

Dissertation  
submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

Put forward by  
M. Sc. Niklas Wahl  
Born in: Heidelberg  
Oral Examination: 11.07.2018



# Analytical Models for Probabilistic Inverse Treatment Planning in Intensity-modulated Proton Therapy

Referees: Prof. Dr. Jürgen Hesser  
Prof. Dr. Oliver Jäkel

## Zusammenfassung:

In der intensitätsmodulierten Protonentherapie ist eine patientenspezifische Analyse und Minimierung von Unsicherheiten unumgänglich. Als Alternative zu Methoden, die mit Fehlerszenarien arbeiten, beschreibt diese Arbeit die Implementierung, Anwendung und konzeptionelle Erweiterung der von Bangert, Hennig und Oelfke (2013) eingeführten Analytischen Probabilistischen Modellierung (APM) von Dosisunsicherheiten. APM setzt dabei auf geschlossene Ausdrücke um die Momente der Wahrscheinlichkeitsverteilung über die Dosis aus Set-Up- und Reichweitenunsicherheit herzuleiten und ermöglicht so auch probabilistische Optimierung.

Zuerst wurde APM in MITKrad, einem von Grund auf neu konzipierten Bestrahlungsplanungs-Plugin für MITK, implementiert. Eine Validierung der Berechnungen von APM gegen Stichprobenstatistik zeigte annähernd perfekte Übereinstimmung. Dabei wurden Laufzeiten von wenigen Minuten für eine probabilistische Optimierung und Evaluation von Bestrahlungsplänen erreicht.

Eine Umformulierung von APM erlaubte lineare Separation der Modelle in zufällige und systematische Unsicherheiten. Dadurch konnten Unsicherheiten über das komplette Fraktionierungsspektrum modelliert und optimiert werden, wozu nur eine Vorberechnung nötig war. Dabei konnte gezeigt werden, dass probabilistische Optimierung die Bestrahlung in Fraktionen ausnutzt um Risikoorgane zu schonen.

APM wurde dann mit probabilistischen Modellen zur Beschreibung von klinisch relevanten Qualitätsindikatoren erweitert. Erwartungswert und Varianz von Dosis-Volumen-Histogrammen sowie mittlerer Organdosis konnten präzise modelliert werden. Näherungen für *equivalent uniform dose* (lit. "homogene Äquivalenzdosis") und minimale sowie maximale Organdosis führten hingegen nicht zu zuverlässigen Ergebnissen.

Schließlich wurden die neuen Modelle genutzt um ein Konzept zur probabilistischen Optimierung mit Nebenbedingungen vorzuschlagen. Neben neuen probabilistische Zielfunktionen konnten auch Nebenbedingungen für Quantile der entsprechenden Qualitätsindikatoren realisiert werden. Aufgrund der gestiegenen Berechnungskomplexität, die die neuen Modelle mit sich bringen, wurde der Machbarkeitsnachweis zunächst anhand eines eindimensionalen Prototypen erbracht.

Zusammenfassend schafft die Erweiterung von APM ein probabilistisches Analogon zu bewährten nominellen Planevaluierungs- und Optimierungskonzepten. Wenn in Zukunft die Berechnungskomplexität weiter reduziert werden kann, könnte klinische Implementierung in greifbare Nähe rücken.

## Abstract:

The sensitivity of intensity-modulated proton therapy to uncertainties requires case-specific uncertainty assessment and mitigation. As an alternative to scenario-based methods, this thesis describes the implementation, application and conceptual extension of the [Analytical Probabilistic Modeling \(APM\)](#) framework introduced by Bangert, Hennig, and Oelfke (2013). APM represents moments of the probability distribution over dose in closed-form, providing a probabilistic analog to nominal pencil-beam dose calculation subject to range and setup uncertainties that further enables probabilistic optimization.

First, APM was implemented in [MITKrad](#), a treatment planning plugin for [MITK](#) built completely from scratch. APM's computations were validated against sample statistics, showing nearly perfect agreement. Run-times within minutes could be realized for uncertainty assessment and probabilistic optimization on patient data.

Reformulation of APM enabled linear separation of the computations into random and systematic uncertainty components. Uncertainty over the full fractionation spectrum could then be modeled and optimized with a single pre-computation. It could be shown that fractionation is exploited in optimization with APM for additional [organ at risk](#) sparing.

APM was then extended to propagation of uncertainties from dose to clinically relevant plan quality metrics. Expectation and variance could be modeled accurately for organ mean dose and [dose-volume histograms](#). However, approximations for [equivalent uniform dose](#) and minimum and maximum dose values did not provide reliable results.

Finally, the closed-form plan metrics were used to conceptualize constrained probabilistic optimization. Besides novel probabilistic objectives, confidence constraints could be established. Due to increased computational complexity of the new models, the proof-of-concept was provided through evaluations on a one-dimensional prototype anatomy.

In conclusion, the herein extended APM framework is able to provide probabilistic analogs to established nominal concepts of dose calculation, plan quality metrics, and constrained optimization. If computational hurdles can be overcome in the future, clinical application would be within reach.



---

# Contents

List of Terms and Acronyms	xii
Recurring Symbols & Notation	xiii
Preface	xv
<b>I Introduction</b>	<b>1</b>
<b>II Background &amp; Literature Review</b>	<b>3</b>
II.1 Proton therapy physics . . . . .	3
II.1.1 Characteristics of proton irradiation . . . . .	3
II.1.2 Pencil-beam dose calculation . . . . .	5
II.2 Inverse treatment planning . . . . .	8
II.2.1 Clinical endpoints & treatment plan quality metrics . . . . .	8
II.2.2 Computational treatment plan optimization . . . . .	9
II.3 Uncertainties in proton therapy and their mitigation . . . . .	13
II.3.1 Sources of uncertainty . . . . .	13
II.3.2 Uncertainty propagation & mitigation . . . . .	15
<b>III The Analytical Probabilistic Modeling Framework</b>	<b>21</b>
III.1 Overview . . . . .	21
III.2 The Gaussian pencil-beam dose model . . . . .	22
III.3 Expectation value . . . . .	23
III.4 Covariance . . . . .	24
III.5 Optimization . . . . .	26
III.6 Uncertainty model . . . . .	27
III.6.1 Magnitude of the uncertainties . . . . .	28
III.6.2 Correlation model . . . . .	28
III.6.3 Fractionation . . . . .	29
<b>IV Analytical Probabilistic Modeling: from Concept to Application</b>	<b>31</b>
IV.1 Overview . . . . .	31
IV.2 Methods . . . . .	33
IV.2.1 MITKrad . . . . .	33

IV.2.2	Dose calculation and APM-related computations . . . . .	35
IV.2.3	Validation . . . . .	40
IV.2.4	Approximative computation techniques . . . . .	43
IV.3	Results . . . . .	45
IV.3.1	Proof-of-work . . . . .	45
IV.3.2	Validation of probabilistic dose calculation . . . . .	49
IV.3.3	Performance . . . . .	51
IV.3.4	Feasibility of approximative computations . . . . .	54
IV.4	Discussion . . . . .	58
<b>V</b>	<b>Probabilistic Treatment Planning Under Fractionation</b>	<b>61</b>
V.1	Overview . . . . .	61
V.2	Methods . . . . .	63
V.2.1	The linear covariance influence model . . . . .	63
V.2.2	Generalization to probabilistic analysis and optimization . . . . .	64
V.2.3	Random sampling study . . . . .	65
V.3	Results . . . . .	66
V.3.1	Uncertainty analysis of the full fractionation spectrum . . . . .	66
V.3.2	Probabilistic fractionated treatment plans . . . . .	69
V.4	Discussion . . . . .	74
<b>VI</b>	<b>Analytical Models for Probabilistic Treatment Plan Quality Metrics</b>	<b>79</b>
VI.1	Overview . . . . .	79
VI.2	Methods . . . . .	81
VI.2.1	The multivariate dose model . . . . .	81
VI.2.2	Analytical uncertainty propagation methods . . . . .	82
VI.2.3	Analytical computations . . . . .	84
VI.2.4	Smooth visualizations . . . . .	93
VI.3	Results . . . . .	94
VI.3.1	Technical evaluation of the analytical models . . . . .	94
VI.3.2	Evaluation of analytical models on all patients . . . . .	99
VI.3.3	Visualization examples . . . . .	103
VI.4	Discussion . . . . .	104
<b>VII</b>	<b>Constrained Probabilistic Optimization</b>	<b>107</b>
VII.1	Overview . . . . .	107
VII.2	Methods . . . . .	108
VII.2.1	The constrained probabilistic optimization problem . . . . .	108
VII.2.2	Probabilistic objectives & constraints . . . . .	110
VII.3	Results . . . . .	116
VII.3.1	Evaluation of probabilistic objectives . . . . .	116
VII.3.2	Optimization with $\alpha$ -DVH constraints . . . . .	119

VII.3.3 Optimization with probabilistic EUD & mean dose constraints . . . . .	122
VII.4 Discussion . . . . .	123
<b>VIII Conclusion</b>	<b>127</b>
VIII.1 Summary . . . . .	127
VIII.2 Innovation . . . . .	128
VIII.3 Limitations & Outlook . . . . .	129
VIII.4 Closing words . . . . .	130
<b>A Mathematical Appendix</b>	<b>131</b>
A.1 Probability distributions . . . . .	131
A.1.1 Marginal distributions . . . . .	131
A.1.2 Log-normal distribution . . . . .	131
A.1.3 Beta distribution . . . . .	132
A.1.4 Gaussian algebra & identities . . . . .	133
A.2 Used Integrals . . . . .	133
<b>B Code</b>	<b>135</b>
<b>C Additional Data &amp; Results</b>	<b>139</b>
C.1 Information on Patient Data . . . . .	139
C.2 Supplementary to Chapter <b>IV</b> . . . . .	140
C.3 Supplementary to Chapter <b>V</b> . . . . .	145
<b>Disclosure of Publications</b>	<b>147</b>
<b>Bibliography</b>	<b>149</b>
<b>List of Used Software</b>	<b>157</b>
<b>List of Figures</b>	<b>159</b>
<b>List of Tables</b>	<b>161</b>



---

# List of Terms and Acronyms

- $\alpha$ -DVH quantile dose-volume histogram. [viii](#), [18](#)
- APM Analytical Probabilistic Modeling. [iv](#), [v](#), [vii](#), [viii](#), [xv](#), [2](#)
- BEV beam's eye view. [7](#)
- CDF cumulative distribution function. [xiii](#), [85](#)
- CPU central processing unit. [37](#)
- CSDA continuous slowing down approximation. [4](#)
- CT computed tomography. [6](#)
- CTV clinical target volume. [16](#)
- CUDA Compute Unified Device Architecture. [38](#)
- DVCM dose-volume coverage map. [18](#)
- DVH dose-volume histogram. [v](#), [viii](#), [xi](#), [8](#)
- EUD equivalent uniform dose. [iv](#), [v](#), [ix](#), [9](#)
- FSDVH fractionated standard deviation volume histogram. [69](#)
- GPGPU general purpose computation on GPU. [32](#)
- GPU graphics processing unit. [xi](#)
- GUI graphical user interface. [33](#)
- HU Hounsfield units. [6](#)
- IMPT intensity-modulated particle therapy. [6](#)
- IMRT intensity-modulated radiation therapy. [16](#)
- L-BFGS Low-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. [10](#)

LUT look-up table. 7

NTCP normal tissue complication probability. 8

OAR organ at risk. v, 6

PCE polynomial chaos expansion. 2

PDF probability density function. 83

PTV planning target volume. 16

QUANTEC Quantitative Analyses of Normal Tissue Effects in the Clinic (Marks et al. 2010). 73

RAM random-access memory. 38

RMSE root-mean-square deviation. 55

RT radiation therapy. 1

SDVH standard deviation volume histogram. 47

SFUD single-field uniform dose. 16

TCP tumor control probability. 8

VOI volume of interest. 6

WEPL water-equivalent path length. 6, 7

---

## Recurring Symbols & Notation

Notation	Meaning
$D$	dose influence matrix
$\mathcal{D}$	expected dose influence matrix $\mathbb{E}[D]$
$\mathcal{C}$	covariance influence tensor
$\mathcal{V}$	variance influence tensor with elements $\mathcal{V}_{ijm} = \mathcal{C}_{ijim}$
$\mathbf{d}$	vectorized dose distribution
$\mathbf{w}$	pencil-beam weight/fluence vector
$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$	A multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\Sigma$
$X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$	random variable $X$ follows a normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$
$\Phi(x)$	CDF of the standard normal distribution $\mathcal{N}(0, 1)$
$\text{diag}(x_1, x_2, \dots, x_n)$	Diagonal matrix of size $n \times n$ with diagonal elements $(x_1, x_2, \dots, x_n)$
$\mathbf{x}_{j:m}$	sub-vector of $\mathbf{x}$ over index range: $\mathbf{x}_{j:m} = (x_j, x_{j+1}, \dots, x_m)^T$
$\mathbf{x}_{j;k;l;\dots}$	sub-vector of $\mathbf{x}$ with specific elements: $\mathbf{x}_{j;k;l;\dots} = (x_j, x_k, x_l, \dots)^T$
$\Sigma_{j:m}$	block out of square covariance matrix $\Sigma$

$$\Sigma_{j:m} = \begin{pmatrix} \Sigma_{jj} & \Sigma_{j(j+1)} & \cdots & \Sigma_{jm} \\ \Sigma_{(j+1)j} & \ddots & & \vdots \\ \vdots & & & \\ \Sigma_{mj} & \cdots & & \Sigma_{mm} \end{pmatrix}$$

$\Sigma_{j;k;l;\dots}$  sub-matrix over index selection  $j; k; l; \dots$  of square covariance matrix  $\Sigma$

$$\Sigma_{j;k;l;\dots} = \begin{pmatrix} \Sigma_{jj} & \Sigma_{jk} & \Sigma_{jl} & \cdots \\ \Sigma_{kj} & \Sigma_{kk} & \Sigma_{kl} & \cdots \\ \Sigma_{lj} & \Sigma_{lk} & & \\ \vdots & \vdots & & \ddots \end{pmatrix}$$



---

# Preface

Never tell me the odds!

---

*Han Solo*

THE EMPIRE STRIKES BACK (1980)

Fortunately, not everybody treats uncertainties like Han Solo. The author is thankful to be able to build on previous work of such more considerate people (Bangert, Hennig, and Oelfke 2013).

**Disclosure:** Chapter III is a recap of the introduced *Analytical Probabilistic Modeling* framework by Bangert, Hennig, and Oelfke (2013). Further scientific findings from this thesis have been published before submission, in accordance with the regulations of the Faculty of Physics and Astronomy of the University of Heidelberg. Chapter IV is largely based on Wahl et al. (2017) and a conference contribution at ICCR 2016 (Wahl, Hennig, and Bangert 2016). Wahl et al. (2018a) is the basis of Chapter V. At the time this manuscript was prepared, the findings of Chapters VI and VII were not yet published in print, but partly presented at ESTRO 2018 (Wahl et al. 2018b; Wahl et al. 2018c). A full disclosure of all publications released during the makings of this dissertation can be found **at the end of this manuscript**.

Furthermore, the findings of this thesis regarding extension of *Analytical Probabilistic Modeling* have been made available as *MATLAB*-code on GitHub as “*APMToolbox*”.<sup>1</sup>

**Animations:** This dissertation features PDF-animations that can be viewed on screen. According to the documentation of the respective *L<sup>A</sup>T<sub>E</sub>X*-package *animate*<sup>2</sup>, these are displaying correctly when viewing the PDF in Adobe Acrobat Reader<sup>3</sup> (non-mobile), PDF-XChange<sup>4</sup> and Foxit Reader<sup>5</sup>. The animations are of illustrative value, thus no significant information will be lost when using other viewers or reading in print. Figures containing animations are annotated accordingly.

---

<sup>1</sup><https://github.com/e0404/APMtoolbox> [accessed 29.04.2018]

<sup>2</sup><https://ctan.org/pkg/animate> [accessed 29.04.2018]

<sup>3</sup><https://get.adobe.com/reader/> [accessed 29.04.2018]

<sup>4</sup><https://www.pdf-xchange.de/> [accessed 29.04.2018]

<sup>5</sup><https://www.foxitsoftware.com/pdf-reader/> [accessed 29.04.2018]



## Introduction

Uncertainty is the motor of physical research. Unavoidable inaccuracies in experimental work nurture the search for better, more exact experiments while at the same time giving rise to new theories that can be later extended or discarded due to improved confidence. In practical applications of physics methodology, however, inevitable uncertainties not only undermine the success of an experiment or confirmation of a hypothesis; here uncertainties may have undesired real life implications that require adequate mitigation themselves.

Radiation therapy (RT) is such an application where precision is of uttermost importance. In RT, ionized radiation is used to destroy localized cancerous tissue. To spare healthy tissue, high target conformity is required. Under presence of uncertainty, however, the actually delivered dose deviates from the planned distribution. This can lead to compromised target coverage or overdosage of critical healthy organs. Hence, knowledge and mitigation of the associated uncertainties is crucial for the clinical outcome of a radiation treatment.

Already Goitein (1985) opted for a case-specific assessment of uncertainties. Conventional RT with photons, however, usually mitigates the associated risks with safety margins. The guidelines for such margins are mostly based on generic probabilistic assumptions (Herk 2004; Herk et al. 2000), and not based on the specific treatment itself. The reason for this is that in most cases, photon dose distributions behave comparably static under, for example, displacements of the patient (Bortfeld, Jiang, and Rietzel 2004). Additionally, case-specific methods increase computational complexity with regard to inverse treatment plan optimization, where the dose distribution—and thus its uncertainty estimate—changes during each iteration in the dose optimization process.

Over the last ten years, however, case-specific methods, mostly based on dose scenario samples, have been on the rise. While this is on the one hand facilitated by an increase in computation power and more complex photon therapy modalities, especially the interest for irradiation with ionized particles drives this development.

Leaving uncertainties aside, particles are ideally suited for use in RT, since their characteristic energy loss in matter results in a dose peak at a certain range. This Bragg-peak, however, due to the strong dose gradient it induces, makes particle therapy particularly prone to uncertainties which cannot be sufficiently quantified using static approximations (Lomax 2008a; Lomax 2008b).

While adaptations of the margin concept (e. g. Knopf et al. 2013) are discussed, research in proton therapy mainly focuses on stochastic and robust optimization (e. g. Fredriksson 2012; Liu

et al. 2012b; Unkelbach et al. 2009). These require the computation of sampled or worst-case dose scenarios, which are then used for uncertainty evaluation and during optimization.

Eventually, such methods provide no insight in the actual model connecting uncertainty over input parameters with uncertainty over dose and derived plan metrics. Consequently, disadvantages of such methods are, for example, the aforementioned requirement of re-computations during optimization, limited choices for optimization routine and objective functions and upper bounds on the complexity of the uncertainty model. Furthermore, at least from an academic point of view, a sound uncertainty propagation model may be preferable as it aids in understanding the intrinsic mechanisms connecting input and output uncertainty in RT.

Few attempts have been made to overcome this lack of model. These approaches do, however, tackle the problem from behind; for example, they reconstruct an uncertainty model based on Gaussian process regression (Sobotta, Söhn, and Alber 2012) or numerical integration for polynomial chaos expansion (PCE) (Perkó et al. 2016). Hence, these approaches also rely on the computation of discrete scenarios at last, yet with more accurate reproductive character and possibly reduced computational complexity.

A forward approach to analytically model the relationship between input uncertainties and uncertainties over dose was proposed by Bangert, Hennig, and Oelfke (2013). Their *Analytical Probabilistic Modeling (APM)* framework inherently re-formulates dose calculation for proton (and photon) therapy in a way that allows for closed-form propagation of uncertainties from treatment plan parameters to dose. Thus, no samples need to be drawn for robustness analysis. Bangert, Hennig, and Oelfke (2013) proposed to use this method for computationally efficient probabilistic analysis and optimization, as well as description of more complex uncertainty models, particularly in the context of fractionation.

This thesis picks up *APM* in the conceptual state from Bangert, Hennig, and Oelfke (2013) and brings the proposed concepts to application on full-fledged 3D patient cases. The proposed methodology is conceptually extended and re-formulated and consequently efficiently implemented and evaluated on patient data. The thesis further proposes extensions to *APM* to analytically model propagation to uncertainties in dose-dependent treatment plan metrics. With derived closed-form expressions, an approach for probabilistically constrained treatment planning is conceptualized. As such, it becomes possible to gradually move between expectation and worst-case robustness by constraining on concrete, desired confidence levels.

The remainder of this thesis is organized in distinct chapters. Chapter II gives a brief introduction into the principles of RT physics and treatment planning as well as literature relevant for this work. The *APM* framework by Bangert, Hennig, and Oelfke (2013) is explained separately in Chapter III. Chapter IV then describes implementation and validation of *APM*. The implementation is used in Chapter V to investigate fractionated probabilistic treatment planning. Chapter VI extends *APM* to model uncertainties in selected treatment plan metrics. Chapter VII takes these models to propose a concept for constrained probabilistic optimization. Thereby, Chapters IV to VII follow similar structure; they begin with a short introduction and literature overview, followed by a description of methodology, consequent presentation of the results and conclude with a discussion. The thesis closes with a summarizing conclusion in Chapter VIII.

# Background & Literature Review

This chapter introduces and briefly explains the main quantities and general environment constituting the basis of the scientific work in this thesis. This includes the physical basis of proton therapy in Section II.1, followed by Section II.2 lining out the treatment planning workflow and Section II.3 concluding this chapter with a discussion of similar academic approaches tackling mitigation of uncertainties in treatment planning.

This chapter does not serve as an in-depth introduction and explanation of the aforementioned much broader fields. The interested reader might be referred to Paganetti (2012a) or Newhauser and Zhang (2015).

## II.1 Proton therapy physics

The physical aspects of proton therapy comprise fundamental physics, i. e., interactions of the protons with the patient tissue, as well modeling and computation of the respective energy dose.

### II.1.1 Characteristics of proton irradiation

First, the fundamental principles underlying treatment with protons will be introduced, i. e., dosimetric characteristics of proton irradiation. This includes the definition of dose and how this dose is distributed due to fundamental interactions.

#### II.1.1.1 Energy dose

The killing of tumor cells and the sparing of healthy tissue is dependent on the energy dose  $d$ . It is given as the differential energy  $dE$  deposited in the infinitesimal mass element  $dm$ , i. e.,

$$d = \frac{dE}{dm} = \frac{1}{\rho} \frac{dE}{dV}, \quad (\text{II.1})$$

where additionally the mass density  $\rho$  in the respective infinitesimal volume element  $dV$  was factored out.

For radiation therapy, computation of Eq. (II.1) hence requires modeling of the energy deposition characteristics of the irradiation beam as well as knowledge over the patient anatomy. The latter is obtained through medical imaging. The energy deposition depends on the type of irradiation, and is, in general, driven by an interplay of multiple interaction processes.

### II.1.1.2 The Bragg-curve

For protons (and other charged heavy particles), the (expected) energy loss along the particle path in z-direction is described by Eq. (II.2) attributed to Bethe (1930) and Bloch (1933),

$$-\frac{dE}{dz} = \frac{4\pi n_e q^2}{m_e c^2 \beta^2} \left( \frac{e^2}{4\pi\epsilon_0} \right)^2 \left[ \ln \left( \frac{2m_e c^2 \beta^2}{I(1-\beta^2)} \right) - \beta^2 \right], \quad (\text{II.2})$$

with electron density of the material  $n_e$ , incoming particle charge  $q$ , electron mass  $m_e$ , speed of light  $c$ , the ratio of particle velocity and speed of light  $\beta = v/c$ , vacuum permittivity  $\epsilon_0$  and ionization potential of matter  $I$ . While Eq. (II.2) describes properties of a particle traversing matter, it can also be interpreted as a material parameter, and is then called the *stopping power*  $S$ .

In Eq. (II.2), the energy loss is proportional to  $v^{-2}$ . Hence, while the particle slows down, it loses more and more energy until it stops. At this point  $v = 0$ , in Eq. (II.2), a singularity is reached. The distance traveled up to this point is called the particles' *range* in that material, which may be approximated with the *continuous slowing down approximation (CSDA)*,

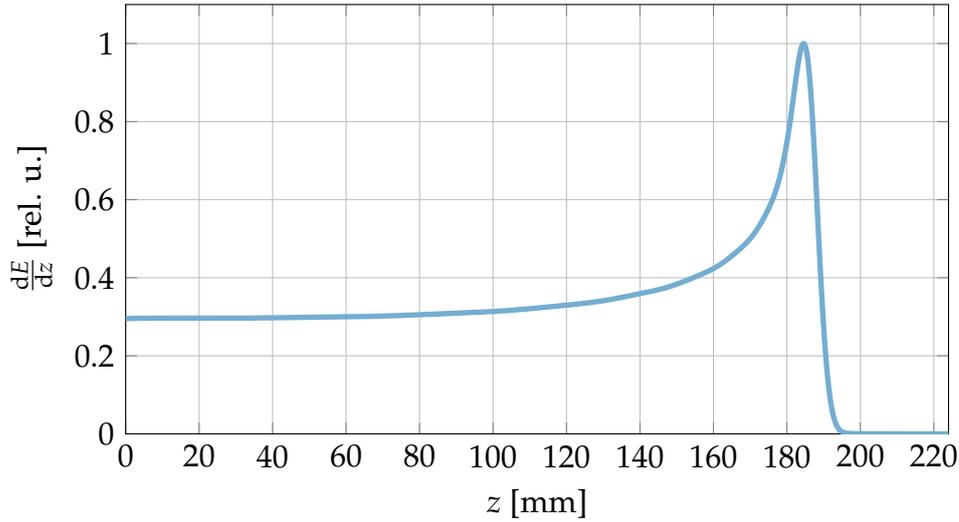
$$R(E_0) = \int_0^{E_0} \frac{1}{-\frac{dE}{dz}} dE, \quad (\text{II.3})$$

where  $E_0$  is the initial energy of the incoming particle. Note that the *CSDA* assumes that the particle is continuously decelerated, which is not the case when observing a single particle. However, as the mean over multiple particles, it can serve to estimate the range of a particle beam with initial energy  $E_0$ . In total, the single particles will exhibit slightly different ranges, an effect known as *range straggling*. Taken together, the expected energy loss of a particle beam along its impinging axis dependent on traveled distance  $z$  will slowly increase, exhibit a maximum shortly before the mean range, and then suddenly fall of to zero. This Bragg-curve (named after its discoverer) is illustrated in Fig. II.1 for a proton beam penetrating a water phantom.

### II.1.1.3 Multiple Coulomb scattering

Particles traversing matter not only loose energy due to inelastic scattering with hull electrons. While inelastic scattering of the incoming particle with material nuclei may also occur, the nuclear interactions are predominantly elastic, and the incoming particles will change their trajectory in each scattering event.

Such single elastic scattering events are well described by Rutherford's theory, but the multiple case is theoretically more complex to evaluate (Newhauser and Zhang 2015). One of the most completely developed theories stems from Molière (for a recent review see Bednyakov 2014), requiring complex and lengthy computations. Since scattering under large angles rarely occurs,



**Figure II.1:** Normalized energy loss  $dE/dz$  of a proton beam in water along the impinging axis  $z$  (i. e., the Bragg-curve), with a range of  $R = 187$  mm and initial energy  $E_0 \approx 170$  MeV. Based on data from `matRad`, which assumes an initial energy spectrum of non-zero width and corresponding range straggling (Wieser et al. 2017b).

Gottschalk et al. (1993) discussed approximations that neglect the respective higher order terms developed by Molière. In particular, Gottschalk et al. (1993) showed that two-dimensional cylindrical Gaussian approximations are suited for representing the angular distribution for small angles  $\theta$  with characteristic angle  $\theta_0$ :

$$f(\theta) = \frac{1}{2\pi\theta_0^2} \exp\left[-\frac{1}{2}\left(\frac{\theta}{\theta_0}\right)^2\right]. \quad (\text{II.4})$$

During the particles journey through the material, the characteristic angle  $\theta_0$  increases. Hence, the energy loss distributes over a widening lateral profile which has approximately cylindrical Gaussian form.

### II.1.2 Pencil-beam dose calculation

For computational treatment planning before patient irradiation, the physical principles briefly outlined above have to be translated into computer algorithms. The dose  $d(\mathbf{r})$  at a coordinate  $\mathbf{r}$  is delivered through external beams created by a particle accelerator, i. e., a synchrotron or a cyclotron. While there are multiple ways to create the desired irradiation fluence, this thesis' work is based on an active scanning technique with full 3D intensity-modulation of the irradiation field (Lomax 1999). This means that the fluence of a beam is generated through  $n_B$  focused, each itself nearly monoenergetic pencil-beams. The particle fluence of these individual pencil-beams is then represented through a fluence vector  $\mathbf{w} \in \mathbb{R}_+^{n_B}$ , where  $\mathbb{R}_+^{n_B}$  is the positive orthant within the  $n_B$ -dimensional real space. This allows representation of the irradiation dose  $d$  at a coordinate  $\mathbf{r}$  as

$$d(\mathbf{r}) = \sum_j D_j(\mathbf{r})w_j, \quad (\text{II.5})$$

where  $D_j(\mathbf{r})$  is the normalized *dose influence* of pencil-beam  $j$  in coordinate  $\mathbf{r}$ .

In a common therapeutic setting, a **computed tomography (CT)** image, where the patient is positioned as intended for treatment, is taken for the computerized planning process. Such a CT image is discretized into small volumetric elements called *voxels* (i. e., the 3D analog to pixels), each providing a photon attenuation value measured in Hounsfield units (HU). Structures relevant for the treatment planning workflow, denoted as **volume of interests (VOIs)**, are segmented by clinical staff and define target structures as well as **organs at risk (OARs)**.

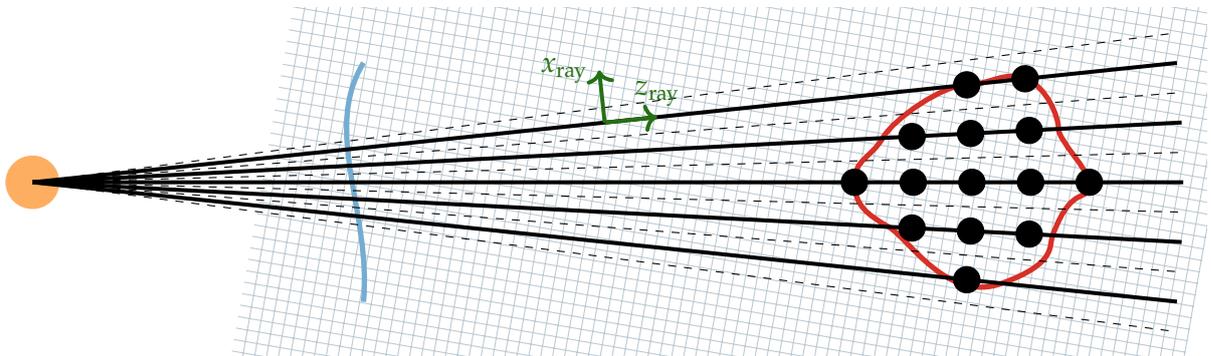
The discretization also translates to dose  $d$  and dose influence  $D_j(\mathbf{r})$ , such that dose is represented as a vector  $\mathbf{d} \in \mathbb{R}_+^{n_V}$  with number of voxels  $n_V$ . This gives  $D \in \mathbb{R}_+^{n_V \times n_B}$  the shape of a matrix, where the dose  $d_i$  in all voxels  $i$  is computed by a matrix multiplication

$$d_i = \sum_j D_{ij} w_j \Rightarrow \mathbf{d} = D\mathbf{w} \quad (\text{II.6})$$

with the fluence vector  $\mathbf{w}$ .

### II.1.2.1 Pencil-beam scanning grid

To compute the dose, steering information for the irradiating machine has to be generated or provided. Besides selection of beam angles, for **intensity-modulated particle therapy (IMPT)** this means the definition of the scanning grid, i. e., the distribution of Bragg-peaks inside the target, by selection of appropriate energies, as depicted in Fig. II.2.



