiltration uncertainty model (see Section 2.2.4), by integrating $\rho(x)$ as follows:

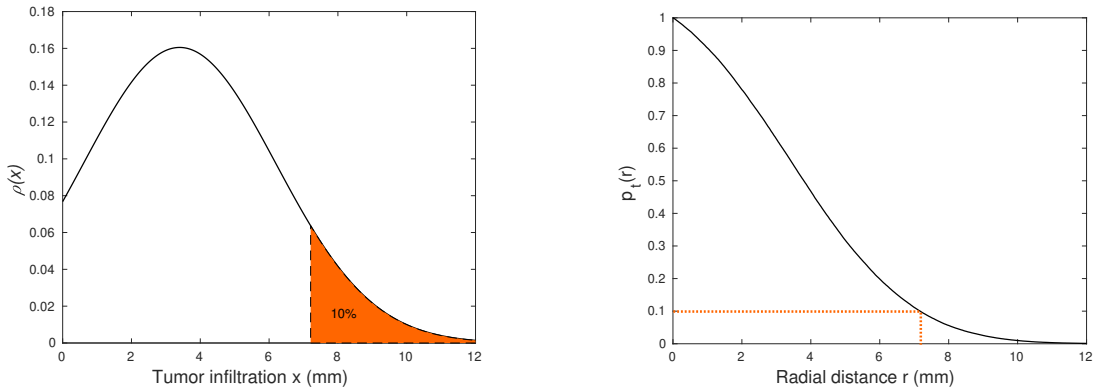
$$p_t(r) = \int_r^\infty \rho(x)dx. \tag{2}$$

In other words, $p_t(r)$ equals the probability that a tumor infiltration X will take a value greater than or equal to r: $p_t(r) = \rho(\mathrm{X} \geq r)$. Note that this approach implicitly assumes that the tumor cell density is constant for a given tumor infiltration. Fig. 1b illustrates $p_t(r)$, together with an example (in orange) of how the integration limits in Fig. 1a yield the corresponding value of $p_t(r)$ in Fig. 1b. A patient-specific CTD then follows from:

(i) generating a 3D-Euclidean distance map where the value of each voxel represents the minimum distance (r) to the GTV (see Fig. 2a),

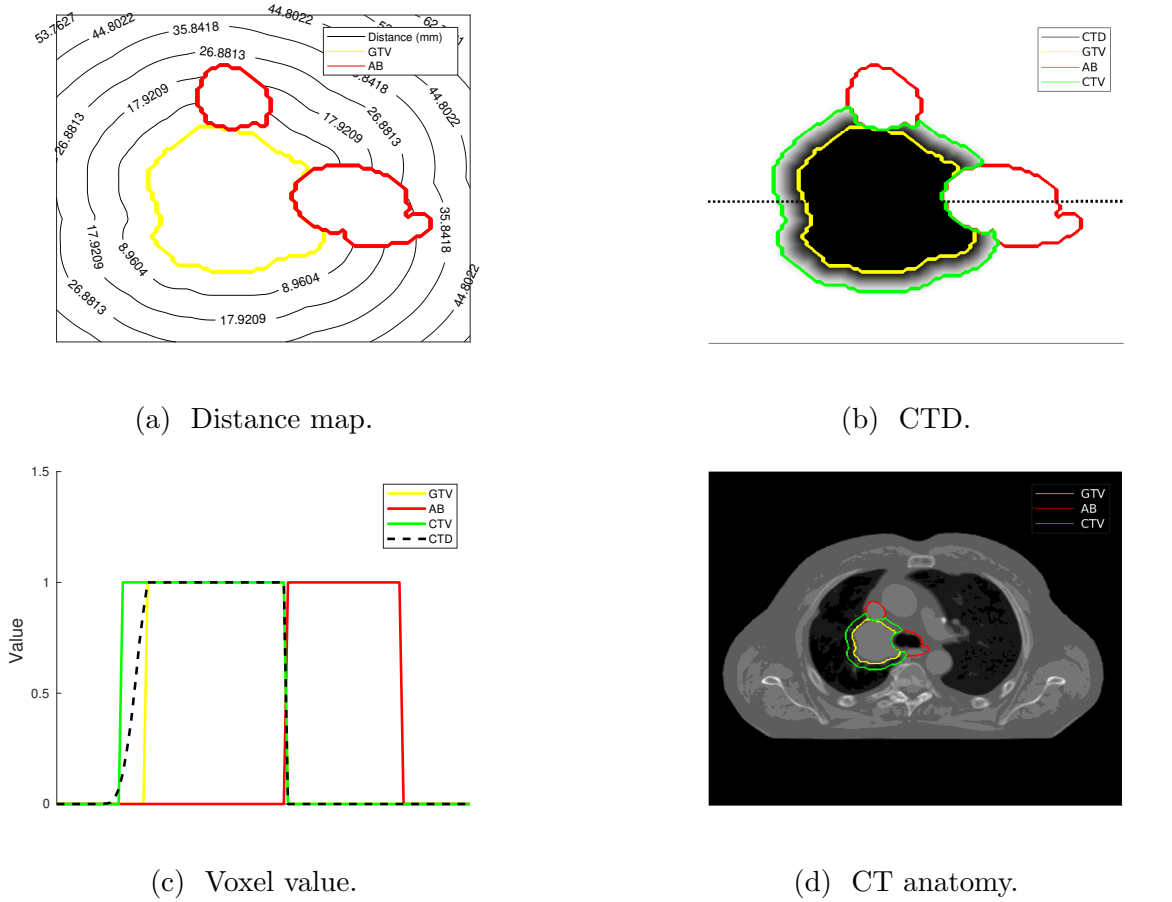
(ii) converting the Euclidean distance map into a probability map by assigning value of $p_{t,i}$ to each target voxel i, using interpolated values of the curve $p_t(r)$, illustrated in Fig. 1b.

Finally, the CTD is corrected for anatomical barriers by setting $p_{t,i} = 0$ for all non-zero voxels that overlap with anatomical barrier masks. Anatomical barriers were identified and delineated for each patient by subtracting an isotropic expansion of the GTV from the clinically-accepted CTV (delineated by an experienced physician). For lung cancer patients, anatomical barriers could for example be the lung wall, the bronchus or the liver. A 2D example of a CTD for a lung tumor case is shown in Fig. 2b. As illustrated in Fig. 2b and 2c, the value of the CTD voxels vary from $p_t = 1$ inside the GTV to decreasing values as one moves away from the GTV edge.



(a) Tumor infiltration uncertainty model. (b) Probability of tumor presence.

Figure 1: (a) Assumed tumor infiltration uncertainty model. The distribution represents a truncated Gaussian probability density function (PDF), with mean 3.4 mm and standard deviation 2.8 mm, analogous to the tumor infiltration data reported in the study of Meng et al. (2012) (see Section 2.2.4). (b) Probability of tumor presence $p_t(r)$ as a function of radial distance r from the GTV edge. $p_t(r)$ is derived by integrating the PDF depicted in (a), according to Eq. 2. For instance, at a distance of 7.2 mm (= the GTV-to-CTV margin), the probability of tumor presence equals 10% which corresponds to the orange area under the curve in (a).



