Analytical probabilistic modeling of dose-volume histograms

Niklas Wahl[a,b,c], Philipp Hennig[d,e], Hans-Peter Wieser[a,b,f,g], Mark Bangert[a,b]

[a]German Cancer Research Center – DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[b]Heidelberg Institute for Radiation Oncology – HIRO, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[c]Department of Physics and Astronomy, Ruprecht Karls University Heidelberg, Grabengasse 1, 69117 Heidelberg, Germany
[d]Probabilistics Numerics, Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany
[e]Chair for the Methods of Machine Learning, Eberhard Karls University Tübingen, 72024 Tübingen, Germany
[f]Medical Faculty, Ruprecht Karls University Heidelberg, Grabengasse 1, 69117 Heidelberg, Germany
[g]Department for Medical Physics, Ludwig-Maximilians-Universität München (LMU Munich), 85748 Garching b. München, Germany

Version 3, typeset July 21, 2020

Corresponding author(s):
Niklas Wahl: n.wahl@dkfz.de

Abstract

Purpose: Radiotherapy, especially with charged particles, is sensitive to executional and preparational uncertainties that propagate to uncertainty in dose and plan quality indicators, e. g., dose-volume histograms (DVHs). Current approaches to quantify and mitigate such uncertainties rely on explicitly computed error scenarios and are thus subject to statistical uncertainty and limitations regarding the underlying uncertainty model. Here we present an alternative, analytical method to approximate moments, in particular expectation value and (co)variance, of the probability distribution of DVH-points, and evaluate its accuracy on patient data.

Methods: We use Analytical Probabilistic Modeling (APM) to derive moments of the probability distribution over individual DVH-points based on the probability distribution over dose. By using the computed moments to parameterize distinct probability distributions over DVH-points (here normal or beta distributions), not only the moments but also percentiles, i. e., α-DVHs, are computed. The model is subsequently evaluated on three patient cases (intracranial, paraspinal, prostate) in 30- and single-fraction scenarios by assuming the dose to follow a multivariate normal distribution, whose moments are computed in closed-form with APM. The results are compared to a benchmark based on discrete random sampling.

Results: The evaluation of the new probabilistic model on the three patient cases

1

against a sampling benchmark proves its correctness under perfect assumptions as well as good agreement in realistic conditions. More precisely, ca. 90 % of all computed expected DVH-points and their standard deviations agree within 1 % volume with their empirical counterpart from sampling computations, for both fractionated and single fraction treatments. When computing α-DVHs, the assumption of a beta distribution achieved better agreement with empirical percentiles than the assumption of a normal distribution: While in both cases probabilities locally showed large deviations (up to ± 0.2), the respective α-DVHs for $\alpha = \{0.05, 0.5, 0.95\}$ only showed small deviations in respective volume (up to ± 5 % volume for a normal distribution, and up to 2 % for a beta distribution). A previously published model from literature, which was included for comparison, exhibited substantially larger deviations.

Conclusions: With APM we could derive a mathematically exact description of moments of probability distributions over DVH-points given a probability distribution over dose. The model generalizes previous attempts and performs well for both choices of probability distributions, i. e., normal or beta distributions, over DVH-points.

2

I. Introduction

Recent years have shown an increased interest in adequate, patient-specific uncertainty quantification and mitigation for radiotherapy treatment planning both academically and clinically (see, e. g., Refs. 1,2, and references therein). This development is, on the one hand, driven by emerging irradiation with particles and their characteristic sensitivity to uncertainties[3,4]. On the other hand, it is facilitated by fast-growing computational capabilities that enable the computation of multiple dose scenarios with acceptable overhead.

Error dose scenarios are either computed as (1) worst case estimates, i. e., extreme realizations of the input uncertainty model which are used for robust optimization (as performed and evaluated within Refs. 5,6,7,8,9,10,11), or (2) random samples from the probability distribution parameterizing the input uncertainty model[12,13,14,15,16,17] for stochastic approaches. An explicit derivation of probabilistic models remains the exception[18,19,20,21,22].

Consequently, the analysis of plan uncertainty is based on the derived worst-case dose distributions or "error bar"-distributions with their respective histograms[6,7,11,23], or statistical moments[13,18,21,22,24] as well as percentiles[14,15,17], according to the very optimization method used in the overall planning workflow. This use of empirical uncertainty estimates, however, exhibits limitations, in particular concerning statistical accuracy and the required recomputations during optimization due to the changing pencil-beam weights. Further, they conceal the inherent mathematical transformation from the input probability space (e. g. setup and range uncertainties) to the probability distribution over dose and the respective plan quality indicator (QI). This aggravates their use in retrospective analyses and puts restrictions on the choice of optimization method and objectives, because the sampling pipeline cannot be inverted and/or efficiently differentiated.

Approaches which rigorously propagate the uncertainty may overcome these limitations. Since they are able to provide analytical objectives, constraints, and derivatives, they can introduce new mathematical simplifications and improve computational efficiency by exploiting the analytical formulation for both uncertainty quantification and mitigation with probabilistic optimization approaches[18,22]. Because such approaches are not susceptible to statistical errors and the "curse of dimensionality" in scenario-based approaches, they deliver stable and reproducible results also for complex uncertainty models. However, derivation of such models is not trivial: Even if a model for dose probability is available, it still needs to be propagated to the derived plan indicators by hand. Other approaches overcome this step by re-sampling based on the derived dose uncertainty model[20,21].

For DVH-points, analytical computation of moments of the probability distribution, given a probability distribution over the dose, has been attempted before[25,26,27]. However, they[25,26] only provide a model for the expected value of DVH-points with an upper bound on the DVHs' standard deviation. Further, only simplified uncertainty models for the underlying dose distribution were assumed: while different shapes of the distributions were evaluated, correlations between voxels were not modeled, even though correlations having crucial impact on the higher moments of the depending probability distribution.

3

To derive a full model including correlations, this work will consequently not build on previous attempts, but provide a fresh start to a general methodology to compute the ν-th moments of the probability distribution over DVH-points. The goal is to derive a generally applicable model for DVH-probabilities allowing arbitrary assumptions on the probability distribution over the dose distribution.

To do so, first a closed-form description for the moments of the probability distribution of DVH-points is derived. Then, these moments parameterize a probability distribution over the respective DVH-point. To evaluate our approach, three patient cases are investigated with statistical reference computations (using a large number of random dose samples from the probability distribution over set-up and range errors in fractionated and non-fractionated treatments). Along the lines of our validation campaign, we illustrate the shortcomings of previous work and highlight where wide approximations regarding the correlation models (as exercised by Refs. 27) render meaningful quantification of DVH-probabilities impossible.

II.   Materials & Methods

II.A.   DVHs under uncertainty

II.A.1.   Nominal computation

DVHs are cumulative histograms over the spatial dose distribution in a volume of interest (VOI) v, here expressed as vector $\boldsymbol{d} \in \mathbb{R}_+^V$ with number of voxels V. Hence, for any given dose parameter $\hat{d}$, a DVH-point $\mathrm{DVH}(\hat{d}; \boldsymbol{d}) \in [0, 1]$ equals the fraction of the volume that receives *at least* dose $\hat{d}$. It can be expressed as averaged Heaviside steps

$$\mathrm{DVH}(\hat{d}; \boldsymbol{d}) = \frac{1}{V} \sum_{i \in v} \Theta(d_i - \hat{d}) \,, \tag{1}$$

meaning that only voxels i with $d_i \geq \hat{d}$ contribute to the sum which is normalized by the total voxel count V in v and thus yielding a fractional volume. Note that (without loss of generality) we assumed that all voxels have similar volume.

II.A.2.   Uncertainty analysis of DVHs

Uncertainty analysis of DVHs is mostly performed on an empirical basis through computation of error dose scenarios (among others Refs. 2,3,6,13,14,15,17,21,22,24,28,29). This enables the computation of a DVH for each dose scenario (which can be either a worst-case scenario or a random sample), from which then worst-case estimates, empirical statistical moments as well as quantiles of the probability distribution over DVH-points are derived.

For the purpose of this work, three forms of "statistical" DVHs will be of importance. First, uncertainty of a DVH can be evaluated through the statistical moments of each DVH-

4

point, for example the expected/mean DVH and its standard deviation, which can be empirically determined from n_s random dose samples $\boldsymbol{d}_s$ as

$$\overline{\mathrm{DVH}}(\hat{d}) = \frac{1}{n_s} \sum_{s=1}^{n_s} \mathrm{DVH}(\hat{d}; \boldsymbol{d}_s) \tag{2}$$

$$\sigma_{\mathrm{DVH}(\hat{d})} = \sqrt{\frac{1}{n_s - 1} \sum_{s=1}^{n_s} \left[\mathrm{DVH}(\hat{d}; \boldsymbol{d}_s) - \overline{\mathrm{DVH}}(\hat{d}) \right]^2}. \tag{3}$$

Secondly we discuss "α-DVHs", which can be compactly expressed as

$$\alpha\text{-DVH}(\hat{d}; \boldsymbol{d}, \alpha) = v_\alpha \Leftrightarrow P(\mathrm{DVH}(\hat{d}; \boldsymbol{d}) \leq v_\alpha) = \alpha \tag{4}$$

where v_α is the volume covered with a probability $P(\mathrm{DVH}(\hat{d}) \geq v_\alpha) = \alpha$. Thus α-DVHs can be used to give percentiles of the probability distribution of each DVH-point and, together with the corresponding $(1\text{-}\alpha)$-DVH, the respective confidence intervals. α-DVHs may be computed with the empirical marginal quantile functions for the respective DVH-points. Alternatively, α-DVHs are equal to iso-probability curves on the respective dose-volume coverage map (DVCM) (as proposed by Refs. 14). Such a DVCM assigns a probability of coverage of each possible volume fraction for any dose threshold $\hat{d}$ and can be defined as

$$\mathrm{DVCM}(\hat{d}, v; \boldsymbol{d}) = P(\mathrm{DVH}(\hat{d}; \boldsymbol{d}) \leq v) = F_{\mathrm{DVH}(\hat{d}, \boldsymbol{d})}(v), \tag{5}$$

where $F_{\mathrm{DVH}(\hat{d}, \boldsymbol{d})}(v)$ is the cumulative distribution function (CDF) of the probability distribution over the respective DVH-point. Equation (5) can then be directly inserted into Equation (4) such that the respective α-DVH is now the iso-curve at DVCM $= \alpha$.