**Figure II.2:** Schematic illustration of the definition of the pencil-beam scanning grid. The goal is to place Bragg-peaks (●) within a target (—). To do so, from the virtual beam source (●) the appropriate range/energy has to be determined by ray-casts (- - -) through the image. With appropriate lateral spacing, a set of central rays (—) is selected, on which multiple energies are selected to create a spread out Bragg-peak with respective pencil-beams. Note that in the “real world” setting, the source is much further away from the patient’s skin (—), and therefore the rays are nearly parallel. Also the coordinate system of the ray is, in general, different from the voxel grid in the patient coordinate system indicated in the background. Adapted from the [matRad Wiki](https://github.com/e0404/matRad/wiki/Dose-influence-matrix-calculation).<sup>1</sup>

### II.1.2.2 Water-equivalent path length

For such a placement of Bragg-peaks within the target, the appropriate range/energy has to be determined. For this, the stopping power of the patient is required, as described above in Sec-

<sup>1</sup><https://github.com/e0404/matRad/wiki/Dose-influence-matrix-calculation> [accessed 28.04.2018]

tion II.1.1.2. The CT scan, in general, only provides HU values for every voxel. Using a look-up table (LUT), based on measurements and calculations, the HU values are usually converted to a relative stopping power compared to water (Newhauser and Zhang 2015).

Performing a ray-cast, i. e., a line integral, through this converted image (Siddon 1985) from the beam's source now generates a water-equivalent path length (WEPL) for each (relevant) voxel per beam. Since the range of a proton beam in water can be computed or measured beforehand, this allows for selection of the required energies as soon as the WEPL within the patient is available. Furthermore, the now available ray-cast images, each of them providing the WEPL as seen from each beam's source, can be used for calculation of the actual dose as described below.

### II.1.2.3 Coordinate separation

The physics outlined in Section II.1 is described within the coordinate system of the incoming particle beam, i. e., the beam's eye view (BEV). This suggests computation of dose directly in the BEV, since then the dose calculation can be separated into computation of a depth-dependent part  $Z$ , and lateral components  $L^x$  and  $L^y$  (Hong et al. 1996). An element of the dose influence matrix  $D_{ij}$  may hence be written as

$$D_{ij} = L_{ij}^x(x_{ij}; z_{ij, \text{WEPL}}; E_{0,j}) L_{ij}^y(y_{ij}; z_{ij, \text{WEPL}}; E_j) Z_{ij}(z_{ij, \text{WEPL}}; E_{0,j}), \quad (\text{II.7})$$

with the lateral coordinates  $x_{ij}$  and  $y_{ij}$  of voxel  $i$  in the respective ray coordinate system (compare Fig. II.2), its WEPL  $z_{ij, \text{WEPL}}$  and the initial energy  $E_{0,j}$  of pencil-beam  $j$ . Note that in Eq. (II.7) also the lateral components depend on the respective effective depth  $z_{ij, \text{WEPL}}$  in voxel  $i$ , as described by Schaffner, Pedroni, and Lomax (1999).

### II.1.2.4 Base data

In Eq. (II.7) all components are parametrized by the initial pencil-beam energy  $E_{0,j}$ . Since the other coordinates are merely of geometrical nature, it is in general possible to use analytical approximations to compute all components  $Z$ ,  $L^x$ ,  $L^y$  (Bortfeld 1997; Gottschalk et al. 1993; Hong et al. 1996).

A in clinical applications more common alternative is to tabulate data gained from measurements or Monte Carlo simulations of the dose deposited by a pencil-beam in water. That is, for each available initial energy, the Bragg-curve (compare Fig. II.1) is stored in a LUT dependent on the WEPL, integrated along the respective lateral spectrum. This allows to represent the lateral components as normal distributions, whose width/variance  $[\lambda_j^{x/y}(z_{i, \text{WEPL}})]^2$  can be similarly stored in a WEPL dependent LUT.

The dose influence element  $D_{ij}$  from Eq. (II.7) is then expressed as

$$D_{ij} = \mathcal{N}(x_{ij}; \mu_j^x, (\lambda_{ij}^x)^2) \mathcal{N}(y_{ij}; \mu_j^y, (\lambda_{ij}^y)^2) Z_{ij}, \quad (\text{II.8})$$

where the lateral position  $\mu_j^{x/y}$  of the central ray corresponding to pencil-beam  $j$  vanishes, if Eq. (II.8) is computed directly in the coordinate system of the respective ray.

That is, the components can be defined based on the combination on voxel index  $i$  and pencil-beam index  $j$ , since the respective geometrical coordinates, WEPLs and pencil-beam energy are known and can be used to query and interpolate values stored in the LUTs.

## II.2 Inverse treatment planning

Modern clinical treatment plans are not completely manually defined. Instead, a set of basic parameters is given by the planner, i. e., beam geometry, modality etc., complemented by a set of objectives and constraints for an inverse planning routine that optimizes the irradiation beam fluences to create therapeutic dose distribution. From a clinical point of view, the goal is to maximize tumor control probability (TCP) while keeping normal tissue complication probability (NTCP) at a minimum. Due to difficulties on modeling consistent and correctly normalized TCP and NTCP, treatment planning relies on mathematical objectives and constraints based on surrogate plan metrics from which NTCP and TCP can be estimated (see e. g. Marks et al. 2010).

In the following, indicators relevant for this thesis will be introduced, and their usage in treatment plan optimization will be outlined shortly, together with a brief introduction into the optimization problem and routine.

### II.2.1 Clinical endpoints & treatment plan quality metrics

Note that in the following, all indicators are usually related to a VOI  $v$ . For the rest of this section,  $\mathbf{d}$  will represent the dose vector over  $n_v$  voxels belonging to an arbitrary VOI.

#### II.2.1.1 Dose-volume histograms

Dose-volume histograms (DVHs) are amongst the most widely used plan metrics used for evaluation of treatment plans. A DVH is a normalized cumulative histogram of the dose within a VOI. Hence, a DVH-point, i. e., a dose-volume DVH( $\hat{d}$ ), represents the volume fraction that is covered by at least dose  $\hat{d}$ . DVH( $\hat{d}$ ) can be represented with the Heaviside step function  $\Theta(x)$  as

$$\text{DVH}(\hat{d}; \mathbf{d}) = \frac{1}{n_v} \sum_{i \in v} \Theta(d_i - \hat{d}). \quad (\text{II.9})$$

Note that the first derivative w. r. t. dose yields a conventional, yet normalized histogram of the dose values  $\mathbf{d}$ . The DVH is of special importance, since all following plan metrics can be also reasonably well approximated from a computed DVH with a sufficient number of histogram bins.

#### II.2.1.2 Extreme values & mean dose

Plan metrics also include trivial statistics over the dose vector  $\mathbf{d}$ . This includes minimum dose  $\min \mathbf{d}$  and maximum dose  $\max \mathbf{d}$  as well as the mean dose  $\bar{\mathbf{d}}$  in the VOI:

$$\bar{\mathbf{d}} = \frac{1}{n_v} \sum_{i \in v} d_i. \quad (\text{II.10})$$

Since minimum and maximum dose are particularly sensitive, often weaker criteria are used which can be deduced from the inverse  $DVH^{-1}(\hat{v}; \mathbf{d})$  as

$$d^{\hat{v}}(\mathbf{d}) = DVH^{-1}(\hat{v}; \mathbf{d}), \quad (\text{II.11})$$

where  $d^{\hat{v}}$  corresponds to the minimal dose received by the volume fraction  $\hat{v} = DVH(d^{\hat{v}}; \mathbf{d})$ . For example,  $d^{98\%}$ , i. e., the minimal dose that is received by 98 % of the volume of the VOI, could serve as a surrogate for  $\min \mathbf{d}$ . For the maximum, one could analogously use the minimal dose received by 2 % of the VOI's volume. In general, these values correspond to order statistics of the dose vector  $\mathbf{d}$  in VOI  $v$ .

### II.2.1.3 Equivalent uniform dose

An alternative to order statistics, i. e., the inverse DVH-points, is given by **equivalent uniform dose (EUD)**. EUD as introduced by Niemierko (1997) aims at defining a VOI-specific dose level  $EUD(\mathbf{d})$  derived from  $\mathbf{d}$  that induces the same radio-biological effect on the tissue when homogeneously irradiated with. For example, a single cold spot in the tumor is extremely significant and may lead to unintended survival of tumor cells, and thus the EUD in the tumor will be lower than the respective mean dose from Eq. (II.10) if such a cold spot is present.

Niemierko (1999) later generalized the concept to be represented by the generalized mean formula

$$EUD^{\kappa}(\mathbf{d}) = \left[ \frac{1}{n_v} \sum_{i \in v} d_i^{\kappa} \right]^{\frac{1}{\kappa}}, \quad (\text{II.12})$$

where the exponent  $\kappa$  can also take non-integer values. While EUD defined by Eq. (II.12) is also often denoted as “gEUD”, the prefix “g” will be omitted and further use of the term EUD in this thesis will refer to the definition from Eq. (II.12).

EUD as defined in Eq. (II.12) yields two interesting limits. If  $\kappa$  approaches positive infinity, i. e.,  $\kappa \rightarrow \infty$ , one has  $EUD^{\kappa}(\mathbf{d}) \rightarrow \max \mathbf{d}$ , and similarly for  $\kappa \rightarrow -\infty$  one has  $EUD^{\kappa}(\mathbf{d}) \rightarrow \min \mathbf{d}$ . Between this extremes, choice of  $\kappa$  continuously determines sensitivity to heterogeneities.

In practice,  $\kappa$  may be referenced as a “tissue-parameter” that describes the sensitivity of distinct organ and tumor types. As such, EUD is of particular importance in modeling NTCP (e. g. Marks et al. 2010).

## II.2.2 Computational treatment plan optimization

For treatment plan optimization, the goal is to shape the dose distribution  $\mathbf{d}$  such that it approximates the prescribed dose  $\mathbf{d}^*$  as good as possible. In general,  $\mathbf{d}^*$  contains the desired tumor dosage in all voxels inside target structures, and is zero otherwise. However, the dose distribution  $\mathbf{d}$  cannot be arbitrarily shaped, but linearly depends on the particle fluence vector  $\mathbf{w}$  (compare Eq. (II.6)). Consequently, healthy tissue is necessarily irradiated when covering the target.

Hence, one attempts to approximate  $\mathbf{d}^*$  by parameterizing a set of objectives and constraints, possibly representing the clinical endpoints described above, to factor in and trade off different tissue and tumor responses.

### II.2.2.1 Optimization problem

An optimal fluence  $\mathbf{w}^*$  can be found by solving the optimization problem

$$\begin{aligned} \mathbf{w}^* = \arg \min_{\mathbf{w}} \quad & \mathcal{F}(\mathbf{d}(\mathbf{w})) = \sum_s p_s \mathcal{F}_s(\mathbf{d}(\mathbf{w})) \\ \text{subject to} \quad & \mathbf{w} \geq \mathbf{0} \\ & c_i(\mathbf{d}(\mathbf{w})) \leq 0 \quad i = 1, \dots, m \\ & h_j(\mathbf{d}(\mathbf{w})) = 0 \quad j = 1, \dots, p \end{aligned} \tag{II.13}$$

with objective function  $\mathcal{F}$ , possibly composed out of sub-objectives  $\mathcal{F}_s$  weighted by a manually defined penalty  $p_s$ .  $\mathbf{0}$  denotes the vector of zeros to bound the fluence  $\mathbf{w}$  to physically sensible positive values during optimization, and  $c_i$  and  $h_j$  are inequality and equality constraints, respectively. Note that within the scope of this thesis, objective and constraints indirectly depend on  $\mathbf{w}$  through  $\mathbf{d}$ , separating the problem into a dose-dependent and fluence-dependent part. Implications of this separation will be addressed further below. Also, equality constraints have limited applications in radiation therapy, but are named for the sake of completeness.

**Optimization technique:** Several techniques exist to solve optimization problem (II.13). In general, optimization of problems with non-linear constraints is based on replacing the constrained problem with an unconstrained one through penalty methods (for inequality constraints) and augmented Lagrangian methods (for equality constraints) (Nocedal and Wright 2006). The scientific work described by this thesis relies on interior-point optimization using the open-source optimizer IPOPT (Wächter and Biegler 2006), and alternatively the `fmincon`<sup>2</sup> routine from the “Optimization Toolbox” of MATLAB.

Interior-point methods base on formulating inequality constraints as logarithmic barrier functions using *slack variables* to create an unconstrained problem whose objective function is defined in an “interior space” (hence the name) bounded by the constraints. The unconstrained problem is then optimized with a Quasi-Newton algorithm using an L-BFGS approximation. For a thorough introduction into the optimization method, the interested reader may be forwarded to Nocedal and Wright (2006) and Wächter and Biegler (2006).

**Quasi-Newton & derivatives:** Where Newton-methods require the first and second derivatives of the objective function, i. e., its gradient and its Hessian w. r. t. the optimization variable, quasi-Newton methods require only the gradient and approximate the Hessian (Nocedal and Wright 2006).

<sup>2</sup><https://de.mathworks.com/help/optim/ug/fmincon.html> [accessed 28.04.2018]

Due to the dependence of objective and constraints on the optimization variable  $w$  through  $d$  in problem (II.13), computation of the gradient  $\nabla_w \mathcal{F}$  separates into

$$\nabla_w \mathcal{F} = \frac{\partial \mathcal{F}}{\partial w} = \left[ \frac{\partial d}{\partial w} \right]^T \cdot \frac{\partial \mathcal{F}}{\partial d} = D^T \frac{\partial \mathcal{F}}{\partial d}, \quad (\text{II.14})$$

by application of the chain rule. Similar separations in  $d$  and  $w$  are applicable for the constraint functions, where the Jacobian  $J_{c,h}$  over all constraint functions is required.

Equation (II.14) is particularly convenient when the objective function is composed of several sub-objectives as in Eq. (II.13). Then, during an iterate in optimization, first the derivatives w. r. t.  $d$  of all objectives can be efficiently evaluated based on the current dose. Afterwards, the computationally expensive product involving the transposed dose influence matrix  $D^T$  in Eq. (II.14) needs only to be exercised once to obtain the derivatives w. r. t.  $w$ .

### II.2.2.2 Objective Functions

In the following, common objectives relevant for this thesis are outlined below. Interested readers might, at this point, also be referred to Wu and Mohan (2000).

**(Piece-wise) Squared Deviation:** The most basic objective function is based on a penalized least-squares fit of the dose distribution  $d$  to the prescribed dose  $d^*$  (Oelfke and Bortfeld 2001).

For a VOI  $v$  with  $n_v$  voxels, for example, the least-squares objective  $\mathcal{F}^{\text{LS}}(d)$  is computed over all voxels  $i \in v$  via

$$\mathcal{F}^{\text{LS}}(d) = \frac{1}{n_v} \sum_i (d_i - d_i^*)^2. \quad (\text{II.15})$$

When composing the overall objective function according to Eq. (II.13) with multiple objectives that are only of type (II.15),  $\mathcal{F}$  can be compactly written in matrix notation as

$$\mathcal{F}^{\text{LS}}(d) = (d - d^*)^T P (d - d^*), \quad (\text{II.16})$$

where  $P = n_v^{-1} \text{diag}(p_1, p_2, \dots, p_{n_v})$  is a diagonal matrix encoding normalized volume based penalties  $p_v$  per voxel.

*Piece-wise definition:* To better include tolerances of OARs to small doses, it may be desirable to relax the objective (II.15) up to a certain tolerance dose value  $d^{\text{max}}$ . Such a relaxation can be incorporated using step-functions as positivity operators, under which Eq. (II.15) becomes

$$\mathcal{F}^{\text{LS}+}(d) = \frac{1}{n_v} \sum_i \Theta(d_i - d_i^{\text{max}}) (d_i - d_i^{\text{max}})^2. \quad (\text{II.17})$$

Objective (II.17) now has no contribution if the dose falls below  $d^{\text{max}}$ .

Note that such relaxations are also possible when exceeding a threshold  $d^{\text{min}}$  by changing the sign of the argument of the Heaviside step  $\Theta(-x)$ .

**Mean dose & EUD:** The open-source treatment planning toolkit `matRad` (Wieser et al. 2017b) includes mean dose and equivalent uniform dose as bare minimizers of dose in healthy tissue. Consequently, they are trivially formulated as

$$\mathcal{F}^{\text{EUD}^\kappa}(\mathbf{d}) = \text{EUD}^\kappa(\mathbf{d}) \quad (\text{II.18})$$

$$\mathcal{F}^{\text{mean}}(\mathbf{d}) = \text{EUD}^{\kappa=1}(\mathbf{d}) = \bar{\mathbf{d}}, \quad (\text{II.19})$$

i. e., directly represented by Eqs. (II.10) and (II.12) used for their computations.

Due to their importance for NTCP-modeling (Marks et al. 2010), more complex objectives, especially based on EUD, exist (Choi and Deasy 2002; Mohan et al. 1992; Wu et al. 2002). However, these are not relevant for the scientific work described by this thesis, and therefore omitted.

**Dose-volume histogram objectives:** Objectives based on DVHs intend to give better control over the dose distribution (Wu and Mohan 2000), particularly avoiding local cold- or hot-spots that can arise when optimizing with objective functions (II.15) & (II.17).

The definition from Wu and Mohan (2000) for a maximum DVH-objective, also used by Wieser et al. (2017b), complements the piece-wise least squares objective from Eq. (II.17) with an additional Heaviside step  $\Theta(d^\hat{v} - d_i)$  where  $d^\hat{v}$  is the (minimal) dose to the prescribed volume fraction  $\hat{v}$  is receiving according to Eq. (II.11):

$$\mathcal{F}^{\text{DV}^+}(\mathbf{d}) = \frac{1}{n_v} \sum_i \Theta(d_i - d_v^*) \Theta(d^\hat{v} - d_i) (d_i - d_v^*)^2. \quad (\text{II.20})$$

When now prescribing a relative volume  $\hat{v}$  and a dose  $d_v^*$  to a VOI, the term  $\Theta(d^\hat{v} - d_i)$  excludes values below  $d^\hat{v}$ . This means that as soon as  $\text{DVH}(d_v^*; \mathbf{d}) \leq \hat{v}$ , the objective function is minimal, namely zero. Similar considerations apply for minimum DVH-objectives by changing the sign of the arguments of the Heaviside-steps.

### II.2.2.3 Constraints

Wieser et al. (2017b) formulate inequality constraints for EUD, mean dose and DVH-points directly based on the formula used to compute the respective plan metric for evaluation (compare Eqs. (II.9), (II.10) and (II.12)). Hence those are not explicitly stated.

For minimum and maximum dose constraints, however, a “LogSumExp”-approximation is used to avoid the definition of voxel-wise constraint functions for computational efficiency,

$$c^{\text{max}d} = \kappa \ln \left[ \sum_i \exp \left( \frac{d_i - d^{\text{max}}}{\kappa} \right) \right] \leq 0, \quad (\text{II.21a})$$

$$c^{\text{min}d} = \kappa \ln \left[ \sum_i \exp \left( \frac{d^{\text{min}} - d_i}{\kappa} \right) \right] \leq 0. \quad (\text{II.21b})$$

where  $\kappa$  is a parameter to control the steepness of the smooth approximation. `matRad`, for example, uses a value of  $\kappa = 1 \times 10^{-3}$ .

## II.3 Uncertainties in proton therapy and their mitigation

Conducting a radiation treatment requires control of several parameters for the patient, the machine and treatment plan computation. Due to the fundamental presence of uncertainty in any experimental or engineered set-up, the actual treatment is merely an unknown realization from the probability distributions of these parameters. This uncertainty in parameters propagates to an uncertainty in dose (Goitein 1985), which needs to be considered and mitigated in any radiation treatment planning, regardless of radiation type.

This section gives a state-of-the-art review of quantifying and mitigating uncertainties in proton therapy. Note that distinct aspects are more or less important for the work with APM in Chapters III to VII, and therefore a more specialized review is given at the beginning of each respective chapter.

### II.3.1 Sources of uncertainty

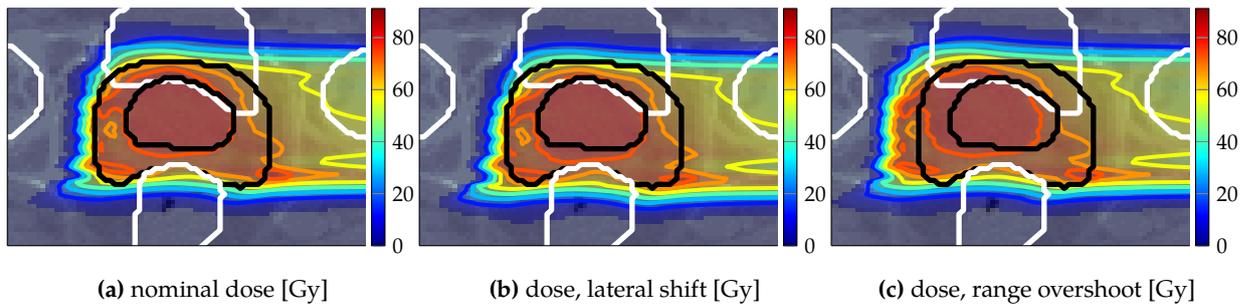
Sources of uncertainty in radiotherapy are omnipresent. Particle therapy, however, is especially sensitive to geometric and calculational uncertainties. There, the Bragg-peak is at the same time a blessing and a curse; the localized energy loss at the range of the particles allows precise shaping of conformal dose distributions, however, inexact knowledge of the tissue traversed up to the peak may induce significant uncertainty in the dose distribution (Lomax 2008b), which is to the same extent not the case in photon therapy. Such inexact knowledge may not only be caused by calculational errors in the treatment planning process, but also displacements and inter-field motions (Lomax 2008a).

This thesis will focus on three aspects of uncertainty, which will be introduced below: (1) spatial set-up uncertainty, i. e., a displacement error of the anatomy relative to the irradiating beam, (2) calculational range uncertainties, i. e., the miscomputation of WEPL, and (3) aspects on both previous uncertainties under fractionation where they separate into systematic and random components. Furthermore, at the end of this section, additional sources of uncertainties are discussed, which are not evaluated in this thesis but play a critical role nevertheless.

#### II.3.1.1 Set-up uncertainties

As explained in Section II.1.2.1, the definition and optimization of a treatment plan is usually based on a pre-treatment planning CT. When the patient is set-up in the treatment room, it has to be ensured that the patient's position is consistent with the CT image. This is usually achieved by skin markers which are matched with laser positioning systems, and/or x-ray images within the treatment room to evaluate anatomic differences and misalignments.

The efforts made show the importance of patient positioning for generation of accurate treatment plans. A misalignment of the patient not only means a static spatial shift of the dose distribution within the patient, but also changes the irradiated anatomy. Especially in particle therapy, this may lead to range changes and thus non-trivial distortions of the dose distribution. Figure II.3 gives an example for such a misalignment.



**Figure II.3:** Exemplary error scenarios for a prostate case irradiated with a single proton beam. (a) shows the originally optimized dose distribution, while in (b) the patient is misaligned by 3 mm in cranial direction, i. e., a lateral displacement relative to the beam axis. (c) simulates a systematic underestimation of WEPL by 3.5 %, resulting in a corresponding range overshoot.

### II.3.1.2 Range uncertainties

Another source for potential errors lies in the conversion from the acquired CT, given in HUs, to particle stopping power (compare Section II.1.2.2). Incorrect knowledge of the exact tissue composition, and therefore in particular of the ionization potential  $I$ , as well as imaging artifacts contribute to a consequent uncertainty over the WEPL.

Lomax (2008b) suggests an uncertainty of  $\pm 3\%$  based on measurements of the calibration curve for converting HU to stopping power (see also Schaffner, Pedroni, and Lomax 1999) and on estimates of CT acquisition uncertainties. A more pessimistic estimate is provided by Paganetti (2012b) who suggests values between 2.7% to 4.6% plus an additional absolute component of 1 mm when analytical pencil-beam models are used, compared to  $\pm 2.4\% + 1$  mm when using a dosimetrically more accurate Monte Carlo simulation of the dose distribution.

While these uncertainties affect the complete path of the particle, they are usually called *range errors*, since they induce the largest dose uncertainty in the vicinity of the Bragg-peak, i. e., near the range of the particle, which corresponds to the region with the largest depth-dose gradients.

An example for the distortion of the dose distribution subject to a 3.5% underestimation of the WEPL is shown in Fig. II.3. Note that such “overshoot” scenarios as in Fig. II.3c assume perfect correlation of the range uncertainty over all all pencil-beams. The validity of this assumption has been questioned by certain authors (e. g. Pflugfelder, Wilkens, and Oelfke 2008; Unkelbach et al. 2009). They suggest alternatives, in which the range errors are only perfectly correlated for pencil-beams traversing the same tissue. Such correlation models will be introduced in more detail in Section III.6 and evaluated in Chapter IV.

### II.3.1.3 Fractionation - random and systematic uncertainties

Radiation treatments are in general applied across several treatment fractions. Hence errors occurring during the daily routine compared to errors that realize once in the computed treatment plan exhibit different autocorrelations across fractions. The first kind, i. e., errors that independently occur during the daily fraction, are subject to no autocorrelation and therefore of *random* nature. *Systematic* errors are formed by the second kind, which exhibit perfect correlation through all treatment fractions. As a consequence, over the course of a fractionated treatment random un-

certainties might cancel out and become less significant compared to systematic errors.

Of course these assumptions are simplifying, yet common in academic literature. Recent scientific publications assessing the interplay of random and systematic uncertainties (like e. g. Kraan et al. 2013; Lowe et al. 2016; Park et al. 2013; Perkó et al. 2016) usually rely on respective autocorrelation assumptions over random and systematic uncertainty. As magnitude for the expected uncertainties Park et al. (2013) assumed 2 mm for a mix of treatment sites (i. e., prostate, lung, and brain cancer) and modalities (i. e., passive scattered and actively scanned beams) for systematic and random error. Perkó et al. (2016) uses similar assumptions for head-and-neck patients. Lowe et al. (2016), focusing only on thoroughly immobilized head and spinal patients, assumed much smaller uncertainties with sub-mm systematic and  $\approx 1$  mm random uncertainties. Kraan et al. (2013) simulated a variety of magnitudes ranging from 1 mm to 3 mm for oropharyngeal cancer.

#### II.3.1.4 Additional sources of uncertainties

In addition to the above discussed uncertainties, which base on geometrically static displacements and relative errors of the *WEPL*, other sources of uncertainty can be identified. These include additional inter-fractional as well as intra-fractional affects, i. e., uncertainties that span multiple fractions and uncertainties that occur during dose delivery, respectively, which are not sufficiently modeled by definition of random and systematic range or set-up uncertainties (Kraan et al. 2013; Thörnqvist et al. 2013).

For example, in-between fractions anatomical deformations might occur, i. e., *OARs* and tumor could move or shrink, or cavities could get filled or emptied. During a fraction, movement interplays with the delivery pattern of the scanning grid (Liu et al. 2012a). In particular, respiratory motion during delivery can lead to heavy distortions in the actually delivered dose distributions (Knopf et al. 2013; Ulrich et al. 2017).

Modeling of these uncertainties inherits difficulties from the respective imaging problem, as for example accurate deformable image registration and dose accumulation on these images. Also, it implicates advanced aspects on treatment delivery which are still in its early stages, as for example treatment plan adaptation using volumetric in-room imaging (Veiga et al. 2016). Consequently this thesis does not model such uncertainties, as the focus lies on more fundamental aspects of probabilistic modeling and optimization.

### II.3.2 Uncertainty propagation & mitigation

The above review of uncertainty sources and their magnitudes not yet captures the difficulty in modeling and mitigating those. While already Goitein (1985) suggested uncertainty modeling specific to the dose distribution at hand, e. g. by recalculation of dose under error scenarios, uncertainty assessment in photon therapy has been based on heuristic assumptions (Herk et al. 2000) that are derived from a static dose-cloud assumption, i. e., the dose distribution does not change its geometrical shape yet its relative location under displacements and deformations (Bortfeld, Jiang, and Rietzel 2004). It was quickly realized that the robustness of a proton treatment is inad-

equately addressed with heuristic margin concepts because the static dose cloud assumption is invalid (Albertini, Hug, and Lomax 2011; Lomax 2008a; Lomax 2008b). Consequently, the dose distribution depends on the patient anatomy at hand and requires a case-specific assessment and mitigation of those uncertainties.

### II.3.2.1 Margin recipes

To keep the familiar margin concept, i. e., the definition of a **planning target volume (PTV)** around the **clinical target volume (CTV)** to capture uncertainties, adaptation techniques have been proposed to define margins that are suitable for use with proton or other particle beams. Knopf et al. (2013), for example, suggested to compute **WEPL** variations based on a **CT** image time series (**4DCT**), and adapt the margin contour accordingly. While their work focused on intra-fractional motions, it would also be applicable to inter-fractional motion and other uncertainty sources, and could rely on sampled **WEPL** computations if no **4DCT** is available.

Such margins were suggested by Park et al. (2012) on beam specific basis. Using re-computations of **WEPL** under lateral shifts and range variations, a margin was constructed based on the local density variations. These showed significantly better robustness of **CTV** coverage (min. 94 %) under simulations of error scenarios than when using the heuristic margin concept (min. 67 %).

Mitigation of the “margin problem” is further attempted through optimization for **single-field uniform dose (SFUD)**. In **SFUD**-optimized treatment plans, the degrees of freedom are limited to separate optimization of beams, such that each each beam homogeneously irradiates the target (Lomax 1999). Such treatment plans are argued to be more robust against uncertainties when relying on a generic safety margins (Albertini, Hug, and Lomax 2010; Liu et al. 2012a; Zhu et al. 2014). However, they can be seen as a sub-set of more advanced techniques which could, in general, add **SFUD** optimization on top of their respective framework.

### II.3.2.2 Robust & stochastic optimization

The adaptation techniques discussed above may, however, be seen as tools to bypass a direct incorporation of uncertainty into the treatment plan optimization procedure. Since they partly rely on re-computations under possible error scenarios, this poses the question why not to use these scenarios directly within treatment optimization.

Over the last years, numerous approaches have been presented to incorporate uncertainties into the optimization of proton treatment plans, mainly differentiating into worst-case approaches (Chen et al. 2012; Fredriksson, Forsgren, and Hårdemark 2011; Liu et al. 2012a; Liu et al. 2012b; Lowe et al. 2017; Pflugfelder, Wilkens, and Oelfke 2008; Steitz et al. 2016), or probabilistic approaches (Unkelbach, Chan, and Bortfeld 2007; Unkelbach et al. 2009). Note that also for photon **intensity-modulated radiation therapy (IMRT)**, numerous techniques have been proposed regarding the worst-case (Bortfeld et al. 2008; Chan, Bortfeld, and Tsitsiklis 2006; Chan and Mišić 2013; Chu et al. 2005) and probabilistic (Bohoslavsky et al. 2013; Sobotta, Söhn, and Alber 2010; Unkelbach and Oelfke 2004) approaches. Such techniques may often generalize to **IMPT** and may be used with respective adaptations.

**Worst-case optimization:** The worst-case approach generally tries to minimize the implications of a pre-defined worst-case over the input uncertainties. For example, the range overshoot of 3.5 % as depicted in Fig. II.3c, might be defined as the worst-case scenario. Worst-case optimization now tries to optimize the plan to find a more robust fluence  $\tilde{w}^*$  such that the implications of the 3.5 % overshoot are decreased. Hence, a less distorted dose distribution than in Fig. II.3c may be achieved under realization of the same worst-case WEPL underestimation.

Worst-case optimization approaches differ in their definition of worst-cases and how they are included in optimization. For example, a number of worst-case scenarios  $d^s(w)$  can be defined for a fluence  $w$ , which then get optimized with a minimax approach (Fredriksson 2012; Fredriksson, Forsgren, and Hårdemark 2011):

$$\tilde{w}^* = \arg \min_w \max_s \mathcal{F}(d^s(w)). \quad (\text{II.22})$$

Hence, after optimization the worst realization over all scenarios should have reached its minimal objective value.