(a) Distance map.

(b) CTD.

(c) Voxel value.

(d) CT anatomy.

Figure 2: (a) Example of a Euclidean distance map for a lung tumor case. (b) The clinical target distribution (CTD) derived from (a) with the CTV as a reference. The shade of the CTD represents the probability of tumor presence (value between 0 (white) and 1 (black)). (c) Values of the voxels for each structure, along the dotted line drawn in (b). (d) The corresponding anatomy with contours of the anatomical barriers (AB) drawn in red.

2.3.2. CTV

Following ICRU recommendations, the CTV is defined as the volume of all voxels with a probability of tumor presence larger than 10%. Therefore, the GTV-to-CTV margin can be derived as the cut-off point where the probability of disease presence reaches 10%. Considering the probability curve of Fig. 1b, a 7.2 mm GTV-to-CTV margin is obtained. The CTV is then constructed by an isotropic expansion of the GTV with the specified margin, followed by a correction for anatomical barriers.‖

‖ OpenReggui is used for this purpose. OpenReggui functions use a spherical kernel element in order to apply a dilation filter on the image.

2.4. Robust optimization methods

In this section, the reference worst-case optimization and the proposed probabilistic optimization method are introduced. Both methods rely on the evaluation of a discrete number of treatment uncertainty scenarios which are selected from a multi-dimensional error-space (the so-called *scenario* space). The differences between the methods are found mainly in: (1) the definition of the objective function, (2) the selection of scenarios in the *scenario* space and (3) the evaluation of the target coverage objectives.

2.4.1. Worst-case optimization

The robust optimization algorithm *minimax*, as proposed by Fredriksson et al. (2011), is used as the reference method. By representing $\mathcal{S}$ as the pre-defined set of uncertainty scenarios s, *minimax* is typically formulated as:

$$\min_{\boldsymbol{w}} \max_{s} f(\boldsymbol{w}, s) \quad \text{s.t.} \begin{cases} \boldsymbol{w} \geq \boldsymbol{0} \\ s \in \mathcal{S}, \end{cases} \tag{3}$$

with f the objective function and $\boldsymbol{w}$ the optimization variable (i.e., spot weight vector) which is subject to (s.t.) a constraint in order to allow only positive solutions. Similar to Fredriksson et al. (2011), an auxiliary variable t is introduced in order to reformulate the *max*-operator in Eq. 3 into an equivalent constrained optimization problem:

$$\min_{\boldsymbol{w}, t} t \quad \text{s.t.} \begin{cases} \boldsymbol{w} \geq \boldsymbol{0} \\ t \geq f(\boldsymbol{w}, s) \ \forall \ s \in \mathcal{S}. \end{cases} \tag{4}$$

The uncertainty set $\mathcal{S}$ is selected in order to encompass a region of *scenario* space, within a certain confidence interval. Similar to previous studies (Buti et al. (2019) and Buti et al. (2020)), we follow this strategy to handle the uncertainties that influence the treatment in a systematic fashion, i.e. systematic setup errors and range errors:

- 90% of the 3D-systematic setup error probability distribution is considered by limiting the magnitude of setup errors in the range of $\pm\alpha_{3D}\Sigma_s$, in each direction x, y and z, with $\alpha_{3D} = 2.5$ (van Herk et al. (2000)). Because intermediate setup errors may also yield dose errors, setup errors are selected at a 2 mm spacing in each direction. Given that $\Sigma_s = 2.4$ mm is assumed, in total, 19 systematic setup errors are selected (nominal scenario together with six error scenarios in $\pm x$, $\pm y$ and $\pm z$ directions).
- 90% of the 1D-range error probability distribution is considered by selecting range errors with a maximum value of $\pm\alpha_{1D}\Sigma_r$, with $\alpha_{1D} = 1.64$. Three range error

scenarios are thus selected: the nominal scenario, an overshoot scenario ($+1.64\Sigma_r$) and an undershoot scenario ($-1.64\Sigma_r$).

Following conventional practice, systematic errors are combined by employing a rectangular sampling of the scenario space. Hence, all possible combinations of the selected systematic setup and range errors are considered, leading to an initial uncertainty set of $19 \times 3 = 57$ scenarios. Following Section 2.2.1, random setup errors of uncertain standard deviation are considered by storing a separate set of scenarios where random setup errors (3 mm standard deviation) are simulated directly in the beamlet dose-influence matrices. Hence, the uncertainty set contains $2 \times 57 = 114$ scenarios (57 scenarios with 0 mm standard deviation and 57 scenarios with 3 mm standard deviation for the random setup errors). As explained in Section 2.2.3, respiratory motion is simulated by evaluating dose distributions with the accumulated 4D-beamlets in all considered scenarios. Hence, the dose evaluations of all 114 scenarios inherently account for respiratory motion.

2.4.2. Probabilistic optimization

In probabilistic optimization, the expected value (EV) of the objective function over the error scenarios s is minimized:

$$\min_{\boldsymbol{w}} \left[\mathcal{E}(\boldsymbol{w}) = \sum_{s \in \mathcal{S}} p(s) f(\boldsymbol{w}, s) \right] \quad \text{s.t.} \begin{cases} \boldsymbol{w} \geq \boldsymbol{0} \\ s \in \mathcal{S}. \end{cases} \tag{5}$$

In this study, an additional penalty term is added to the objective function of Eq. 5 that minimizes the standard deviation (SD) over the scenarios:

$$\min_{\boldsymbol{w}} \left[(1-\lambda)\mathcal{E}(\boldsymbol{w}) + \lambda \sqrt{\sum_{s \in \mathcal{S}} p(s) \left(f(\boldsymbol{w}, s) - \mathcal{E}(\boldsymbol{w}) \right)^2} \right] \quad \text{s.t.} \begin{cases} \boldsymbol{w} \geq \boldsymbol{0} \\ s \in \mathcal{S}, \end{cases} \tag{6}$$

with $\mathcal{E}(\boldsymbol{w})$ defined in Eq. 5 and λ a user-defined parameter that defines the importance of the standard deviation term in the objective function. Similar approaches can be found in financial asset optimization theory where a mean-variance framework is employed in order to capture the trade-off between the expected return (the mean) and risk (the variance) (Markowitz (1952)). Moreover, the standard deviation term functions similarly to the L1 regularization norm, which is commonly applied in machine learning optimization problems, with λ analogous to the regularization rate. Regularization terms have also been introduced to proton therapy optimization, most notably in the

proton arc optimization study of Gu et al. (2020). In probabilistic planning, the aim of the expected value-standard deviation (EV-SD) approach is to provide an effective way to control the degree of robustness in probabilistic optimization.

In Eq. 6, $f(\boldsymbol{w}, s)$ is weighted by a probability factor $p(s)$, representing the scenario probability. Since all uncertainty sources are considered mutually independent, $p(s)$ is computed as the product of the probability of each individual uncertainty:

$$p(s) = p_s(s)p_r(s), \tag{7}$$

where $p_s(s)$ and $p_r(s)$ are evaluated with the setup and range error uncertainty distributions defined in Section 2.2. Normalization is applied such that the total sum of the probabilities over the scenarios equals one.