Note that such α-DVHs or DVCMs do not yield confidences or probabilities for the *full* DVH but only over single DVH-points (i.e., they represent marginal quantiles and CDFs), and hence do not generally represent naturally occuring DVH-scenarios.

II.B.   Moments of the probability distribution over dose-volume histograms

II.B.1.   Analytical integration

If the probability distribution over the dose $\boldsymbol{d}$ has the multivariate CDF $F_{\boldsymbol{d}}$, the ν-th moment of the probability distribution of a transformation $I(\boldsymbol{d})$ can be computed via integration

$$\mathbb{E}\left[I(\boldsymbol{d})^\nu\right] = \int_{\mathbb{R}^V} I(\tilde{\boldsymbol{d}})^\nu \mathrm{d}F_{\boldsymbol{d}}(\tilde{\boldsymbol{d}}) \tag{6a}$$

$$= \int_{\mathbb{R}^V} I(\tilde{\boldsymbol{d}})^\nu f_{\boldsymbol{d}}(\tilde{\boldsymbol{d}}) \mathrm{d}\tilde{\boldsymbol{d}}. \tag{6b}$$

5

Moments of the probability distribution over a DVH may thus be explicitly calculated by solving Equation (6) for $I(\boldsymbol{d}) = \mathrm{DVH}(\hat{d}; \boldsymbol{d})$. For the first non-central moment, this yields

$$\mathbb{E}\left[\mathrm{DVH}(\hat{d}; \boldsymbol{d})\right] = \int_{\mathbb{R}^V} \frac{1}{V} \sum_{i \in v} \Theta(\tilde{d}_i - \hat{d}) f_{\boldsymbol{d}}(\tilde{\boldsymbol{d}}) \mathrm{d}\tilde{\boldsymbol{d}} \tag{7a}$$

$$= \frac{1}{V} \sum_i \int_{-\infty}^{\infty} \Theta(\tilde{d}_i - \hat{d}) f_{d_i}(\tilde{d}_i) \mathrm{d}\tilde{d}_i \tag{7b}$$

$$= \frac{1}{V} \sum_i \int_{\hat{d}}^{\infty} f_{d_i}(\tilde{d}_i) \mathrm{d}\tilde{d}_i \tag{7c}$$

$$= \frac{1}{V} \sum_i \left[1 - F_{d_i}(\hat{d})\right] . \tag{7d}$$

Similar steps lead to the mixed non-central moment $\mathbb{E}\left[\mathrm{DVH}(\hat{d}_p; \boldsymbol{d})\mathrm{DVH}(\hat{d}_q; \boldsymbol{d})\right]$:

$$\mathbb{E}\left[\mathrm{DVH}(\hat{d}_p; \boldsymbol{d})\mathrm{DVH}(\hat{d}_q; \boldsymbol{d})\right] = \int_{\mathbb{R}^V} \frac{1}{V^2} \sum_{il \in v} \Theta(\tilde{d}_i - \hat{d}_p)\Theta(\tilde{d}_l - \hat{d}_q) f_{\boldsymbol{d}}(\hat{d}) \mathrm{d}\tilde{\boldsymbol{d}} \tag{8a}$$

$$= \frac{1}{V^2} \sum_{il \in v} \int_{\mathbb{R}^2} \Theta(\tilde{d}_i - \hat{d}_p)\Theta(\tilde{d}_l - \hat{d}_q) f_{\boldsymbol{d}_{i;l}}(\tilde{\boldsymbol{d}}_{i;l}) \mathrm{d}\tilde{\boldsymbol{d}}_{i;l} \tag{8b}$$

$$= \frac{1}{V^2} \sum_{il \in v} \int_{\hat{d}_p}^{\infty} \int_{\hat{d}_q}^{\infty} f_{\boldsymbol{d}_{i;l}}(\tilde{\boldsymbol{d}}_{i;l}) \mathrm{d}\tilde{d}_l \mathrm{d}\tilde{d}_i . \tag{8c}$$

For the second non-central moment with $p = q$, i.e., $\mathbb{E}\left[\mathrm{DVH}(\hat{d}_p; \boldsymbol{d})^2\right]$, Equation (8c) can be expressed with the marginal bivariate cumulative distribution function $F_{\boldsymbol{d}_{i;l}}$ as

$$\mathbb{E}\left[\mathrm{DVH}(\hat{d}; \boldsymbol{d})^2\right] = \frac{1}{V^2} \sum_{il \in v} \left[1 - F_{\boldsymbol{d}_{i;l}}\left(\hat{d}\mathbf{1}_2\right)\right] , \tag{9}$$

where $\mathbf{1}_2 = (1, 1)^T$.

Together Equations (7), (8c) and (9) then give the (co)variance of DVH-points at $\hat{d}_p$ and $\hat{d}_q$ using $\mathrm{Cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$, i.e.,

$$\mathrm{Cov}\left[\mathrm{DVH}(\hat{d}_p; \boldsymbol{d}), \mathrm{DVH}(\hat{d}_q; \boldsymbol{d})\right]$$
$$= \mathbb{E}\left[\mathrm{DVH}(\hat{d}_p; \boldsymbol{d})\mathrm{DVH}(\hat{d}_q; \boldsymbol{d})\right] - \mathbb{E}\left[\mathrm{DVH}(\hat{d}_p; \boldsymbol{d})\right]\mathbb{E}\left[\mathrm{DVH}(\hat{d}_q; \boldsymbol{d})\right] \tag{10}$$

which, in case of the variance of a DVH-point at $\hat{d}$, consequently reduces to

$$\mathrm{Var}\left[\mathrm{DVH}(\hat{d}; \boldsymbol{d})\right] = \mathbb{E}\left[\mathrm{DVH}(\hat{d}; \boldsymbol{d})^2\right] - \mathbb{E}\left[\mathrm{DVH}(\hat{d}; \boldsymbol{d})\right]^2 . \tag{11}$$

6

Hence, for each point of a DVH for a VOI v, Equations (7) to (11) allow explicit computation of the expected value and variance of a DVH as well as the covariance between all DVH points, valid for any probability distribution over the dose $\boldsymbol{d}$ in v as long as its univariate and bivariate marginal CDF can be evaluated.

While not explicitly evaluated in this work, similar steps can be taken to compute higher moments to more accurately parameterize the underlying probability distribution. This requires an expansion of the respective power of the sums of Heaviside steps with the multinomial theorem and evaluation of multivariate probabilities of higher dimensionality. We provide such a generalization in Appendix C.

II.B.2. Summary of previous work

A result similar to Equation (7) was already documented in literature[25], where it was obtained by interpreting the computation of the DVH according to Equation (1) as a sum of Bernoulli experiments: Each voxel i falls into the current DVH bin at $\hat{d}$ with a probability $p_i = P(d_i > \hat{d}) = 1 - F_{d_i}(\hat{d})$, where F_{d_i} is the marginal CDF for d_i, and does not fall into the bin with a probability $1 - p_i$. With the linearity of the expectation value one then can directly derive Equation (7d).

The result proved to be applicable with different families of assumed probability distributions (i. e., Gaussian, triangular and rectangular/uniform)[25,26]. However, due to the simplified uncertainty model using constant relative standard deviation and no correlation between voxels, an exact computation of higher moments of a DVH-point's probability distribution was not attempted.

II.C.   Confidence bounds for DVH-points

II.C.1.   Parameterization of the DVH probability distribution

Since Equations (7) to (11) provide expected value and covariance of any DVH-point, one could possibly directly parameterize the probability distribution over the full DVH with a multivariate normal distribution. This parameterization is, however, unphysical; since DVH-points represent the fraction of a volume, their values are confined to the interval $[0, 1]$ in contrast to the infinite support of the multivariate normal distribution. Hence, the probability distribution of a DVH point might be more "physically" represented by a distribution supported only in the interval $[0, 1]$, such as a beta distribution $\mathcal{B}(a, b)$ with shape parameters a and b. The beta distribution is shortly characterized in Appendix A. However, lacking a generalized multivariate form (for recent approaches on constructing bivariate beta distributions see, e. g., Refs. 30,31), $\mathcal{B}(a, b)$ may only be used to parameterize the marginal distribution over a single DVH-point, not the full multivariate DVH.

Current approaches applying DVH confidences define these on a marginal by-point basis[14,15,17,29]. Hence, quantifying probabilities over marginal DVH-points is in line with

7

literature and enables comparability. Therefore, in this work, marginal probabilities will be evaluated, based on the (unphysical) parameterization with normal distributions as well as the more physical approach using a beta distribution whose shape parameters a and b are obtained from the respective DVH-point's expectation value and variance with Equations (23) and (24).

Using either a normal or beta distribution, one can directly compute DVCMs according to Equation (5) or an α-DVH using

$$
\begin{aligned}
\alpha\text{-DVH}(\hat{d}; \boldsymbol{d}, \alpha) &= F^{-1}_{\text{DVH}(\hat{d};\boldsymbol{d})}(\alpha) \\
&= \begin{cases} \mathbb{E}\left[\text{DVH}(\hat{d}; \boldsymbol{d})\right] + \sqrt{2\,\text{Var}\left[\text{DVH}(\hat{d}; \boldsymbol{d})\right]}\,\text{erf}^{-1}(2\alpha - 1) & |\text{ normal} \\ I^{-1}_{\alpha}(a, b) & |\text{ beta} \end{cases},
\end{aligned} \tag{12}
$$

where erf^{-1} denotes the inverse error function and I^{-1}_{α} represents the inverse of the regularized incomplete beta-function.