An alternative approach developed by Pflugfelder, Wilkens, and Oelfke (2008) and picked up by Lomax (2008a) and Lomax et al. (2004) relies on definition of a voxel-wise worst-case distribution  $d^{\text{WC}}$ .  $d^{\text{WC}}$  contains, for each voxel  $i$ , the minimum observed dose for target voxels or the maximal observed dose in OARs across the respective worst-case scenarios  $d^s$ :

$$d_i^{\text{WC}} = \begin{cases} \min d_i^s & i \in \text{target} \\ \max d_i^s & i \in \text{OAR} \end{cases}. \quad (\text{II.23})$$

Such a worst-case distribution might then be incorporated into optimization (compare Pflugfelder, Wilkens, and Oelfke 2008; Steitz et al. 2016) via a (here simplified) composite objective function

$$\tilde{\mathcal{F}}(d, d^{\text{WC}}) = \mathcal{F}(d) + p^{\text{WC}} \mathcal{F}(d^{\text{WC}}). \quad (\text{II.24})$$

In addition to the objective function value  $\mathcal{F}(d)$  evaluated on  $d$  (compare Section II.2), the function value  $\mathcal{F}(d^{\text{WC}})$  evaluated on  $d^{\text{WC}}$  is added and penalized with a worst-case penalty  $p^{\text{WC}}$ .

**Stochastic optimization:** Stochastic optimization, also referred to as probabilistic optimization (Unkelbach et al. 2009), does not rely on worst-case scenarios, but on the full probability distribution over the input uncertainties. Objectives are then optimized based on their probability distribution, which is determined by stochastic scenario sampling from the distribution over the input parameters. Unkelbach, Chan, and Bortfeld (2007) and Unkelbach et al. (2009), for example, optimize an expected value  $\mathbb{E}[\mathcal{F}]$  of the objective  $\mathcal{F}$  using stochastic gradient descent. Their work focuses on minimization of the expectation of the least-squares objective from Eq. (II.15).

Other works by Sobotta, Söhn, and Alber (2010) parametrize different objectives, for example EUD from Eq. (II.18), with a normal distribution based on stochastic estimates and then minimize the width of a respective confidence interval to achieve robustness.

**Limitations:** Both approaches suffer from specific advantages and limitations (Casiraghi, Albertini, and Lomax 2013; Fredriksson 2012). Construction of either worst-case scenarios or samples for stochastic optimization is expensive and has to be performed at each optimization step after change of the fluence vector  $w$ . Unkelbach et al. (2009), for example, report run-times of several hours for a stochastic treatment plan optimization. Further limitations concern the concepts themselves. In the worst-case approach it is not clear what physical meaning, for example, the voxel-wise worst-case distribution has. Also, it may yield overly conservative treatment plans based on the defined worst-case. On the other other side, stochastic optimization might not be conservative enough, since optimization of the expected value may not adequately ensure the robustness against worst-cases. Eventually, both worst-case and stochastic approaches depicted above will make treatment plans more robust, while however the degree of robustness is only indirectly quantifiable.

### II.3.2.3 Alternative approaches

To overcome the limitations of worst-case and stochastic optimization, different intermediate approaches have been proposed (Fredriksson 2012; Gordon and Siebers 2009; Gordon et al. 2010; Mescher, Ulrich, and Bangert 2017; Moore et al. 2009). Fredriksson (2012) for example, proposed conditional value-at-risk optimization (Rockafellar and Uryasev 1997), which allows to scale the optimization problem between expected value optimization and worst-case optimization. Gordon and Siebers (2009), Gordon et al. (2010), and Moore et al. (2009), for photon IMRT propose to optimize coverage criteria based on dose-volume coverage maps (DVCMs). DVCMs represent the cumulative probability distribution over all DVH-points, i. e., the probability  $P(v \leq \text{DVH}(\hat{d}))$  that the volume  $v$  lies below the respective DVH-point at dose  $\hat{d}$ :

$$\text{DVCM}(v, \hat{d}) = P(v \leq \text{DVH}(\hat{d})). \quad (\text{II.25})$$

From these DVCMs, one can deduct quantile dose-volume histograms ( $\alpha$ -DVHs), which correspond to lines of iso-probability  $\alpha$  within the DVCM:

$$\alpha\text{-DVH}(\hat{d}) = v \Leftrightarrow P(v \leq \text{DVH}(\hat{d})) = \alpha. \quad (\text{II.26})$$

Hence,  $\alpha$ -DVHs represent the quantile functions of the distributions over all DVH-points.

Optimizing probabilities on a DVCM or  $\alpha$ -DVH-points (e. g. similar to a DVH objective from Eq. (II.20)) now gives a more direct control over the desired robustness. Mescher, Ulrich, and Bangert (2017) extended this concept to enable DVCM constraints for IMRT for photons, giving the possibility to prescribe a certain probability to a DVH point.

The discussed approaches, however, also rely on discretely computed scenarios, which are either sampled randomly from an underlying probability distribution or placed on an evenly spaced grid. This induces problems regarding accuracy and computational efficiency for optimization. Grid-based sampling suffers from the curse of dimensionality, and thus may be only efficiently used within low-dimensional uncertainty models or as worst-case estimates. Compu-

tation of random samples allows to mitigate the problem with higher-dimensional multivariate models, but still require a lot of samples to capture tails over the probability distribution of the outcome. This is especially a problem in optimization of fractionated treatments, under which sample approaches get computationally too expensive (Kraan et al. 2013; Park et al. 2013), and simplifications have to be made (Lowe et al. 2017) for optimization.

There exists a limited amount of approaches to reduce the computational requirements of scenario sampling approaches while maintaining or improving accuracy. For example, Perkó et al. (2016) build a PCE of the probability distribution over dose using numerical integration on sparse grids. This allows to achieve high accuracy of the probability distribution over dose with few hundred samples incorporating the interplay of random and systematic errors, however only for low dimensionality of the uncertainty model. Also, a generalization to optimization is difficult, since the full expansion would need to be re-constructed as soon as the dose changes. Perkó et al. (2016) report construction times of approximately 1 h. Consequently, PCE might serve as an accurate tool for uncertainty analysis, but does not improve the situation regarding re-computations during optimization.

Sobotta, Söhn, and Alber (2012), on the other hand, propose to model the response of treatment plan metrics using Gaussian processes. They report a reduction of sample numbers by approximately one magnitude to achieve similar accuracy as with random sampling. Still, the required re-computations when the dose distribution changes during optimization, demand patience when waiting for an optimization result.

Both alternative approaches, however, still rely on sampled dose scenarios at their core, yet they are drawn in a “smarter” way. The question is if it is possible to not use scenario samples at all. That means a mathematical mapping from the input, i. e., the fluence generating the dose distribution and the input uncertainties, to the output, i. e., uncertainty in dose and, more general, treatment plan quality, has to be found. An approach introduced by Bangert, Hennig, and Oelfke (2013) follows this idea, and since it forms the backbone of this thesis’ work, will be thoroughly described in the following chapter.



## The Analytical Probabilistic Modeling Framework

This chapter describes *Analytical Probabilistic Modeling (APM)* for radiation therapy, a concept introduced by Bangert, Hennig, and Oelfke (2013). Instead of using statistics from scenario sampling, *APM* uses approximations to a pencil-beam dose calculation algorithm in order to derive closed-form approximations of expectation value and covariance of the dose distribution subject to uncertainties over treatment parameters. Since *APM* lays the foundations for this dissertation, this chapter recaps the derivations by Bangert, Hennig, and Oelfke (2013) and adjusts them to the notation used throughout this manuscript. Note that during the preparation of this thesis, the in the following described functionalities of *APM* were published by the author and Hans-Peter Wieser and Mark Bangert within the *APMToolbox*.

### III.1 Overview

At  $p$  given uncertain parameters/coordinates, denoted as random vector  $\mathbf{R} \in \mathbb{R}^p$ , the  $\nu$ -th moment of the probability distribution over dose  $d(\mathbf{R})$ , where  $f_{\mathbf{R}}$  is the probability density over  $\mathbf{R}$ , can be obtained by solving the volume integral

$$\mathbb{E} [d(\mathbf{R})^\nu] = \int_{\mathbb{R}^p} d(\mathbf{r})^\nu f_{\mathbf{R}}(\mathbf{r}) d\mathbf{r}. \quad (\text{III.1})$$

Attempting to solve Eq. (III.1) reveals multiple pitfalls: First of all, the random parameter set  $\mathbf{R}$  is composed of non-trivial quantities, i. e., *WEPL*, coordinates on a deformable anatomy etc. Thus its uncertainty model is, in general, difficult to define and quantify (compare Section II.3). Furthermore, the dose  $d(\mathbf{r})$  is a function which is difficult to represent analytically, and usually computed numerically based on tabulated data values (compare Section II.1.2). Hence, to find an analytical solution, a representation of the parameter space needs to be found in which an uncertainty model  $f_{\mathbf{R}}(\mathbf{r})$  and an analytically tractable representation of dose  $d(\mathbf{r})$  can be defined.

Bangert, Hennig, and Oelfke (2013) approach the integral directly in the pencil-beam coordinate frame used for computation of the dose influence matrix  $D$  (compare Section II.1.2) with discrete spatial coordinates, i. e., voxel indices. Their uncertainty model reduces to lateral shifts of pencil-beams, described by the random vectors  $\Delta^{X/Y} \in \mathbb{R}^{n_B}$ , and offsets to the computed *WEPL*  $\Delta^Z \in \mathbb{R}^{n_B}$ , where  $n_B$  is the number of pencil beams. The uncertainty over those offsets is described

by a multivariate distribution

$$f_{\Delta^{x/y/z}}(\Delta^x, \Delta^y, \Delta^z) = f_{\Delta^x}(\Delta^x) f_{\Delta^y}(\Delta^y) f_{\Delta^z}(\Delta^z) \quad (\text{III.2})$$

factorizing into the spatial components.

In this parametrization, using  $\mathbf{d} = D\mathbf{w}$ , integral (III.1) can be written as

$$\begin{aligned} \mathbb{E} \left[ \left( D(\Delta^x, \Delta^y, \Delta^z) \mathbf{w} \right)^\nu \right] \\ = \iiint_{\mathbb{R}^{3 \times n_B}} (D(\Delta^x, \Delta^y, \Delta^z) \mathbf{w})^\nu f_{\Delta^{x/y/z}}(\Delta^x, \Delta^y, \Delta^z) d\Delta^x d\Delta^y d\Delta^z. \end{aligned} \quad (\text{III.3})$$

Following the spatial factorization from Eqs. (II.7) and (III.2), integral (III.3) may be solved for each spatial dimension independently.

A closed-form solution now requires an approximation of  $D(\Delta^x, \Delta^y, \Delta^z)$  to be integrable against the probability density  $f_{\Delta^{x/y/z}}(\Delta^x, \Delta^y, \Delta^z)$ . Bangert, Hennig, and Oelfke (2013) chose a multivariate normal distribution for  $f_{\Delta^{x/y/z}}$ , i. e.,

$$f_{\Delta^{x/y/z}}(\Delta^x, \Delta^y, \Delta^z) = \mathcal{N}(\Delta^x; \mathbf{0}, \Sigma^x) \mathcal{N}(\Delta^y; \mathbf{0}, \Sigma^y) \mathcal{N}(\Delta^z; \mathbf{0}, \Sigma^z) \quad (\text{III.4})$$

with zero means and covariance matrices  $\Sigma^x, \Sigma^y, \Sigma^z \in \mathbb{R}^{n_B \times n_B}$ . Then they constructed  $D(\Delta^x, \Delta^y, \Delta^z)$  completely from (superpositions of) Gaussian functions along each spatial dimension, allowing the use of Gaussian algebra (compare Appendix A.1.4) to solve integral (III.3).

Section III.2 will explain this approximation, and Sections III.3 and III.4 will summarize the solutions of Eq. (III.3) for the first two moments  $\nu = 1$  and  $\nu = 2$ , respectively. Section III.6 will then introduce uncertainty models to parametrize  $f_{\Delta^{x/y/z}}$ .

## III.2 The Gaussian pencil-beam dose model

Considering the lateral spread  $L^{x/y}$  of a pencil-beam, analytical pencil-beam dose calculation algorithms model  $L^{x/y}$  as (a superposition of) Gaussians anyway (compare Eq. (II.8)). However, the depth-dose component  $Z$ , i. e., the Bragg-curve, is often just stored as a LUT of depth-dose values at indexed depths, where such LUTs are then computed for set of energies that can be provided by the accelerator (see Section II.1.2). While analytical models for the Bragg-curve do exist (Bortfeld 1997), their analytical integration against Gaussian probabilities is not straightforward, especially when higher moments ( $\nu > 1$ ) shall be computed. For this reason Bangert, Hennig, and Oelfke (2013) suggested to model the Bragg-curve with a superposition of Gaussian functions up to desired accuracy, allowing them to solve all necessary integrals analytically with the same principle along all three spatial dimensions.

For the Gaussian model of the Bragg-curve, Bangert, Hennig, and Oelfke (2014) empirically determined that with ten Gaussian components sufficient accuracy is obtained; then they “observe less than 0.3 % average deviation of the Gaussian parametrization [i. e., Eq. (III.5)] from the [...] proton depth dose curve for proton ranges  $R \leq 35$  cm.” The resulting depth-dose contribution

$Z_{ij}$  of a pencil-beam  $j$  to voxel  $i$  using a superposition of ten Gaussian components is then given by

$$Z_{ij} = \sum_{k=1}^{10} \omega_{jk} \mathcal{N} \left( z_{ij}; \mu_{jk}^z, \delta_{jk}^2 \right), \quad (\text{III.5})$$

where  $z_{ij}$  is the WEPL from the source of pencil-beam  $j$  to voxel  $i$ .  $\mu_{jk}^z$  and  $\delta_{jk}^2$  are mean and variance of the  $k$ -th Gaussian component weighted with  $\omega_{jk}$ , which are obtained by a fit to reference pencil-beam base data.

This approximations is accompanied by the Gaussian lateral components

$$L_{ij}^x = \mathcal{N} \left( x_{ij}; \mu_j^x, \lambda_{ij}^2 \right) \quad (\text{III.6})$$

as in Eq. (II.8), where  $x_{ij}$  is the geometrical distance of voxel  $i$  to the central beam-axis along the  $x$ -dimension in BEV, and  $\mu_j^x$  refers to the respective lateral position of the central pencil-beam ray.  $\lambda_{ij}^2$  describes the broadening of the pencil-beam  $j$  in depth and is assumed to be the same along both lateral coordinates  $x$  and  $y$ . Consequently, it increases with the respective WEPL  $z_{ij}$ . Please note Eq. (III.6) only depicts lateral profiles shaped by single Gaussian components. However, since all following computations generalize to superpositions of Gaussian components as used in Eq. (III.5), also multiple Gaussian components could be used to provide a more accurate lateral beam model (Bellinzona et al. 2015).

Using the Gaussian model, the random dependence of the dose influence  $D(\Delta^x, \Delta^y, \Delta^z)$  on the spatial offsets  $\Delta^{x/y/z}$  can now be easily incorporated into all spatial components from Eqs. (III.5) and (III.6) as

$$Z_{ij}(\Delta_j^z) = \sum_k \omega_{jk} \mathcal{N} \left( z_{ij} + \Delta_j^z; \mu_{jk}^z, \delta_{jk}^2 \right), \quad (\text{III.7})$$

$$L_{ij}^x(\Delta_j^x) = \mathcal{N} \left( x_{ij} + \Delta_j^x; \mu_j^x, \lambda_{ij}^2 \right). \quad (\text{III.8})$$

The computations for  $L^y$  are analogous to Eq. (III.8).

### III.3 Expectation value

Computation of the expectation value requires the analytical evaluation of Eq. (III.3) for  $\nu = 1$ , i. e., computation of the first raw moment. For the lateral component, this results in

$$\begin{aligned} \mathcal{L}_{ij}^x &\stackrel{!}{=} \mathbb{E} \left[ L_{ij}^x(\Delta_j^x) \right] = \int_{-\infty}^{\infty} L_{ij}^x(\Delta_j^x) f_{\Delta_j^x}(\Delta_j^x) d\Delta_j^x \\ &= \int_{-\infty}^{\infty} \mathcal{N} \left( x_{ij} + \Delta_j^x; \mu_j^x, \lambda_{ij}^2 \right) \mathcal{N} \left( \Delta_j^x; 0, \Sigma_{jj}^x \right) d\Delta_j^x \\ &= \mathcal{N} \left( x_{ij}; \mu_j^x, \lambda_{ij}^2 + \Sigma_{jj}^x \right) \int_{-\infty}^{\infty} \mathcal{N} \left( \Delta_j^x; \tilde{\mu}, \tilde{\sigma}^2 \right) d\Delta_j^x \end{aligned} \quad (\text{III.9})$$

where, in step two, Eq. (A.12) was applied.

The expectation value  $\mathcal{Z}_{ij} = \mathbb{E}[Z_{ij}]$  is computed analogously for all weighted components reading

$$\mathcal{Z}_{ij} = \sum_{k=1}^{10} \omega_{jk} \mathcal{N} \left( z_{ij}; \mu_{jk}^z, \delta_{jk}^2 + \Sigma_{jj}^z \right) \quad (\text{III.10})$$

This yields an expected dose influence matrix  $\mathcal{D} \in \mathbb{R}^{n_V \times n_B}$  with elements

$$\mathcal{D}_{ij} = \mathbb{E} [D_{ij}] = \mathcal{L}_{ij}^x \mathcal{L}_{ij}^y \mathcal{Z}_{ij}. \quad (\text{III.11})$$

Hence,  $\mathcal{D}$  is of similar structure as  $D$ , with the difference of slightly broader Gaussian components. Conveniently, this enables use of similar algorithms for expected as well as nominal dose computations.

For instance, the expected dose influence matrix  $\mathcal{D}$  maps the pencil-beam weights  $w$  to the expected dose vector  $\mathbb{E}[\mathbf{d}]$  via the linear transformation

$$\mathbb{E} [\mathbf{d}] = \mathcal{D}w, \quad (\text{III.12})$$

analogous to the nominal dose  $\mathbf{d} = Dw$ .

## III.4 Covariance

Computation of the second moment follows a similar principle like computation of the expectation value explained in Section III.3. Solving Eq. (III.3) for  $\nu = 2$  gives the second raw moment  $\mathbb{E}[D_{ij}^2]$  of a dose influence matrix element  $D_{ij}$ , allowing the computation of the second central moment, i. e., the variance

$$\text{Var} [D_{ij}] = \mathbb{E} [D_{ij}^2] - \mathbb{E} [D_{ij}]^2 = \mathbb{E} [D_{ij}^2] - \mathcal{D}_{ij}^2 \quad (\text{III.13})$$

using the results from Eq. (III.11). To generalize computation to the covariance, however, instead of Eq. (III.13) the derivations to come will find expressions for  $\mathbb{E}[D_{ij}D_{lm}]$  to obtain

$$\text{Cov} [D_{ij}, D_{lm}] = \mathbb{E} [D_{ij}D_{lm}] - \mathbb{E} [D_{ij}] \mathbb{E} [D_{lm}] = \mathbb{E} [D_{ij}D_{lm}] - \mathcal{D}_{ij}\mathcal{D}_{lm}. \quad (\text{III.14})$$

Since spatial independence had been assumed (compare Eq. (III.4)), computation of  $\mathbb{E}[D_{ij}D_{lm}]$  can again be factorized into its spatial components, requiring an integral against the bivariate

marginal distribution  $f_{\Delta_{j;m}^x}$  describing the uncertain positioning of pencil-beams  $j$  and  $m$ :

$$\begin{aligned} Y_{ijlm}^x &\stackrel{!}{=} \mathbb{E} \left[ L_{ij} \left( \Delta_j^x \right) L_{lm} \left( \Delta_m^x \right) \right] \\ &= \int_{\mathbb{R}^2} L_{ij}^x \left( \Delta_j^x \right) L_{lm}^x \left( \Delta_m^x \right) f_{\Delta_{j;m}^x} \left( \Delta_{j;m}^x \right) d\Delta_{j;m}^x \\ &= \int_{\mathbb{R}^2} \mathcal{N} \left( \mathbf{x}_{(ij);(lm)} + \Delta_{j;m}^x; \boldsymbol{\mu}_{j;m}^x, \text{diag} \left( \lambda_{ij}^2, \lambda_{lm}^2 \right) \right) \mathcal{N} \left( \Delta_{j;m}^x; \mathbf{0}_2, \Sigma_{j;m}^x \right) d\Delta_{j;m}^x \end{aligned} \quad (\text{III.15})$$

Similar to Eq. (III.9), the integrand was expressed as a product of two multivariate normal distributions—this time with two dimensions. Notation for marginals follows definition at the beginning of this thesis, i. e.,  $\mathbf{x}_{(ij);(lm)}$  corresponds to the vector with elements  $(x_{ij}, x_{lm})^T$ ,  $\boldsymbol{\mu}_{j;m} = (\mu_j, \mu_m)^T$  etc. After carrying out the product with Eq. (A.12), the integral can be solved and a bivariate normal distribution remains:

$$Y_{ijlm}^x = \mathcal{N} \left( \mathbf{x}_{(ij);(lm)}; \boldsymbol{\mu}_{j;m}^x, \text{diag} \left( \lambda_{ij}^2, \lambda_{lm}^2 \right) + \Sigma_{j;m}^x \right). \quad (\text{III.16})$$

The computations for the depth component are similar, yet more cumbersome due to the superimposing Gaussian components, and yield

$$\begin{aligned} \Xi_{ijlm} &\stackrel{!}{=} \mathbb{E} \left[ Z_{ij} \left( \Delta_j^z \right) Z_{lm} \left( \Delta_m^z \right) \right] \\ &= \sum_{kn} \omega_{jk} \omega_{wn} \mathcal{N} \left( \mathbf{z}_{(ij);(lm)}; \boldsymbol{\mu}_{j;m}^z, \text{diag} \left( \delta_{jk}^2, \delta_{mn}^2 \right) + \Sigma_{j;m}^z \right). \end{aligned} \quad (\text{III.17})$$

Combining all spatial components, one can obtain a four-dimensional *covariance influence “tensor”*<sup>1</sup>  $\mathcal{C} \in \mathbb{R}^{n_V \times n_B \times n_V \times n_B}$

$$\mathcal{C}_{ijlm} = \underbrace{Y_{ijlm}^x Y_{ijlm}^y \Xi_{ijlm}}_{\mathcal{C}_{ijlm}} - \mathcal{D}_{ij} \mathcal{D}_{lm} \quad (\text{III.18})$$

in analogy to Eq. (III.11).  $\tilde{\mathcal{C}}_{ijlm}$  will be called the “raw contribution” to  $\mathcal{C}_{ijlm}$ , as it describes the influence on the raw second moment of dose.

Hence, a tensor element  $\mathcal{C}_{ijlm}$  can be understood as the contribution of the marginal bivariate probability distribution—dependent on the uncertainty model and respective correlation assumptions—over the coordinates of pencil-beams  $j$  and  $m$  to the dose covariance in voxels  $i$  and  $l$ . Contracting  $\mathcal{C}$  two times with the pencil-beam weight vector  $\boldsymbol{w}$  therefore maps them to the covariance in dose  $\boldsymbol{d}$ , giving an element of the dose covariance matrix  $\Sigma^{\boldsymbol{d}}$  as

$$\Sigma_{il}^{\boldsymbol{d}} = \sum_{jm} \mathcal{C}_{ijlm} w_j w_m. \quad (\text{III.19})$$

<sup>1</sup>The term “tensor” has been put in quotation marks, since  $\mathcal{C}$  is quite different from common physical tensors taking geometrical vectors in space or spacetime as arguments. Since the indices  $i$  and  $l$  denote points in space for which covariance shall be computed,  $\mathcal{C}$  could be more accurately interpreted as a tensor field of second order (0,2)-tensors on the domain of a dose covariance function  $k_{\boldsymbol{d}}(\boldsymbol{x}, \boldsymbol{x}') = \mathcal{C}_{mj}(\boldsymbol{x}, \boldsymbol{x}') w^j w^m$  (using Einstein notation). Similar,  $\mathcal{D}$  and  $\mathcal{D}$  can be understood as covector fields  $D_j(\boldsymbol{x})$  and  $\mathcal{D}_j(\boldsymbol{x})$ , respectively. Denoting  $\mathcal{C}$  as a fourth order tensor does, however, allow connection of tensor operations to quantities with physical meaning.

(a) lateral profile

(b) depth profile

(This figure is animated in the electronic PDF version of this thesis.)

**Figure III.1:** Visualization of the dose approximation and moment computation (expectation value and standard deviation) for exemplary pencil-beam components  $L$  and  $Z$ . (a) illustrates the lateral component under uncertainty, while (b) displays the Bragg-curve and the respective behavior of the expectation value of the ten Gaussian components shown as gray lines.

While the variance is now given as the diagonal entries of  $\Sigma^d$ , it can also be directly computed as

$$\text{Var} [d_i] = \sum_{jm} \mathcal{C}_{ijim} w_j w_m, \quad (\text{III.20})$$

without explicit computation of elements  $i \neq l$ , which also allows designation of a sub-tensor  $\mathcal{V}_{ijm} = \mathcal{C}_{ijim}$  with  $\mathcal{V} \in \mathbb{R}^{n_V \times n_B \times n_B}$  as *variance influence tensor*.

In Fig. III.1 results of the computation of expectation value and standard deviation of the spatial components with APM are illustrated based on a single pencil-beam.

## III.5 Optimization

Bangert, Hennig, and Oelfke (2013) also demonstrated feasibility of probabilistic optimization, i. e., optimization of the expectation value of an objective function  $\mathcal{F}$ , to find an optimal fluence  $w^*$ :

$$\begin{aligned} w^* = \arg \min_w \quad & \mathbb{E} [\mathcal{F} (d(w))] \\ \text{subject to} \quad & w \geq \mathbf{0} \end{aligned} \quad (\text{III.21})$$

The derivations of Bangert, Hennig, and Oelfke (2013) are restricted to the expectation value of the penalized least-squares objective  $\mathbb{E}[\mathcal{F}^{\text{LS}}]$  from Eq. (II.16). Assuming  $d - d^*$  follows a multivariate normal distribution,  $\mathcal{F}^{\text{LS}}$  follows a generalized  $\chi^2$ -distribution with expectation value

$$\mathbb{E} \left[ \mathcal{F}^{\text{LS}}(\mathbf{d}(\mathbf{w})) \right] = \text{tr} \left( P \Sigma^d(\mathbf{w}) \right) + (\mathbb{E}[\mathbf{d}(\mathbf{w})] - \mathbf{d}^*)^T P (\mathbb{E}[\mathbf{d}(\mathbf{w})] - \mathbf{d}^*), \quad (\text{III.22})$$

where  $P$  is the diagonal voxel penalty matrix from Eq. (II.16).

Using the expected dose influence matrix  $\mathcal{D}$  derived in Section III.3, evaluation of the penalized squared-deviation term  $(\mathbb{E}[\mathbf{d}(\mathbf{w})] - \mathbf{d}^*)^T P (\mathbb{E}[\mathbf{d}(\mathbf{w})] - \mathbf{d}^*)$  is of the same computational complexity as evaluation of Eq. (II.16). This is, at first glance, not the case for  $\text{tr}(P \Sigma^d(\mathbf{w}))$  in Eq. (III.22). Yet Bangert, Hennig, and Oelfke (2013) pointed out that it may be rewritten as

$$\begin{aligned} \text{tr} \left( P \Sigma^d(\mathbf{w}) \right) &= \sum_i p_i \sum_{jm} \mathcal{C}_{ijlm} w_j w_m \\ &= \sum_{jm} w_j w_m \underbrace{\sum_i p_i \mathcal{C}_{ijim}}_{\stackrel{!}{=} \Omega_{jm}} \\ &= \mathbf{w}^T \Omega \mathbf{w} \end{aligned} \quad (\text{III.23})$$

for optimization of problem (III.21) with a positive-definite (since  $\text{Var}[\mathbf{d}] > \mathbf{0}$ ) matrix  $\Omega \in \mathbb{R}^{n_B \times n_B}$ .  $\Omega$  can be interpreted as the penalized integral variance induced by co-varying pencil-beam combinations. Derivation of the gradient w. r. t.  $\mathbf{w}$  is straightforward:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathbb{E} \left[ \mathcal{F}^{\text{LS}} \right] &= \nabla_{\mathbf{w}} (\mathbb{E}[\mathbf{d}(\mathbf{w})] - \mathbf{d}^*)^T P (\mathbb{E}[\mathbf{d}(\mathbf{w})] - \mathbf{d}^*) + \nabla_{\mathbf{w}} \text{tr} \left( P \Sigma^d \right) \\ &= 2\mathcal{D}^T P (\mathcal{D}\mathbf{w} - \mathbf{d}^*) + 2\Omega \mathbf{w}. \end{aligned} \quad (\text{III.24})$$

Hence precomputation of  $\Omega$  allows to avoid storage of  $\mathcal{C}$  during optimization which in turn avoids re-computation of  $\Sigma^d$  due to changes in  $\mathbf{w}$  in-between successive iterations of the numerical optimization routine.

## III.6 Uncertainty model

Section II.3.1 explained that the definition of uncertainty models in radiation therapy is not trivial. For range uncertainties, literature statements vary and give values from approximately 2% to 5% (Lomax 2008b; Paganetti 2012b), and for set-up uncertainties values depend on the applied heuristics (see e. g. Lomax 2008a; Lowe et al. 2016). Furthermore, validity of correlation assumptions of the respective uncertainties are debated (Unkelbach et al. 2009).

In general, APM is able to include a variety of uncertainty assumptions, and substantial amounts of efforts could be put in an accurate definition of those. This thesis, however, does not focus on finding the best suited uncertainty model, since it demonstrates first applications of APM and further extensions. For this scope, the thesis relies on suggestions from literature to be represented within the APM framework.

### III.6.1 Magnitude of the uncertainties

Literature statements (compare Section II.3.1) were boiled down to using a relative range uncertainty of a few percent, and set-up uncertainties within few millimeters. Usually, magnitude of the uncertainties is assumed to be constant across all pencil-beams. That is, all pencil-beams are subject to range uncertainties of similar relative magnitude and to set-up uncertainties with similar absolute magnitude. Relativeness of range error can be justified with the particle traversing the tissue, where uncertainty WEPL piles up along the particle's path. The absolute lateral uncertainty corresponds to absolute displacements.

Assumptions used in Fig. III.1 reflect their magnitude, where relative range uncertainty is given by  $\sigma_z = \pm 3.5\%$  of the WEPL, and a lateral set-up uncertainty of  $\sigma_{x/y} = \pm 2.5$  mm was assumed. Note that for the depth component, Eq. (III.10) and Eq. (III.17) do not explicitly model a relative uncertainty. The respective adaptations to the equations are, however, straightforward and omitted.

In addition to the magnitude, assumptions over the correlations across pencil-beams and different fractions have to be made, as discussed in the following sections.