Similar to worst-case optimization, $f(\boldsymbol{w}, s)$ consists of quadratic dose fidelity terms. However, in probabilistic optimization, the target coverage objectives include the probability of tumor presence p_t. This is achieved by weighting each target voxel i with $p_{t,i}$, as defined by the CTD:

$$f(\boldsymbol{w}, s) \propto \sum_{i \in T} p_{t,i} \left(\max\{0, d_{presc} - d_i(\boldsymbol{w}, s)\}\right)^2, \tag{8}$$

with d_{presc} the prescription and T the volume of all voxels with non-zero value in the CTD. Eq. 8 represents the minimum target coverage objective, with a similar expression existing for the maximum target coverage.

In principle, in probabilistic optimization, the objective function should be integrated over the entire space of uncertainties. However, because a discrete representation of the *scenario* space is considered, a rectangular sampling of the scenario space is employed in order to select the scenarios:

- Systematic setup errors are selected at a 2 mm spacing within the [-3 Σ_s, +3 Σ_s] interval (minimum setup error of 2 mm), in each direction. Hence, 19 systematic setup errors are selected (nominal scenario together with six scenarios in each $\pm x$, $\pm y$ and $\pm z$ directions).
- Five range error scenarios are selected within the [-2Σ_r, +2Σ_r] interval (nominal scenario, $\pm 1\Sigma_r$ and $\pm 2\Sigma_r$ scenarios).

By taking the combinations of all aforementioned errors, an initial uncertainty set of $19 \times 5 = 95$ scenarios is defined. Analogous to the worst-case method, a separate set of scenarios is stored in order to account for the random setup errors whilst respiratory motion is simulated in all scenarios. Therefore, the fully probabilistic method utilizes a final uncertainty set of $2 \times 95 = 190$ scenarios. This particular uncertainty set is chosen because of the following two reasons: first, each considered uncertainty distribution is now approximated by coarse and discrete distribution, and second, scenarios are considered with a greater error magnitude in the *scenario* space with respect to worst-case optimization. Even though the combined probabilities of such extreme scenarios

are relatively low, the convergence towards a particular solution will determine whether these scenarios will influence the final dose distribution.

The uncertainty set $\mathcal{S}$ is fixed throughout the optimization process. Hence, the optimization problems defined by Eqs. 5 and 6 are solved by evaluating $f(\boldsymbol{w}, s)\ \forall\ s \in \mathcal{S}$ at each iteration of the optimization. Given that the uncertainty set $\mathcal{S}$ is fixed over time, it follows that the optimization variable $\boldsymbol{w}$ is updated in a deterministic way.

2.5. Evaluation

Plan robustness is evaluated with the Monte Carlo dose engine MCsquare (Souris (n.d.)) through the open-source platform OpenReggui (Janssens (n.d.)). Details on the comprehensive robustness evaluation procedure employed by MCsquare can be found in Souris et al. (2019) and Sterpin et al. (2021). In short, at least 250 *evaluation* scenarios are randomly sampled from the respective uncertainty distributions, which include uncertainties of setup errors (both systematic and random), range errors and respiratory motion. The dose distribution is recomputed for each scenario, with a statistical noise level below 2%. For each treatment plan evaluation, the nominal dose distribution is normalized with a correction factor such that GTV $D_{95} = d_{presc}$. This correction factor is subsequently applied to the evaluation scenarios by multiplying the dose distributions with the correction factor.

Two types of evaluation metrics are calculated: (1) dose-volume histogram (DVH) metrics in the worst case *evaluation* scenario, computed after discarding the 10% worst scenarios (based on the target D_{95}), and (2) DVH metrics based on the average *evaluation* scenario, taking into account all sampled scenarios.

The evaluation procedure, accessed through OpenReggui, is modified from the available open-source version, in order to evaluate tumor infiltration uncertainties. Rather than using a fixed target volume in each *evaluation* scenario, the effects of geometric errors and tumor infiltration errors are combined as follows: in each *evaluation* scenario, a tumor infiltration is randomly sampled from the assumed tumor infiltration distribution. A target volume "realization" is defined by following a similar procedure to the construction of the CTV (see section 2.3.2): the GTV is isotropically dilated with a target margin that is equal to the sampled tumor infiltration, followed by a correction for anatomical barriers. The target coverage metrics (D_{95} and D_5) are subsequently calculated by evaluating the dose in the obtained target volume realization. As a result, each evaluation scenario features a target volume with a specific target margin that depends on the sampled tumor infiltration error.

2.6. Patient cases

Five lung tumor cases are used to test and compare the optimization methods. All patients had a prescription of 60 Gy (delivered in 30 fractions of 2 Gy), treated with the IMPT-PBS modality. An overview of the patient characteristics (GTV size, motion amplitude and tumor position) and treatment plan features (beam angles) are listed in

Table 1. The OARs considered in this study are the lungs-GTV volume, esophagus, heart and spinal cord.

The goal of the present study is to demonstrate the features of the optimization methods when conflicts are present in the objective function. Therefore, the reference dose level is set at 0 Gy for all OAR planning objectives, so that all non-zero dose in an OAR voxel is penalized. Following clinical practice, the target and serial OAR (spinal cord, oesophagus and bronchus) objectives are robustified whilst the parallel OARs (heart and lungs) are treated in the nominal scenario only.

To compare the optimization methods consistently, the treatment plans are designed to have similar target coverage between different methods while limiting the OAR doses as much as possible. For each patient, this is achieved by adjusting the target coverage objective weights of each method whilst keeping the OAR objective weights identical. Acceptability for target coverage is defined as D_{95} must be at least $95\%d_{presc}$ ($D_{95} \geq 57$ Gy) and D_5 may not exceed $105\%d_{presc}$ ($D_5 \leq 63$ Gy), in the worst *evaluation* scenario.

Table 1: Patient (P1-5) and treatment plan characteristics.

	GTV size	**Motion amplitude**			**Tumor position**	**Beam angles**
		LR	AP	SI		
	[cm^3]	[mm]	[mm]	[mm]		[°]
P1	75.4	4.2	2.1	3.1	RML	0, 270, 310
P2	61.0	3.1	2.9	3.7	LLL	90, 135, 180
P3	16.0	1.4	2.9	0.8	RUL	180, 225, 270
P4	31.9	0.8	1.2	0.5	LUL	90, 135, 180
P5	68.9	2.2	1.8	6.6	RUL	180, 225, 270

Tumor motion amplitude (in left-right (LR), anterior-posterior (AP) and superior-inferior (SI) directions). Tumor positions (right-middle lobe (RML), right-upper lobe (RUL), right-lower lobe (RLL), left-lower lobe (LUL) and left-upper lobe (LUL)).

3. Results

In this section, the results of the worst-case (WC) optimization method are compared to two probabilistic optimization methods: expected value (EV) optimization ($\lambda = 0$ in Eq. 6) and expected value-standard deviation (EV-SD) optimization ($\lambda > 0$ in Eq. 6). The results of the EV-SD method are presented for a λ value equal to 0.5, thereby giving equal importance to the mean and and standard deviation terms. In Section 3.1, the performance of each method is assessed in terms of the obtained plan quality and robustness. Section 3.2 reports the computation cost of the plan optimization algorithms.