II.C.2.   Summary of previous work

A subsequent work[27] (to Refs. 25,26) already formulated an analytical calculation of α-DVHs as defined in Equation (4).

In summary, their model is based on the definition of a "binary random variable"[27]

$$
T_i^{\alpha,\hat{d}} = \begin{cases} 1 & P(d_i \geq \hat{d}) > 1 - \alpha \\ 0 & P(d_i \geq \hat{d}) \leq 1 - \alpha \end{cases} \tag{13}
$$

interpreted as "the volume receiving a dose greater than $[\hat{d}]$ with a probability greater than $1 - \alpha$"[27]. This interpretation then leads them to define $T_i^{\alpha,\hat{d}} = \Theta\left(1 - \alpha - F_{d_i}(\hat{d})\right)$ which, in analogy to Equation (4), translates to

$$
\alpha\text{-DVH}^{\text{HC}}(\hat{d}; \boldsymbol{d}, \alpha) = \frac{1}{V}\sum_{i \in v} T_i^{\alpha,\hat{d}} = \frac{1}{V}\sum_{i \in v} \Theta\left(1 - \alpha - F_{d_i}(\hat{d})\right). \tag{14}
$$

Equation (14) substantially differs from our result in Equation (12). This may be attributed to disregarding correlation across voxels in their derivations or inconsistencies in the definition of $T_i^{\alpha,\hat{d}}$, which apparently does not describe a random variable per se, but rather includes the evaluation of a probability—in this case the probability $P(d_i - \hat{d})$ of uncertain dose d_i exceeding $\hat{d}$—which is not a random but a fixed value obtained from the CDF over d_i. Since Equation (14) has only been tested with an assumed uncertainty model and not been benchmarked against sample statistics[27], we incorporate Equation (14) in our evaluations and discuss the Equation further in Section IV..

8

II.D. Dose Uncertainty Model

Evaluation of Equations (7), (10) and (11) or, in general, Equation (25) requires a model for the probability distribution over dose $\boldsymbol{d}$, such that its CDF can be evaluated. Note that empirical CDFs as well as analytical/parameterized CDFs can be used.

II.D.1. Gaussian model for the probability distribution over dose

For a first evaluation and validation of the new probabilistic computations, we assume

$$\boldsymbol{d} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \tag{15}$$

i.e., the dose follows a multivariate normal distribution with mean dose $\boldsymbol{\mu}$ and covariance Σ. This choice of probability distribution cannot represent the *true* underlying probability distribution: First, the multivariate normal is supported on the full multidimensional real space, while physical dose is bound to the positive orthant. And second, empirical evidence (e.g. Refs. 6,24) as well as heuristic considerations show that the respective distribution exhibits considerable skewness and is consequently not part of the symmetric Gaussian family. Alternatives to assumption (15) will be discussed in Section IV.. As a first order approximation, however, Equation (15) is well suited to study the probabilistic DVH-model, because on the one hand, its univariate and bivariate probabilities can be calculated[32], which is sufficient to compute an expected DVH and its (co)variance. On the other hand, evaluation on patient cases with assumption (15) implicitly studies the impact of the inaccurate Gaussian dose model on the evaluation of VOI-based dose statistics like DVHs under uncertainty. Further, while for a single fraction treatment the non-Gaussian shape of the probability distribution over dose is to be expected, under multiple fractions a more Gaussian-like shape may form (for examples of voxel dose probability distributions see Refs. 6,24).

II.D.2. Computation of dose uncertainty

The Gaussian dose model from Section II.D.1. requires mean $\boldsymbol{\mu}$ and covariance Σ for evaluation. E.g., these could be empirically estimated with sample mean and covariance of a set of discrete error scenarios.

Since the motivation of this work is to build a fully analytical model (which also facilitates future use in optimization), we rely on computation of $\boldsymbol{\mu}$ and Σ through APM (as introduced by Refs. 18). In previous works we could already show that APM accurately models $\boldsymbol{\mu}$ and $\boldsymbol{\sigma} = \sqrt{\text{diag}(\Sigma)}$[33], and that efficient application to patient data—especially in the context of fractionation[22]—is possible. Extension to biological optimization demonstrated applicability of APMs to intensity-modulated carbon ion therapy planning.[34]

APM acts as a probabilistic pencil-beam dose calculation algorithm inherently enabling computation of moments of the probability distribution over the resulting dose. More exact, APM represents the constituents of a pencil-beam algorithm as superpositions of Gaussian functions (including the integrated depth dose, i.e., the Bragg peak), enabling propagation

of uncertainties through the dose calculation in closed-form via analytical integration (for a detailed explanation see Refs. 18).

The accuracy of this approximation of the pencil-beam algorithm can, in principle, be arbitrarily chosen by varying the number of Gaussian components, thus providing also a nominal dose calculation algorithm for inverse treatment planning that is similar in quality to common pencil-beam algorithms. As such, it is able to provide a dose influence matrix $D \in \mathbb{R}_+^{V \times B}$ with number of voxels V and number of pencil-beams B generating the dose $\boldsymbol{d} \in \mathbb{R}_+^V$ from the fluence vector $\boldsymbol{w} \in \mathbb{R}_+^B$ via the linear transformation

$$d_i = \sum_j D_{ij} w_j \,. \tag{16}$$

In Equation (16), i indexes voxels in the patient while j indexes pencil-beams.

Now, in addition to Equation (16), APM provides probabilistic analogs to Equation (16) for moments of the probability distribution over $\boldsymbol{d}$ by enabling element-wise computation of expectation value and (co)variance of elements of the dose influence matrix D. This allows to represent the expected value of dose $\mathbb{E}\left[\boldsymbol{d}\right]$ as a linear transformation

$$\mathbb{E}\left[d_i\right] = \sum_j \underbrace{\mathbb{E}\left[D_{ij}\right]}_{\mathcal{D}_{ij}} w_j = \sum_j \mathcal{D}_{ij} w_j \tag{17}$$

and the covariance in dose with a quadratic form

$$\mathrm{Cov}\left[d_i, d_l\right] = \sum_{jm} \underbrace{\mathrm{Cov}\left[D_{ij}, D_{lm}\right]}_{\mathcal{V}_{ijlm}} w_j w_m = \sum_{jm} \mathcal{V}_{ijlm} w_j w_m \,. \tag{18}$$

Hence, one can denote $\mathcal{D} \in \mathbb{R}_+^{V \times B}$ and $\mathcal{V} \in \mathbb{R}_+^{V \times B \times V \times B}$ as *expected dose influence matrix* and *covariance influence tensor*, respectively. While $\mathcal{V}$ is, in general, too large to be stored in memory, the element-wise computation with APM allows on-the-fly evaluation of dependent quantities, e. g. the variance or covariance of dose.

II.E.   Validation and application of the model

The analytical probabilistic DVH-model will be evaluated on three patient cases, i. e., an intracranial, a paraspinal, and a prostate case. Parameters used for planning and uncertainty computations are laid out in Table 1. These cases were already evaluated in our previous works [22,33,34]. Furthermore, a detailed comparison of α-DVHs and DVCMs is exercised to validate empirical percentiles against results from the respective quantile functions from Equation (4) and the previous works [27].

II.E.1.   Application on all cases using the fractionated treatment samples

For all three patient cases, empirical estimates $\overline{\mathrm{DVH}}$, $\bar{\sigma}_{\mathrm{DVH}}^2$ are obtained by computing 100×30 dose scenarios, i. e., 100 sampled treatment scenarios with 30 fraction each, based

10

on a Gaussian uncertainty model using the assumed setup and range errors from Table 1 as respective standard deviation with zero mean (compare to the description of the APM uncertainty model in Refs. 18). Additionally, expected dose $\boldsymbol{\mu}$ and covariance Σ within the respective VOIs was also computed with APM (see Section II.D.2.) for the same input uncertainty model, and then fed into the herein presented DVH-models (by assuming a multivariate normal distribution as described in Section II.D.). This enabled comparison of the sample statistics to a fully analytical method and serves as proof-of-concept of the derived model.

II.E.2.   Full validation on the intracranial case with 5000 samples

As the intracranial case is the smallest one with lowest computational overhead, we further computed 5000 realizations of single-fraction treatments. These will be used to nearly eliminate statistical inaccuracy for benchmarking APM.

To further validate the analytical computation itself excluding inherent mismatch of modeled and real probability distribution over dose, the analytically computed $\boldsymbol{\mu}$ and Σ are additionally used to create 5000 new dose samples *under the assumption that dose* actually *follows a multivariate normal distribution*, i. e., their samples are drawn from the distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. From these samples, a second statistical estimate of the DVH is obtained. This can be used to validate if the analytical computation is actually correct under the multivariate normal assumption from Equation (15). Further, DVCMs and α-DVHs are computed under the assumption of marginally normally distributed DVH-points and marginally beta distributed DVH-points. These are compared to the respective statistical estimates from the 5000 scenario samples.

III.   Results

We evaluated the described methodology on three patient cases – an intracranial, paraspinal, and prostate patient. Information about the datasets, treatment plans and the assumed input uncertainty model can be found in Table 1.

III.A.   Proof of work – computation on patient data

Figure 1 compares sample mean and standard deviation of DVHs to the respective analytical computations with Equations (7) to (11) for treatments with 30 fractions.