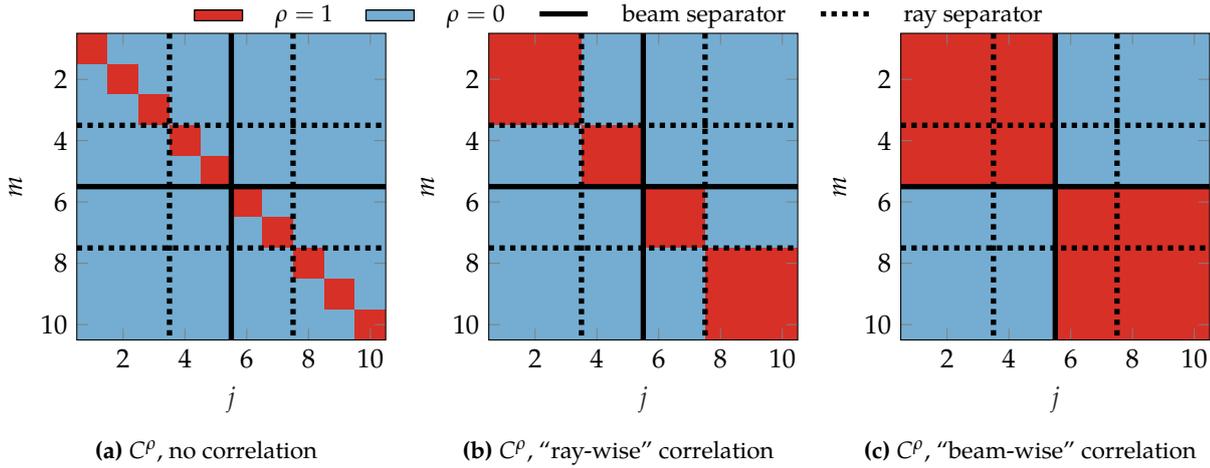
### III.6.2 Correlation model

The moment computations derived with APM allow arbitrary correlation models for the multivariate normal density in Eq. (III.4). Moreover, the computational complexity does not depend on the correlation model, yet optimizations are possible for certain patterns as explained in Chapter IV. For example, if all spots of the same beam are expected to be shifted similarly in an error scenario, this can be encoded in block structured covariance matrices  $\Sigma^{\Delta x/y}$ . Alternatively, other proposed uncertainty models (Bangert, Hennig, and Oelfke 2013; Pflugfelder, Wilkens, and Oelfke 2008; Unkelbach et al. 2009) can be assumed. Examples for correlation models are illustrated based on their respective correlation matrix  $C^p$  in Fig. III.2.

Note that even singular covariance matrices (i. e., perfect correlation) can be defined, since in the resulting Gaussians from Eqs. (III.16) and (III.17) the respective marginal matrices are added to the diagonal matrix composed of the width of the Gaussian components of the pencil-beam algorithm. While this would, in theory, a priori require a more careful approach on handling the degenerate multivariate distribution within the Gaussian integrals (i. e., a suitable subspace measure with existing density has to be defined), this manuscript, as well as Bangert, Hennig, and Oelfke (2013), omits detailed discussion of the respective computations leading to similar results. Furthermore, the validity of the expressions (III.16) and (III.17) for singular covariance matrices will be empirically tested through validation of the APM method in Chapter IV.

The APM framework would also allow dropping the spatial factorization in Eq. (III.4) to implement correlation models across spatial dimensions. This facilitates, for example, more complex displacement patterns that do not neatly factorize in the BEV. The Computations in Sections III.3 and III.4 would then become integrals against three-dimensional Gaussians and six-dimensional Gaussians, respectively. While this will increase computational cost due to the more challenging covariance matrix inversion compared to the trivial one- and two-dimensional case, the general

algebraic formulation remains.



**Figure III.2:** Exemplary correlation matrices of proposed correlation models for 10 pencil beams along a single spatial dimension. The solid line separates pencil-beams from two different beams, e. g. pencil-beams 1 to 5 come from gantry angle  $\alpha$  and pencil-beams 6 to 10 come from gantry angle  $\beta$ . The dashed line separates pencil-beams into rays, i. e., the respective pencil-beams share the same lateral positioning yet different energies (compare Fig. II.2). (a) corresponds to no correlation at all, and is always the correlation structure of the  $\Sigma^{\text{uncorr}}$  covariance matrix used for modeling fractionation. (b) depicts a "ray-wise" correlation matrix, in which all pencil-beam lying on the same lateral position, i. e., the same ray, are perfectly correlated (Unkelbach et al. 2009, range correlation "Model 3"; Pflugfelder, Wilkens, and Oelfke 2008, range correlation model). In (c) all pencil-beams applied within the same beam are perfectly correlated (typical lateral correlation model, also range correlation from Unkelbach et al. 2009, "Model 2").

### III.6.3 Fractionation

In fractionated treatments, the input uncertainties are commonly separated into random and systematic components (compare Section II.3.1.3). A systematic error realizes once throughout a fractionated treatment, i. e., it is the same for all applied fractions. This could, for example, be miscalculated WEPLs on a single treatment planning CT. Random errors, e. g. offsets in the daily patient position, realize independently per fraction. Bangert, Hennig, and Oelfke (2013) prototyped the incorporation of fractionation effects on dose uncertainty subject to such random and systematic errors uncertainties within the APM framework.

In principle, APM could incorporate all kinds of linear correlation assumptions connecting uncertainties over fractions. However, from the definition of random and systematic errors, random errors can be modeled as uncorrelated over all fractions, while systematic errors are then perfectly correlated over all fractions. As a consequence, the APM framework then allows for a full incorporation of random and systematic errors at limited computational overhead.

For example, in a fractionated treatment with  $n_f$  fractions the lateral component  $L$  accumulates the components  $L_f$  of each fraction  $f$ , which is subject to systematic and random errors realizing as  $\Delta^{\text{sys}}$  and  $\Delta_f^{\text{rand}}$ , respectively:

$$L = \sum_{f=1}^{n_f} \frac{1}{n_f} L_f \left( \Delta^{\text{sys}} + \Delta_f^{\text{rand}} \right). \quad (\text{III.25})$$

That is, in each fraction,  $L_f$  depends on the same the systematic displacement  $\Delta^{\text{sys}}$  and an independently realized  $\Delta_f^{\text{rand}}$ . Note that for clarity the superscript of  $L$  denoting the lateral dimension has been dropped.

Assuming the probability distribution over  $\Delta_f^{\text{rand}}$  is the same for each fraction, Eq. (III.9) can be extended to model systematic and random uncertainties through the respective covariance matrices  $\Sigma^{\text{sys}}$  and  $\Sigma^{\text{rand}}$  over the lateral position. For the lateral component this becomes

$$\mathcal{L}_{ij} = \mathcal{N} \left( x_{ij}; \mu_j, \lambda_{ij}^2 + \Sigma_{jj}^{\text{sys}} + \Sigma_{jj}^{\text{rand}} \right), \quad (\text{III.26})$$

while resulting in

$$\mathcal{Z}_{ij} = \sum_{k=1}^{10} \omega_{jk} \mathcal{N} \left( z_{ij}; \mu_{jk}^z, \delta_{jk}^2 + \Sigma_{jj}^{z,\text{sys}} + \Sigma_{jj}^{z,\text{rand}} \right), \quad (\text{III.27})$$

for the depth component.

For the covariance influence, where correlations between pencil-beams come into play, two different cases have to be evaluated for handling random uncertainties. Within a fraction, the assumed spatial correlation model for our random uncertainties is valid (e. g. “beam-wise” correlations as in Fig. III.2c), and the respective covariance matrix will be denoted as  $\Sigma^{\text{corr}}$ . Across fractions, the random uncertainties over the pencil-beams are uncorrelated, and thus correspond to a diagonal matrix  $\Sigma^{\text{uncorr}} = \text{diag}(\Sigma^{\text{corr}})$  (compare Fig. III.2a). Using  $\Sigma^{\text{corr}}$  and  $\Sigma^{\text{uncorr}}$ , the lateral contribution to the covariance influence from Eq. (III.16) can be now written as

$$\begin{aligned} Y_{ijlm} = & \frac{1}{n_f} \left[ \underbrace{\mathcal{N} \left( \mathbf{x}_{(ij);(lm)}; \boldsymbol{\mu}_{j;m}, \text{diag}(\lambda_{ij}^2, \lambda_{lm}^2) + \Sigma_{j;m}^{\text{corr}} + \Sigma_{j;m}^{\text{sys}} \right)}_{Y_{ijlm}^{\text{corr}}} \right. \\ & \left. + (n_f - 1) \underbrace{\mathcal{N} \left( \mathbf{x}_{(ij);(lm)}; \boldsymbol{\mu}_{j;m}, \text{diag}(\lambda_{ij}^2, \lambda_{lm}^2) + \Sigma_{j;m}^{\text{uncorr}} + \Sigma_{j;m}^{\text{sys}} \right)}_{Y_{ijlm}^{\text{uncorr}}} \right]. \end{aligned} \quad (\text{III.28})$$

Intuitively,  $Y_{ijlm}^{\text{corr}}$  describes the influence of uncertainties on the raw moment assuming a single fraction, while  $Y_{ijlm}^{\text{uncorr}}$  corrects for fraction number dependent effects as the canceling out of independent error scenarios. For the second lateral component as well as the depth component, the computations are analogous and therefore omitted.

Equations (III.26) and (III.28) demonstrate that APM enables computation of the first two moments of the probability distribution over dose for *any* fraction number  $n_f$  within constant time complexity, incorporating the non-trivial interplay of random and systematic errors. Chapter V will elaborate on the implications of these possibilities within the APM framework, and extend functionality.

Please bear in mind that the here demonstrated uncertainty computations for fractionated treatments are only valid for the physical dose, which is sufficient for the scope of this manuscript. For extensions of APM to uncertainties in biological dose or effect, the reader shall be referenced to Wieser et al. (2017a).

# Analytical Probabilistic Modeling: from Concept to Application

This chapter discusses the implementation of the *APM* concept developed by Bangert, Hennig, and Oelfke (2013) and its application within a realistic treatment planning workflow on three-dimensional patient datasets. Besides a general proof of work on patient data, it focuses upon computational performance aspects as well as validation of the moment computations described in Chapter III.

## IV.1 Overview

The analytical computations of the *APM* framework laid out in Chapter III operate at the heart of dose calculation. *APM* transforms conventional dose calculation for a single scenario to a probabilistic dose calculation that simultaneously considers a continuum of all possible scenarios as specified by a Gaussian probability density. This imposes that for application of *APM* to patient data, a full pencil-beam dose calculation algorithm has to be implemented—or at least adapted—following the exclusively Gaussian dose model explained in Section III.2. Furthermore, the newly conceptualized moment computations from Sections III.3 and III.4 require a new kind of input, i. e., a multivariate uncertainty model. They increase the computational complexity from linearity in number of voxels  $n_V$  and number of pencil-beams  $n_B$  for nominal and expected dose calculation, i. e.,  $\mathcal{O}(n_V n_B)$ , to  $\mathcal{O}(n_V^2 n_B^2)$  for covariance (and to  $\mathcal{O}(n_V n_B^2)$  for variance) calculations.

These requirements prohibit the use of a closed source, commercial treatment planning system for implementation of a full-fledged 3D *APM* prototype. In such commercial systems, internal dose computation routines are not accessible, and interfacing often has to be done through script interfaces. Additionally, since the field of particle therapy is comparably new in radiation therapy, access to manipulable source code is scarce. Only few projects, as for example the ongoing developments in the MATLAB based *matRad* (Wieser et al. 2017b) or latest activities regarding the C++ toolkit “SlicerRT” (Desplanques 2015; Pinter et al. 2012) steer in the direction of open-source treatment planning with particles.

MATLAB, whilst well suited for prototyping and optimized for standard linear algebra prob-

lems arising in conventional treatment planning, was ruled out as candidate for efficient implementation of *APM*. One of the reasons was the increased dimensional complexity of computations required by *APM*. Usage of SlicerRT (which at the time investigated also did not include a full-fledged proton dose calculation) was overruled by the advantages of using *MITK* (Nolden et al. 2013; Wolf et al. 2005), which as in-house developed open-source image-processing software facilitated shorter communication paths regarding development support. Since some *RT* development for *MITK* was already going on its branch “*DIPP*”, the decision was made to use *DIPP* / *MITK* as a backbone for the development of *APM*’s algorithms. Developments were bundled in a treatment planning plugin called *MITKrad*, with *MATLAB* and *matRad* aiding in prototyping and data analysis.

Due to the expertise available through *matRad*, the implementation of the necessary background algorithms, i. e., conventional pencil-beam dose calculation, ray-casting, optimization etc., was comparably straightforward apart from programming technicalities. Unlike the dose calculation algorithms of *matRad*, the corresponding twin implementations in *MITKrad* have not been thoroughly benchmarked against a validated treatment planning system as this was not the scientific scope of this thesis. Given substantial overlap between the two implementations, however, this can be made up for in the future.

Validation of a novel probabilistic dose calculation engine like *APM*, however, is a non-trivial task. While Bangert, Hennig, and Oelfke (2014) already quantified the *APM* pencil-beam algorithm to be accurate compared to conventional pencil-beam dose calculation, this does not imply accuracy of the resulting expectation value and covariance of dose, even under the assumption that the pencil-beam dose computation accurately models reality. This is mainly due to the non-trivial relationship between *WEPL* changes under lateral shifts.

Further, no academic let alone clinical protocol exists to validate probabilistic dose calculations, apart from single academic attempts on quantifying the accuracy of dose uncertainty estimates. Casiraghi, Albertini, and Lomax (2013), for example, benchmark the accuracy of a voxel-wise worst-case estimate from few shifts with selected random samples. Perkó et al. (2016) explicitly validate samples from their *PCE* model with a  $\gamma$ -analysis (see Section IV.2.3.3). Due to being based on dose scenarios, these approaches are, however, specific to the method and do not cover *APM*’s computational capabilities which do not rely on scenario samples.

Therefore, this work proposes a method to directly evaluate the distributions of expectation value and standard deviation, based on a “ground truth” obtained through sample statistics calculated from a large number of random samples of the input uncertainty space. *APM*’s computations within *MITKrad* are evaluated against this benchmark, in comparison with weighted statistics obtained through common grid sampling approaches.

In addition, this chapter features an analysis of the performance of the implemented algorithms. For computation of the standard deviation with *APM*, an additional *GPGPU* approach is proposed. Furthermore, approximation techniques are conceptualized and evaluated. These allow to explicitly explore the trade-off between accuracy and computational speed within the *APM* framework, both for probabilistic treatment plan analysis and probabilistic treatment plan optimization.

The remainder of this chapter is organized as follows. Section IV.2 will introduce methodology, by first explaining the MITKrad development and the APM algorithms, followed by a description of the validation benchmark and the approximation techniques. A proof-of-work will be presented in Section IV.3, together with quantitative validation as well as demonstration and evaluation of the approximation techniques. A discussion concludes this chapter in Section IV.4.

## IV.2 Methods

This section begins with a description of the MITKrad plugin of MITK and DIPP. Readers who are more interested in the physical aspects of this section, are invited to skim over the following technical part mainly focusing on implementation details, and get back to a more careful reading from Section IV.2.3 on.

### IV.2.1 MITKrad

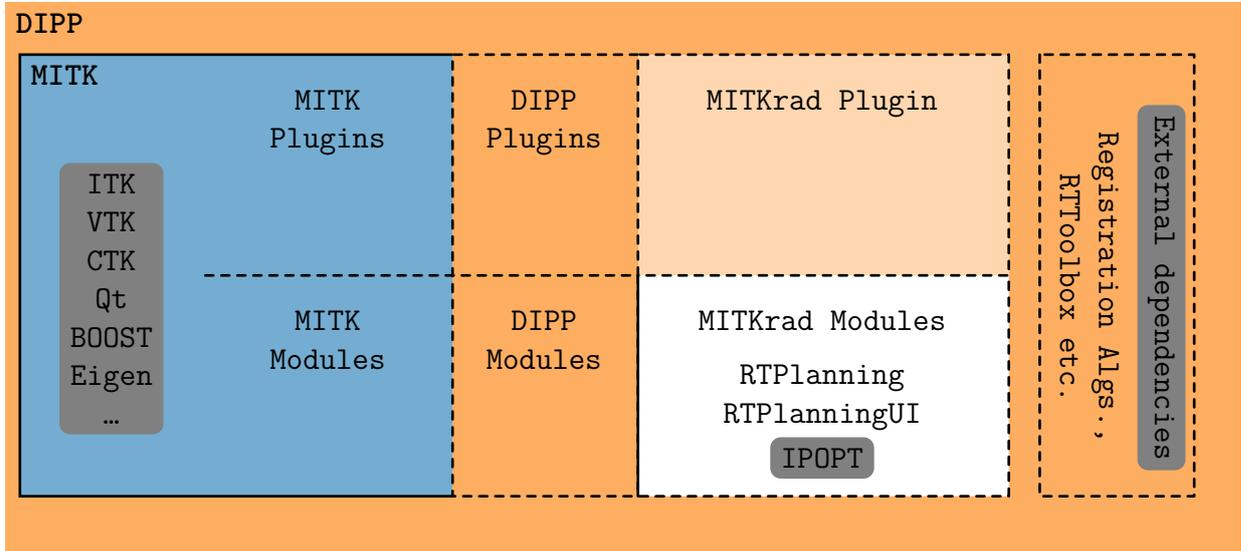
Since MITK and DIPP lacked functionality for full-fledged external beam therapy, MITKrad's treatment planning functionality was built completely from scratch to support

- 1) interactive definition of an external beam geometry and segmented patient data through the MITK graphical user interface (GUI), as well as GUI-based workflow execution,
- 2) a ray-casting algorithm adapted from Heinrich et al. (2014), Siddon (1985), and Siggel et al. (2012),
- 3) definition of the pencil-beam scanning grid, i. e., energy selection and geometrical positioning of Bragg-peaks for IMPT (compare Section II.1.2.1),
- 4) a particle pencil-beam dose calculation algorithm using a generic base-data set, comparable to implementations based on Hong et al. (1996) and Schaffner, Pedroni, and Lomax (1999),
- 5) full nominal and probabilistic dose computation with APM based on the analytical formulations in Chapter III,
- 6) nominal and probabilistic treatment plan optimization using IPOPT (Wächter and Biegler 2006), and
- 7) random and grid sampling based uncertainty analysis using the implemented dose calculation engines.

#### IV.2.1.1 Integration of MITKrad within DIPP/MITK

At the time this manuscript was prepared, MITKrad was developed as a branch of DIPP using MITK's plugin and module mechanism. Figure IV.1 visualizes the role of MITKrad within the complete framework. Efforts were made to decouple the algorithmic side of MITKrad from the MITK GUI; MITKrad's algorithms therefore exhibit less external dependencies (some of them shared

with MITK), but not on MITK modules itself. Hence, separating MITKrad’s algorithms from DIPP and bundling them in a self-contained library would be possible at limited overhead. In the long run, this could ease the development of interfaces for other platforms as for example matRad through MATLAB’s C++ interface (MathWorks 2017), or publishing of an open-source treatment planning library itself.



**Figure IV.1:** Illustration of the DIPP/MITK structure. DIPP as application contains MITK and extends it with external software packages and additional plugins and modules that can communicate with the integrated MITK plugins and modules. MITKrad implements a plugin for the communication with MITK workbench, the modules containing algorithmic development are, in general, independent of other MITK or DIPP modules, while making use of the integration of dependencies like e. g. ITK or Eigen in MITK.

At the core of MITKrad lies the treatment planning plugin RTPanning, comprising all algorithmic developments as well as structures for handling and storing treatment plan information. The module is accompanied by a GUI-module RTPanningUI, which provides Qt based widgets for user interaction regarding workflow, e. g. definition of optimization objectives. Both modules work independently of the MITK/DIPP workbench. Communication between modules and the workbench is handled by the MITKrad plugin, which prepares and transfers data (references) between workbench and modules, and also provides visualization of defined beam configurations. Through the workbench, other functionalities of MITK and DIPP can be accessed, e. g. image processing algorithms or segmentation tools. The availability of such tools within MITK and DIPP as well as the management of important dependencies like ITK or Eigen were the main reason for development directly within the DIPP application.

#### IV.2.1.2 MITKrad design

Additional efforts were put into software design considerations of MITKrad. The goal was to create a framework in which, in the long run, additional treatment algorithms, e. g. dose calculation engines for more modalities, could be implemented easily, encouraging code re-usage and hence increase maintainability. While MITKrad currently implements only the aforementioned

algorithms required for this project, in parallel an abstract class hierarchy was put into place, such that the workflow can be handled using the abstract interface with the correct implementations selected at run-time. The software design, including some implementations which will be described in Section IV.2.2, is illustrated in Fig. IV.2.

When using MITKrad through the user interface, the data and workflow management is mainly handled within the RTPPlanningManager. Through factory instantiation, matching algorithms can be selected at run-time, e. g. a pencil-beam particle dose computation engine is only available with the corresponding base-data and precomputed steering information. Multiple, virtual inheritance is possible; APM, for instance, provides pencil-beam dose calculation as well as functionalities for robustness analysis (i. e., dose standard deviation and expectation value computation). For optimization, the abstract hierarchy with a defined interface (e. g. providing objective function value and gradient) allows straightforward implementation of new objectives and constraints.

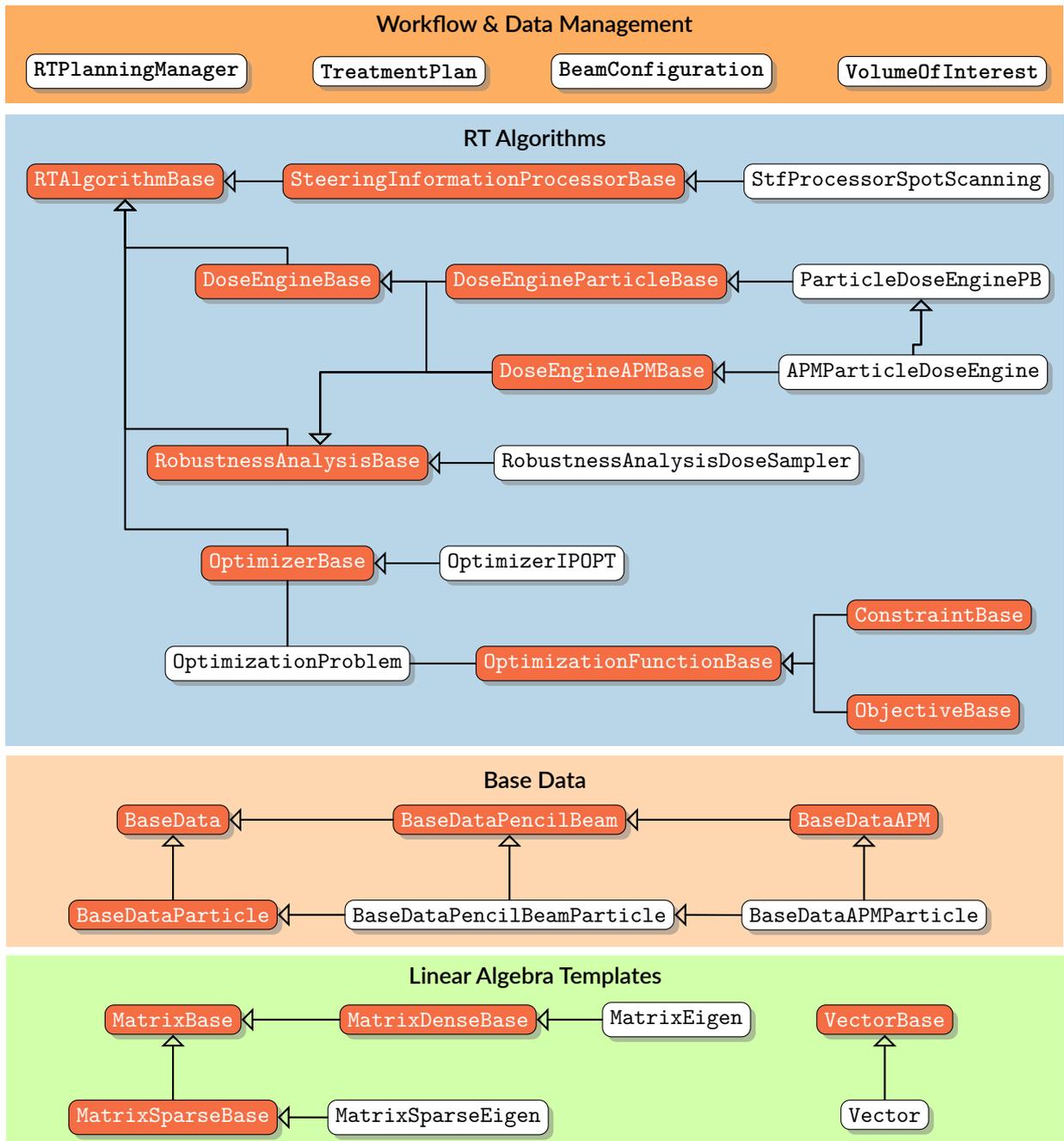
## IV.2.2 Dose calculation and APM-related computations

The probabilistic dose calculation is the heart of APM. Hence, its implementation (including the nominal pencil-beam dose calculation with APM) is outlined below.

**Nominal dose influence computation:** Within MITKrad, the responsibilities for computing the dose influence  $D_{ij}$  of a pencil-beam  $j$  in a voxel  $i$  are shared between the base-data and dose engine implementation. Since pencil-beam algorithms are able to provide the dose influence in any voxel  $i$  just given its WEPL as well as geometrical coordinates within the coordinate system of pencil-beam  $j$  and its parameters, in MITKrad the base-data needs to implement an interface returning the respective dose influence values. The pencil-beam dose calculation engine, on the other hand, is responsible for providing the respective coordinates, WEPL, and pencil-beam parameters, i. e., energy in the case of particles, for every combination  $ij$ .

*Base-data:* APM's base-data contains integrated depth-dose profiles up to a range of 350 mm WEPL described with the 10 Gaussian depth components, which were fitted to the generic proton base-data set contained in matRad (Wieser et al. 2017b). The generic base-data, i. e., using tabulated depth dose curve, is also available within MITKrad. The base-data is indexed with range and energy, whereas the range resolution is 1 mm of WEPL. The lateral contributions are modeled with single Gaussian components, which are, just as for many conventional pencil-beam algorithms, tabulated dependent on WEPL, also with a resolution of 1 mm of WEPL. This applies for the APM as well as the generic base-data set.

*Cut-off:* In order to reduce memory requirements and computational load, MITKrad only computes dose where there are significant contributions from individual pencil-beams. As the protons stop within the patient and scattering is predominantly restricted to small angles, computations can be cut-off both in lateral direction and in depth. In MITKrad, the threshold for lateral contributions defaults to  $3\lambda_{ij}$ , i. e., at less than 0.27 % of the maximum of the dose at the respective



**Figure IV.2:** (Reduced) class diagram of MITKrad illustrating algorithmic implementations as well as data and workflow management. The algorithmic side of MITKrad, e.g. dose calculation, uses a hierarchical system to maximize code re-usability and facilitate new implementations. Hence, also the base-data follows a hierarchical system with virtual inheritance to define interfaces based on treatment modality to enable implementation of custom base-data. For linear algebra, e.g. products involving the dose influence matrix and weight vector, MITKrad defines wrappers, which in the current implementation interface to Eigen, but enable incorporation of other optimized functionality (compare, for example, Ziegenhein et al. (2008)). For vectors, also memory maps (e.g. to common container types) are implemented, yet not shown in the diagram.

depth is neglected and values are set to 0. For the depth component, the cut-off is hardwired into the base-data due to range of the tabulated data.

*Ray-casting:* While the computation of the geometrical transformations is trivial, calculation of the correct WEPL is not, as explained in Section II.1.2.2. Within MITKrad, dose calculation requires a WEPL value per beam for each voxel  $i$  from the respective beam's source (Schaffner, Pedroni, and Lomax 1999). This allows for a single precomputation of a WEPL cube per beam. The respective ray-cast algorithm is implemented based on Siddon (1985) and Siggel et al. (2012) through code adaption from Heinrich et al. (2014). The selection of rays ensures that each voxel is visited by the algorithm at least once, and selects the closest ray in case multiple rays cast through the same voxel. Additionally, MITKrad provides an abstract ray-casting interface that enables implementation (or wrapping) of other ray-casting algorithms (e. g. provided by ITK).

**Uncertainty model:** The uncertainty models described in Section III.6 and illustrated in Fig. III.2 are represented as individual objects, separating into random and systematic component. The corresponding classes enable the definition of arbitrary, linearly correlated uncertainty models. For simplicity, pre-defined implementations rely on representations of the correlation matrices from Fig. III.2 and set all lateral pencil-beam positions with equal absolute variance as well as similar relative variance for WEPL (compare Section III.6).

**Expected dose influence computation:** The computation of the expected dose influence is similar to the nominal APM pencil-beam dose calculation described before. Basically, the only difference lies in the addition of the variance to the squared width of the spatial Gaussian components (compare Eqs. (III.26) and (III.27)). In principle the computation of the expected dose influence matrix  $\mathcal{D}$  with APM is possible at minimal overhead compared to its nominal dose calculation. The additional broadening of the Gaussian components requires, however, readjustment of the lateral cutoff with the respective standard deviation  $\sigma_j$  to  $3(\lambda_{ij}^2 + \sigma_j^2)^{\frac{1}{2}}$ . This increases computation time (since more values need to be computed) as well as memory consumption.

Further, in anticipation of the coming (co)variance computations, one can use the expected value computation to cache certain quantities that will be required, as explained below. This adds significant overhead to the computation, which however stays within worst-case complexity of voxels times pencil beams  $\mathcal{O}(n_V n_B)$ . This is an advantage over computing these quantities once again in the covariance computation with its increased dimensionality and therefore complexity.

**(Co)variance computation:** The computations required for APM laid out in Chapter III are on the lowest level element-wise Gaussian kernels executed on combinations of voxels and pencil-beams. The required input variables (i. e., geometrical information and particle base-data) and kernel output are clearly separated. Also, the kernels exhibit no interdependence. Hence, the problem of evaluating all Gaussian kernels can be seen as an “embarrassingly parallel” (Herlihy and Shavit 2012) computation problem. This straightforward parallelism was exploited on CPU and GPU architecture.

For common data sizes of patient cases (e. g. relevant  $n_V \approx 1 \times 10^6$ ,  $n_B \approx 1 \times 10^5$ ), the full tensors  $\mathcal{C}$  or even  $\mathcal{V}$  are, however, too large to be stored despite their sparsity. Hence, the variance or covariance are computed on the fly. In the same pass, it is possible to compute the matrix  $\Omega$  proposed by Bangert, Hennig, and Oelfke (2013) (compare Section III.5). The computations work with data cached from the calculation of  $\mathcal{D}$  explained above. This cache is indexed by voxel index and contains, for each voxel  $i$  expecting non-zero dose, a list of impinging pencil-beams  $j$  including their geometrical distance to the voxel  $i$  in their respective coordinate frame, their energy, and the expected dose influence value  $\mathcal{D}_{ij}$ .

With this information, (co)variance may be calculated as a loop over the voxel indices while avoiding slow access patterns (e. g. matching voxels to correlated pencil-beams, slow access of individual elements of a sparse matrix etc.).

*Computation on the CPU:* On the CPU, computations are parallelized by dynamically assigning chunks of voxels to multiple threads. This parallelization is straightforward; writing the variance to a voxel  $i$  or the covariance to a voxel combination  $il$  is independent from the other threads. Updates of the matrix  $\Omega$  according to Eq. (III.23) are then, however, not thread-safe per se. The updates can be, however, performed as atomic operations, i. e., indivisible addition that can not be interrupted by a concurrent process.

All remaining operations on shared data are read-only and necessary computations can be performed within the thread scope, implying good scalability of the algorithm. Parallelization is realized programmatically using OpenMP-directives. Hence, parallel computation is, so far, only scaling on shared-memory systems. For distributed memory systems, adaptations using, for example, the MPI standard would be required. However, the goal of the implementation is to be feasible on common desktop computer hardware (as used in the clinics), and thus distributed computing systems exceed the scope of this manuscript.

*Computation on the GPU with CUDA:* The computations of a single covariance influence element given the indices  $i$ ,  $l$ ,  $j$  and  $m$  directly translates to a program that returns a desired set of values for a specific index combination through the same algorithm. This structure is ideally suited for implementation as a kernel for GPGPU (Cook 2013).

In such GPGPU programs, first the necessary data will be copied from the host system's memory (i. e., RAM) to the respective GPU devices. Then, multiple instances of the kernel are arranged on a multi-dimensional grid intrinsically defining the index combination to be evaluated. After concurrent evaluation of the kernel grid, the results are copied back to the host.

GPGPU programming with CUDA does, however, only support three dimensional kernel launches—that is, only three indices are intrinsically available in a CUDA program. While this is sufficient to define a kernel for operation on all elements of  $\mathcal{V}$ , at least full covariance computations have to be chunked into several kernel launches.