3.1. Dosimetric results

The DVH bands illustrated in Fig. 3 display the results of the evaluation procedure for patients P1-5. The relevant evaluation metrics are summarized in Table 2 and Table 3. Table 2 reports the target coverage metrics (D_{95} and D_5) in the worst *evaluation* scenario, together with the target DVH bandwidths (ΔD_{95} and ΔD_5). The DVH bandwidths are computed as the difference between the highest and lowest dose level, within the 90% confidence interval. DVH bandwidths provide a measure of the plan robustness, i.e. the narrower the band, the lower the sensitivity to the uncertainties. Note that the target volume is not fixed in each *evaluation* scenario. Rather, as explained in Section 2.5, the target DVH metrics D_{95} and D_5 are calculated by evaluating the dose in a variable target volume "realization" that depends on the sampled tumor infiltration error.

Except for a slightly elevated target D_5 for patient P1 in the worst-case method (0.2 Gy over the constraint), the generated treatment plans have a target coverage within the acceptability criteria, as defined in Section 2.6. The target robustness for all patients between the methods is comparable, illustrated by a similar DVH bandwidths at both the D_{95} and D_5 dose level.

Table 3 reports the OAR DVH metrics in both the worst and average *evaluation* scenarios. For OARs that received meaningful dose levels, the EV and EV-SD methods reduced the D_2 dose of the esophagus for all patients. Most notably a reduction of worst case D_2 dose is observed for patients P2-5 of 2.8 Gy, 7.7 Gy, 7.2 Gy and 1.2 Gy for the EV method and 3.0 Gy, 7.6 Gy, 11.0 Gy and 2.1 Gy for the EV-SD method. Moreover, for patients P2-5, the reduction of average D_2 esophagus dose was 3.9 Gy, 5 Gy, 4.5 Gy and 3.8 Gy for the EV method and 5.0 Gy, 5.0 Gy, 5.4 Gy and 4.0 Gy for the EV-SD method. Minor differences are observed for the heart dose with a decrease of respectively 0.8% and 0.7 Gy for worst V_{15} and D_{mean}, for patient P1 for the EV-SD method. For patient P2, the EV-SD method showed a similar heart dose as the worst case method whilst the EV method had a slightly increased worst D_{mean} of 0.5 Gy. The EV and EV-SD methods reduced the worst case spinal cord D_2 by 4.3 Gy and 2.9 Gy, respectively for patient P2. Whilst an increase of 5.2 Gy and 4.9 Gy D_2 spinal cord dose is observed for patient P5. In terms of bronchus dose, which often acts as a anatomical barrier for lung tumor patients, no significant differences are present between the studied methods. However, the lung dose was higher for both probabilistic methods for all patients, with maximum differences of 1.0 Gy and 1.1 Gy worst case mean lung dose for the EV and EV-SD method, respectively.

Examples of the planned dose distributions produced by each method are illustrated in Fig. 4. Taking the dose distributions of P4 as representative example, the high-dose isodose lines are closer to near the most proximal OAR (the esophagus) for the worst-case method whilst a sharper dose fall-off is observed towards the OAR for the probabilistic methods (most notably the esophagus). However, the probabilistic methods display an increased total dose volume as compared to the worst-case method.

Table 2: Target DVH metrics for plans of patients P1-5, obtained using worst-case (WC), expected value (EV) and expected value-standard deviation (EV-SD) methods. Target coverage: D_{95} and D_5 computed in the worst *evaluation* scenario; and target robustness: ΔD_{95} and ΔD_5, as the DVH bandwidths at the D_{95} and D_5 dose level, respectively.

ROI	Metric	Method	P1	P2	P3	P4	P5
Target	D_{95} [Gy]	WC	58.6	57.6	57.8	57.4	57.5
	(worst)	EV	58.4	57.7	57.6	57.2	57.5
		EV-SD	58.4	57.5	57.5	57.5	57.3
	D_5 [Gy]	WC	63.2	63.0	62.6	63.0	62.9
	(worst)	EV	62.9	62.9	62.8	62.9	63.0
		EV-SD	63.0	63.0	62.7	62.8	62.9
	ΔD_{95} [Gy]	WC	1.7	2.6	2.2	2.9	2.8
		EV	1.7	2.4	2.5	2.9	2.6
		EV-SD	1.8	2.5	2.5	2.6	2.6
	ΔD_5 [Gy]	WC	0.8	0.5	0.8	1.0	0.9
		EV	0.6	0.7	0.9	1.2	1.2
		EV-SD	0.7	0.7	0.8	1.0	1.1

Table 3: Organ-at-risk DVH metrics (lungs-GTV, esophagus, heart, spinal cord and bronchus) for plans of patients P1-5, obtained using worst-case (WC), expected value (EV) and expected value-standard deviation (EV-SD) methods. Metrics are reported in the worst and average (avg.) *evaluation* scenarios.

ROI	Metric	Method	P1	P2	P3	P4	P5
Lungs-GTV	V_{20} [%] (worst)	WC	12.6	10.4	8.7	9.8	12.1
		EV	14.6	12.4	9.9	10.7	14.0
		EV-SD	14.5	12.5	10.2	10.1	13.9
	V_{20} [%] (avg.)	WC	9.9	9.5	7.7	8.1	11.1
		EV	12.5	11.3	8.6	9.0	12.2
		EV-SD	12.3	11.3	8.6	8.4	12.3
	D_{mean} [Gy] (worst)	WC	6.1	4.9	4.4	5.0	5.9
		EV	6.8	5.9	5.0	5.5	6.8
		EV-SD	6.8	6.0	5.1	5.1	6.8
	D_{mean} [Gy] (avg.)	WC	5.1	4.5	4.0	4.2	5.4
		EV	6.0	5.4	4.4	4.7	6.1
		EV-SD	5.9	5.4	4.4	4.3	6.1
Esophagus	D_2 [Gy] (worst)	WC	10.2	27.3	16.7	29.8	57.6
		EV	9.7	24.5	9.0	22.6	56.4
		EV-SD	6.5	24.3	9.1	18.8	55.5
	D_2 [Gy] (avg.)	WC	3.2	17.0	11.5	12.9	43.3
		EV	3.4	13.1	6.5	8.4	39.5
		EV-SD	3.1	12.0	6.5	7.5	39.3
Heart	V_{15} [%] (worst)	WC	8.9	10.2	0.0	0.0	4.3
		EV	8.4	11.5	0.0	0.0	4.8
		EV-SD	8.1	10.2	0.0	0.0	4.5
	V_{15} [%] (avg.)	WC	6.6	7.2	0.0	0.0	2.5
		EV	5.7	6.8	0.0	0.0	2.8
		EV-SD	5.8	6.5	0.0	0.0	2.7
	D_{mean} [Gy] (worst)	WC	4.2	4.8	0.0	0.0	2.1
		EV	3.8	5.3	0.0	0.0	2.4
		EV-SD	3.8	4.7	0.0	0.0	2.2
	D_{mean} [Gy] (avg.)	WC	3.0	3.4	0.0	0.0	1.3
		EV	2.5	3.1	0.0	0.0	1.4
		EV-SD	2.6	3.0	0.0	0.0	1.3
Spinal cord	D_2 [Gy] (worst)	WC	0.3	0.1	33.4	2.2	16.1
		EV	0.8	0.1	29.1	3.9	21.3
		EV-SD	0.4	0.1	30.5	3.3	21.0
	D_2 [Gy] (avg.)	WC	0.1	0.1	19.8	0.6	8.3
		EV	0.1	0.1	17.0	0.9	11.5
		EV-SD	0.1	0.1	17.5	0.9	11.5
Bronchus	D_2 [Gy] (worst)	WC	60.8	61.2	58.1	3.2	61.6
		EV	60.3	61.8	59.0	3.4	61.6
		EV-SD	60.1	61.7	58.3	4.3	61.7
	D_2 [Gy] (avg.)	WC	58.4	59.0	49.2	0.9	59.6
		EV	56.5	59.3	47.3	1.1	59.2
		EV-SD	56.4	58.9	47.5	1.1	59.2