For the prostate case, the sampled and analytically computed mean DVH and its standard deviation yield good agreement for both target and OAR. For the intracranial case, especially the curves illustrating standard deviation seem to exhibit larger differences. However, a closer look reveals that the differences originate from both discrepancies of the mean and the standard deviation estimates.
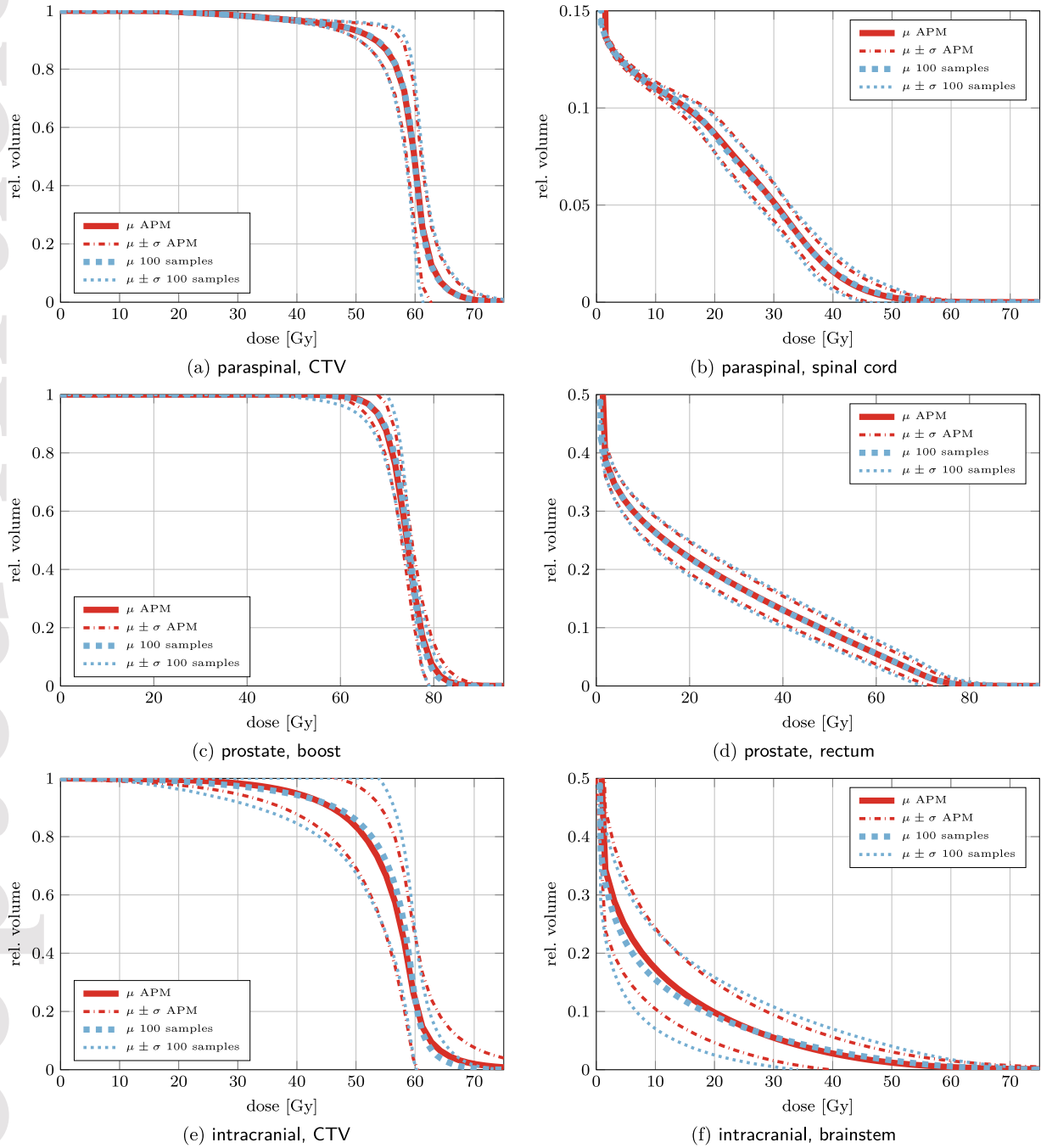
11

Figure 1: Analytically computed expectation value and standard deviation of DVHs of a target volume and organ at risk (OAR) for each of the three patient cases, compared to the respective sample mean and standard deviation. For the sampling benchmark, 100 treatments were simulated by multivariate normal sampling using the systematic errors from Table 1 as standard deviation, while for each treatment taking 30 fraction samples based on the random component.

To better quantify the differences between the analytical computation and the sample reference, Figure 2 summarizes the absolute difference in relative volume for all patients grouped by (1) mean and standard deviation, (2) targets and OARs, and (3) 1 and 30 fraction treatments. Differences between analytical and sample computations are, in general, larger for targets than for OARs. For the OARs, the evaluation for multiple fractions shows an increase in accuracy. This does not seem to transfer to targets, where, although the number of points with minimal difference (< 0.001) also increases, a stronger tail to higher differences is present. In general, more outliers, i.e., single DVH-points with large difference, can be observed when performing the calculations for a treatment in 30 fractions.
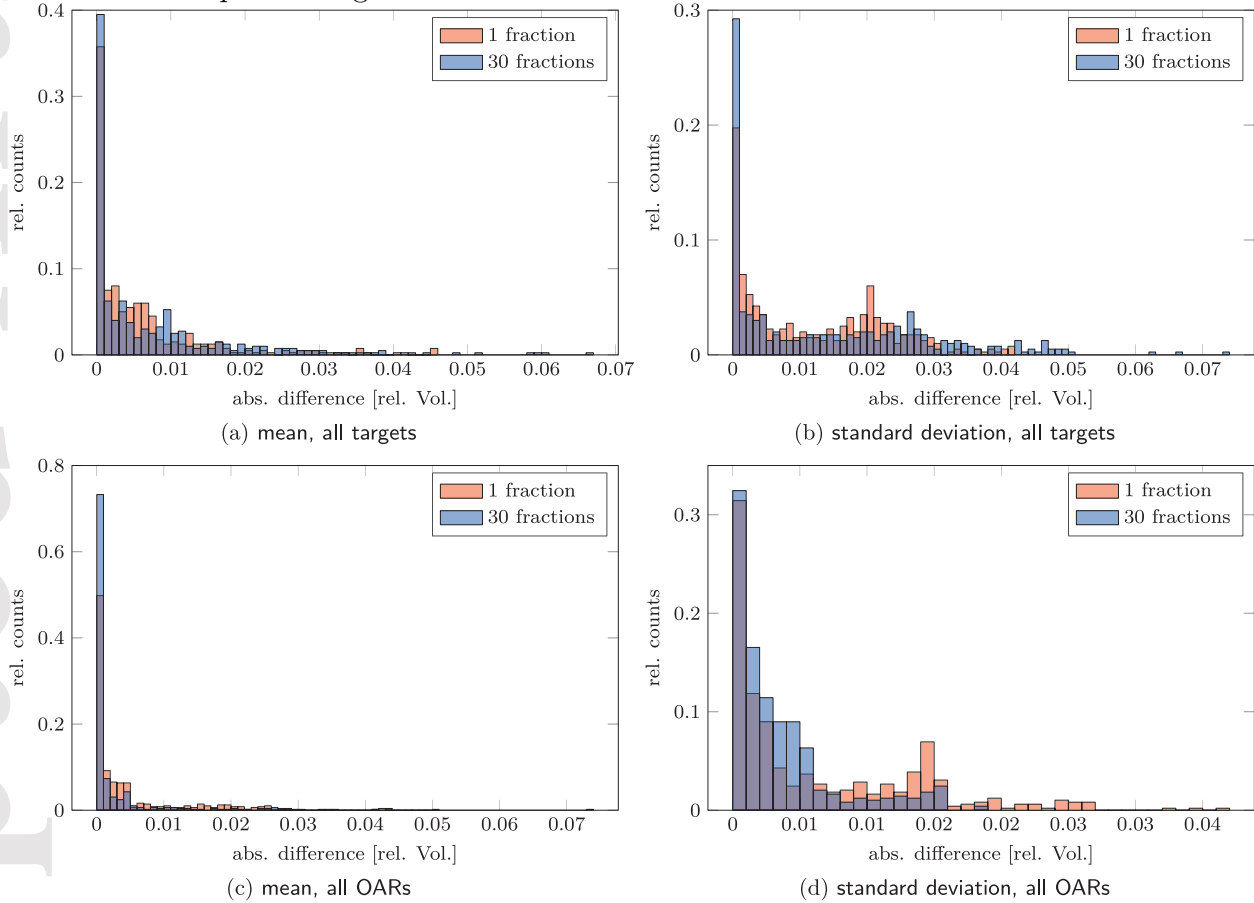

Figure 2: Histograms (bin width: 0.001) of the absolute differences observed between the mean and standard deviation of all DVH-points computed analytically and via random sampling for all patients. (a,b) show the analysis for all evaluated targets in all patients, while (c,d) display it for all evaluated OARs.