Memory requirements, additional limits on kernel dimensionality and kernel execution time-out (on desktop machines with single GPU) may interfere with execution of the complete three-dimensional kernel. Furthermore, not always all pencil-beam combinations  $jm$  have a significant

contribution to the (co)variance in a voxel (combination). Consequently, the actual implementation launches a 2D `CUDA` kernel over only the subset of significant pencil-beam combinations for each voxel (combination). The (simplified) kernel code is appended to this manuscript in Code Listing B.1. In its most memory saving form, the kernel only requires the cached geometrical information of the current voxel in the device memory, also generalizing to large cases. Heuristic performance tests also showed that preparation of the next voxel cache for the device can be efficiently handled on the host during the current kernel execution. Hence, the transfer of the consequent voxel cache alongside the next kernel launch imposes minimal overhead compared to preparing the full cache on the device beforehand.

### IV.2.2.1 Optimization

Optimization of the conventional problem (II.13) and the probabilistic problem (III.21) is implemented via an interface to `IPOPT` (Wächter and Biegler 2006). `IPOPT` is an interior-point optimizer allowing to solve large-scale non-linear (sparse) continuous problems, accepting inequality and equality constraints, and supporting the `L-BFGS`-algorithm for low-memory Hessian approximations. The availability of a `C++` interface and experience with its `MATLAB` interface from use in `matRad` made `IPOPT` the ideal choice for the `APM` implementation in `MITKrad`.

**Adaptation of the pencil-beam scanning grid:** The pencil-beam scanning grid generated for conventional dose calculation and treatment plan optimization usually does not (or rather only in rare exceptions) consider pencil-beams with Bragg-peaks outside of the target volume. This is justified by their major dose contribution lying outside of the target volume, hence it is expected that their weights will be optimized to approximate zero anyway.

For probabilistic optimization, however, due to the additional uncertainty assumptions, pencil-beams not predominantly contributing dose to the target volume in the nominal case may yet have influence on the probability distribution of dose in the target. Therefore, in steering file generation, additional pencil-beams must be placed, with Bragg-peaks located within a margin around the target volume. For the patient cases investigated within this manuscript, the size of these margins was heuristically chosen based on the uncertainty magnitudes. They are listed in Table C.1,

**Structure based  $\Omega$ -matrices:** As explained in Section III.5, Bangert, Hennig, and Oelfke (2013) suggested the use of a precomputed helper matrix  $\Omega$  (compare Eq. (III.23)). When using volume-based penalty factors  $p^v$ , i. e., they are the same for all voxels belonging to `VOI`  $v$ , a `VOI`-based  $\Omega^v$  matrix can be computed:

$$\Omega_{jm}^v = \sum_{i \in v} \mathcal{C}_{ijim} = \sum_{i \in v} \mathcal{V}_{ijm} \quad (\text{IV.1})$$

This  $\Omega^v$  matrix is independent of the optimization penalty, and Eq. (III.23) can then be rewritten as a sum over the volume-based contributions:

$$\text{tr} \left( P \Sigma^d \right) = \sum_v p^v \mathbf{w}^T \Omega^v \mathbf{w}. \quad (\text{IV.2})$$

Equation (IV.2) then allows subsequent probabilistic optimization runs with changing penalties without re-computation of  $\Omega^v$ . Additionally, this accompanies a modular, VOI-based implementation approach of the optimization, where VOI-based objectives are encapsulated in objects which can be dynamically added and removed from the optimization problem.

## IV.2.3 Validation

### IV.2.3.1 Random sampling benchmark

To establish a ground truth for the expectation value and variance of the dose for a certain treatment setup and corresponding uncertainty model, sample statistics from excessive random sampling can be used. Since the sample mean and sample variance are, by nature, only estimators for the true expectation value and variance, it has to be ensured that their estimates are accurate enough to allow a validation with limited noise. This may be done using estimators for the errors of the sample mean, sample standard deviation and/or sample covariance. The number of samples to calculate may then be conditioned on their relative magnitude.

**Standard error of the sample mean:** The accuracy of the sample mean  $\bar{d}$  can be quantified by the commonly known *standard error of the sample mean*  $\sigma_{\bar{d}}$ , which decreases with increasing sample number  $n_S$ ,

$$\sigma_{\bar{d}} = \frac{\hat{\sigma}_d}{\sqrt{n_S}}, \quad (\text{IV.3})$$

where  $\hat{\sigma}_d$  is the sample standard deviation.

**Standard error of the sample covariance:** One could also quantify the standard error for the sample covariance of dose, i. e., the scatter matrix  $\hat{\Sigma}^d$ . If  $\hat{\Sigma}^d$  is estimated from  $n_S$  samples from a multivariate normal distribution  $\mathcal{N}(\mathbf{0}, \Sigma^d)$ ,  $\hat{\Sigma}^d$  follows a Wishart-distribution (Muirhead 1982), i. e.,  $(n_S - 1)\hat{\Sigma}^d \sim W_{n_V}(n_S - 1, \Sigma^d)$ , with  $n_S - 1$  degrees of freedom and with variance

$$\text{Var} \left[ (n_S - 1)\hat{\Sigma}_{ij}^d \right] = (n_S - 1) \left[ \Sigma_{ij}^{d^2} + \Sigma_{ii}^d \Sigma_{jj}^d \right]. \quad (\text{IV.4})$$

The Wishart-distribution is the matrix-variate generalization of the  $\chi^2$ -distribution. This means that in the special case of sample variance  $\hat{\sigma}_{d_i}^2 = \hat{\Sigma}_{ii}^d$  in a voxel  $i$  one has  $(n_S - 1)\hat{\sigma}_{d_i}^2 \sim \chi_{n_S - 1}^2$  where  $\chi_{n_S - 1}^2$  has variance  $2(n_S - 1)$ . Hence, for the sample variance  $\hat{\sigma}_d^2$  of the whole distribution

$d$ , a standard error  $\sigma_{\hat{\sigma}_d^2} / \sigma_d^2$  relative to the true variance  $\sigma_d^2$  can be given as

$$\frac{\sigma_{\hat{\sigma}_d^2}}{\sigma_d^2} = \sqrt{\frac{2}{n_S - 1}}. \quad (\text{IV.5})$$

**Standard error of the sample standard deviation:** For the quantification of “error bars” the standard deviation is commonly used. Since the sample standard deviation  $\hat{\sigma}_d = \sqrt{\hat{\sigma}_d^2}$  follows a chi-distribution  $\sqrt{n_S - 1} \hat{\sigma}_d / \sigma_d \sim \chi_{n-1}$ , its standard error exhibits a more complicated form compared to Eq. (IV.5).

The corresponding derivations shall be omitted here, since for large sample numbers  $n$ , its relative standard error converges to

$$\frac{\sigma_{\hat{\sigma}_d}}{\hat{\sigma}_d} \approx \frac{1}{\sqrt{2(n_S - 1)}} \quad \text{for large } n_S. \quad (\text{IV.6})$$

A compact overview of the necessary analytical computations is given in Ahn and Fessler (2003) for the interested reader.

**Choosing the sample number  $n_S$ :** In summary, relative standard errors for sample variance and standard deviation have a compact form while at the same time only depending on the sample number  $n_S$ . For sample mean and covariance, the relative error additionally depends on other quantities than  $n_S$ , i. e., the standard deviation and variance, respectively, which makes conditioning on their accuracy solely depending on the sample number hardly possible. Furthermore, by Eqs. (IV.3) and (IV.4) it is ensured that the relative error of sample mean and elements of the covariance matrix is always smaller or equal to the relative error of sample standard deviation and variance, respectively. This encourages conditioning of the sample number  $n_S$  on the accuracy of standard deviation and/or variance.

A relative accuracy of  $\sigma_{\hat{\sigma}_d} / \hat{\sigma}_d \leq 1\%$ , for example, is achieved when using  $n_S > 5,000$  random samples to compute the sample statistics. Therefore, for the validation of the expectation value and standard deviation computations of APM, the sample statistics of  $n_S = 5,000$  samples were used as the “true” reference for APM’s moment computations.

### IV.2.3.2 Grid sampling approaches

Instead of random sampling, grid-based importance sampling approaches are often used to quantify uncertainty, commonly motivated with specific error or worst case scenarios. These methods define a (usually evenly spaced) grid in the space of uncertain parameters. Due to the “curse of dimensionality” (Bellman 1961), sufficient grid sampling over a high-dimensional parameter space is more or less infeasible. Also, correlation models in general introduce complex combinatorics. Therefore, methods based on sampled grids often only use independent points along the respective parameter axis, i. e., selective spatial shifts of the patient or range over- and undershoots (e. g. Lomax 2008a).

To compare APM's computations against these common grid sampling approaches, the standard deviation and expectation value is explicitly computed for three grid sampling methods with different coverage of the multivariate sample space. Using APM's "beam-wise" uncertainty model, i. e., spatial and range uncertainties are correlated within a beam and uncorrelated otherwise, the corresponding sample grid has  $d \cdot n_B$  dimensions where  $n_B$  is the number of beams and  $d$  is the number of spatial dimensions. Then, the complexity of the three models can be defined by the resulting sample number  $n_S$  as follows, when  $n_P$  is the number of grid sample points per dimension:

W<sup>I</sup>: Full negligence of combinatorics, i. e., error scenarios are computed independently per dimension;  $n_S = (n_P - 1) \cdot n_B d + 1$

W<sup>II</sup>: Combination of spatial dimensions, but not beams;  $n_S = (n_P^d - 1) \cdot n_B + 1$

W<sup>III</sup>: Full combinatorics across spatial dimensions and beams;  $n_S = n_P^{d \cdot n_B}$

In the beam coordinate frame used for defining APM's uncertainty model, the dimensionality is given by  $d = 3$  ( $x$ ,  $y$  and WEPL). All sample numbers  $n_S$  include the nominal scenario once.

For each dose sample  $\mathbf{d}_s$  on the grid, one can define the respective weight  $w_s$  as the value of the probability density (reduced to the subspace representing the respective dimensionality) at the grid point's coordinates. Estimates of the weighted arithmetic mean  $\bar{\mathbf{d}}_W$  and weighted covariance matrix  $\hat{\Sigma}_W^{\mathbf{d}}$  can then be computed through

$$\bar{\mathbf{d}}_W = \frac{1}{\sum_{s=1}^{n_S} \omega_s} \cdot \sum_{s=1}^{n_S} \omega_s \mathbf{d}_s, \quad (\text{IV.7})$$

$$\hat{\Sigma}_W^{\mathbf{d}} = \frac{\sum_{s=1}^{n_S} \omega_s}{\left(\sum_{s=1}^{n_S} \omega_s\right)^2 - \sum_{s=1}^{n_S} \omega_s^2} \cdot \sum_{s=1}^{n_S} \omega_s (\mathbf{d}_s - \bar{\mathbf{d}}_W)^T (\mathbf{d}_s - \bar{\mathbf{d}}_W), \quad (\text{IV.8})$$

respectively (Galassi et al. 2016). The factors preceding the sums over the samples normalize the weights and, in Eq. (IV.8), apply Bessel's correction to obtain an unbiased estimate. If all weights are equal, Eqs. (IV.7) and (IV.8) become the common expressions for sample mean and unbiased sample covariance.

### IV.2.3.3 $\gamma$ -analysis

In dosimetric analyses, the  $\gamma$ -analysis introduced by (Low et al. 1998) is commonly used in the treatment planning workflow to quantify the agreement between two, e. g. simulated and measured, dose distributions. In comparison to other tests like measuring absolute or relative deviations, the concept of  $\gamma$ -analysis relies on "distance-to-agreement"; instead of comparing two doses corresponding to the same exact location in both distributions, a pre-defined neighborhood is searched for best agreement.

**Definition:** In its original definition (Low et al. 1998), the analysis quantifies this agreement by returning a distribution  $\gamma(\mathbf{r})$  at the respective coordinates  $\mathbf{r}$  given by

$$\gamma(\mathbf{r}) = \min \left\{ \sqrt{\frac{|\mathbf{r}_c - \mathbf{r}|^2}{\Delta\tilde{r}^2} + \frac{(d_c(\mathbf{r}_c) - d(\mathbf{r}))^2}{\Delta\tilde{d}^2}}, \forall \mathbf{r}_c \right\} \quad (\text{IV.9})$$

where  $d_c$  and  $\mathbf{r}_c$  correspond to the dose and coordinates to compare to, respectively, and  $\Delta\tilde{r}$  and  $\Delta\tilde{d}$  are pre-defined criteria describing the acceptance tolerance in distance and dose, respectively. Geometrically, the criteria  $\Delta\tilde{r}$  and  $\Delta\tilde{d}$  are half-axes of an ellipse in the space of distance  $|\mathbf{r}_c - \mathbf{r}|^2$  vs. dose deviation  $(d_c(\mathbf{r}_c) - d(\mathbf{r}))^2$ , with points lying within or on the ellipse *passing* and points lying outside the ellipse *failing* a  $\gamma$  analysis:

$$\begin{cases} \gamma(\mathbf{r}) \leq 1, & \text{passed} \\ \gamma(\mathbf{r}) > 1, & \text{failed.} \end{cases} \quad (\text{IV.10})$$

It is possible to distinguish between *local* or *global*  $\gamma$ -tests; local tests define a relative dose criterion  $\Delta\tilde{d}$  based on the local dose value, whereas global tests base  $\Delta\tilde{d}$  on the maximum global dose value to be robust against failing voxels with very low dose.

**Implementation:** For computation of the  $\gamma$ -index from Eq. (IV.9) in practice one has to consider two main pitfalls. First, dose values are stored in voxels at discretized spatial coordinates, which can be arbitrarily intersected by the corresponding acceptance ellipsoids. This leads to a dependence of the  $\gamma$ -values on voxel resolution. Second, Eq. (IV.9) compares to all coordinates in the comparison distribution, meaning that trivial application of Eq. (IV.9) would be of complexity  $\mathcal{O}(n_V^2)$  with number of voxels  $n_V$ .

In the following analyses using  $\gamma$ -distributions, the  $\gamma$ -index implementation from `matRad` was used (Wieser et al. 2017b). In this implementation, the resolution dependence was reduced by interpolation of the comparison image, i. e., while the  $\gamma$ -distribution gets computed at original resolution of the distribution to evaluate, comparisons are performed against the interpolated reference image with higher resolution. Note that this approach can only keep or increase the acceptance rate—or, respectively, keep or decrease the  $\gamma$ -values—since in Eq. (IV.9) the minimum of all comparisons is selected. Second, the implementation uses a patch-based scan of the image, i. e., restricting the search to a confined neighborhood enclosing the agreement distance parameter, reducing the computation time. Pass rates are not affected by this reduction of the comparison to the respective neighborhood patch, but in some cases  $\gamma$ -values of failing voxels might be overestimated.

#### IV.2.4 Approximative computation techniques

The dose variance as well as the matrix  $\Omega$  are computed as large sums of element-wise operations. In the case of variance, for each voxel a quadratic form needs to be computed (compare Eq. (III.20)). For computation of  $\Omega$ ,  $\mathcal{V}$  is summed up over all voxels (compare Eqs. (III.23)

and (IV.1)). Instead of computing the complete sum, estimates of arbitrary accuracy may be computed by only sampling a subset of the respective sums. In the following, this approach will be outlined for obtaining an approximate standard deviation and  $\Omega$ -matrix.

#### IV.2.4.1 Approximating $\sigma_d$ with pencil-beam sub-sampling

Since the standard deviation of dose  $\sigma_{d_i}$  in a voxel  $i$  is computed through variance, at first a strategy will be derived for approximating the variance. To do so, a set  $JM := \{j, m \mid j > m\}$  is defined containing all pencil beam combinations where  $j > m$  (since  $\mathcal{V}_{ijm} = \mathcal{V}_{imj}$ ). Evaluating all pencil-beams with index combinations  $jm \in JM$  in Eq. (III.20) is (exactly) of complexity  $\mathcal{O}((n_B^2 - n_B) / 2)$ . Evaluation of the remaining diagonal elements  $j = m$ , which always have a contribution to variance, is of linear complexity  $\mathcal{O}(n_B)$ . Hence, the evaluation over the index set  $JM$  represents the computationally expensive part of the variance calculation, and may be approximated by a randomly sampled subset  $\widehat{JM} \subset JM$  on top of evaluation of the diagonal elements.

An approximation  $\hat{\sigma}_{d_i}^2$  of the variance from Eq. (III.20) can then be obtained through

$$\hat{\sigma}_{d_i}^2 = \sum_j w_j^2 \mathcal{V}_{ijj} + 2 \cdot \frac{n_{JM}}{n_{\widehat{JM}}} \sum_{jm \in \widehat{JM}} w_j w_m \mathcal{V}_{ijm} \approx \underbrace{\sum_{jm} w_j w_m \mathcal{V}_{ijm}}_{\text{Eq. (III.20)}} \quad (\text{IV.11})$$

where  $n_{JM}$  and  $n_{\widehat{JM}}$  are the number of combinations in the population and the sample, respectively.

Due to sampling error negative estimates for  $\hat{\sigma}_d^2$  can occur. Here, these are simply projected to zero when computing the final estimate of the standard deviation

$$\hat{\sigma}_{d_i} = \begin{cases} \sqrt{\hat{\sigma}_{d_i}^2} & \hat{\sigma}_{d_i}^2 \geq 0 \\ 0 & \hat{\sigma}_{d_i}^2 < 0 \end{cases} \quad (\text{IV.12})$$

The noise in  $\hat{\sigma}_d^2$  could be reduced by application of a denoising filter.

Theoretically, simultaneous computation of an estimate of  $\Omega$  is feasible using weighted sampling statistics. This approach was, however, not pursued since the random selection of index combinations  $jm \in \widehat{JM}$  does not guarantee the estimate of  $\Omega$  to be positive semi-definite, which is required for optimization. While it could, in principle, be possible to correct such an approximated  $\Omega$  matrix (e. g. by finding the nearest symmetric positive definite matrix as described by Higham (1988)), a more straightforward method based on voxel subsets will be presented below.

#### IV.2.4.2 Approximating $\Omega$ with voxel sub-sampling

Given a voxel population  $I$ , e. g. all voxels belonging to some VOI, the corresponding  $\Omega$ -matrix may be estimated evaluating a sample  $\hat{I}$  from the sum  $\Omega_{jm}^I = \sum_{i \in I} p_i \mathcal{V}_{ijm}$ . The estimated  $\hat{\Omega}^I$  is

then given by

$$\hat{\Omega}_{jm}^I = \frac{n_I}{n_{\hat{I}}} \sum_{i \in \hat{I}} p_i \mathcal{V}_{ijm} \approx \sum_{i \in I} p_i \mathcal{V}_{ijm} \quad (\text{IV.13})$$

where  $n_I$  and  $n_{\hat{I}}$  are the number of voxels in the population and the sample, respectively.

## IV.3 Results

To evaluate the described `MITKrad` implementation, three patient datasets were imported into `MITK`; a prostate case, a paraspinal case, and an intracranial case. Treatment plans were defined to embody different set-ups in which errors are expected to have varying impacts.

The prostate case is prone to large range uncertainties at the distal tumor edge due to the large distance particles have to travel through the pelvis, while the most important `OARs`, i. e., bladder and rectum, are positioned laterally and therefore subject to lateral set-up errors.

For the intracranial case, with the brainstem directly located next to the `CTV`, a set-up to induce dose uncertainty mainly subject to range errors within the `OAR` was chosen. Therefore, a beam geometry was defined with both beams' distal fall-off lying within the brainstem. While in the plan a `PTV` margin was delineated, treatment planning was performed on the `CTV`.

The paraspinal case was chosen, on the one hand, for comparability of results to works by (Unkelbach, Chan, and Bortfeld 2007; Unkelbach et al. 2009). Also, it can be seen as a “worst-case” anatomy regarding range and set-up uncertainty, with the significant `OAR`, i. e., the spinal cord, being surrounded by the target to be irradiated.

Information on all three datasets is given in Table C.1, containing the respective irradiation geometry, assumed uncertainties, dose prescription and complexity, i. e., voxel resolution as well as number of pencil-beams.

The results shown slightly differ from published results (Wahl et al. 2017). The present and up-to-date results in this section feature no fractionation, and were computed on newer hardware with an updated implementation.

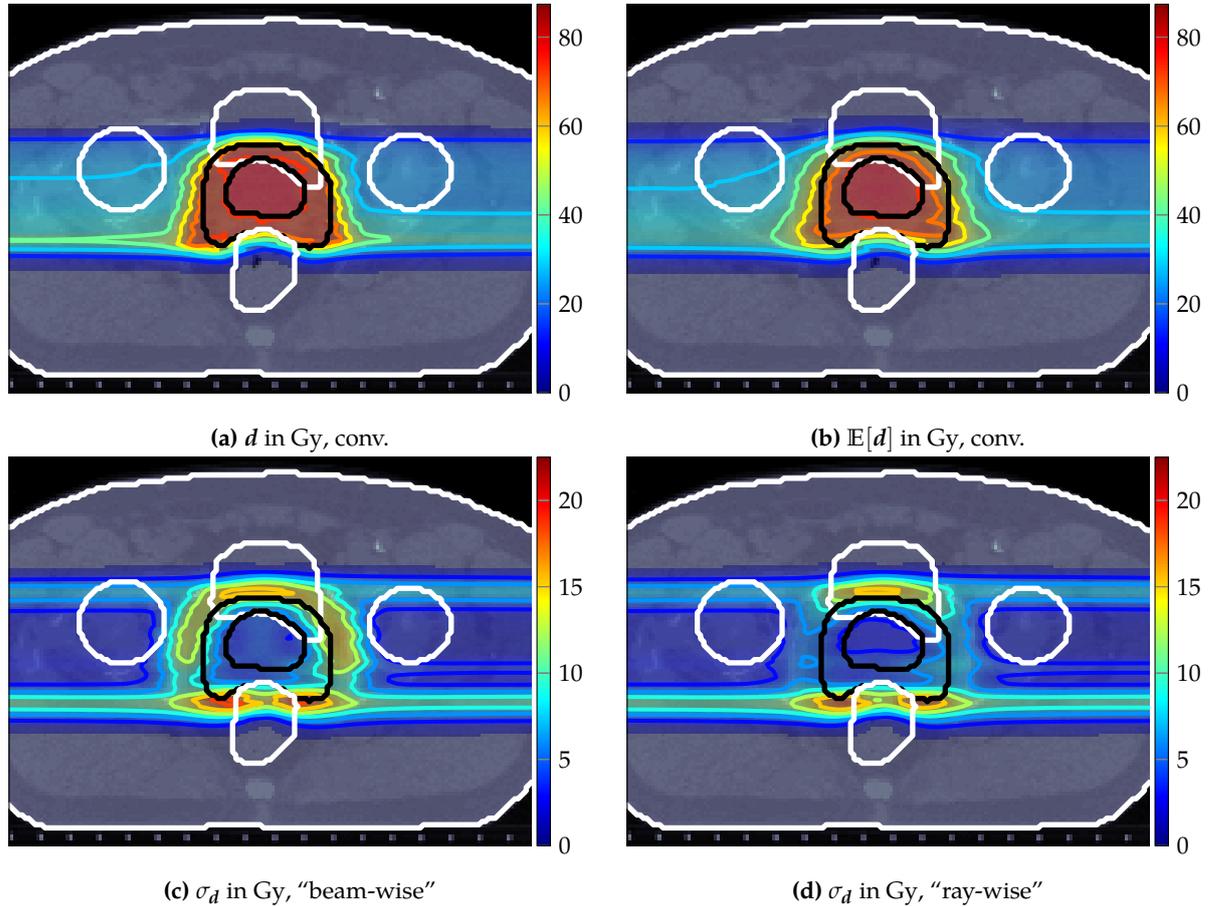
### IV.3.1 Proof-of-work

`MITKrad` was used to create a set of conventional as well as probabilistic treatment plans for three patient cases; an intracranial case, a paraspinal case and a prostate case. While lateral set-up uncertainties were always assumed to be perfectly correlated within a beam, corresponding to lateral shifts of the patient in the `BEV` as depicted in Fig. III.2c, two range correlation models were evaluated for all patient cases: First, likewise to the set-up uncertainties, a “beam-wise” correlation was assumed. Second, perfect correlation “ray-wise” as illustrated in Fig. III.2b was evaluated.

One standard deviation for range uncertainties was quantified with  $\pm 3.5\%$  systematic and  $\pm 1$  mm random component. Set-up uncertainties were expected to have a bigger random component of  $\pm 2$  mm compared to their systematic component  $\pm 1$  mm. The uncertainty assumptions

are also stated in Table C.1, since they apply throughout the whole manuscript. Despite separating the uncertainty model into systematic and random components, only single fraction cases are presented within this section, with fractionation being thoroughly evaluated and discussed in Chapter V.

Figure IV.3 displays nominal and probabilistic APM computations by means of an exemplary slice of the prostate case after a *conventional* optimization, i. e., optimization with the nominal dose influence matrix with objective function (II.16).



**Figure IV.3:** Exemplary axial slice of (a) a conventionally optimized treatment plan using a dose influence matrix computed with APM's dose approximation. Further, the expected dose could be computed (b) which is independent of the correlation model. Finally, the standard deviation of dose was computed, once assuming "beam-wise" correlations (c) and once assuming "ray-wise" correlations (d) for the range errors for a single fraction.

The expected dose distribution in Fig. IV.3b appears as the "blurred out" nominal dose distribution from Fig. IV.3a; a result from the broadening of the Gaussian dose influence computation components through addition of component width and variance as given in Eqs. (III.9) and (III.10). As the expectation value of dose only depends on the variances defined in the uncertainty model, no dependency on the correlation assumptions exists.

For the computation of standard deviation, however, these correlations factor in and lead to drastically different standard deviation distributions for the different correlation models explained in Section III.6. If range uncertainties are correlated per beam, as defined in the "beam-wise" uncertainty model, at the respective distal edges of the tumor high variance is observed (Fig. IV.3c).

At this point, particles have traveled a comparably far distance through the tissue which, due to the relative nature of the assumed range uncertainties (i. e., 3.5%), may be subject to significant over- or undershoots of the complete beam. For the “ray-wise” correlation model, while the relatively expected range over- or undershoots are still of the same magnitude, they occur for pencil-beams on the same path independently, and may therefore cancel each other out through lateral scatter contributions from neighboring beams, resulting in reduced dose variance at the distal edges of the tumor.

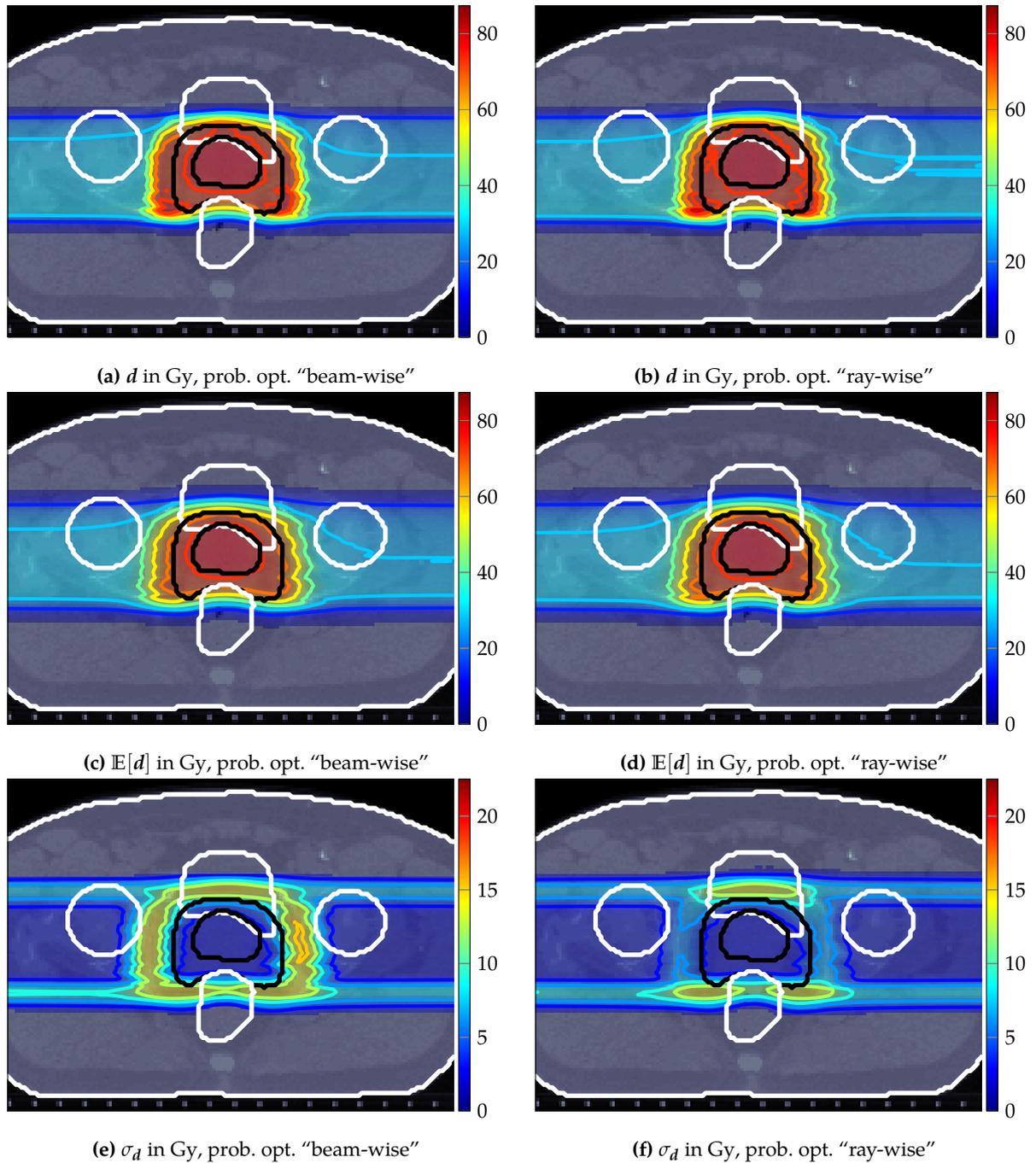
Similar considerations apply for the remaining paraspinal and intracranial case. Exemplary slices for both cases can be found in Figs. C.1 and C.3, respectively. During optimization, the variance contribution in the unclassified tissue had been neglected, due to its low optimization penalty.

Figure IV.4 displays the same axial slice as in Fig. IV.3, this time after probabilistic optimization as proposed in Eq. (III.22), i. e., the expected objective function as presented in Bangert, Hennig, and Oelfke (2013). For the intracranial and paraspinal case, similar illustrations can be found in Figs. C.2 and C.4. Two plans were optimized using both the “ray-wise” and “beam-wise” correlation model, respectively.

For both plans, the nominal dose distributions in Figs. IV.4a and IV.4b exhibit a growth of the targets’ high dose regions (recognizable from the isodose lines), similar to the definition of a margin. For the “beam-wise” model, the high dose region appears to be stretched in the direction of the beams’ distal tumor edges in Fig. IV.4a, which is not so pronounced for the “ray-wise” model in Fig. IV.4b. This behavior comes from the “beam-wise” model inducing a much larger uncertainty on dose at the distal tumor edges due to the perfectly correlated range error, which needs to be compensated by pushing the dose-fall of away from the target. In the “ray-wise” model, where contributions of neighboring pencil-beams may cancel out, this uncertainty is reduced, and its compensation may also be achieved by re-modulation of the respective pencil-beams.