3.2. Computational cost of the optimization algorithms

Table 4 reports the number of beamlets, maximum number of iterations and the total optimization time (t_{opt}) of the WC, EV and EV-SD methods:

Table 4: Treatment plan features including the number of beamlets, maximum number of iterations and total optimization time for worst-case (WC), expected value (EV) and expected value-standard deviation (EV-SD) methods (patients P1-5).

ROI	Method	P1	P2	P3	P4	P5
No beamlets	WC	6339	5870	3855	3407	6183
	EV	10483	9655	4994	6197	9675
	EV-SD	10487	8857	5035	5861	9704
Max iterations	WC	1768	750	616	609	1083
	EV	350	350	350	300	350
	EV-SD	350	350	350	300	350
$\mathbf{t}_{opt}$ [hours]	WC	110	36	19	12	87
	EV	48	54	41	20	39
	EV-SD	45	52	45	18	38

The treatment plans of the WC method features more beamlets than the EV and EV-SD methods. Note that the number of beamlets between EV and EV-SD methods varies slightly due to a spot filtering step applied by MIROpt after the optimization process. This consists of removing the low MU spots (MU threshold = 0.011) from the treatment plan. The plan optimization was faster for three out of five patients (P2, P3 and P4) for the worst-case method. However, the significantly higher maximum number of iterations for patients P1 and P5 resulted in a higher optimization time for these cases, as compared to the EV and EV-SD method.

4. Discussion

This study introduces a probabilistic approach that considers microscopic tumor infiltration uncertainty in three areas of the treatment planning process: (1) the definition of the target, (2) the robust optimization process, and (3) the plan evaluation procedure.

4.1. Probabilistic target

The CTD represents a 3D-probability map of tumor presence for a given patient. This enables the target voxels to be weighted in the objective function according to their

assumed probabilities, as opposed to the binary CTV, which weights each target voxel equally. The probability of tumor presence was derived from a population-based tumor infiltration uncertainty distribution found in the literature (Meng et al. (2012)). Studies reporting biological models of tumor infiltration uncertainty are relatively sparse and generally depend on various histopathological characteristics such as the tumor site and tumor progression (Apolle et al. (2017)). In this study, however, the proposed CTD procedure is independent of the assumed uncertainty model. Hence, any uncertainty model can be inserted, e.g. anisotropic or manual models, given that the distribution of tumor infiltrations over a patient population is known.

It must be noticed that the total volume of non-zero voxels in the CTD is significantly larger than the CTV. The drawback being that CTD-based treatment plans feature more beamlets than CTV-based plans (see Table 4). This results in a longer dose-influence matrix calculation process for the probabilistic methods and higher memory consumption in the optimization. Similarly, the computational cost of an iteration is higher for the probabilistic methods, given their increased number of beamlets. However, this is partly compensated by the lower number of iterations necessary to produce a probabilistic method treatment plan, indicating a faster convergence rate as compared to the WC method.

The CTD could be further improved by refining the proposed procedure: in Section 2.3.1, the probability of disease presence was associated with the Euclidean distance from the GTV edge. However, this approach does not consider the path that the tumor cells travel around the anatomical barriers. An alternative to the Euclidean distance has been proposed by Shusharina et al. (2020), who employed a 'shortest path' algorithm in order to compute distance maps that take into account anatomical barriers.

4.2. Probabilistic formulation of the objective function

Tumor infiltration uncertainties are combined with the geometric treatment uncertainties by incorporating the CTD into a probabilistic optimization algorithm. The results are subsequently compared with an approach that utilizes the CTV in a worst-case optimization algorithm. These options are the most statistically consistent if the targets are defined according to a tumor infiltration probability distribution: on the one hand, the CTD represents the probability of tumor presence and therefore needs to be evaluated in a probabilistic objective function. On the other hand, the CTV is especially appropriate for a worst-case approach, as it can be interpreted as a worst-case volume that encompasses 90% to 95% of tumor infiltrations in patient population. Mixing both frameworks, e.g. using the CTD in a worst-case algorithm, is statistically inconsistent and should therefore be avoided.

In worst-case optimization, the plan's robustness is defined *a priori* by the choice of the uncertainty set $\mathcal{S}$ which determines the fraction of covered scenarios in the *scenario* space (Fredriksson (2013)). Hence, the degree of robustness can be defined by specifying the *scenario* space integration limits, α_{3D} and α_{1D}, which in turn establishes a confidence

interval (Buti et al. (2019)). Unfortunately, in probabilistic optimization, the robustness can not be quantified in a similar way as the objective function is integrated over the entire space of uncertainties. Therefore, in this study, an additional term – the standard deviation – is introduced with the goal of controlling the degree of robustness. By increasing its relative importance (λ in Eq. 6) in the objective function, more emphasis is placed minimizing the spread of the objectives from the mean and hence the final solution will become more robust to uncertainties. Although this approach lacks the quantitative nature found in the worst-case optimization, the standard deviation term can be considered a useful tool to manage the degree of robustness in probabilistic optimization.

For consistent comparison of the methods, the treatment plans were designed to have similar target coverage and robustness. Given that the IMPT plans value target coverage most of all, the DVH bandwidths are similar between the EV and EV-SD methods. This indicates that the standard deviation term in the EV-SD method had limited impact for the studied patient cases. Without the constraint of equal target coverage, we expect a larger difference between the EV and EV-SD method. In that case, λ becomes a meta-parameter of the algorithm that needs to be optimized. Further research is needed to investigate the clinical value of such a probabilistic framework as there could be a trade-off between plan quality and robustness.

Balancing the trade-off between minimizing OAR exposure and achieving sufficient target coverage is the main conflict that an optimizer needs to solve. When tumor infiltration uncertainties are considered explicitly in the optimization process, a fully probabilistic approach provides an alternative method to redefine the trade-off preferred by the more conservative worst-case implementation. Probabilistic optimization allows the optimizer to mitigate conflicts in the objective function due to its following two features: first, the decrease in probability with distance will have as a result that target voxels near the surrounding ROIs, are weighted less in the objective function, as compared to the CTV case. Second, less importance is given to improbable scenarios where typically most conflicts are found. Results of Section 3 indicate that the fully probabilistic methods are able to generate treatment plans that reduce exposure of OARs that are located near the target (usually the esophagus and heart for lung tumor patients), whilst ensuring acceptable target coverage. The effect is negligible for organs that nearly overlap with the GTV, such as the bronchus, which often acts as an anatomical barrier in the studied patient cases.