III.B. Validation of model and analytical computations

III.B.1. Distribution of single DVH-points

Figure 3 shows normalized histograms of two DVH-points (one evaluated at 57 Gy for the target and one at 30 Gy for the brainstem of the intracranial case), comparing the samples from the dose scenarios and from the multivariate normal approximation. Their respective

13

approximations with a normal distribution visualize differences between the respective moments of the DVH-point's probability distribution: in the clinical target volume (CTV), the re-sampled mean underestimates the DVH at $57\,\mathrm{Gy}$ by a volume fraction of $4\,\%$, whereas in the brainstem difference between the mean / expected values is negligible. The opposite holds true for the computed standard deviation.
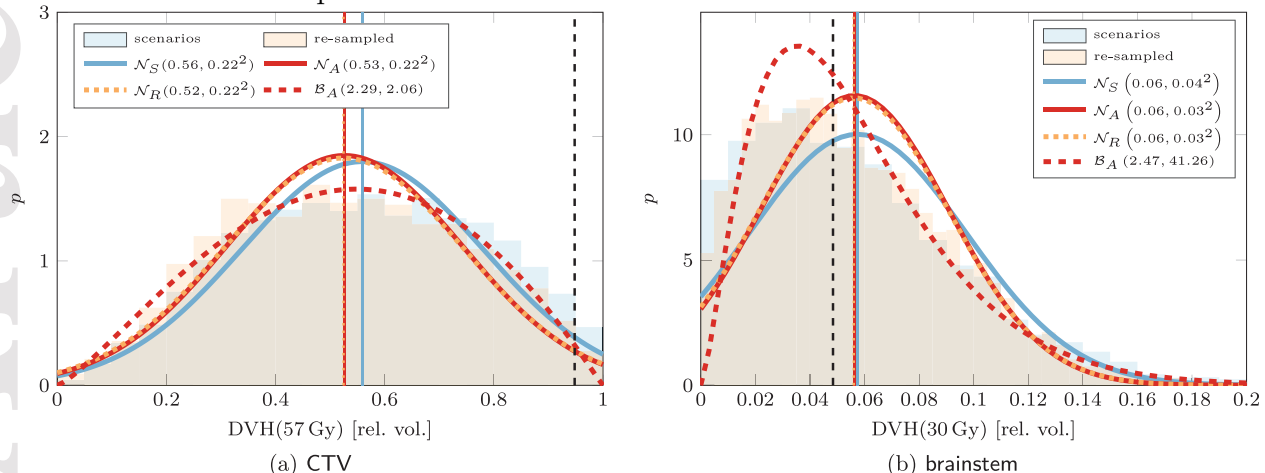

(a) CTV

(b) brainstem

Figure 3: Probability distribution over DVH-points evaluated at $\hat{d} = 57\,\mathrm{Gy}$ in the CTV (a) and at $30\,\mathrm{Gy}$ in the brainstem (b) of the intracranial case. The histograms show the distribution from the 5000 dose scenarios and the 5000 dose distributions re-sampled under assumption (15). $\mathcal{N}_{\mathcal{S}}$ represents a normal distribution parameterized from sample mean and variance from the 5000 DVHs obtained from the scenario samples, $\mathcal{N}_{\mathcal{R}}$ has ben similarly computed from the re-sampled scenarios. $\mathcal{N}_{\mathcal{A}}$ parameterizes a normal distribution based on the analytical (APM) computation of DVH-point expectation and variance, and $\mathcal{B}_{\mathcal{A}}$ uses the same values to parameterize a beta distribution. The vertical lines indicate the respective expected/mean values, with the dashed black line giving the nominal value.

The analytically computed expectation value and variance of the respective DVH-points shows no significant difference to the statistical moments obtained from the re-sampled data. This is expected, since the analytical computations are mathematically exact and only negligible numerical inaccuracy is introduced when evaluating the univariate and bivariate normal CDF.

The Gaussian approximation is not bound to the volumetric interval $[0, 1]$, and thus would assign non-zero probability to non-existing, e.g., negative, volume fractions. Figure 3 thus shows corresponding approximations with beta distributions, whose shape parameters are obtained from Equations (23) and (24) using analytically computed expectation and variance of the DVH-point. This leads to a physically more reasonable distribution which is further backed by the Q-Q plots comparing the quantiles of the normal and beta approximation to the empirical quantiles in Figure 4.

Figures 4a and 4b underline the problem of the infinite support of the normal distribution, i.e., unphysical quantiles exist in the theoretical normal model. This goes hand in hand with an especially pronounced disagreement between theoretical and empirical quantiles approaching the boundaries of the interval $[0, 1]$, but also concerns possible skewness (compare Figures 3b and 4b) or excess kurtosis of the distribution (Figures 3a and 4a). These disagreements are reduced using a beta distribution. Especially the evaluated DVH-point
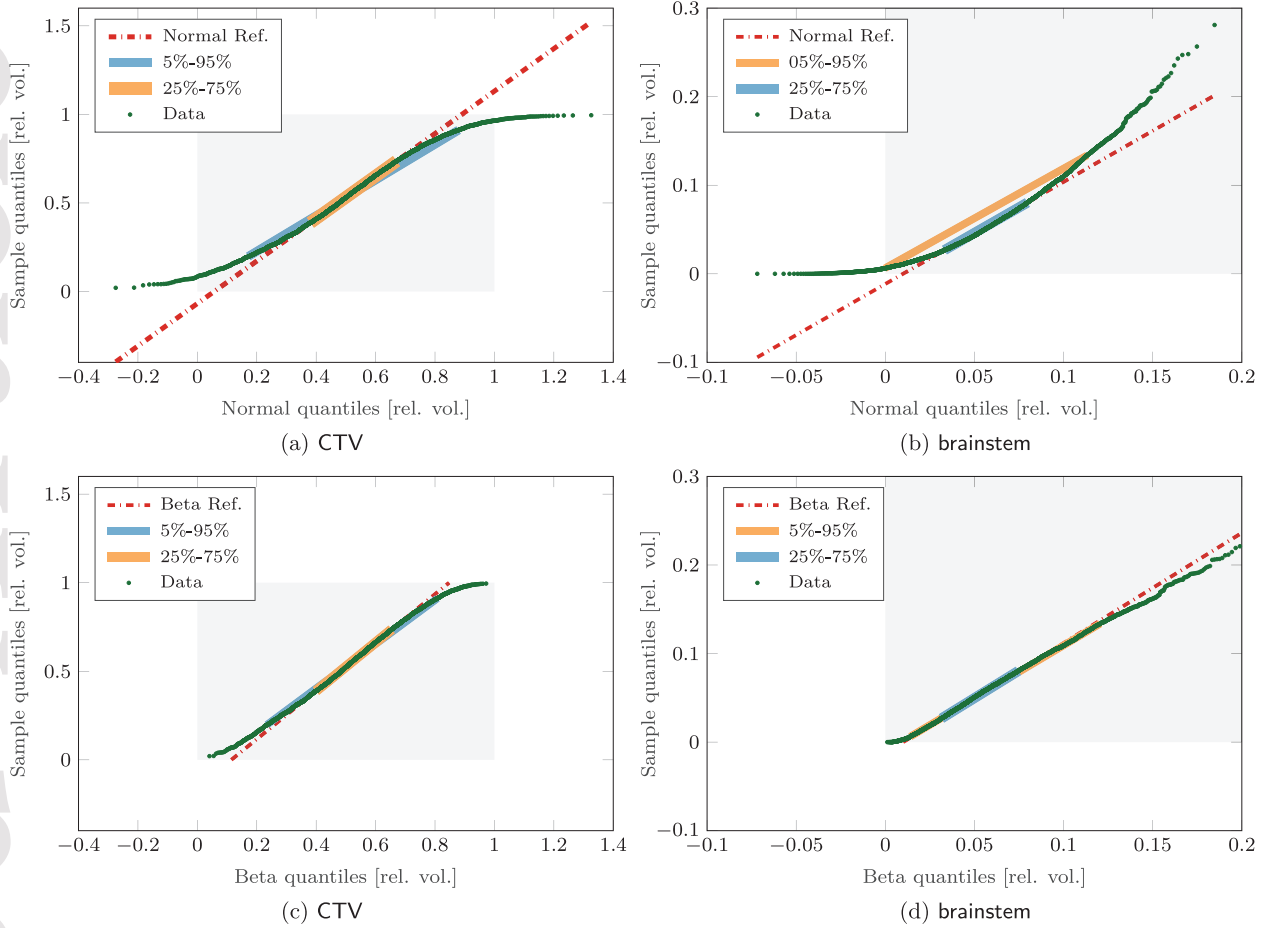
Figure 4: Quantile-quantile plots comparing empirical quantiles (y-axis) for the data (green) to quantiles from the hypothesized normal (a,b) and beta (c,d) distributions (x-axis) obtained from analytical moment computations (red). The thicker green and yellow lines span the $5\,\%$–$95\,\%$ and $25\,\%$–$75\,\%$ quantiles, respectively. In each plot the gray area additionally enclose "physically feasible" volumes in $[0, 1]$. CTV and brainstem of the intracranial case are shown for the same DVH-points as in Figure 3. The data is based on the 5000 dose scenario samples for a single fraction.

in the brainstem in Figure 4d shows near perfect agreement with the empirical quantiles. For all four evaluations in Figure 4, however, near perfect agreement is achieved for "inner" quantiles, i.e., the first and third quartile.

III.B.2. Evaluation of cumulative probabilities

Next, we assess the accuracy of complete α-DVHs by comparing α-DVHs computed from the quantile functions of the respective probability distribution parameterized from the analytical computations with empirical quantiles over the full DVH. Furthermore we compare to the previous attempt of analytical computation of α-DVHs[27] as laid out in Section II.C.2..

Figure 5 shows the respective comparisons of analytically computed DVCMs to DVCMs obtained from sample statistics (compare Equation (5)). The α-DVHs in Figures 5a and 5b computed with Equation (14) (i.e., the method from Refs. 27), show significant differences

15

to the corresponding reference α-DVHs from sampling.