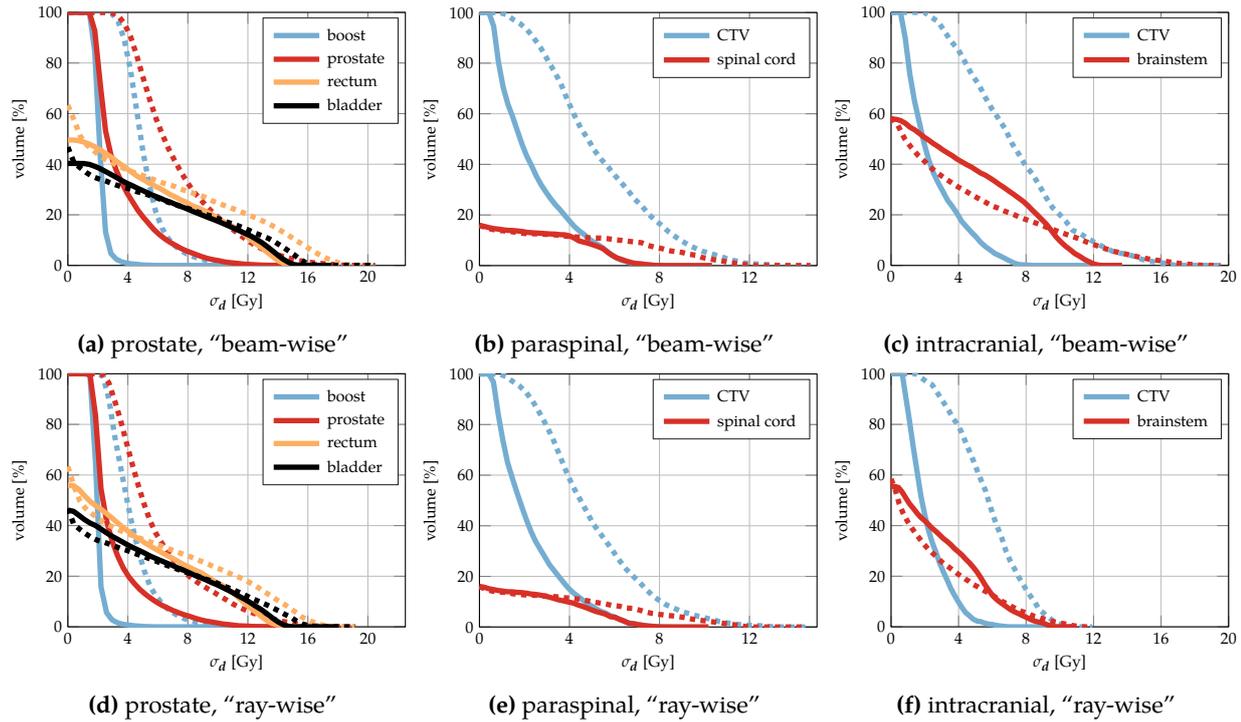
The expectation value of dose distributions in Figs. IV.4c and IV.4d now show better coverage of the targets, which is more comparable to the nominal dose coverage after conventional optimization in Fig. IV.3a. This is accompanied by an overall reduction of the integrated variance for both correlation models (Figs. IV.4e and IV.4f) and especially within the targets (as they are bound to the highest optimization penalty).

This variance reduction can also be identified in the standard deviation volume histograms (SDVHs) presented in Fig. IV.5, this time for all three investigated patient cases. SDVHs are analogons to conventional DVHs from Eq. (II.9) for standard deviation instead of dose distributions. In the targets, the integrated standard deviation is roughly reduced about a factor of two in all cases. In particular, volumes of high as well as low dose variability get reduced, although the reduced standard deviation can still reach substantial maximum values of up to 13% of the prescribed dose. In the OARs, especially the standard deviation peaks are reduced, while the volume exhibiting low dose variability actually increases. This can be argued with the optimization prioritizing homogeneous expected dose with low variance in the target over OARs. Hence, dose gradients are pushed out into the healthy tissue—i. e., a margin around the target is produced—which pushes dose variability into OARs. This effect can only be partly mitigated by additional, non-trivial



**Figure IV.4:** Exemplary axial slices of distributions of nominal (a-b), expected (c-d) and standard deviation (e-f) of dose resulting from probabilistic optimization. For the plan shown in the left column (a,c,e), the "beam-wise" correlation model was assumed for probabilistic optimization. The right column (b,d,f) illustrates a probabilistically optimized plan under assumption of the "ray-wise" correlation model. Both treatment plans were optimized under 1 fraction.

modulations of the fluence.



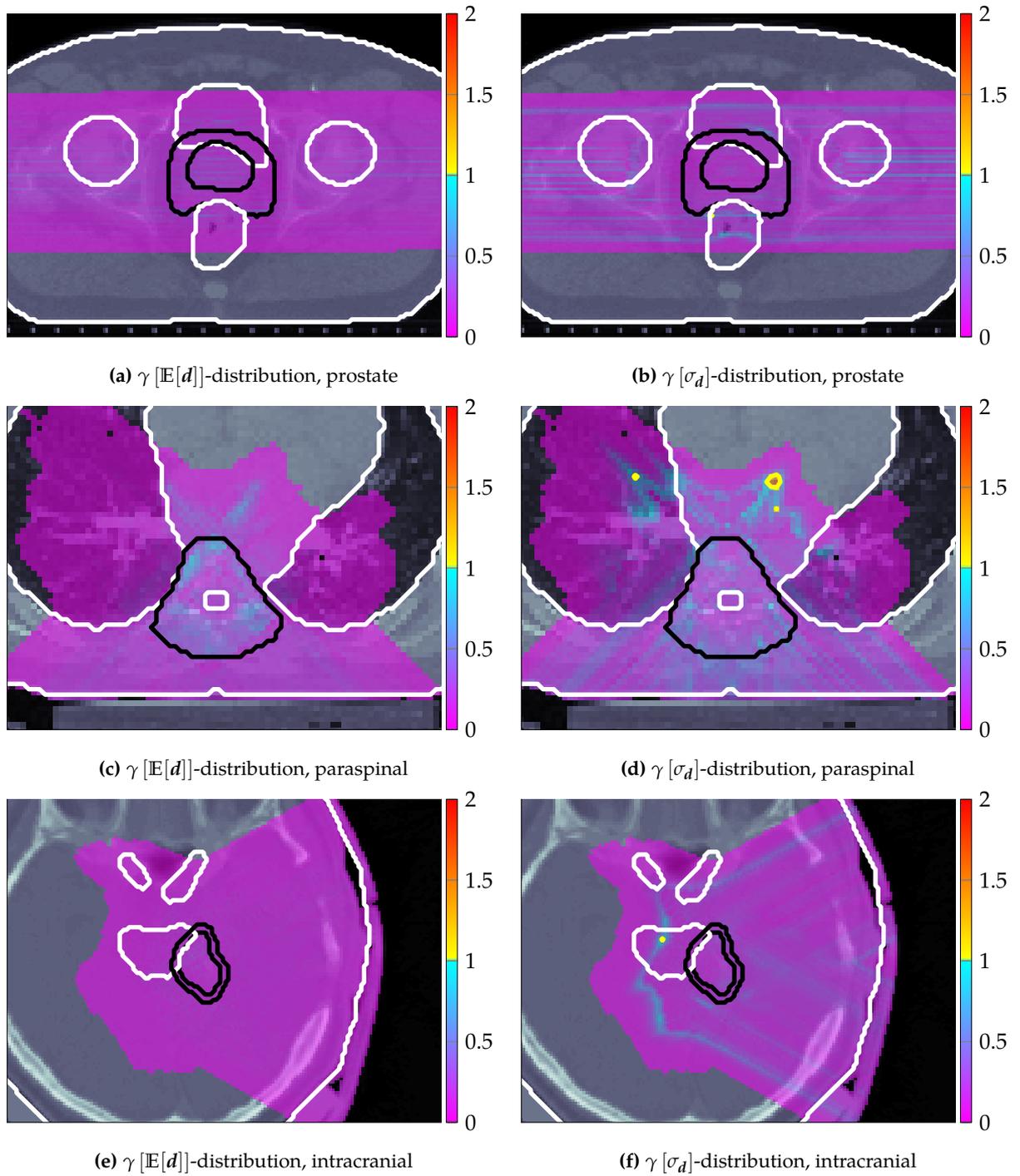
**Figure IV.5:** SDVHs after conventional optimization (dotted lines  $\cdots$ ) and probabilistic optimization (solid lines  $\text{—}$ ) for all three patient cases. In (a-c), the “beam-wise” correlation model was assumed while (d-f) modeled “ray-wise” correlations.

### IV.3.2 Validation of probabilistic dose calculation

For validation of *APM*’s computation of expectation value and standard deviation of dose, 5,000 multivariate random dose samples (i. e., 5,000 treatment scenarios with a single fraction) were computed using the “beam-wise” uncertainty model as a benchmark. At this sample number the standard error of the sample standard deviation can be estimated to be approximately 1% (as explained in Section IV.2.3.1), with even higher precision of the expectation value.

Distribution of sample mean and standard deviation from the sampling benchmark were used as reference for the distributions computed with *APM*, using the  $\gamma$ -criterion with two distance-to-agreement parameters of 3%/3 mm and 2%/2 mm. Three interpolation points were used to compensate for image resolution. Exemplary axial slices for  $\gamma$ -distributions using the weaker 3%/3 mm are shown in Fig. IV.6. Table IV.1 provides the  $\gamma$ -pass rates corresponding to Fig. IV.6 and adds mean absolute error, i. e., the bias, to the analysis as well as comparison to the grid sampling approaches introduced in Section IV.2.3.2.

The global  $\gamma$ -analyses show that validation of the expected value results in better  $\gamma$ -pass rates than the validation of the standard deviation. This may be, on the one hand, explained through the benchmark standard deviation being less accurate (relatively) than the expectation values; an expected dose value of  $60 \pm 10$  Gy, for example, will have a relative standard error of 0.2% compared to the standard deviation’s 1%, inflicting less noise on the  $\gamma$ -computation. On the other



**Figure IV.6:**  $\gamma_{3\%/3\text{mm}}$ -distributions validating APM's computations of expectation value (left column) and standard deviation (right column) against the respective sampled moment from the random sampling benchmark for exemplary axial slices of all three patient cases. The acceptance threshold of  $\gamma = 1$  is indicated as a color jump in the used colormap and through display of isolines at  $\gamma = 1$ . Bear in mind that due to good agreement not all depicted slices show threshold violations.

hand, the standard deviation computations are geometrically more complex as pair-wise combinations of pencil-beams are evaluated with additional dependence on the correlation model. Therefore, possible deficiencies intrinsic to the APM framework are expected to condense mainly in inaccurate standard deviation, consequently leading to lower  $\gamma$ -pass rates.

Table IV.1 also includes  $\gamma$ -pass rates and bias compared to the random sampling benchmark for multiple weighted sampling approaches as discussed in Section IV.2.3.2. For all evaluations, at the basis a positive and a negative shift of  $\pm 1.5\sigma^{x/y/z}$  along each spatial dimension (and, respectively, their combinations) was considered, i. e.,  $d = 3$  and  $n_p = 3$ , requiring  $6n_B + 1$  scenarios for  $W^I$ ,  $26n_B + 1$  scenarios for  $W^{II}$  and  $27^{n_B}$  scenarios for  $W^{III}$ , depending on the number of beams  $n_B$ . For the paraspinal case, evaluation of  $W^{III}$  would thus already require approximately 20,000 scenarios, which is why a  $W^{III}$ -analysis on the case was omitted. The  $\gamma$ -tests confirm that grid sampling consistently underestimates standard deviation (i. e., negative bias) when not considering the full combinatorics. Even when relying on the complete combinatorial space, the agreement of grid sampled moments is worse than for the moments computed analytically with APM. The weighted sample mean on the other hand, which is not dependent on possible correlations and therefore combinatorics, yields good agreement near  $\gamma$ -pass rates of 100%. Despite the good agreement, APM is also consistently subject to a negative bias. This slight underestimation of the respective moments comes from the the cut-off used to restrict computations to significant voxels.

Table IV.2 details the  $\gamma$ -validation of APM by providing additional pass rates, broken down by VOIs and separated into global and local analyses with 3%/3 mm and 2%/2 mm criteria. Even with the more strict 2%/2 mm, global  $\gamma$ -pass maximally drop by 5.5 percentage points in the right lung of the paraspinal case as an exception, but are in general stable within approximately two percentage points.

The local analysis shows good agreement between APM computations and sampled moments within the target ( $> 95\%$ ) for local  $\gamma$ -analyses. For OARs, the pass rates are, in general, tenth of percents smaller. While this may look alarming, the effect may be explained by the aforementioned cut-offs applied in the nominal and probabilistic dose computation. In the sampling scenario, a single, improbable sample is enough to contribute to the sample moments within a voxel which is, on the other hand, neglected by APM. In the local analysis, this voxel is then most likely doomed to fail the  $\gamma$ -test. In Tables IV.3a to IV.3c, this effect reflects in lowest  $\gamma$ -pass rates in OARs which are only affected in a few voxels, e. g. the lungs in the paraspinal case, and higher pass rates in OARs containing a larger fraction of significant voxels. While this effect can be factored out, it was decided to stick with the results to be consistent in all  $\gamma$ -analyses and, as quantified in Table IV.1, neglecting significant voxels via a cut-off induces a negative bias into APM computations which should be kept in mind.

### IV.3.3 Performance

#### IV.3.3.1 Computational performance of APM within MITKrad

The computational performance of nominal and probabilistic dose calculation and optimization using the APM framework within MITKrad was evaluated on a desktop machine running Win-

**Table IV.1:** Global  $\gamma_{3\%/3\text{mm}}$ -pass rates and bias of moment computations with APM and grid-based sampling.

patient & moment		$\gamma$ pass rates [%]				mean error / bias [Gy]			
		APM	W <sup>I</sup>	W <sup>II</sup>	W <sup>III</sup>	APM	W <sup>I</sup>	W <sup>II</sup>	W <sup>III</sup>
intracranial	$\mathbb{E}[\mathbf{d}]$	100.0	100.0	100.0	100.0	-0.06	0.05	0.02	-0.03
	$\sigma_d$	99.8	66.7	87.6	99.1	-0.14	-1.00	-0.48	0.31
paraspinal	$\mathbb{E}[\mathbf{d}]$	99.9	100.0	100.0	-	-0.24	-0.05	-0.05	-
	$\sigma_d$	99.0	29.9	47.6	-	-0.20	-1.67	-1.17	-
prostate	$\mathbb{E}[\mathbf{d}]$	100.0	99.8	100.0	100.0	-0.20	0.03	0.01	0.03
	$\sigma_d$	100.0	63.0	80.8	98.8	-0.11	-1.44	-0.74	0.31

**Table IV.2:**  $\gamma$ -pass rates [%] of APM validation for all cases including sub-analyses for relevant VOIs.

(a) prostate

VOI	global analysis				local analysis			
	3 %/3 mm		2 %/2 mm		3 %/3 mm		2 %/2 mm	
	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$
all	100.0	100.0	100.0	99.8	89.2	75.7	82.5	69.2
boost	100.0	100.0	100.0	100.0	100.0	99.9	100.0	99.2
prostate	100.0	100.0	100.0	99.9	100.0	99.9	100.0	99.5
rectum	100.0	99.8	99.7	97.3	92.5	70.7	84.5	62.8
bladder	100.0	100.0	100.0	99.9	95.7	68.0	87.7	61.5
left femur	100.0	100.0	100.0	100.0	99.3	94.0	98.3	89.8
right femur	100.0	100.0	100.0	100.0	98.8	93.2	97.1	89.0

(b) paraspinal

VOI	global analysis				local analysis			
	3 %/3 mm		2 %/2 mm		3 %/3 mm		2 %/2 mm	
	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$
all	99.9	99.0	99.5	97.0	69.4	54.3	61.8	47.5
CTV	99.2	100.0	97.2	99.3	99.1	99.4	97.1	95.5
spinal cord	100.0	100.0	100.0	100.0	86.1	81.4	80.4	75.5
left lung	100	99.4	100	98.2	50.6	29.0	42.4	24.2
right lung	100	96.8	99.9	91.3	50.0	32.2	42.6	27.2

(c) intracranial

VOI	global analysis				local analysis			
	3 %/3 mm		2 %/2 mm		3 %/3 mm		2 %/2 mm	
	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$	$\mathbb{E}[\mathbf{d}]$	$\sigma_d$
all	100.0	99.8	99.9	99.3	86.5	64.3	78.3	58.8
CTV	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
brainstem	100.0	99.8	100.0	97.5	99.2	67.8	96.6	62.2
optic chiasma	100.0	100.0	100.0	98.3	100.0	60.8	98.3	45.3

dows 7 on a Intel® Core™ i7-2600 CPU @ 3.4 GHz able to run 8 parallel logical threads and with a single nVidia GeForce GTX 970 GPU.

**Runtimes:** Table IV.4 displays runtime analyses on all three patient cases. For the standard deviation, the given runtimes are differentiated into architecture, i. e., GPU and CPU, as well as incorporated voxels, i. e., all voxels full  $\sigma_d$ -distributions and only voxels in targets and OARs needed for  $\Omega$ -matrices for optimization. As before, a single fraction was assumed.

**Table IV.4:** Runtimes (in seconds) for nominal and probabilistic dose/moment calculation and optimization for fractionated treatment plans of all three patient cases. The table differentiates between CPU and GPU implementations for the full  $\sigma_d$ -calculation and the  $\Omega$ -matrix calculation without normal tissue for optimization (in parentheses). (“—”) indicates that values are, within measurement accuracy, the same as the value above.

patient & model		$d$	$\mathbb{E}[d]$	$\sigma_d (\Omega)$		$\mathcal{F}$ opt.	$\mathbb{E}[\mathcal{F}]$ opt.
				CPU	GPU		
intra-cranial	“ray-wise”	10.1	32.1	390.7 ( 64.3)	83.9 ( 14.6)	17.6	9.6
	“beam-wise”	—	—	2,218.8 ( 300.1)	—	—	10.1
para-spinal	“ray-wise”	16.2	40.3	976.5 ( 451.4)	222.9 (107.2)	30.1	47.1
	“beam-wise”	—	—	4,292.8 (1,852.6)	—	—	43.2
prostate	“ray-wise”	31.2	111.2	1,040.2 ( 386.1)	201.5 ( 74.8)	41.8	33.7
	“beam-wise”	—	—	5,365.4 (1,633.1)	—	—	36.3

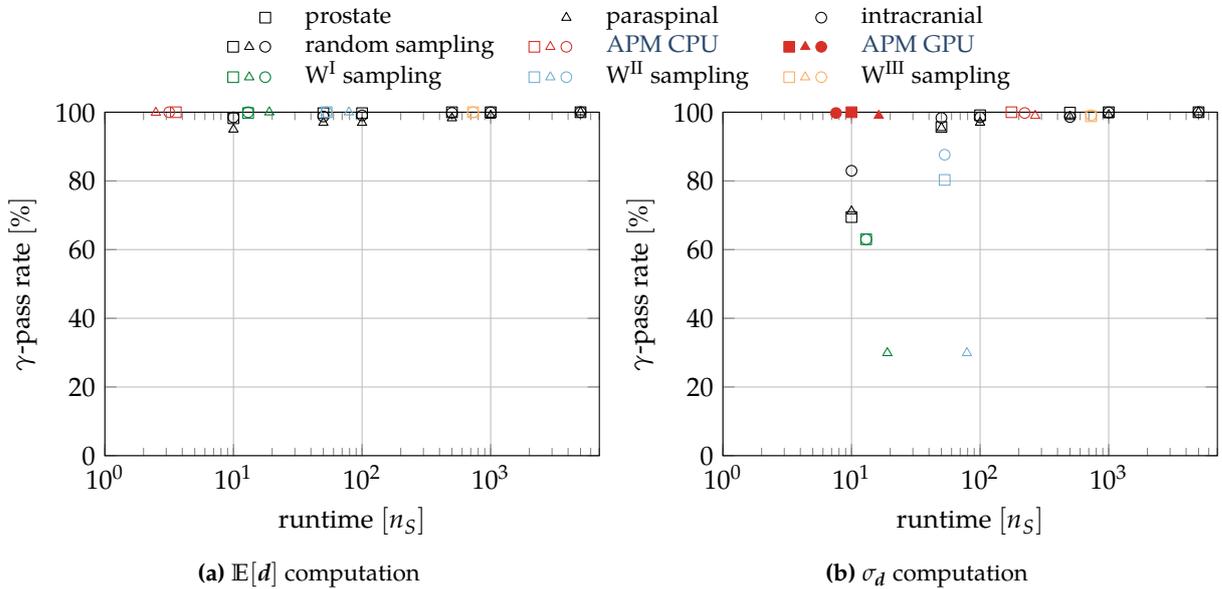
A consequence of neglecting the unclassified normal tissue for optimization is the reduction in computation time by approximately a factor three to eight. It should be noted that this reduction does not directly correspond to the fraction of voxels that is discarded. The APM implementation within MITKrad only computes the contribution of the relevant subset of pencil-beams impinging on the respective voxels, which varies extensively from voxel to voxel, especially in multiple beam plans.

Table IV.4 also proves that APM is particularly suited for massive parallel GPU computations, which can be more than a magnitude faster than on the CPU. Furthermore, GPU computations are not depending on the correlation model in the used kernel implementation (compare Code Listing B.1).

**Accuracy vs. performance analysis:** The runtime analysis in Table IV.4 can be combined with the accuracy analysis in Table IV.1. Figure IV.7 shows the accuracy compared to the runtime equivalent of number of computed dose scenarios  $n_S$ . While the expected dose is accurately modeled by all sampling approaches, APM’s expected value computation clearly outperforms other methods regarding computation time. This is due to the low overhead in APM’s computation of  $\mathcal{D}$  which stays in computational complexity of  $\mathcal{O}(n_V n_B)$ , while computation of the sample mean is of complexity  $\mathcal{O}(n_V n_B n_S)$ .

For the standard deviation in Fig. IV.7b, APM is not, in general, the fastest method (when evaluated on the CPU), however proves to be the fastest in obtaining accuracies near 100%  $\gamma$ -pass rate. Evaluation on the GPU, however, has run-times equivalent to approximately 10 dose calculations, and therefore nearly outperforms even the sparsest and most inaccurate grid sampling approach  $W^I$ . It should be noted that the analyses in Fig. IV.7 have only been performed with the

“beam-wise” correlation model. Using the “ray-wise” correlation substantially decreases runtime of APM (on the CPU) while increasing complexity and therefore runtime of the grid-based sampling approaches. As such, results presented in Fig. IV.7b can be considered a worst case estimation of the speed to accuracy trade-off for APM.



**Figure IV.7:** Accuracy of uncertainty propagation with APM and sampling methods vs. the runtime equivalent of  $n_S$  that need to be calculated. The analysis is shown for the expectation value (a) and the standard deviation of dose (b).

To put the accuracy into more context regarding random sampling, additional  $\gamma$ -tests were performed to obtain the accuracy achieved when using lower numbers of multivariate random samples. Random sampling approaches exhibited nearly perfect  $\gamma$ -pass rates for both expectation and standard deviation when using  $\gtrsim 100$  samples, comparably to computation with APM on the GPU. One has to consider, however, that APM precomputed the helper matrix  $\Omega$  at the same time, allowing optimization with minimal overhead. To achieve the same accuracy with random samples, the dose scenarios need to be re-computed in every iteration. This is computationally inefficient, especially when the dose influence for every scenario cannot be held in RAM due to memory limitations.

### IV.3.4 Feasibility of approximative computations

#### IV.3.4.1 Quality of approximated $\sigma_d$ -distributions

First, the approximative standard deviation distributions  $\hat{\sigma}_d$  are evaluated for different sizes of the pencil-beam combination subsample. Combining methodology used in the analyses in Sections IV.3.2 and IV.3.3 before, the generated approximative distributions are evaluated based on the  $\gamma$ -technique and with runtime measurements.

To illustrate the noise induced by sub-sampling the sum over the pencil-beam combinations explained in Eq. (IV.11), exemplary slices for all cases are shown in Fig. IV.8. Additional to the raw, noisy approximate distributions, they were, as a proof-of-concept, denoised using the total

variation denoising filter implemented in MITK<sup>1</sup> using only 2 iterations and  $\lambda = 0.1$ . In Table IV.5

**Table IV.5:** RMSE [Gy] and  $\gamma$ -pass rates [%] for approximated  $\sigma_d$ -distributions of all cases

patient	processing	$s = 50\%$		$s = 25\%$		$s = 10\%$		$s = 5\%$	
		RMSE	$\gamma$	RMSE	$\gamma$	RMSE	$\gamma$	RMSE	$\gamma$
intracranial	raw	0.15	99.7	0.26	99.0	0.44	97.3	0.64	95.0
	filtered	0.64	94.9	0.64	95.4	0.66	95.0	0.68	94.4
paraspinal	raw	0.20	99.3	0.35	97.1	0.61	93.0	0.84	89.1
	filtered	0.70	95.0	0.70	95.0	0.73	93.7	0.76	92.6
prostate	raw	0.91	85.6	1.30	78.6	1.90	70.5	2.50	64.5
	filtered	0.75	94.3	0.82	92.6	0.97	89.9	1.10	81.0

the root-mean-square deviation (RMSE), used as a noise measure, increases with decreasing sample size, as expected. For the smallest sample size  $s = 5\%$ , the RMSE reaches relative values of up to  $\approx 3 - 4\%$  of the maximum standard deviation. Throughout Table IV.5 the RMSE is highest in the prostate case, and smallest in the intracranial case. This indicates correlation between the spacing of the Bragg-peak grid (compare Table C.1) and RMSE; the accuracy of the sub-sampling increases in cases where more pencil-beams contribute to a voxel.

$\gamma$ -pass rates are comparable or better than the ones achieved with importance sampling approaches in Section IV.3.2. The  $\gamma$  values of the noisy distributions have, however, to be interpreted carefully, since usage of the interpolation method enables coverage a large range of values. This explains the high pass rates of the noisy distribution in, for example, the intracranial case, with different results being published in Wahl et al. (2017) using no interpolation.

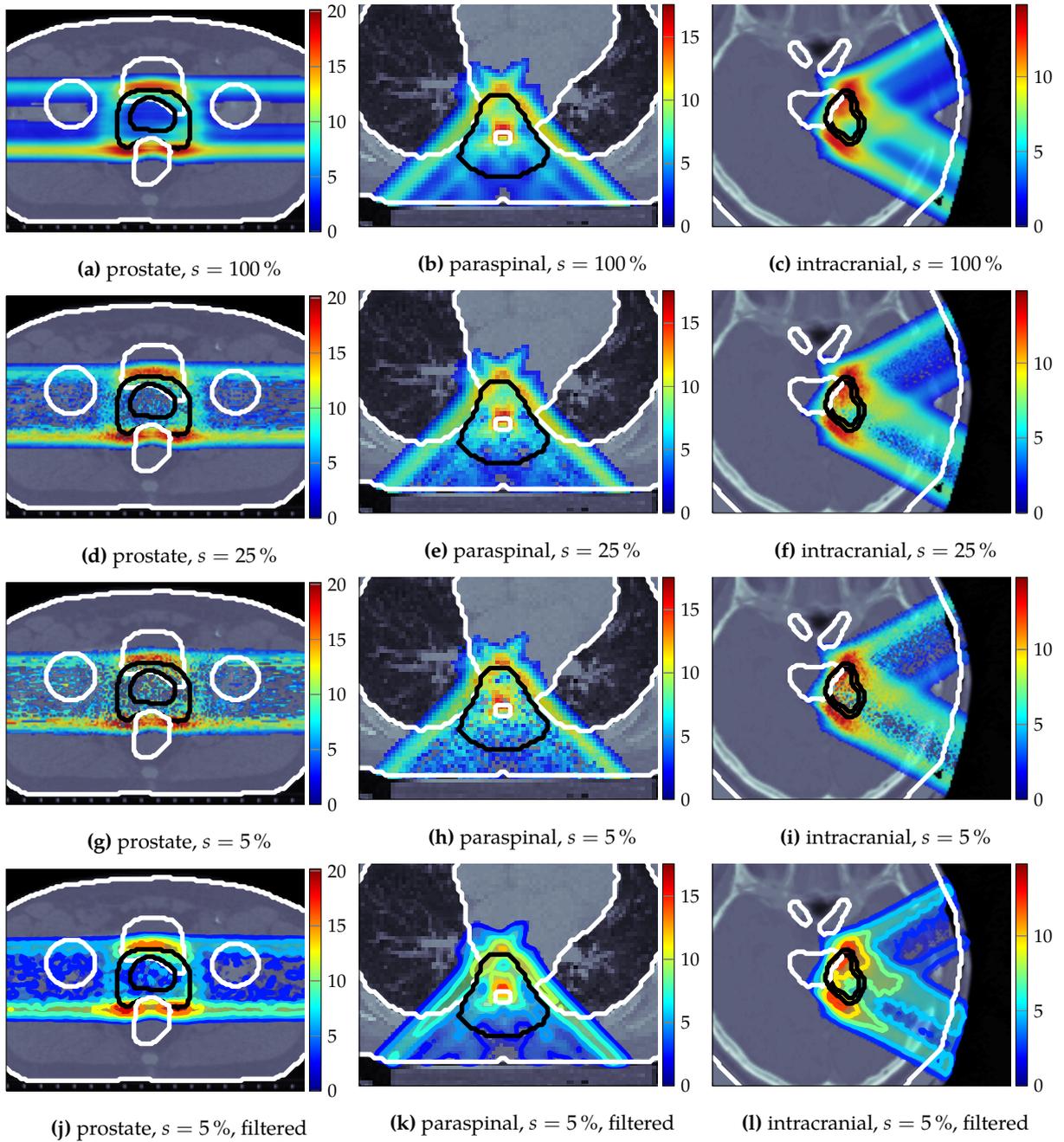
#### IV.3.4.2 Quality of probabilistic plans optimized with approximate $\hat{\Omega}$ -matrix

For plans optimized with the estimated  $\hat{\Omega}$ -matrix from sum sub-sampling, a statistical assessment, e. g. based on RMSE like in Section IV.3.4.1, is inadequate since noise in  $\hat{\Omega}$  systematically influences the optimization routine through its iterations, consequently leading to systematically different treatment plans. Hence, the treatment plans were qualitatively analyzed based on exemplary absolute difference slices, and quantitatively examining difference of resulting plan variability through SDVHs. The analysis is illustrated in Fig. IV.9 comparing plans optimized with exact  $\Omega$  and estimated  $\hat{\Omega}$  at a sampling probability  $s \approx 1.6\%$ .

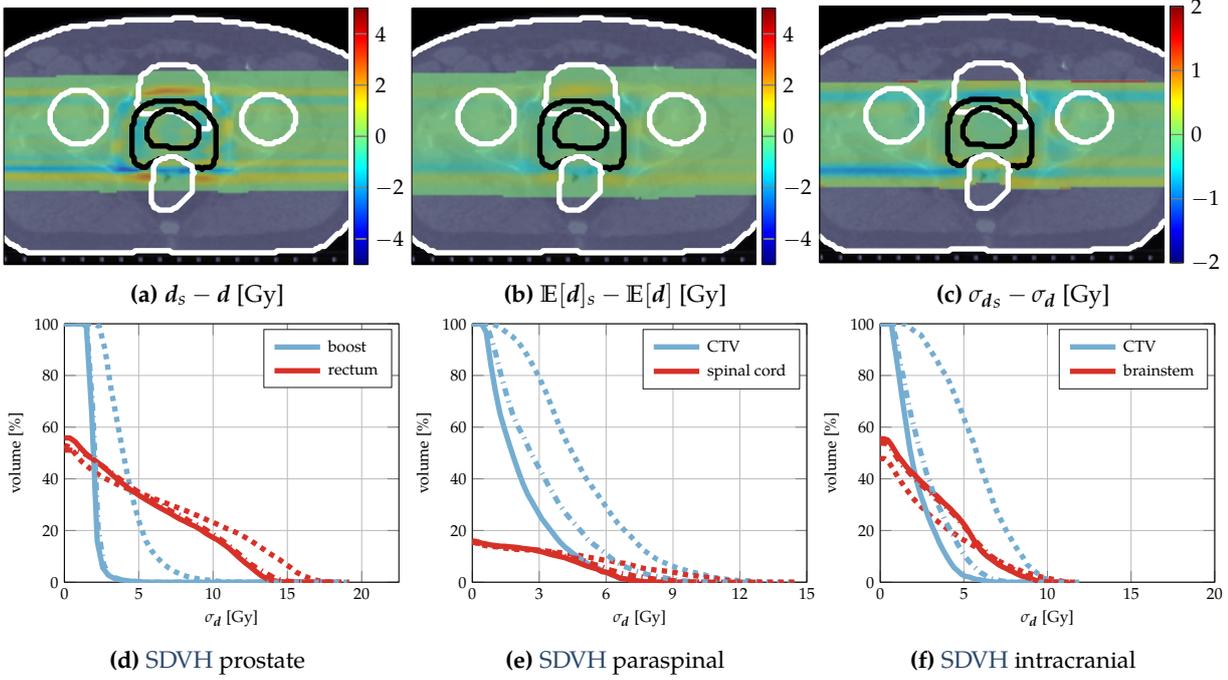
Figures IV.9a to IV.9c show the absolute difference in the nominal, expected and standard deviation of dose distributions for the prostate case, optimized with  $\Omega$  and estimated  $\hat{\Omega}$ . Positive as well as negative differences of up to  $\approx \pm 5\%$  of the prescribed dose manifest in nominal dose through single pencil-beams being differently weighted after optimization with approximated  $\hat{\Omega}$ . Differences are smaller for the expected value. For the standard deviation, differences are below  $\pm 1$  Gy.