Without the presence of dose limiting structures around the target, a probabilistic optimizer can increase the extent of the high-dose volume without cost. Therefore, the irradiated volume can potentially increase as compared to CTV-based worst-case optimization. Similarly, because fully probabilistic optimization explores more extreme scenarios, high-doses regions can be extended in these scenarios, if planning objectives are not conflicting. These features of probabilistic optimization have the potential drawback to yield treatment plans with increased integral dose. The dosimetric results demonstrate that in lung tumor cases, where the target is embedded in the lung

structure, the mean lung dose is elevated for most patient cases as compared to the worst-case method. Moreover, the dose can also potentially slightly increase for OARs that located far away from the target, as seen in the spinal cord dose for patient P5. The high-dose region of the dose distribution could potentially be controlled by introducing a dose limiting structure around the GTV or including a dose fall-off function in the objective function.

The above-mentioned findings were derived from the analysis of a set of five lung tumor cases. Future case-studies should involve a larger patient cohort and a variety of tumor locations. This study can serve as a framework for research on the possible clinical benefits of probabilistic treatment planning.

4.3. Evaluating tumor infiltration uncertainties

The aim of the evaluation procedure has been to evaluate the effects of the tumor infiltration uncertainties combined with the geometric uncertainties in a scenario-based approach. This was achieved by sampling realizations of the target volume from the assumed tumor infiltration uncertainty model. Therefore, the worst-case *evaluation* scenario does not necessarily feature the same target volume. Rather, the evaluation of the target coverage in the worst case scenario depends on how the target volume uncertainty combines with the other geometric uncertainties. By employing this evaluation procedure, the tumor infiltration uncertainties are treated together with other assumed uncertainties during plan evaluation. Therefore, any bias is removed from assumptions made in the plan optimization stage. In Shusharina et al. (2018), an alternative method is suggested when evaluating CTD-based plans, namely to compute evaluation metrics based on the expected target volume (that is, computing a DVH by weighting each voxel with the CTD probability). Although such an approach is valid for the evaluation of CTD coverage, it can be considered less adequate for evaluating treatment plans that use the CTV as the target, which is a part of this study.

5. Conclusion

This study proposes a fully probabilistic approach that incorporates a probabilistic target, i.e. the *clinical target distribution* (CTD), in a robust optimization process. The CTD explicitly features the probability of tumor presence in its target definition and is derived directly from a reported probability distribution of tumor infiltrations. By applying a probabilistic formulation of the objective function, the CTD is combined with other treatment uncertainties in a statistically sound framework. The method has been tested on five lung tumor patients and was benchmarked against CTV-based worst-case optimization. Results indicate that for the studied lung tumor patients, a fully probabilistic approach favours the reduction of dose levels in the esophagus, and compensates by extending the high-dose region in a low conflicting organ such as the lung. These findings demonstrate that a fully probabilistic approach can be considered

a promising alternative when including tumor infiltration uncertainties explicitly in the treatment planning process.

Acknowledgements

Gregory Buti is supported by the Télévie Grant from the Belgian 'Fonds National pour la Recherche Scientifique' F.R.S-FNRS (Grant No. 7453918F). Kevin Souris and Ana M. Barragán Montero are funded by the Walloon region with MECATECH/BIOWIN (Grant No. 8090) and PROTHERWAL/CHARP (Grant No. 7289), respectively. John A. Lee is a Senior Research Associate with the F.R.S.-FNRS. Computational resources have been provided by the supercomputing facilities of the Université Catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the F.R.S.-FNRS under convention 2.5020.11.

References

Apolle, R., Rehm, M., Bortfeld, T., Baumann, M. and Troost, E. G. (2017). The clinical target volume in lung, head-and-neck, and esophageal cancer: Lessons from pathological measurement and recurrence analysis, *Clinical and Translational Radiation Oncology* **3**: 1–8.
URL: *https://doi.org/10.1016/j.ctro.2017.01.006*

Barragán-Montero, A. M. (2017). *Robust, accurate and patient-specific treatment planning for proton therapy*, PhD thesis, UCL-Université Catholique de Louvain.

Barragán-Montero, A. M. (n.d.). Miropt - http://www.openmiropt.org/ accessed november 2019.

Barragán-Montero, A. M., Souris, K., Sanchez-Parcerisa, D., Sterpin, E. and Lee, J. A. (2017). Performance of a hybrid monte carlo-pencil beam dose algorithm for proton therapy inverse planning, *Medical Physics* **45**(2): 846–862.
URL: *https://doi.org/10.1002/mp.12688*

Barragán-Montero, A. M., Souris, K., Sterpin, E. and Lee, J. (2016). OC-0265: Efficient implementation of random errors in robust optimization for proton therapy with monte carlo, *Radiotherapy and Oncology* **119**: S123–S124.
URL: *https://doi.org/10.1016/s0167-8140(16)31514-6*

Bohoslavsky, R., Witte, M. G., Janssen, T. M. and van Herk, M. (2013). Probabilistic objective functions for margin-less IMRT planning, *Physics in Medicine and Biology* **58**(11): 3563–3580.
URL: *https://doi.org/10.1088/0031-9155/58/11/3563*

Bortfeld, T., Shusharina, N. and Craft, D. (2021). Probabilistic definition of the clinical target volume—implications for tumor control probability modeling and optimization,

Physics in Medicine & Biology **66**(1): 01NT01.
URL: *https://doi.org/10.1088/1361-6560/abcad8*

Buti, G., Souris, K., Barragán-Montero, A. M., Cohilis, M., Lee, J. A. and Sterpin, E. (2020). Accelerated robust optimization algorithm for proton therapy treatment planning, *Medical Physics* .
URL: *https://doi.org/10.1002/mp.14132*

Buti, G., Souris, K., Barragán-Montero, A. M., Lee, J. A. and Sterpin, E. (2019). Towards fast and robust 4d optimization for moving tumors with scanned proton therapy, *Medical Physics* **46**(12): 5434–5443.
URL: *https://doi.org/10.1002/mp.13850*

Fredriksson, A. (2012). A characterization of robust radiation therapy treatment planning methods-from expected value to worst case optimization, *Medical Physics* **39**(8): 5169–5181.
URL: *https://doi.org/10.1118/1.4737113*

Fredriksson, A. (2013). *Robust optimization of radiation therapy accounting for geometric uncertainty*, PhD thesis, KTH Royal Institute of Technology.

Fredriksson, A. and Bokrantz, R. (2014). A critical evaluation of worst case optimization methods for robust intensity-modulated proton therapy planning, *Medical Physics* **41**(8Part1): 081701.
URL: *https://doi.org/10.1118/1.4883837*

Fredriksson, A. and Bokrantz, R. (2016). The scenario-based generalization of radiation therapy margins, *Physics in Medicine and Biology* **61**(5): 2067–2082.
URL: *https://doi.org/10.1088/0031-9155/61/5/2067*

Fredriksson, A., Forsgren, A. and Hårdemark, B. (2011). Minimax optimization for handling range and setup uncertainties in proton therapy, *Medical Physics* **38**(3): 1672–1684.
URL: *https://doi.org/10.1118/1.3556559*

Gu, W., Ruan, D., Lyu, Q., Zou, W., Dong, L. and Sheng, K. (2020). A novel energy layer optimization framework for spot-scanning proton arc therapy, *Medical Physics* **47**(5): 2072–2084.
URL: *https://doi.org/10.1002/mp.14083*

Janssens, G. (n.d.). Openreggui - https://openreggui.org/.