In Figures 5c and 5d, the probabilities within the empirical DVCMs show large differences when compared to DVCMs computed with the CDF from the Gaussian approximation, especially near full volume coverage and near zero volume coverage, where the approximated CDF exhibits differences of up to ± 0.2. This is to be expected due to the infinite support of the normal distribution, and therefore the agreement is much better with the beta approximation in Figures 5e and 5f (overall the deviation is more than halved compared to the normal approximation). This transfers to the computation of α-DVHs based on the quantile function of the beta distribution, similarly showing better agreement than with the assumption of normally distributed DVH-points. Overall, our explicit parametrization and evaluation of quantile functions of either normal or especially beta distributions outperforms the previously published method[27] (compare Section II.C.2.).
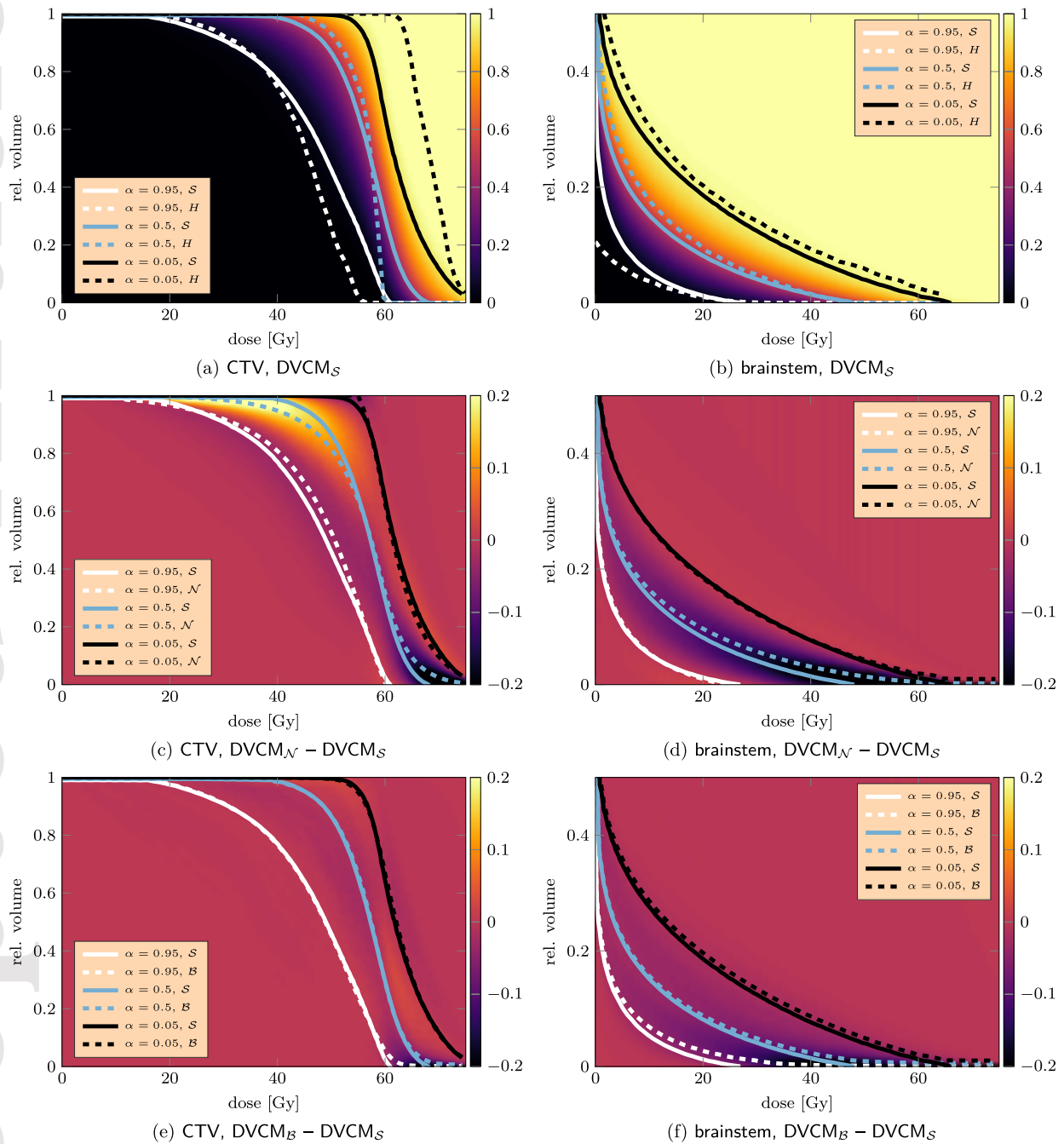
16

Figure 5: Evaluation of probabilities and α-DVHs obtained through scenario sampling and analytical computations. (a) and (b) show empirical DVCMs (indexed with $\mathcal{S}$) obtained from the 5000 dose scenarios, where the color indicates the local value of the CDF of the respective DVH-point. Corresponding α-DVHs were derived, based on sampling ($\mathcal{S}$) and on Ref. 27 ($\mathcal{H}$). For (c) and (d) the difference of DVCM$_\mathcal{N}$, a DVCM constructed from the normal parameterization, to DVCM$_\mathcal{S}$ has been evaluated with corresponding α-DVHs obtained from the normal quantile function. (e,f) provide a similar analysis using DVCM$_\mathcal{B}$ and α-DVHs ($\mathcal{B}$) with the beta parameterization.

17

IV. Discussion

The very core of this work is the description of an analytical model that can compute statistical moments of DVH-points for arbitrary probability distributions over the dose distribution. The only requirement is that the dose distribution follows/obeys a probability distribution function where the marginal CDFs can be evaluated. Hence, we provide a mathematically exact formulation of the moments which eliminates statistical uncertainty (in particular in combination with APM-based uncertainty quantification[18,22]); if the probability distribution over dose is known, the respective moments can be exactly computed.

In our initial analysis, a multivariate normal distribution was assumed based on expected dose and its covariance obtained from a probabilistic dose calculation employing APM, which we compared to a sampling benchmark. Most of the analytically computed expected and standard deviations of the investigated DVHs lied within $\pm 1.5\%$ difference in volume from the sampling benchmark. Evaluating the same model for a larger number of fractions showed that especially for the OARs, more values cluster around the small differences.

Those differences between analytical model and sampling benchmark mainly stem from the assumed probability distribution over the dose distribution. It is clear that the multivariate normal assumption for dose uncertainty used throughout this work does not reflect every detail of reality perfectly. This is directly obvious because dose cannot be negative while the normal distribution has infinite support. But also deformations and anatomical variations, the non-linearity of dose deposition, and the so far not captured complexity of input uncertainty suggest that "real" dose uncertainty will most likely manifest as non-Gaussian probability distributions differing across heavily interdependent voxels. For parametrizing such models, an appropriate family of probability distributions that reasonably models variation in voxel doses would have to be selected for which correlations could then, for example, be modeled with copulas. While finding a suitable probabilistic model accurately describing dose distributions under uncertainty is beyond the scope of this work, our method would also be compatible with such advanced models, since it only requires that some form of multivariate CDF can be computed. This includes non-Gaussian analytical as well as empirical CDFs. In doing so our model has the advantage to preserve the mathematical connection between dose DVH-variation which would not be accessible when sampling DVH-scenarios themselves.

Despite the inaccuracy inherent to our choice of a multivariate normal distribution, it served as good initial proof of concept for the presented method. The multivariate normal parametrization allowed straightforward resampling of dose scenarios to demonstrate the exactness of the closed-form solution. The validation with results from sampled dose cubes and DVHs on the three patient cases showed that despite using this physically flawed model, we still obtain reasonable results, especially for the prostate and paraspinal case. The larger deviations in the intracranial case may be attributed to the smaller VOIs compared to the other two cases. The above discussed observation that fractionation seems to decrease the difference in most voxels might be attributed to a more Gaussian shaped probability distribution of dose uncertainty in the context of fractionation (as also indicated within

18

Refs. [7]).

Our model describes moments of the DVH-points' probability distributions and not the probability distribution over the full DVHs. Therefore it is not directly possible to exactly compute confidences on actual realizations of the full DVH. It is, however, possible to use the computed moments to parameterize distributions over DVH points. In our evaluations based on expected DVH-points and their standard deviation, usage of normal distributions was a first obvious choice. To some extent it may be surprising that this yielded acceptable results within few volume percent difference to the sampling benchmark: The choice of a normal distribution (with infinite support) is in this case of a volume fraction, which can take only values in the interval $[0, 1]$, similarly physically unreasonable as in the case of the dose distribution. A more plausible (concerning the support interval) parameterization was found using a beta distribution. However, both distributions do not represent a mathematically exact model. Interpreting the calculation of a DVH-point as a series of Bernoulli trials (similar to Refs. [25], see Section II.B.2.), suggests parameterization with a Binomial distribution in the case of independently and identically distributed individual voxel doses. Still, assuming independence of voxels would not be realistic. More suitable correlated binomial models[35] make specific assumptions and are computationally demanding, rendering them not applicable for our purpose. Nevertheless, assuming a beta distribution (or even a normal distribution), which is parameterized by mean and standard deviation for the DVH-point, could facilitate uncertainty propagation through models that build on DVHs itself, e.g. in deriving biologically effective dose[36], or refining the statistical models for optimization purposes[19].

In comparison to the previous works[25,26,27], our methodology directly reproduced their model for the expectation value of DVH-points[25]. Our approach using integration directly generalizes to higher moments like covariance, while their model only yielded upper bounds on the variance because correlations between voxels were not explicitly included. Yet, despite the lack of an uncertainty model considering covariance in dose, they formulated confidence bounds, i.e., α-DVHs[27]. Those were, however, in stark disagreement with the sampling benchmark (and thence our model). This may mainly be attributed to neglecting the contribution of (in)dependencies between voxels in the derivations. The difference of Equation (14) to a "true" confidence of dose over coverage of a volume fraction may be shown by the following gedankenexperiment: Let us assume all V voxels in a VOI v are independently normally distributed with a mean value of $\hat{d}$ and identical variances, and therefore exhibiting $F_{d_i}(\hat{d}) = 0.5$ in all voxels $i \in v$. In this setup, one can see that the median DVH-point at $\hat{d}$ takes the value 0.5. Yet, plugging this into Equation (14) with again $\alpha = 0.5$ yields $\alpha\text{-DVH}^{\text{HC}}(\hat{d}) = \frac{1}{V} \cdot V \cdot \Theta(1 - 0.5 - 0.5) = 1 \neq 0.5$ (depending on the definition of the Heaviside-step). Furthermore, in this case Equation (14) exhibits a sharp "step" around $\hat{d}$: For smaller doses $\hat{d} - \epsilon$, the argument of the step function in Equation (14) becomes negative and therefore resulting in $\alpha\text{-DVH}(\hat{d} - \epsilon) = 1$, while for larger doses it becomes $\alpha\text{-DVH}(\hat{d} - \epsilon) = 0$. For the assumed independently distributed model, however, a smooth decrease of the median DVH around $\hat{d}$ is expected, whereas the step would be correctly reproduced when assuming perfectly correlated voxels. Instead we suggest to interpret their result as the "fraction of voxels whose probability of exceeding $\hat{d}$ *independently from each*

19

other is larger than $1 - \alpha$". Such a quantity, however, does not have a palpable clinical interpretation.