In a global assessment of the plan variability using the SDVHs in Figs. IV.9d to IV.9f, global plan variability is slightly higher, especially in the target, when the approximate  $\hat{\Omega}$  is used. For OARs, the effects are negligible. Results published in Wahl et al. (2017), which are based on a

<sup>1</sup>MITK implements the itkTotalVariationDenoisingImageFilter based on Chan, Osher, and Shen (2001).



**Figure IV.8:** Axial slices of  $\hat{\sigma}_d$  distributions (in Gy) for all three cases from exact (a-c) and sampled (d-l) calculations at relative sample sizes  $s$ . In (j-l) the approximated distributions were post-processed with total variation filtering. Note that no isolines are displayed in (a-i) due to the noise in the sampling methods. A single fraction and “ray-wise” correlation was assumed.



**Figure IV.9:** Probabilistic treatment plans optimized with estimated  $\hat{\Omega}$  vs. exact  $\Omega$ . Difference within exemplary distributions slices for the prostate case are shown for nominal dose (a), expected dose (b) and standard deviation (c). The relative sample size was 10% for a single fractionated treatment. For all three cases, (d-f) compare plan variability based on SDVHs when optimized with exact  $\Omega$  (solid — / —) and with estimated  $\hat{\Omega}$  (dash-dotted -·-· / -·-·) to the conventional plan (dotted ···· / ····).

treatment plan optimized under 5 fractions, show that differences vanish when multiple fractions are assumed. This is reasonable, since for a single fraction the variance is higher, and therefore noise in  $\hat{\Omega}$  has a larger impact on the optimization outcome. Hence, for treatment plans subject to large variance, usage of the approximations requires larger sub-samples than for plans subject to low variance to obtain robustness similar to a plan generated with exact computations.

#### IV.3.4.3 Performance of sub-sampling approximations

Using the approximation  $\hat{\Omega}$  from sum sub-sampling, the runtime is expected to decrease linearly with as the relative sample size gets smaller. Hence, instead of providing numerical runtime values, it is evaluated if the runtime  $t_s$  needed to compute the approximation for relative sample size  $s$  follows the linear model

$$s \stackrel{!}{=} a \frac{t_s}{t_{\text{exact}}} + b, \quad (\text{IV.14})$$

where  $t_{\text{exact}}$  is the runtime needed for exact computation, i. e.,  $s = 1$ . Ideal values in Eq. (IV.14) are  $a = 1$  and  $b = 0$ . The same model was evaluated for the approximation of the standard deviation approximation, since the sample is chosen from the off-diagonal pencil-beam combinations  $j \neq m$ , and thus also linearly decreases complexity of evaluating the respective sum. This scenario, however, expects  $b > 0$ , since the diagonal entries  $j = m$  are always evaluated. The resulting value of  $b$  is then a good estimate of the relative run-time needed to evaluate the diagonal, and

can thus be used also used as an estimator of the time that would be required to evaluate *APM* on a completely uncorrelated uncertainty model (compare Fig. III.2a).

Linear regression of measured relative run-times comparing to the sample size, averaged over all cases, found

$$\Omega\text{-approximation: } a = 0.99 \quad b = 1.02 \times 10^{-2},$$

$$\sigma_d\text{-approximation: } a = 0.93 \quad b = 7.33 \times 10^{-2},$$

i. e., nearly linear decrease when sub-sampling voxels, and more significant run-time overhead when sub-sampling pencil-beam combinations.

## IV.4 Discussion

This chapter described the successful implementation and evaluation of a probabilistic pencil-beam dose calculation engine for proton therapy based on the *APM* framework. As backbone for the implementation of *APM*, a module for *MITK*—called *MITKrad*—has been implemented from scratch, providing essential treatment planning functionality. *MITKrad*'s structure was designed in an effort to maximize code re-usability and encourage sustainable use and extendability. Relying on optimized linear algebra structures and operations from the *Eigen* package and the optimizer *IPOPT* yielded competitive run-times for dose calculation as well as optimization (compare to e. . Wieser et al. 2017b). *MITKrad*'s objective oriented design allowed straightforward implementation of the *APM* dose calculation and base data. For example, the nominal dose calculation can be performed with the same dose engine for common tabulated depth dose data and *APM*'s representation, for which only the base data interface needed to be re-implemented. Most scientific programs used across this thesis for treatment plan evaluation were hard-coded using *MITKrad*'s interface for sub-programs. However, basic generation of treatment plans and robustness analysis with *APM* is possible through a basic GUI widget allowing to set treatment plan parameters such as beam geometry and optimization objectives.

*MITKrad* enabled efficient implementation of *APM*'s analytical probabilistic dose calculation, which would have hardly been realizable within *MITKrad*'s *MATLAB*-“twin” *matRad* (compare, for example, to the *MATLAB* run-times for *APM* of several hours reported by Wieser et al. (2017b)). This is due to the increased complexity of the computations of the second moment going beyond the thoroughly optimized linear algebra operations with matrices and vectors in *MATLAB*. The “embarrassingly parallel” nature of the element-wise Gaussian kernels could be exploited, especially within *GPGPU*, achieving run-times equivalent to a few nominal dose calculations. Together with complementary results from Wahl et al. (2017), good scalability of the implementations is attested, encouraging future use in distributed computer systems—maybe even based on commercial cloud systems. Note that through modification of the *CUDA*-kernel in Code Listing B.1 to only accepting standard data-types, it could be called directly from *MATLAB* using its `CUDAKernel` interface. Alternatively, the *MEX*-interface may be used to build an interface between, for example, *matRad* and faster implementations from *MITKrad*.

From the physical side, the probabilistic plans shown in Section IV.3.1 which were generated

with the *APM* implementation in *MITKrad* are in accordance with literature results discussing similar probabilistic treatment plan analysis and optimization approaches (i. e., Unkelbach, Chan, and Bortfeld 2007; Unkelbach et al. 2009). The *APM* computations proved to be accurate, achieving global  $\gamma$ -pass rates of nearly 100 % in  $\gamma$ -validation tests at criteria of 3 %/3 mm and 2 %/2 mm of expectation value and standard deviation of dose. The numbers presented in Section IV.3.2 are, at least for the standard deviation, at the limit of the relative accuracy of the “ground truth”, which was obtained using 5,000 multivariate normal random samples using the “beam-wise” correlation model. The work of Wieser et al. (2017a), extending *APM* to biological dose in carbon therapy, confirm these accuracy measurements, suggesting that the *APM* framework can be seen as providing highly accurate estimates of the first two moments of the dose distribution. Consequently, it might be even considered as a benchmark for prototyping other, new uncertainty propagation methods using a pencil-beam dose computation engine.

*APM*’s strength particularly manifests in the accuracy to runtime trade-off. An analysis in Fig. IV.7 shows that computation of expected value as well as standard deviation is magnitudes faster than the computation of a comparably accurate sample statistics, and even outperforms sparse grid sampling approaches as used in worst case approaches that are at the very heart inaccurate through neglecting spatial combinatorics. However, further comparisons with sampling relying on *GPGPU* accelerated dose calculation algorithms might mitigate the observed run-time benefits of using *APM*.

In this context, comparison to a more accurate reference, as for example a Monte Carlo dose calculation algorithm, would also be of interest. While pencil-beam algorithms and Monte Carlo algorithms are often compared regarding nominal accuracy and then discouraged (e. g. Taylor, Kry, and Followill 2017), it is unclear whether their inferiority to Monte Carlo amplifies or dissolves when uncertainty shall be estimated. Monte Carlo implementations are, however, comparably slow, such that computation of sample numbers necessary for benchmarking would be an ordeal.

Since especially the *GPGPU* implementation of the standard deviation computation is independent of the underlying correlation assumptions in the uncertainty model, highly complex uncertainty models may be accurately evaluated, which could barely be evaluated using a combinatorial grid sampling approach. On the other hand, *APM*’s uncertainty model is, in its current implementation, restricted to lateral shifts in the *BEV* and errors in *WEPL*. Modeling of other kinds of uncertainties, i. e., deformable anatomies, would require a different approach, presumably in the patient coordinate frame. While this could be, in theory, also realized within the *APM* framework, the computational implications can, at this point, not be explicitly quantified.

As discussed and shown before, *APM*’s implementation already exhibits reasonable run-times on a desktop computer using a common end-user *GPU*. Nevertheless, additionally two approximation techniques based on sum sub-sampling were proposed to linearly trade runtime against accuracy. These enable estimation of standard deviation as well as the helper matrix  $\Omega$  used for optimization.

The estimated standard deviation distributions exhibit visible noise, but may be used for a qualitative assessment of regions with high uncertainty in dose. This is particularly convenient,

as with more dense placement of Bragg-peaks the sub-sampling accuracy increases, which may allow further reduction of sample size in these especially computationally expensive scenarios. As proof-of-concept, also denoised distributions were produced using total variation filtering. Since the magnitude of the noise can, however, be quantified through the sample and population size, more sophisticated denoising methods may be used explicitly working with the noise model, to achieve even better approximations.

Using the estimated  $\hat{\Omega}$ -matrices in optimization generated plans with different modulation, while achieving similar to slightly worse plan variability, for single fractioned treatments. These results may be discussed complementary to the results from Wahl et al. (2017), where treatments with 5 fractions were investigated while using the same uncertainty model, with even smaller difference between plans optimized with  $\hat{\Omega}$  and  $\Omega$ . Hence, generation of probabilistic treatment plans using  $\Omega$  estimates from APM may be exploited within standard uncertainty assumptions for fractionated treatments. However, in a case where only a single fraction should be optimized, or the plan is subject to uncommonly large uncertainty, perfect computations should be preferred. To suppress random modulation effects, which might lead to under- or overdosage in undesired areas, voxels could be sub-sampled on, for example, regular grids. Alternatively, additional objectives or non-linear (in)equality constraints are possible within the optimization framework using IPOPT, to impose restrictions on the expected or nominal dose distribution.

The previous paragraphs already addressed important issues; fractionation as well as extension of the optimization. The effect of fractionation on dose uncertainty and probabilistic optimization will be thoroughly examined in the following chapter. The transition to analytical probabilistic modeling of advanced planning metrics and objective and constraint functions is then the topic of Chapters VI and VII.

# Probabilistic Treatment Planning Under Fractionation

In this chapter *APM*'s capabilities of incorporating the non-trivial interplay of random and systematic setup and range errors in fractionated treatments is investigated. More than a plain implementation of the analytic computations explained in Section III.6.3, the model is refined to completely separate covariance influence into independent random and systematic components, generalizing to uncertainty analysis and optimization. Further, a simulation study is conducted to quantify the benefits of using probabilistic optimization with explicit consideration of the number of fractions.

In general, the methodology and the results discussed in this chapter have already been published in Wahl et al. (2018a). Differences between this chapter and the respective publications arise due to re-computation of results for consistency across this manuscript, and due to supplementary results omitted or not yet presented in Wahl et al. (2018a).

Also note that the methodology introduced in Section V.2 has been made publicly available as *MATLAB* code within the *APMToolbox*.

## V.1 Overview

Chapter IV demonstrated that the *APM* framework is suited for computations on patient data sets, featuring competitive run-times and accuracy encouraging its use within or on top of typical treatment planning workflows to facilitate probabilistic plan analysis and optimization.

A key attribute of *APM*'s computations thereby is that random and systematic uncertainties may be included into the uncertainty model at constant computational complexity, as explained in Section III.6.3. This inclusion of (arbitrary) interplay of random and systematic errors in the uncertainty model is a non-trivial problem investigated in academia and usually worked around, due to the overhead for conventional sampling approaches (compare Section II.3.2). Difficulties arise especially when using fractionation in optimization, where often re-computations of the respective models are necessary within optimization under fractionation. Therefore, worst-case or probabilistic optimization is often only tackled for a single fraction (e. g. Unkelbach et al. 2009). If fractions are considered, simplifications on the uncertainty model are imposed, as e. g. the afore-

mentioned strict separation by Lowe et al. (2016) or assumption of infinite fractions. The latter is based on the rapid decrease of the random component in dose uncertainty with increasing fraction number (see e. g. Park et al. 2013; Perkó et al. 2016), however requiring additional countermeasures against the resulting dose heterogeneity when used in optimization (Chan, Bortfeld, and Tsitsiklis 2006; Fredriksson 2012; Unkelbach and Oelfke 2004). For example, Lowe et al. (2016) and Lowe et al. (2017) separate by dimension into the spatial error components being *either* of random *or* of systematic nature. Eventually, the studies actually incorporating a larger number of multiple randomly sampled combinations align themselves with “Big Data” approaches to derive general insights on specific treatment sites (Kraan et al. 2013; Park et al. 2013).

Probabilistic approaches, on the other hand, exhibit the possibility to intrinsically incorporate this interplay. Perkó et al. (2016) managed to “compress” a PCE to incorporate random and systematic errors. However, still several hundred scenario computations are required while constructing the PCE with numerical integration, yet without generalizing to optimization. Also the methodology proposed by Sobotta, Söhn, and Alber (2010) and Sobotta, Söhn, and Alber (2012) may allow intrinsic incorporation. Their generalization to probabilistic or robust optimization under fractionation is, however, not trivial.

With APM, this generalization can directly be made for optimization with the expected least-squares objective from Eq. (III.22) using the precomputed helper matrix  $\Omega$  as in Eq. (III.23). The latter is precomputed based on the covariance influence elements, for which Bangert, Hennig, and Oelfke (2013) and Wahl et al. (2017) already demonstrated feasibility of computing and optimizing uncertainties for a desired fraction number at constant time complexity. Yet the formulations by Bangert, Hennig, and Oelfke (2013), recapped in Section III.6.3, are not the end of the road, and further optimizations of the model are possible.

As a matter of fact, this chapter will introduce a reformulation of the fraction number dependence allowing a separation of the covariance influence into a linear combination of random and systematic covariance influence components in Section V.2. While these reformulations stay in constant time complexity at none to limited overhead, they enable computation of the *complete* fractionation spectrum of the covariance influence, with generalization to dose (co)variance and the helper matrix  $\Omega$ . To evaluate these computations, a random sampling study under fractionation (similar to Section IV.2.3.1 but with fewer treatment samples) is set up to quantify the effect of consideration of the fraction number on the outcome of probabilistic optimization. Within Section V.3 the uncertainty spectrum of the non-trivial interplay of random and systematic uncertainties under fractionation is then evaluated. The results of the random sampling studies are used to quantify the benefit of considering the fraction number over just applying a probabilistic plan for a single fraction in multiple fractions. Finally, Section V.4 discusses the results and concludes this chapter.

## V.2 Methods

### V.2.1 The linear covariance influence model

Section III.6.3 shows that for incorporation of fractionation effects, the raw contribution to the spatial covariance influence elements need to be decomposed into a “correlated” and “uncorrelated” component, modeling correlations across fractions and only within a single fraction, respectively.

For a covariance influence element  $\mathcal{C}_{ijlm}$ , the raw contribution  $\tilde{\mathcal{C}}_{ijlm}$  subject to a number of fractions  $n_f$  decomposes into

$$\mathcal{C}_{ijlm} = \underbrace{\frac{\tilde{\mathcal{C}}_{ijlm}^{\text{corr}} + (n_f - 1)\tilde{\mathcal{C}}_{ijlm}^{\text{uncorr}}}{n_f}}_{\tilde{\mathcal{C}}_{ijlm}} - \mathcal{D}_{ij}\mathcal{D}_{lm}. \quad (\text{V.1})$$

Therein,  $\tilde{\mathcal{C}}^{\text{corr}}$  contains the full correlation model that holds within a single fraction. The remaining fractions are described by  $\tilde{\mathcal{C}}^{\text{uncorr}}$  which considers the full correlation model only for systematic errors while treating random errors as completely uncorrelated. Both compose the raw contribution  $\tilde{\mathcal{C}}$  to the full covariance influence  $\mathcal{C}$  as defined in Eq. (III.18). The last term in Eq. (V.1) then subtracts the mixed expectation value of dose influence  $\mathcal{D}_{ij}\mathcal{D}_{lm}$ .

Theoretically,  $\tilde{\mathcal{C}}^{\text{corr}}$  and  $\tilde{\mathcal{C}}^{\text{uncorr}}$  would only require a single precomputation, from which  $\mathcal{C}$  could then be computed dynamically for any number of fractions  $n_f$  with available expected dose influence matrix  $\mathcal{D}$ .

However, when expanding Eq. (V.1) to

$$\mathcal{C}_{ijlm} = \frac{1}{n_f} \cdot \left[ \tilde{\mathcal{C}}_{ijlm}^{\text{corr}} - \tilde{\mathcal{C}}_{ijlm}^{\text{uncorr}} \right] + \left[ \tilde{\mathcal{C}}_{ijlm}^{\text{uncorr}} - \mathcal{D}_{ij}\mathcal{D}_{lm} \right], \quad (\text{V.2})$$

denoting  $\mathcal{C}_{ijlm}^{\text{rand}} = \tilde{\mathcal{C}}_{ijlm}^{\text{corr}} - \tilde{\mathcal{C}}_{ijlm}^{\text{uncorr}}$  and  $\mathcal{C}_{ijlm}^{\text{sys}} = \tilde{\mathcal{C}}_{ijlm}^{\text{uncorr}} - \mathcal{D}_{ij}\mathcal{D}_{lm}$ , those can be linearly combined in the much more clean formulation

$$\mathcal{C}_{ijlm} = \frac{1}{n_f} \cdot \mathcal{C}_{ijlm}^{\text{rand}} + \mathcal{C}_{ijlm}^{\text{sys}}. \quad (\text{V.3})$$

In Eq. (V.3),  $\mathcal{C}^{\text{rand}}$  now describes the covariance influence caused by random errors, while  $\mathcal{C}^{\text{sys}}$  describes the systematic contributions. This can be deducted from above definitions of  $\tilde{\mathcal{C}}^{\text{uncorr}}$  and  $\tilde{\mathcal{C}}^{\text{corr}}$ ; in  $\mathcal{C}^{\text{rand}} = \tilde{\mathcal{C}}^{\text{corr}} - \tilde{\mathcal{C}}^{\text{uncorr}}$ , all systematic components subtract to zero, while  $\tilde{\mathcal{C}}^{\text{uncorr}}$  already considers only systematic correlation assumptions.

While this reformulation may seem merely as a mathematical technicality, it facilitates APM’s computations for fractionated treatments in certain aspects. Besides obtaining a more illustrative description of the impact of random and systematic uncertainties, for example, one can directly investigate the infinite fraction assumption  $n_f \rightarrow \infty$ . Also generalization to the dose covariance and to optimization is straightforward, as explained below in Section V.2.2.

## V.2.2 Generalization to probabilistic analysis and optimization

The linear model derived in Eq. (V.3) separates into two distinct summands. This generalizes to the sum from Eq. (III.19) used to compute (co)variance components

$$\begin{aligned} \text{Cov}[d_i, d_l](n_f) &= \sum_{jm} w_j w_m \left[ \frac{1}{n_f} \mathcal{C}_{ijlm}^{\text{rand}} + \mathcal{C}_{ijlm}^{\text{sys}} \right] = \frac{1}{n_f} \underbrace{\sum_{jm} w_j w_m \mathcal{C}_{ijlm}^{\text{rand}}}_{\Sigma_d^{\text{rand}}} + \underbrace{\sum_{jm} w_j w_m \mathcal{C}_{ijlm}^{\text{sys}}}_{\Sigma_d^{\text{sys}}} \\ \Rightarrow \Sigma^d(n_f) &= \frac{1}{n_f} \Sigma^{\text{rand}} + \Sigma^{\text{sys}} \end{aligned} \quad (\text{V.4a})$$

$$\Rightarrow \sigma_d^2(n_f) = \frac{1}{n_f} (\sigma_d^{\text{rand}})^2 + (\sigma_d^{\text{sys}})^2, \quad (\text{V.4b})$$

such that random and systematic covariance components  $\Sigma^{\text{rand}}$  and  $\Sigma^{\text{sys}}$  and variance components  $(\sigma_d^{\text{rand}})^2$  and  $(\sigma_d^{\text{sys}})^2$ , respectively, can be obtained.

Similarly, the sum to compute the helper matrix  $\Omega^v$  described in Eqs. (III.23) and (IV.1) factorizes into its the random and systematic component  $\Omega^{v,\text{rand}}$  and  $\Omega^{v,\text{sys}}$ :

$$\begin{aligned} \Omega_{jm}^v(n_f) &= \sum_{i \in v} \left[ \frac{1}{n_f} \mathcal{C}_{ijim}^{\text{rand}} + \mathcal{C}_{ijim}^{\text{sys}} \right] = \frac{1}{n_f} \underbrace{\sum_{i \in v} \mathcal{V}_{ijim}^{\text{rand}}}_{\Omega_{jm}^{v,\text{rand}}} + \underbrace{\sum_{i \in v} \mathcal{V}_{ijim}^{\text{sys}}}_{\Omega_{jm}^{v,\text{sys}}} \\ \Rightarrow \Omega^v(n_f) &= \frac{1}{n_f} \Omega^{v,\text{rand}} + \Omega^{v,\text{sys}}. \end{aligned} \quad (\text{V.5})$$

When using APM for uncertainty propagation, the separations exercised from Eqs. (V.1) to (V.5) are especially convenient; separated terms can be computed independently and on-the-fly for any requested set of voxels or pencil-beams. Hence, from a single run of APM's computations of expected value and (co)variance, the full fractionation spectrum of (co)variance is available. Using then Eq. (V.4a) or Eq. (V.4b), (co)variance for any fraction number—including the assumption of infinite fractions—may be queried and used for further analysis.

Using  $\Omega^v(n_f)$  in probabilistic optimization of the expected objective function in Eq. (III.22) induces dependence on the number of fractions  $n_f$  in the variance term (see Eq. (III.23)). This means that for a change in  $n_f$  Eq. (III.22) needs to be re-optimized, but since  $\Omega^v(n_f)$  can be separated with Eq. (V.5), no re-computation of the uncertainty mapping is needed between optimization runs.

### V.2.2.1 Use within other methods

Equations (V.1) to (V.5) apply equally to other methods capable of evaluating elements of  $\mathcal{C}$  for the same correlation model for fractionation. If instead of  $\mathcal{C}$  only a derived quantity like  $\sigma_d^2(n_f)$  is available for several fraction numbers, Eq. (V.4b) would still be valid and the respective systematic and random components could be inferred from the derived quantities.

For example, assume the case of common estimation of  $\sigma_d^2$  through sampling methods, when

$\sigma_d^2(n_f = n_f^p)$  is available for  $P$  fractionation schemes indexed by  $p$ . Looking at individual voxels  $i$ , the relationship of estimates  $\sigma_{d_i}^2(n_f^p)$  and its random and systematic components can be represented as linear system

$$\begin{pmatrix} \sigma_{d_i}^2(n_f^1) \\ \sigma_{d_i}^2(n_f^2) \\ \vdots \\ \sigma_{d_i}^2(n_f^P) \end{pmatrix} = \begin{pmatrix} 1 & (n_f^1)^{-1} \\ 1 & (n_f^2)^{-1} \\ \vdots & \vdots \\ 1 & (n_f^P)^{-1} \end{pmatrix} \begin{pmatrix} \sigma_{d_i}^{\text{sys}2} \\ \sigma_{d_i}^{\text{rand}2} \end{pmatrix} + \epsilon \quad (\text{V.6})$$

with error  $\epsilon \in \mathbb{R}^P$ .

A straightforward approach to solve model (V.6) is linear regression by, for example, least squares/minimization of  $|\epsilon|^2$  or matrix inversion algorithms. Exercised for each voxel  $i$ , an estimate for  $(\sigma_d^{\text{sys}})^2$  and  $(\sigma_d^{\text{rand}})^2$  may be obtained.

Note that this approach will just be used qualitatively for comparison, and is thus not further developed; a quantitative approach would, at least, need to constrain the random and systematic comparison to the positive domain. Further implications will be discussed in Section V.4

### V.2.3 Random sampling study

To quantitatively analyze the methodology derived above, a random sampling study was set up, similar to the random sampling benchmark in Section IV.2.3.1, but with reduced sample number. Using less samples is a consequence of the increased complexity under fractionation; each treatment sample is now the accumulated dose of in turn randomly sampled fractions  $n_f$ . This now requires  $n_S = n_T \cdot n_f$  sample dose calculations in total, where  $n_T$  is the number of desired (fractionated) treatment samples.

In the investigated scenario, i. e., random and systematic uncertainties for lateral pencil-beam displacement and WEPL under a static anatomy, the accumulated dose for a treatment sample  $d_t$  is simply given by

$$d_t = \sum_{f=1}^{n_f} d_{t,f} \quad (\text{V.7})$$

with the respective fraction dose of the treatment sample  $d_{t,f}$ . The fraction dose samples  $d_{t,f}$  are computed by first drawing a displacement sample from the multivariate distribution describing systematic uncertainties, proceeding from which then  $n_f$  samples are drawn from the respective distribution describing the random uncertainties. Sample statistics for the fractionated treatment can then be deduced from all sampled treatment doses  $d_t$  similar to Section IV.2.3.1.

The sampling study shall aim at investigating the behavior of dose uncertainty for (a) a given treatment dose that may be applied in different fractionation schemes, (b) treatment plans that are probabilistically optimized under fractionation, (c) probabilistic treatment plans optimized for a single fraction, but applied in multiple fractions. To derive this information, one conventional treatment plan shall be generated, as well as probabilistic treatment plans for a set of  $p$  increasing

fraction numbers  $\{n_f^1, n_f^2, \dots, n_f^p\}$ , such that  $n_f^p$  is the maximum number of investigated fractions. For each single one of this treatment plans, now treatment samples with  $n_f^p$  fractions are computed. This enables the analysis of all  $p + 1$  created treatment plans considering fractionation schemes from the interval  $[1, n_f^p]$ . It should be pointed out that this approach entails autocorrelations in the fraction samples, since the accumulated dose up to a fraction depends on the previous fractions. Between sufficiently small subsets of fractions, this autocorrelation might be avoided by choosing the subsets from different fraction samples.

## V.3 Results

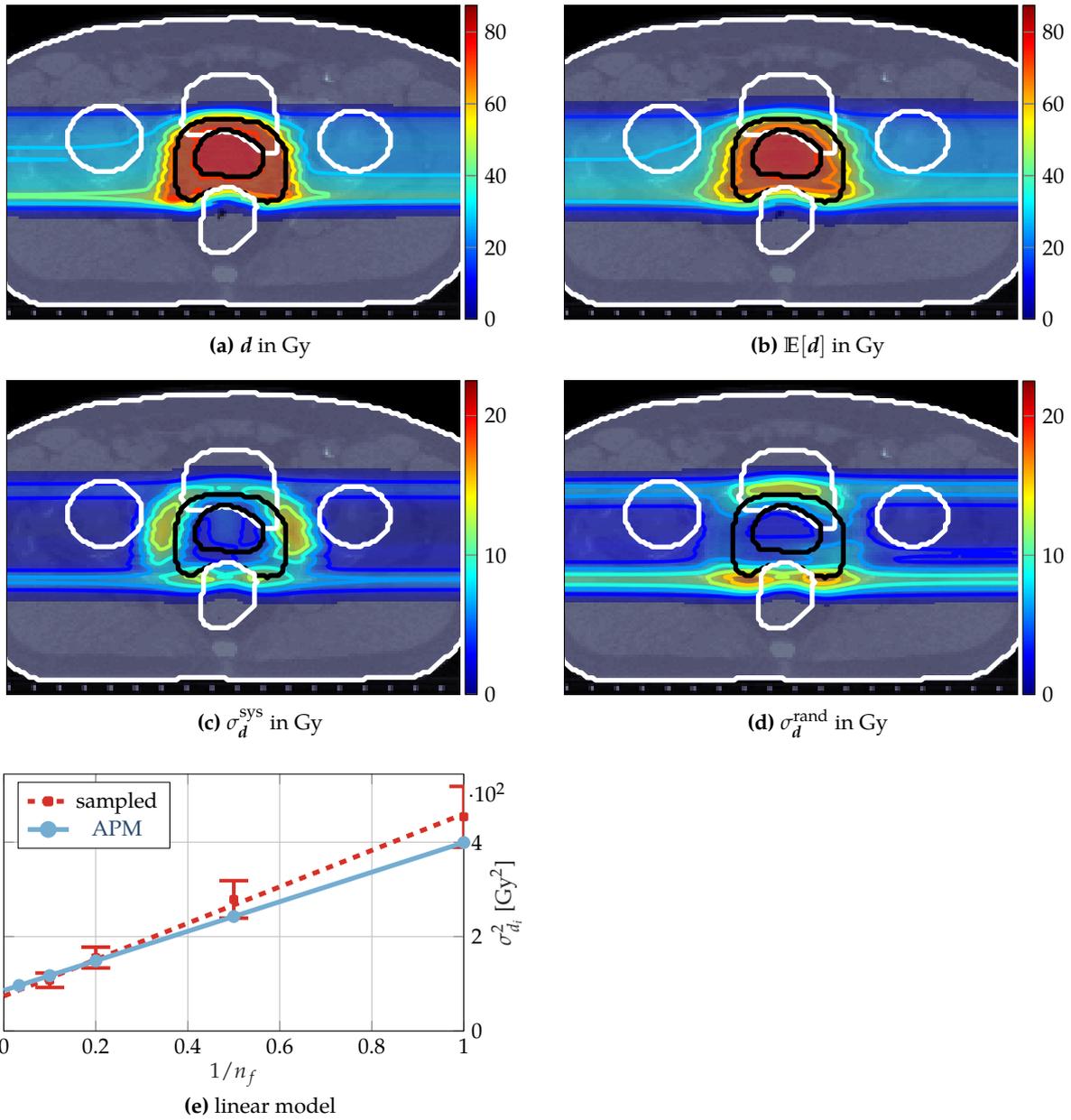
This section presents results based on similar uncertainty assumptions as used for the validation in Chapter IV. That is, range and set-up errors are perfectly correlated for pencil-beams belonging to the same beam, i. e., impinging from the same gantry angle, and uncorrelated otherwise. The fractionation model was added, i. e., random errors are completely uncorrelated across different fractions, while systematic errors exhibit perfect correlation. The magnitude of the assumed uncertainties as well as parameters of the patient data and treatment plans is given in Table C.1.

### V.3.1 Uncertainty analysis of the full fractionation spectrum

The separated variance from Eq. (V.4b) could be computed within a single run of APM's covariance computation routine. As exemplary illustration, Fig. V.1 depicts the prostate case, similar to Fig. IV.3, but this time constructing the random and systematic component of standard deviation. Further, in Fig. V.1e the linear model is illustrated (in the maximum variance voxel), comparing APM's perfectly linear variance modeling with the result from linear regression (compare Eq. (V.6)) using  $n_T = 100$  treatment samples according to Section V.2.3. APM's linear variance model is then used to construct fractionated dose cubes, shown in Fig. V.1f.