Liu, W., Zhang, X., Li, Y. and Mohan, R. (2012). Robust optimization of intensity modulated proton therapy, *Medical Physics* **39**(2): 1079–1091.
URL: *https://doi.org/10.1118/1.3679340*

Markowitz, H. (1952). PORTFOLIO SELECTION∗, *The Journal of Finance* **7**(1): 77–91.
URL: *https://doi.org/10.1111/j.1540-6261.1952.tb01525.x*

Meng, X., Sun, X., Mu, D., Xing, L., Ma, L., Zhang, B., Zhao, S., Yang, G., Kong, F.-M. S. and Yu, J. (2012). Noninvasive evaluation of microscopic tumor extensions

using standardized uptake value and metabolic tumor volume in non-small-cell lung cancer, *International Journal of Radiation Oncology∗Biology∗Physics* **82**(2): 960–966.
URL: *https://doi.org/10.1016/j.ijrobp.2010.10.064*

Paganetti, H. (2012). Range uncertainties in proton therapy and the role of monte carlo simulations, *Physics in Medicine and Biology* **57**(11): R99–R117.
URL: *https://doi.org/10.1088/0031-9155/57/11/r99*

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, *Physics in Medicine and Biology* **53**(6): 1689–1700.
URL: *https://doi.org/10.1088/0031-9155/53/6/013*

Shusharina, N., Craft, D., Chen, Y.-L., Shih, H. and Bortfeld, T. (2018). The clinical target distribution: a probabilistic alternative to the clinical target volume, *Physics in Medicine & Biology* **63**(15): 155001.
URL: *https://doi.org/10.1088/1361-6560/aacfb4*

Shusharina, N., Söderberg, J., Edmunds, D., Löfman, F., Shih, H. and Bortfeld, T. (2020). Automated delineation of the clinical target volume using anatomically constrained 3d expansion of the gross tumor volume, *Radiotherapy and Oncology* **146**: 37–43.
URL: *https://doi.org/10.1016/j.radonc.2020.01.028*

Souris, K. (n.d.). Mcsquare - http://www.openmcsquare.org/ accessed june 2019.

Souris, K., Barragán-Montero, A. M., Janssens, G., Perri, D. D., Sterpin, E. and Lee, J. A. (2019). Technical note: Monte carlo methods to comprehensively evaluate the robustness of 4d treatments in proton therapy, *Medical Physics* .
URL: *https://doi.org/10.1002/mp.13749*

Sterpin, E., Rivas, S. T., den Heuvel, F. V., George, B., Lee, J. A. and Souris, K. (2021). Development of robustness evaluation strategies for enabling statistically consistent reporting, *Physics in Medicine & Biology* **66**(4): 045002.
URL: *https://doi.org/10.1088/1361-6560/abd22f*

Stroom, J., Gilhuijs, K., Vieira, S., Chen, W., Salguero, J., Moser, E. and Sonke, J.-J. (2014). Combined recipe for clinical target volume and planning target volume margins, *International Journal of Radiation Oncology∗Biology∗Physics* **88**(3): 708–714.
URL: *https://doi.org/10.1016/j.ijrobp.2013.08.028*

The International Commission on Radiation Units and Measurements (2010). *Journal of the ICRU* **10**(1): NP.2–NP.
URL: *https://doi.org/10.1093/jicru/ndq001*

Unkelbach, J., Alber, M., Bangert, M., Bokrantz, R., Chan, T. C. Y., Deasy, J. O., Fredriksson, A., Gorissen, B. L., van Herk, M., Liu, W., Mahmoudzadeh, H., Nohadani, O., Siebers, J. V., Witte, M. and Xu, H. (2018). Robust radiotherapy planning, *Physics in Medicine & Biology* **63**(22): 22TR02.
URL: *https://doi.org/10.1088/1361-6560/aae659*

673 Unkelbach, J., Bortfeld, T., Cardenas, C. E., Gregoire, V., Hager, W., Heijmen, B.,
674 Jeraj, R., Korreman, S. S., Ludwig, R., Pouymayou, B., Shusharina, N., Söderberg,
675 J., Toma-Dasu, I., Troost, E. G. and Osorio, E. V. (2020). The role of computational
676 methods for automating and improving clinical target volume definition, *Radiotherapy*
677 *and Oncology* **153**: 15–25.
678 **URL:** *https://doi.org/10.1016/j.radonc.2020.10.002*

679 Unkelbach, J., Bortfeld, T., Martin, B. C. and Soukup, M. (2008). Reducing the
680 sensitivity of IMPT treatment plans to setup errors and range uncertainties via
681 probabilistic treatment planning, *Medical Physics* **36**(1): 149–163.
682 **URL:** *https://doi.org/10.1118/1.3021139*

683 van Herk, M., Remeijer, P., Rasch, C. and Lebesque, J. V. (2000). The probability of
684 correct target dosage: dose-population histograms for deriving treatment margins
685 in radiotherapy, *International Journal of Radiation Oncology∗Biology∗Physics*
686 **47**(4): 1121–1135.
687 **URL:** *https://doi.org/10.1016/s0360-3016(00)00518-6*

688 Wächter, A. and Biegler, L. T. (2005). On the implementation of an interior-
689 point filter line-search algorithm for large-scale nonlinear programming, *Mathematical*
690 *Programming* **106**(1): 25–57.
691 **URL:** *https://doi.org/10.1007/s10107-004-0559-y*

692 Wanet, M., Sterpin, E., Janssens, G., Delor, A., Lee, J. A. and Geets, X. (2014).
693 Validation of the mid-position strategy for lung tumors in helical TomoTherapy,
694 *Radiotherapy and Oncology* **110**(3): 529–537.
695 **URL:** *https://doi.org/10.1016/j.radonc.2013.10.025*