The applicability of a method that propagates uncertainty from dose to DVHs in terms of treatment planning might not be directly obvious. As discussed in Section I., uncertainty quantification usually relies on sampled (stochastic approach) or selected (worst-case approach) dose scenarios, which can directly be used to obtain similar uncertainty information about the DVH. However, the analytical probabilistic method presented here provides a closed-form, continuous relationship between dose uncertainty and DVH-uncertainty, which can be useful in treatment planning.

For optimization with probabilistic constraints, the method facilitates the use of continuous differentiable functions, which "fill the gap" between empirical samples, and could possibly enable exact definition of the "allowed" probability that certain clinical constraints are failing. There, using the parameterization of the probability distribution over a DVH-point, we now have functional access to the probability that a specific dose level is covered with a certain probability via the α-DVH and its analytical dependence on mean and covariance of the dose (instead of nominal dose as for conventional non-probabilisitc objectives).[37,38] Using this information for planning would extend widely used margin principles[39,40], which are themselves defined based on demanding tumor coverage with a desired probability for a given dose level. Compared to common robust approaches where the worst case is dynamically found from a discrete pre-defined uncertainty subset in the input variables, an analytical confidence constraint based on α-DVHs captures the full uncertainty space and allows for definition of the desired robustness on the relevant endpoint. For example, 90 % probability to cover a tumor with the 95 % of the prescribed dose could reproduce common margin assumptions also for anatomically complex cases where margins break down, while at the same time other probabilistic or nominal endpoints could be optimized. Since these objectives (or also constraint functions) continue to be functions of mean and covariance of dose, optimization would remain possible using conventional (quasi-)Newton optimizers. This, however, comes at the cost of deriving the arguably complex gradients (w.r.t. dose mean & covariance), which could be mitigated by using automatic differentiation[41], accompanied by a computationally challenging, yet "embarrassingly parallel" (co)variance computation.[37,38] Further, the role of model inaccuracy in such a probabilistic optimization approach would need further investigation—previous investigations, however, suggest that the assumption of normal distributions in probabilistic optimization still enables mitigation of dose uncertainties even if the "real" distribution differs substantially[42]. A full analysis of computational performance and resulting dose distributions of such an optimization approach would exceed this manuscript and may be part of a future study.

Further, the probabilistic model only requires the dose's probability distribution and is independent of the method used to obtain it. This allows its general application, e. g., in (retrospective) plan analyses or in combination with sampling-based stochastic optimization approaches. For plan analysis and comparison, the closed-form connection of dose and DVH-uncertainty could be used to observe how higher dose uncertainty (simulated by scaling covariance in desired voxels, for example) would impact the DVH and therefore target coverage. When relying on sample statistics for the dose instead of a probabilstic dose calcu-

20

lation, the parameterizations can be useful to extrapolate to areas not covered by samples. Additionally, the method could also be used in scenarios where computation of uncertainties statistically is no longer possible (e. g. when analysing an older cohort) based on assumptions over the dose probability distributions while still retaining the mathematical link between dose and DVH probabilities, which may be useful for machine learning applications investigating plan uncertainties.

And last but not least, the general concept does not only apply to the computation of DVH points. As already indicated within Ref. 37, the concept may—given mathematical efforts—be extended to other plan quality metrics as mean dose or equivalent uniform dose (EUD), or distinct treatment planning objective and constraint functions.

V. Conclusion

We presented a method to calculate moments of the probability distribution over DVH-points given a known probability distribution over the dose distribution. The resulting analytical model corrects and generalizes previous attempts and can be readily combined with every method able to quantify a probability distribution over dose of either empirical or probabilistic nature.

We successfully benchmarked the model against excessive sampling (with and without fractionation), proving mathematical correctness and good agreement even with unrealistic but common assumptions for probability distributions within the computational pipeline. The methodology can serve as blueprint for future models on other QIs and can provide a generalizable framework for confidence-constrained probabilistic treatment plan optimization.

Acknowledgments

References

[1]   J. Unkelbach, M. Alber, M. Bangert, R. Bokrantz, T. C. Y. Chan, J. O. Deasy, A. Fredriksson, B. L. Gorissen, M. van Herk, W. Liu, H. Mahmoudzadeh, O. Nohadani, J. V. Siebers, M. Witte, and H. Xu, Robust radiotherapy planning, Physics in Medicine and Biology **63**, 22TR02 (2018).

21

2 A. Fredriksson, A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization, Medical Physics **39**, 5169–5181 (2012).

3 A. J. Lomax, Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties., Physics in Medicine and Biology **53**, 1027–1042 (2008).

4 A. J. Lomax, Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions, Physics in Medicine and Biology **53**, 1043–1056 (2008).

5 D. Pflugfelder, J. J. Wilkens, and U. Oelfke, Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy., Physics in Medicine and Biology **53**, 1689–700 (2008).

6 M. Lowe, F. Albertini, A. Aitkenhead, A. J. Lomax, and R. I. MacKay, Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors, Physics in Medicine and Biology **61**, 413–429 (2016).

7 M. Lowe, A. Aitkenhead, F. Albertini, A. J. Lomax, and R. I. MacKay, A robust optimisation approach accounting for the effect of fractionation on setup uncertainties, Physics in Medicine and Biology **62**, 8178–8196 (2017).

8 A. Fredriksson, A. Forsgren, and B. Hårdemark, Minimax optimization for handling range and setup uncertainties in proton therapy, Medical Physics **38**, 1672–1684 (2011).

9 A. Fredriksson and R. Bokrantz, The scenario-based generalization of radiation therapy margins, Physics in Medicine and Biology **61**, 2067–2082 (2016).

10 M. Casiraghi, F. Albertini, and A. J. Lomax, Advantages and limitations of the 'worst case scenario' approach in IMPT treatment planning., Physics in Medicine and Biology **58**, 1323–1339 (2013).

11 W. Liu, X. Zhang, Y. Li, and R. Mohan, Robust optimization of intensity modulated proton therapy, Medical Physics **39**, 1079–1091 (2012).

12 J. Unkelbach, T. C. Y. Chan, and T. Bortfeld, Accounting for range uncertainties in the optimization of intensity modulated proton therapy., Physics in Medicine and Biology **52**, 2755–2773 (2007).

13 J. Unkelbach, T. Bortfeld, B. C. Martin, and M. Soukup, Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning., Medical Physics **36**, 149–163 (2009).

14 J. J. Gordon and J. V. Siebers, Coverage-based treatment planning: optimizing the IMRT PTV to meet a CTV coverage criterion., Medical Physics **36**, 961–973 (2009).

22

[15] J. J. Gordon, N. Sayah, E. Weiss, and J. V. Siebers, Coverage optimized planning: probabilistic treatment planning based on dose coverage histogram criteria., Medical Physics **37**, 550–563 (2010).

[16] R. Bohoslavsky, M. G. Witte, T. M. Janssen, and M. van Herk, Probabilistic objective functions for margin-less IMRT planning, Physics in Medicine and Biology **58**, 3563–3580 (2013).

[17] H. Mescher, S. Ulrich, and M. Bangert, Coverage-based constraints for IMRT optimization, Physics in Medicine and Biology **62**, N460–N473 (2017).

[18] M. Bangert, P. Hennig, and U. Oelfke, Analytical probabilistic modeling for radiation therapy treatment planning., Physics in Medicine and Biology **58**, 5401–5419 (2013).

[19] B. Sobotta, M. Söhn, and M. Alber, Robust optimization based upon statistical theory, Medical Physics **37**, 4019–4028 (2010).

[20] B. Sobotta, M. Söhn, and M. Alber, Accelerated evaluation of the robustness of treatment plans against geometric uncertainties by Gaussian processes, Physics in Medicine and Biology **57**, 8023–8039 (2012).

[21] Z. Perkó, S. R. van der Voort, S. van de Water, C. M. H. Hartman, M. Hoogeman, and D. Lathouwers, Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion., Physics in Medicine and Biology **61**, 4646–4664 (2016).

[22] N. Wahl, P. Hennig, H.-P. Wieser, and M. Bangert, Analytical incorporation of fractionation effects in probabilistic treatment planning for intensity-modulated proton therapy, Medical Physics **45**, 1317–1328 (2018).

[23] S. E. McGowan, F. Albertini, S. J. Thomas, and A. J. Lomax, Defining robustness protocols: a method to include and evaluate robustness in clinical plans., Physics in Medicine and Biology **60**, 2671–2684 (2015).

[24] P. C. Park, J. P. Cheung, X. R. Zhu, A. K. Lee, N. Sahoo, S. L. Tucker, W. Liu, H. Li, R. Mohan, L. E. Court, and L. Dong, Statistical assessment of proton treatment plans under setup and range uncertainties, International Journal of Radiation Oncology*Biology*Physics **86**, 1007–1013 (2013).

[25] F. Cutanda Henríquez and S. Vargas Castrillón, A Novel Method for the Evaluation of Uncertainty in Dose-Volume Histogram Computation, International Journal of Radiation Oncology*Biology*Physics **70**, 1263–1271 (2008).