Additional figures

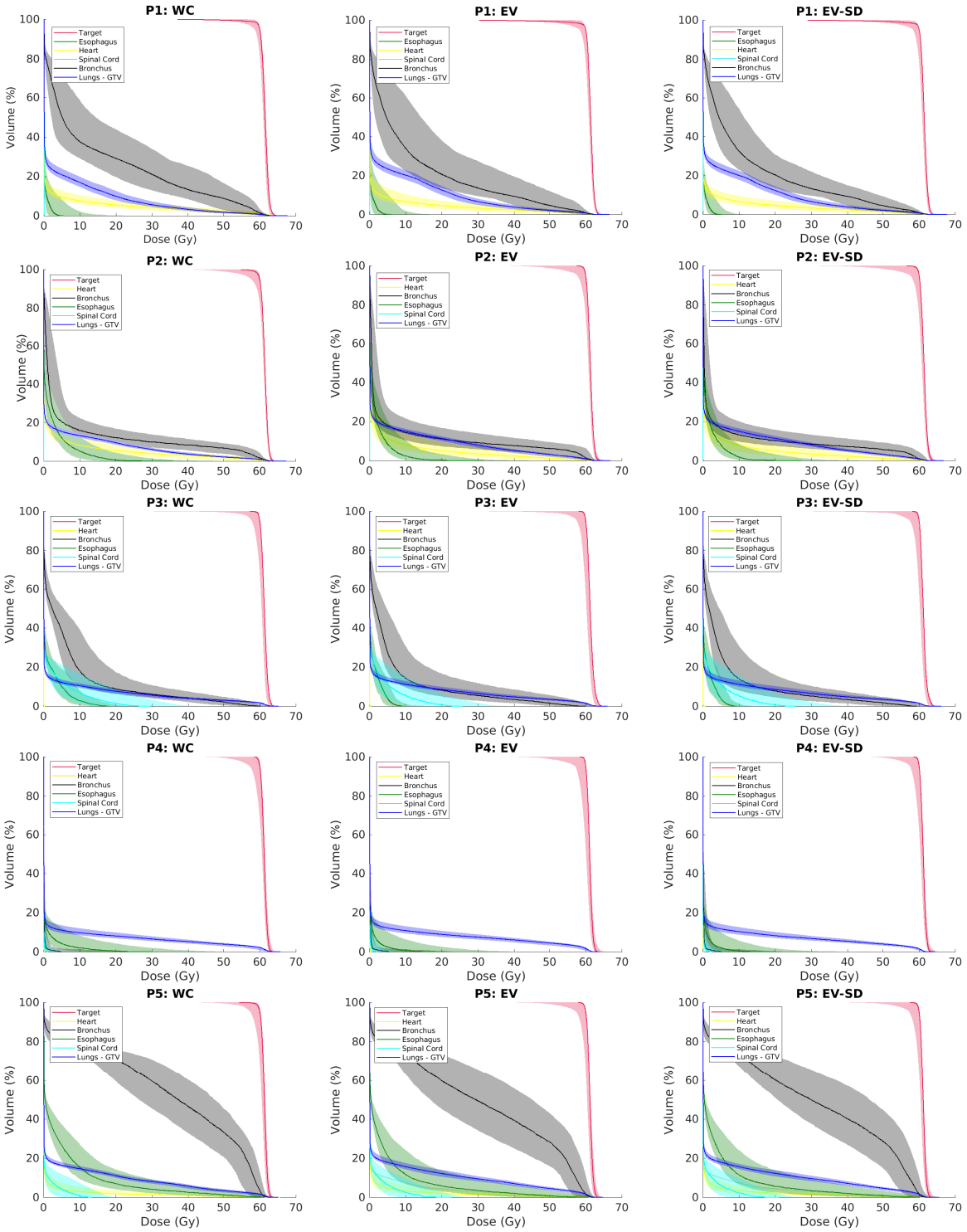


Figure 3: Dose-volume histogram (DVH) bands for treatment plans produced by the worst-case (WC), expected value (EV) and expected value-standard deviation (EV-SD) methods, for patients P1-5. Solid lines represent the DVH of the nominal scenario.

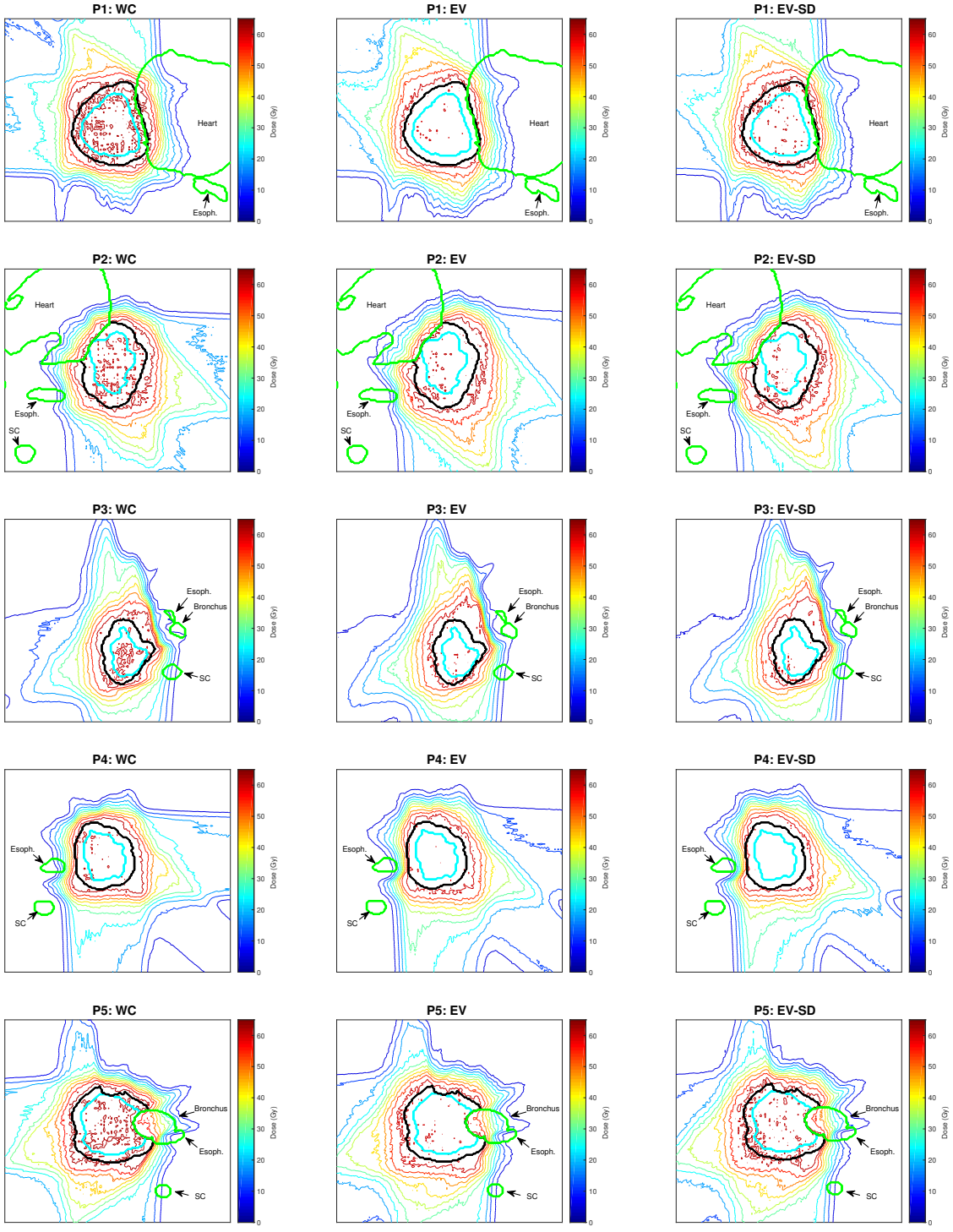


Figure 4: Planned dose distributions produced by the worst-case (WC), expected value (EV) and expected value-standard deviation (EV-SD) methods, for patients P1-5. The GTV, CTV and OARs contours are shown in turquoise, black and green, respectively.