[26] F. Cutanda Henríquez and S. Vargas Castrillón, The effect of the different uncertainty models in dose expected volume histogram computation, Australasian Physical and Engineering Sciences in Medicine **31**, 196–202 (2008).

[27] F. Cutanda Henríquez and S. Vargas Castrillón, Confidence intervals in dose volume histogram computation, Medical Physics **37**, 1545–1553 (2010).

23

[28] A. C. Kraan, S. Van De Water, D. N. Teguh, A. Al-Mamgani, T. Madden, H. M. Kooy, B. J. M. Heijmen, and M. S. Hoogeman, Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors, International Journal of Radiation Oncology*Biology*Physics **87**, 888–896 (2013).

[29] J. A. Moore, J. J. Gordon, M. S. Anscher, and J. V. Siebers, Comparisons of treatment optimization directly incorporating random patient setup uncertainty with a margin-based approach., Medical Physics **36**, 3880–3890 (2009).

[30] I. Olkin and T. A. Trikalinos, Constructions for a bivariate beta distribution, Statistics & Probability Letters **96**, 54–60 (2015).

[31] S. Nadarajah, S. H. Shih, and D. K. Nagar, A new bivariate beta distribution, Statistics **51**, 455–474 (2017).

[32] A. Genz, Numerical computation of rectangular bivariate and trivariate normal and t probabilities, Statistics and Computing **14**, 251–260 (2004).

[33] N. Wahl, P. Hennig, H.-P. Wieser, and M. Bangert, Efficiency of analytical and sampling-based uncertainty propagation in intensity-modulated proton therapy, Physics in Medicine and Biology **62**, 5790–5807 (2017).

[34] H.-P. Wieser, P. Hennig, N. Wahl, and M. Bangert, Analytical probabilistic modeling of RBE-weighted dose for ion therapy, Physics in Medicine and Biology **62**, 8959–8982 (2017).

[35] M. Hisakado, K. Kitsukawa, and S. Mori, Correlated Binomial Models and Correlation Structures, Journal of Physics A: Mathematical and General **39**, 15365–15378 (2006).

[36] T. E. Wheldon, C. Deehan, E. G. Wheldon, and A. Barrett, The linear-quadratic transformation of dose–volume histograms in fractionated radiotherapy, Radiotherapy and Oncology **46**, 285–295 (1998).

[37] N. Wahl, *Analytical Models for Probabilistic Inverse Treatment Planning in Intensity-modulated Proton Therapy*, Dissertation, Ruprecht-Karls Universität Heidelberg, Heidelberg, 2018.

[38] N. Wahl, P. Hennig, H.-P. Wieser, and M. Bangert, Confidence constraints for probabilistic radiotherapy treatment planning, in *19th International Conference on the Use of Computers in Radiation Therapy (ICCR)*, 2019.

[39] M. van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque, The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy, International Journal of Radiation Oncology*Biology*Physics **47**, 1121–1135 (2000).

[40] M. van Herk, Errors and margins in radiotherapy, Seminars in Radiation Oncology **14**, 52–64 (2004).

24

[41] A. Griewank and A. Walther, *Evaluating Derivatives*, Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, 2008.

[42] H.-P. Wieser, C. P. Karger, N. Wahl, and M. Bangert, Impact of Gaussian uncertainty assumptions on probabilistic optimization in particle therapy, Physics in Medicine and Biology (2020).

Appendices

A   Beta distribution

Suppose a random variable X follows a beta distribution, i.e., $X \sim \mathcal{B}(a,b)$ with shape parameters a and b.

Within the interval $x \in [0,1]$ (for $a > b > 1$, $x \in (0,1)$ otherwise) its probability density function (PDF) $f_X(x)$ is given by

$$f_X(x) = \frac{1}{\mathrm{B}(a,b)} x^{a-1}(1-x)^{b-1} \tag{19}$$

where the normalization $\mathrm{B}(a,b)$ is the beta-function.

The CDF

$$F_X(x) = I_x(a,b) \tag{20}$$

requires evaluation of the regularized incomplete beta-function I_x.

Expectation and variance of X are then given by

$$\mathbb{E}\left[X\right] = \frac{a}{a+b}\,, \tag{21}$$

$$\mathrm{Var}\left[X\right] = \frac{ab}{(a+b+1)(a+b)^2} = \frac{\mathbb{E}\left[X\right]^2 b}{a^2 + ab + a}\,. \tag{22}$$

The shape parameters a and b can be inferred from sample statistics, i.e., the sample mean $\bar{x}$ and the sample variance $\bar{\sigma}^2$, with the method of moments based on Equations (21) and (22) for $\bar{\sigma}^2 < \bar{x}(1-\bar{x})$:

$$\hat{a} = \bar{x}\left(\frac{\bar{x}(1-\bar{x})}{\bar{\sigma}^2} - 1\right)\,, \tag{23}$$

$$\hat{a} = (1-\bar{x})\left(\frac{\bar{x}(1-\bar{x})}{\bar{\sigma}^2} - 1\right) = \bar{x}(1-\bar{x})\hat{a}\,. \tag{24}$$

B Patient Data Information

Table 1: Information on the three patient datasets used for evaluation (similar to Ref. [22]).

patient	intra-cranial	para-spinal	prostate
beam angles	60°, 120°	135°, 180°, 225°	90°, 270°
prescribed dose	60 Gy	60 Gy	70 Gy (76 Gy)
beamlet distance	3 mm	4 mm	5 mm
#beamlets	1705	13274	6803
resolution	$(1.2 \times 1.2 \times 3)\,\text{mm}^3$	$(3 \times 3 \times 3)\,\text{mm}^3$	$(2 \times 2 \times 3)\,\text{mm}^3$
setup error [std. dev.]	$(1\,\text{mm})^{\text{sys}} + (2\,\text{mm})^{\text{rand}}$	$(1\,\text{mm})^{\text{sys}} + (2\,\text{mm})^{\text{rand}}$	$(1\,\text{mm})^{\text{sys}} + (3\,\text{mm})^{\text{rand}}$
range error [std. dev.]	$(3.5\,\%)^{\text{sys}} + (1\,\text{mm})^{\text{rand}}$	$(3.5\,\%)^{\text{sys}} + (1\,\text{mm})^{\text{rand}}$	$(3.5\,\%)^{\text{sys}} + (1\,\text{mm})^{\text{rand}}$

C Generalized model for the ν-th moment of a probability distribution over a DVH-point

Using multi-index notation with the multi-index $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_V) \in \mathbb{N}_0^V$ and by definition of a multi-indexed Heaviside step $\Theta^{\boldsymbol{\kappa}}(\tilde{\boldsymbol{d}} - \hat{d}) = \prod_{i=1}^n \Theta(\tilde{d}_i - \hat{d})^{\kappa_i}$, one can provide a compact general formula to compute the ν-th non-central moment of the probability distribution of a DVH-point, i.e.,

$$
\begin{aligned}
\mathbb{E}\left[\text{DVH}(\hat{d}; \boldsymbol{d})^{\nu}\right] &= \int_{\mathbb{R}^n} \frac{1}{V^{\nu}} \sum_{|\boldsymbol{\kappa}|=\nu} \binom{\nu}{\boldsymbol{\kappa}} \Theta^{\boldsymbol{\kappa}}(\tilde{\boldsymbol{d}} - \hat{d}) f_{\boldsymbol{d}}(\tilde{\boldsymbol{d}})\,\mathrm{d}\tilde{\boldsymbol{d}} \\
&= \frac{1}{V^{\nu}} \sum_{|\boldsymbol{\kappa}|=\nu} \binom{\nu}{\boldsymbol{\kappa}} \left[1 - F_{\boldsymbol{d}_{\kappa}}(\hat{d} \cdot \mathbf{1}_{\nu})\right]
\end{aligned}
\tag{25}
$$

where $F_{\boldsymbol{d}_{\kappa}}(\hat{d} \cdot \mathbf{1}_{\nu})$ corresponds to the evaluation of a ν-variate marginal probability and $\mathbf{1}_{\nu} \in \mathbb{R}^{\nu}$ is a vector with each of the ν components equal to 1. For example, in the case of $\nu = 3$ and $V = 4$, given an index combination $\boldsymbol{\kappa} = (2, 0, 1, 0)$ (satisfying the sum condition $|\boldsymbol{\kappa}| = \nu = 3$), the trivariate probability $F_{\boldsymbol{d}_{1;1;3}}((\hat{d}, \hat{d}, \hat{d})^T)$ needs to be evaluated. Note that the possible "doubling" of an index (i.e., $\kappa_i > 1$) can also be eliminated in the underlying integral using $\Theta(x)^{\kappa_i} = \Theta(x)$. This reduces the given example to an evaluation of a bivariate probability $F_{\boldsymbol{d}_{1;3}}((\hat{d}, \hat{d})^T) = F_{\boldsymbol{d}_{1;1;3}}((\hat{d}, \hat{d}, \hat{d})^T)$.

patient	intra-cranial	para-spinal	prostate
beam angles	60°, 120°	135°, 180°, 225°	90°, 270°
prescribed dose	60 Gy	60 Gy	70 Gy (76 Gy)
beamlet distance	3 mm	4 mm	5 mm
#beamlets	1705	13274	6803
resolution	$(1.2 \times 1.2 \times 3)\ mm^3$	$(3 \times 3 \times 3)\ mm^3$	$(2 \times 2 \times 3)\ mm^3$
setup-error [std. dev.]	$(1\ mm)^{sys} + (2\ mm)^{rand}$	$(1\ mm)^{sys} + (2\ mm)^{rand}$	$(1\ mm)^{sys} + (3\ mm)^{rand}$
range-error [std. dev.]	$(3.5\ \%)^{sys} + (1\ mm)^{rand}$	$(3.5\ \%)^{sys} + (1\ mm)^{rand}$	$(3.5\ \%)^{sys} + (1\ mm)^{rand}$