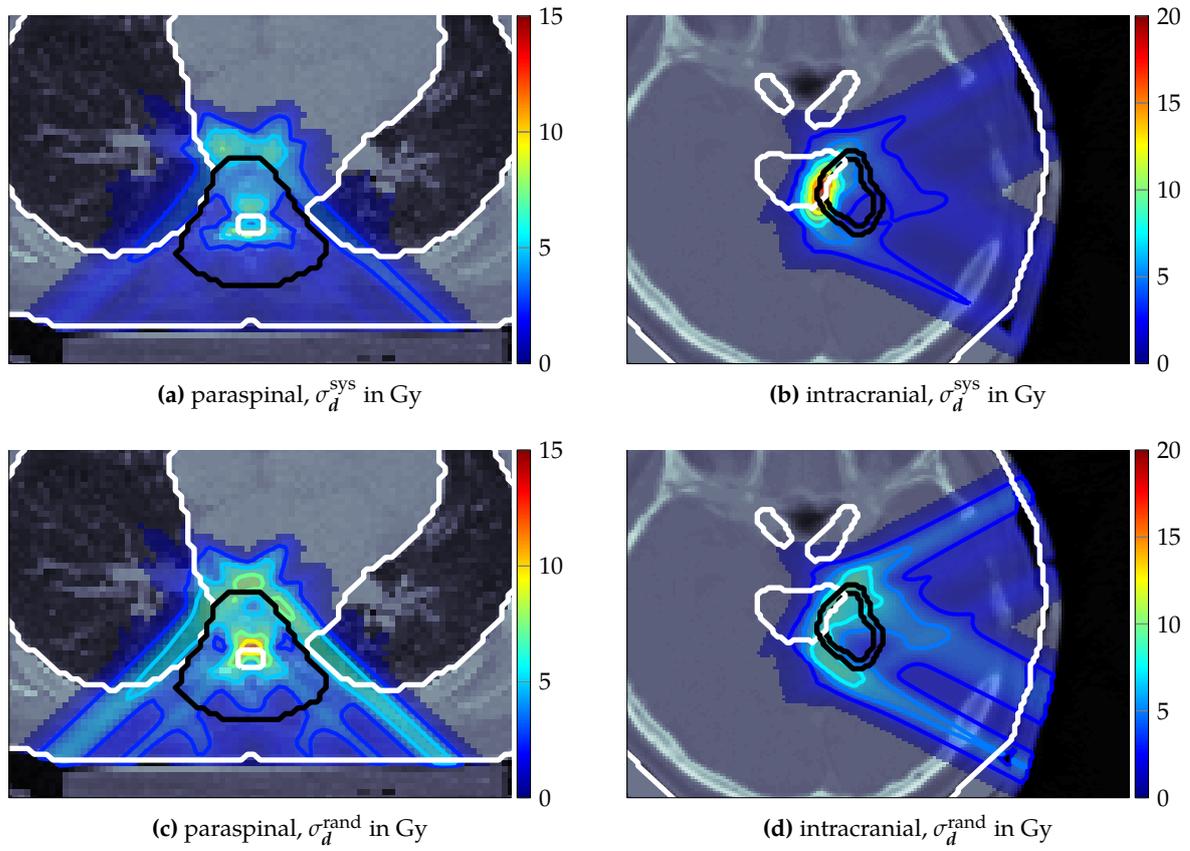
The separated components in Figs. V.1c and V.1d exhibit different structure, according to the respective assumption made in the input uncertainty space. Systematic range uncertainties dominate the picture of Fig. V.1c, manifesting in the peak uncertainties at the distal tumor edge. Random lateral set-up errors lead to a tubular high-variance enclosure of the beams' dose, especially pronounced near the rectum and bladder where the dose is high. Figure V.1f shows that under fractionation, with increasing fraction number the random error quickly loses impact due to possible cancellation of random error scenarios across multiple fractions.

The relative contribution of random and systematic dose uncertainty to uncertainty of the full fractionated treatment is different across patient anatomies and beam geometries. To illustrate this behavior, the (separated) standard deviation is also shown in Fig. V.2 for the paraspinal and the intracranial case by means of exemplary slice distributions. For the intracranial case, the systematic error has a large impact on the composed full standard deviation, which can be attributed to both beams impinging from the same lateral half circle with only  $60^\circ$  separation, i. e., "sharing" their distal tumor edge. The paraspinal case exhibits comparable influence from both components.



(This figure is animated in the electronic PDF version of this thesis.)

**Figure V.1:** Exemplary axial slice of nominal dose (a), expectation value (b) and corresponding separated systematic (c) and random (d) component of the standard deviation. With the linear variance model—depicted in (e) for the maximum variance voxel—it is possible to deduce the standard deviation for the desired number of fractions ( $f$ ). The maximum variance voxel is indicated by a square marker in the distributions. In (e) the with APM computed variance for  $n_f$  fractions (blue points) is compared against the sample variance (orange squares, error bars correspond to  $\pm 1\sigma$ ). Additionally, the respective linear models computed on-the-fly by APM (blue line) and based on linear regression on the sample statistics (orange dashed line) are shown.

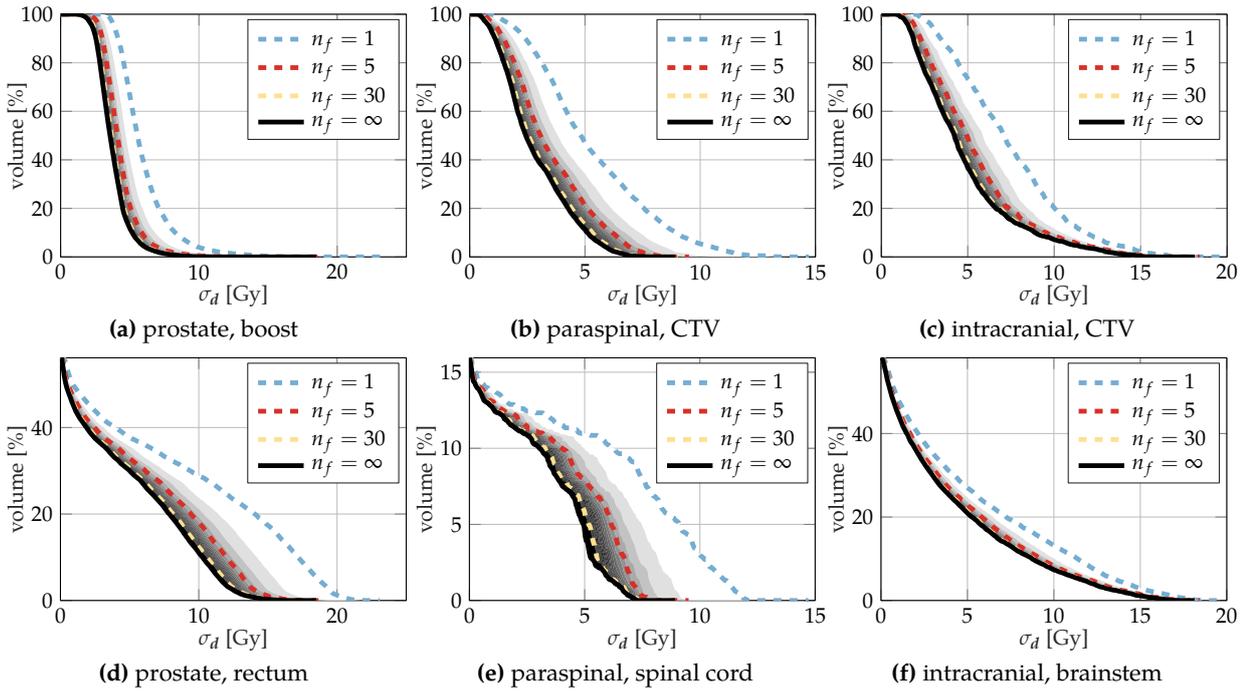


(This figure is animated in the electronic PDF version of this thesis.)

**Figure V.2:** Exemplary slices of separated random and systematic standard deviation of dose for the paraspinal (a,c,e) and the intracranial case (b,d,f). Similarly presented in Wahl et al. (2018a).

A more quantitative analysis of the behavior of the standard deviation of dose with changing fraction number can be done by using fractionated standard deviation volume histograms (FSDVHs). Figure V.3 shows FSDVHs for a target and an OAR for all three patient cases across the complete fractionation spectrum. The analysis underpins the first insights gained from the qualitative view at exemplary slices in Figs. V.1 and V.2. With increasing fraction number, the overall standard deviation decreases rapidly towards the systematic contribution, such that a treatment with 30 fractions has almost no sensitivity to random errors, i. e., is nearly similar to a treatment with infinite fractions.

This confirms other analyses of the behavior of dose variability subject to random errors by Lowe et al. (2016), Park et al. (2013), and Perkó et al. (2016).



**Figure V.3:** FSDVHs for 1, 5, 30 and theoretically  $\infty$  fractions of all cases for a target (top row) and OAR (bottom row), respectively. Furthermore, the complete fractionation spectrum is illustrated by the gray-shaded area, in which color transitions to a darker shade correspond to the SDVHs of a treatment with one more fraction.

### V.3.2 Probabilistic fractionated treatment plans

As the next step, the APM framework was used to optimize the expected objective function (III.23) to generate treatment plans for  $n_f = \{1, 5, 30\}$  fractions, using the helper matrix  $\Omega^v$ . Based on the separation of  $\Omega^v$  (for the respective VOIs  $v$ ) given in Eq. (V.5), only a single computation of the random and systematic components  $\Omega^{v,\text{sys}}$  and  $\Omega^{v,\text{rand}}$ , respectively, was required to perform all three probabilistic optimizations. After the probabilistic optimizations, for each plan  $\mathbb{E}[d]$  as well as the standard deviation components  $\sigma_d^{\text{rand}}$  and  $\sigma_d^{\text{sys}}$  were computed.

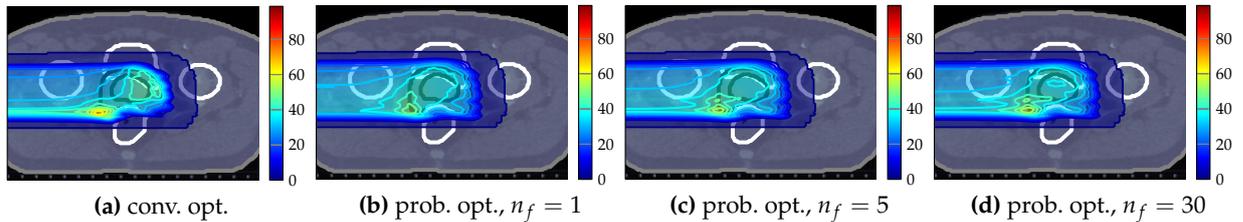
Furthermore, for all treatment plans, i. e., the conventional and the three probabilistic ones, a random sampling study was performed as explained in Section V.2.3 with  $n_T = 100$  treatment

samples. For all sampled treatments, the respective DVHs were computed to derive  $\alpha$ -DVHs (compare Eq. (II.26)).

### V.3.2.1 Effect of fractionation on fluence modulation

Compared to Section IV.3.1, where plans optimized for  $n_f = 1$  have been shown, optimization with  $n_f > 1$  decreases contribution of the variance term in Eq. (III.22), while especially the random component's relative contribution decreases with increasing  $n_f$  in Eq. (V.5). As random components are uncorrelated through the treatment, this enables independent modulation of  $w$  in optimization of Eq. (III.22) without substantial increase in variance.

Figure V.4 demonstrates this effect by showing the dose contribution for the 270° beam for the prostate case. In conventional optimization, the beam is sharply modulated, especially in the



**Figure V.4:** Comparison of prostate beam modulations achieved by conventional and probabilistic optimization under fractionation. The 270° beam's nominal dose [Gy] for the conventional plan (a) and probabilistically optimized plans (b-d) is shown with  $n_f = \{1, 5, 30\}$  fractions. Similarly presented in Wahl et al. (2018a).

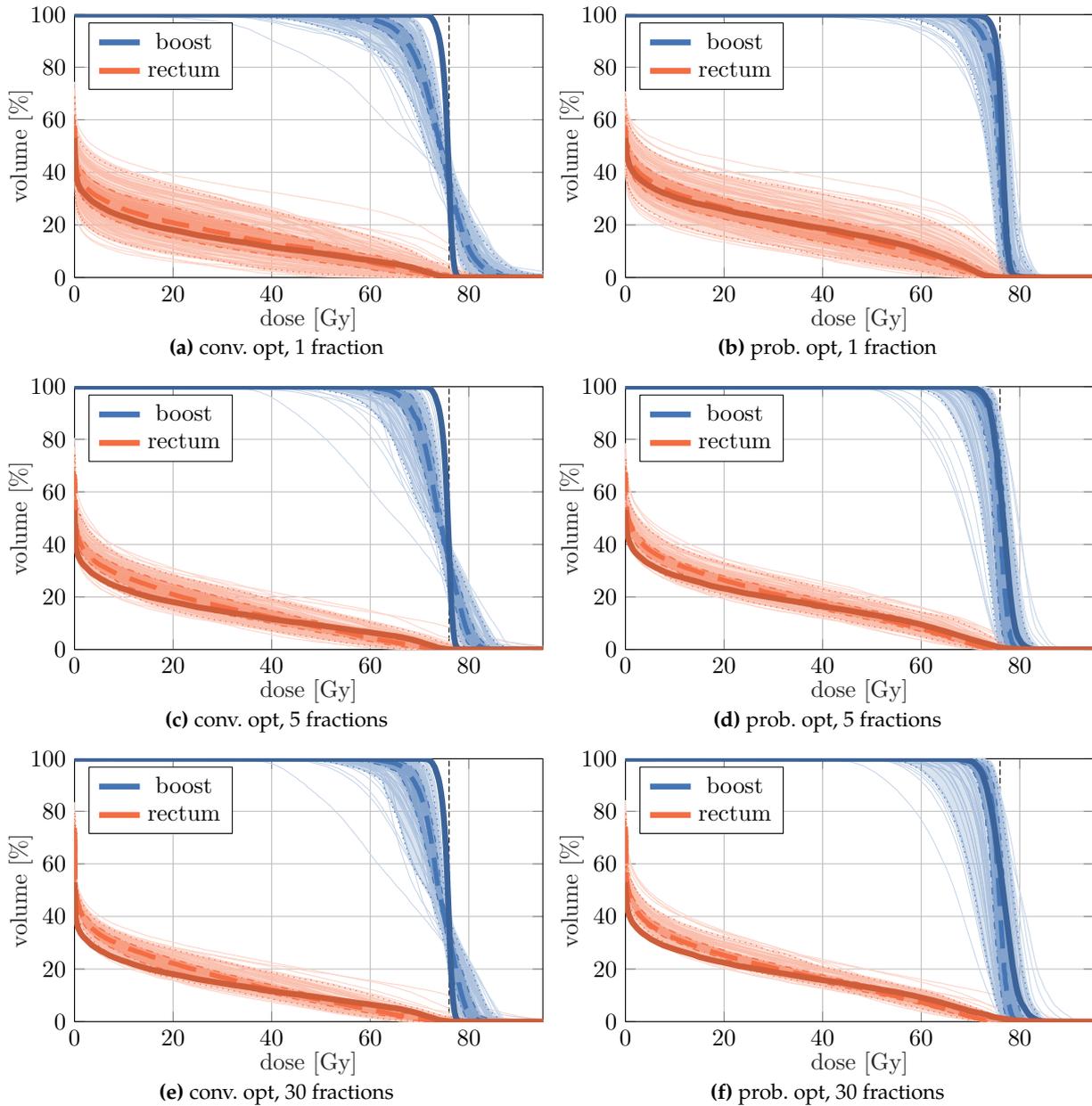
vicinity of the rectum. The resulting high dose gradients are especially prone to the input uncertainties, and therefore probabilistic optimization creates smoother beam fields. With increasing fraction number, modulation increases since random errors are able to cancel each other out. The sharp modulation from the conventional plan is, however, not recreated, since the systematic contribution persists independently of the number of fractions.

### V.3.2.2 DVH statistics

For the prostate case, the sampled DVHs and  $\alpha$ -DVHs are shown in Fig. V.5. Similar illustrations of the DVH-statistics for the paraspinal and the intracranial case are given in the section in Figs. C.6 and C.7, respectively.

The DVH-statistics show, in general, different effects in targets and in OARs. For OARs, the median DVHs are close to the nominal DVH. In the prostate and the intracranial case, the median DVH shows a slightly steeper descent, and on average the low-dose volume is increased. For the targets, the conventional plans exhibit steep nominal DVHs, which is, however, extremely unlikely to occur; the median DVH is significantly less steep, i. e., even the 5 % and 95 %  $\alpha$ -DVHs do not enclose the nominal DVH. This effect is compensated with probabilistic optimization, producing a nominal scenario with a DVH that is much more close to its median and, in general, enclosed by the 5 % and 95 %  $\alpha$ -DVHs.

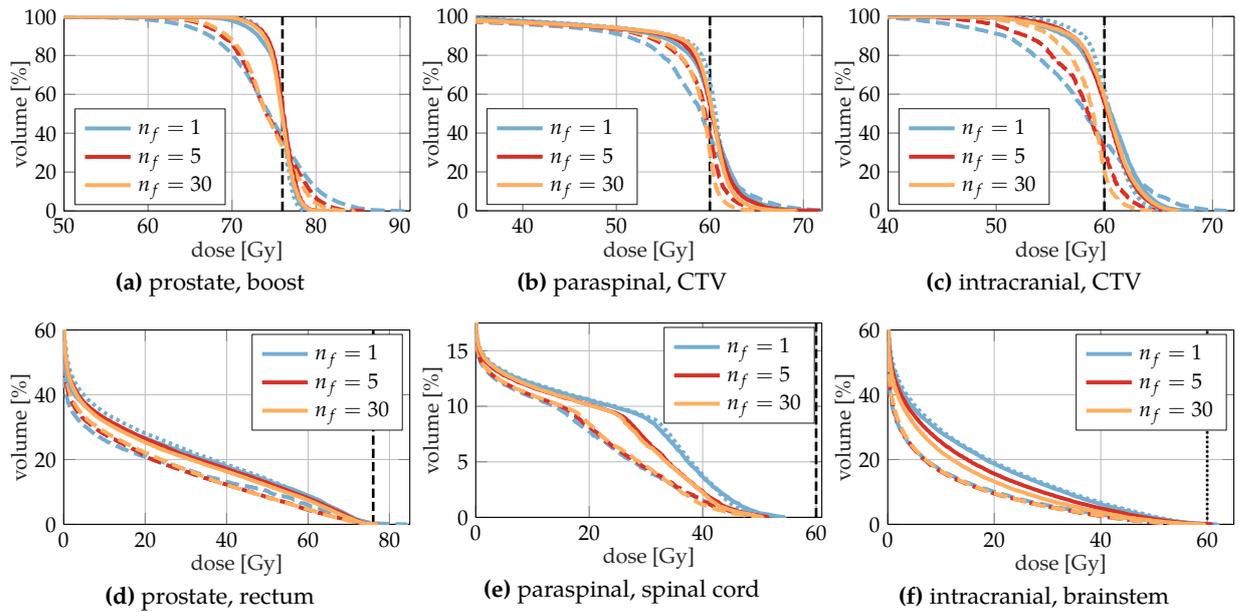
Regarding fractionation, the DVHs for the conventional treatment plan exhibit the aforementioned rapid decrease in variability with increasing fraction number. This decrease in variability



**Figure V.5:** Sampling based DVH-statistics for fractionated conventional and probabilistic treatment plans for the prostate case. (a,c,e) show the statistics for the conventionally optimized treatment planned applied in 1, 5 and 30 fractions, while (b,d,f) show the corresponding probabilistically optimized treatment plans. Nominal DVHs are represented by solid lines (—/—) while dashed lines (---/---) represent median DVHs (i. e., 50%  $\alpha$ -DVHs). The thin dotted (...../.....) and dash-dotted lines (-.-.-/-.-.-) enclosing the shaded areas refer to the 5% to 95%  $\alpha$ -DVHs and 25% to 75%  $\alpha$ -DVHs quantiles, respectively. The underlying DVHs for the treatment samples are indicated by the thin grayed out solid lines (—/—) in the background. The vertical dashed black line (-.-.-) indicates the prescribed dose to the respective target. Similarly presented in Wahl et al. (2018a).

does, however, not compensate for the observed difference between median and nominal DVH in the target. For the probabilistic treatment plans, the optimizer is able to compensate for this difference in the target to a similar extent across all investigated fractionation schemes. Within OARs, the variability persists, which can be attributed to the optimization penalty, favoring expected target coverage over OAR robustness and sparing during optimization.

Figures V.5, C.6 and C.7 are not ideal for a quantitative investigation of the differences between expected treatments with either conventional or probabilistic plans. Therefore, Fig. V.6 directly compares only the median DVHs for the different plans. For each of the three patient cases, respective targets (Figs. V.6a to V.6c) and OARs (Figs. V.6d to V.6f) are investigated.



**Figure V.6:** Median DVHs from sampling statistics for conventional and probabilistic treatment plans under fractionation using 1, 5 and 30 fractions. For each case, a target (a-c) and an OAR (d-f) is investigated. The dashed lines (— / - - / - · - ·) show the median DVHs for the conventional treatment plans, the solid lines (— / — / —) represent the median DVHs for the respective probabilistic treatment plan. The additional dotted line (· · · ·) shows the median DVH for a plan probabilistically optimized for a single fraction, but applied in 30 fractions. The dashed black line (— · — ·) indicates the prescribed dose to the respective target. Similarly presented in Wahl et al. (2018a).

From conventional to probabilistic plans a general trend towards higher integral doses in probabilistic plans is evident, for both targets and OARs. For targets, the higher dose combined with a steeper descent of the median DVH leads to a higher target coverage probability and a median scenario comparably close to the seen nominal scenario. For a treatment with larger fraction number, in generally slightly steeper median DVHs are observed in the target, however, this effect is minimal. For the OARs, the probabilistic treatment plan optimized for a single fraction experiences highest median dose coverage, which is reduced by optimization for a larger number of fractions. This is due to the random component's contribution to the variance term in the expected objective function (III.22), shifting focus on the term minimizing the squared deviation of the expected dose to the prescribed dose.

### V.3.2.3 Generalization of probabilistic treatment plans across fraction numbers

In Fig. V.6 the additional median DVHs for a probabilistic treatment plan optimized for a single fraction, but applied in 30 fractions, are shown. Compared to the respective probabilistic treatment plans directly optimized and consequently applied in 30 fractions, they indicate worse OAR sparing and none to only slight improvement of the median target DVH. This effect is visible in the spinal cord, but also apparent in the brainstem. For the prostate case, only a slightly increased median dose in the rectum can be observed.

Figure V.7 analyzes statistics of selected target and OAR quality metrics, the latter based on indicators used in QUANTEC (Marks et al. 2010). For the targets, the  $D^{95\%}$  indicator, i. e., the minimal dose covering 95 % of the targets' volume, is evaluated as surrogate for the minimum dose. For the paraspinal and the intracranial case,  $D^{5\%}$  surrogates the maximum dose in spinal cord and brainstem, and for the rectum the volume covered by 50 Gy, i. e.,  $V^{50\text{Gy}}$ , is chosen.

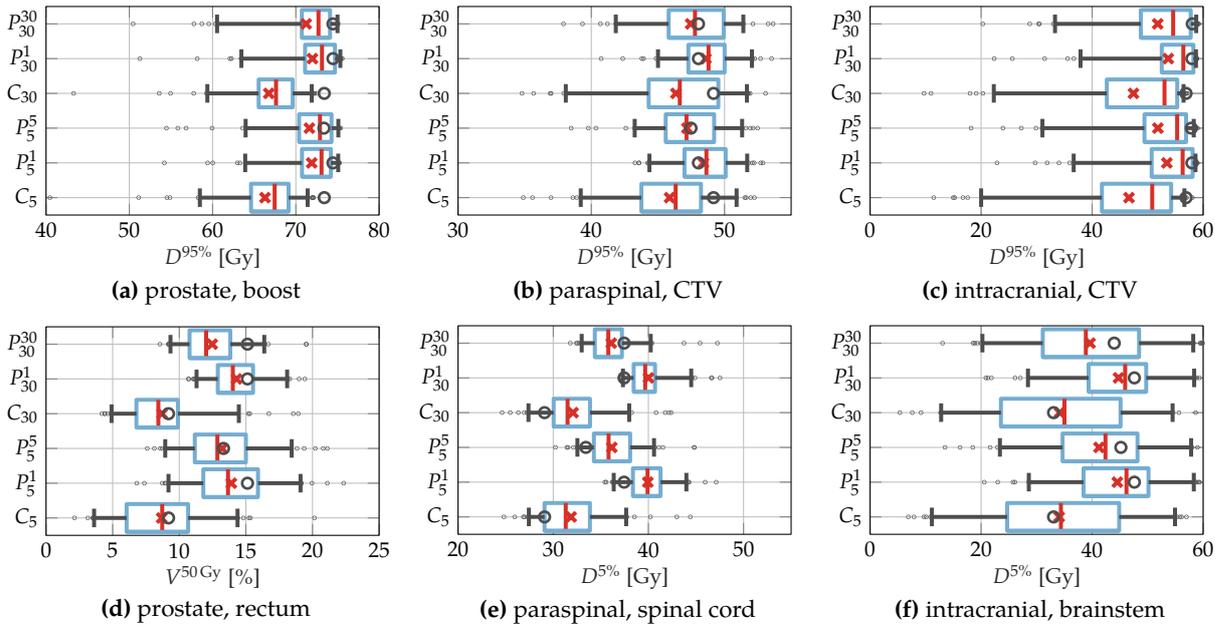
For the targets, probabilistic optimization shifts the probability mass to higher doses, i. e., increasing target coverage probability. Behavior of the more extreme quantiles i. e., the 5 % and 95 % percentiles, is inconsistent, which can be attributed to the low treatment sample number of  $n_T = 100$ . From the 25 % and 75 % percentiles, a reduction of the variance can be deduced compared to the conventional plan. A plan optimized for a single fraction, but applied in multiple fractions is, in general, slightly less variant and has slightly better target coverage than a plan optimized for the respective fraction number.

For the OARs, maximum dose/covered volume are expected to increase after probabilistic optimization, i. e., OAR sparing is traded against target robustness. Similar to the targets, probabilistic optimization therefore shifts the probability mass to higher dose(-volume)s. Probabilistic optimization neglecting multiple fractions is confirmed to result in expected higher maximum dose or dose-volume, as observed above.

This effect of probably overdosing the OARs while getting slightly better expected target coverage is caused by the importance of the variance term during optimization of Eq. (III.23). When optimizing for a single fraction, the variance term has higher contribution, which gives the optimizer less incentive to conform the expected dose than when optimizing with the reduced variance contribution in a multi-fractioned treatment. Paired with the observations so far that the median dose does, in general, not change significantly when changing the fraction number (for the same planned fluence), this leads to increased expected OAR dose. This is underpinned by Fig. V.8, showing FSDVHs for multiple combinations of optimized and applied fraction numbers.

Figure V.8 shows that an increased variance contribution in the optimization increases the incentive for the optimizer to reduce the integral variance in the respective VOI. This effect is mostly pronounced for targets. There, application of a probabilistic treatment plan, which was initially optimized for a single fraction, in multiple fractions consequently leads to a higher reduction of variance than in the plan specifically optimized for this number of fractions. Vice versa a plan optimized for 30 fractions and then applied in only a single fraction underestimates the variance contribution for optimization, and finally leads to the highest variability of target dose.

For the OARs, this effect is reduced up to nearly similar FSDVHs under changes the fraction



**Figure V.7:** Statistical analysis of probabilistic treatment plan quality of conventional ( $C_{n_f}$ ) and probabilistic treatment plans  $P_{n_f}^{n_o}$  using box plots.  $n_f$  is the number of fractions the treatment plan is applied in.  $n_o$  equals the number of fractions used for probabilistic optimization. The  $D^{95\%}$  is shown as surrogate for minimal target dose in (a-c). (d-f) analyze OARs based on suggested QUANTEC indicators. In analogy to Figs. V.5 and C.6 the whiskers enclose the 5% and 95% quantiles while the box encloses 25% and 75%. The vertical line represents the median value while the cross gives the arithmetic mean. Additionally, the corresponding value from the nominal plan is indicated by a circle. Similarly presented in Wahl et al. (2018a).

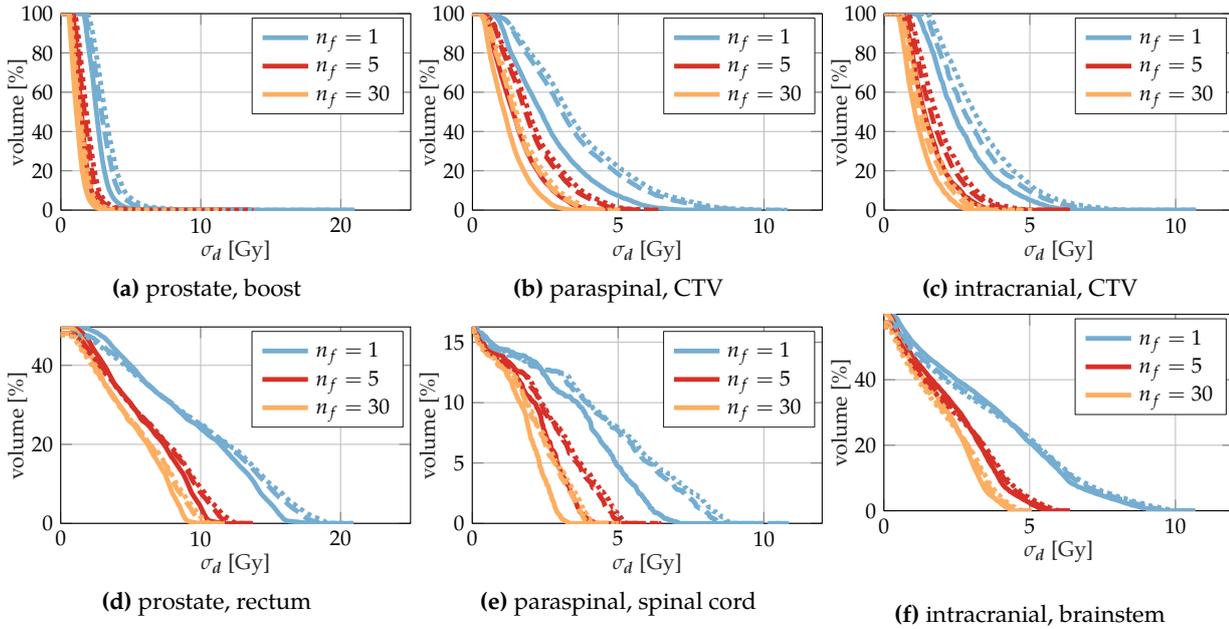
number in application. The spinal cord in the paraspinal case stands out, which can be attributed to it being located in the center of the target.

In summary, optimization for a single fraction—followed by application in multiple fractions—trades expected OAR sparing against only slight improvement on expected target coverage, compared to optimization directly for the applied fraction number.

## V.4 Discussion

With a slight reformulation of the covariance influence computation within the APM framework, the linear model of variance subject to random and systematic errors under fractionation can be computed at constant time complexity. This allows for analysis of expected value and standard deviation of the full fractionation spectrum by separation into the random and systematic components of standard deviation. With APM, this linear model can be incorporated into probabilistic optimization, i. e., optimization of the expected value of a penalized least squares objective function (Eq. (III.23)), using precomputed separated helper matrices (compare Eq. (V.5)).

The linear variance model is a direct consequence of the definition of random and systematic uncertainties common in literature (Fredriksson 2012; Kraan et al. 2013; Lowe et al. 2016; Lowe et al. 2017; Park et al. 2013; Perkó et al. 2016; Unkelbach et al. 2009). Hence, the linear model itself is generally not a novel “discovery” and also other probabilistic methods (e. g. Perkó et al. 2016; Sobotta, Söhn, and Alber 2012) are, in general, suited for its computation. Comparing to previous



**Figure V.8:** FSDVHs for treatment plans analyzed under combinations of  $n_f = \{1, 5, 30\}$  fractions used for probabilistic optimization and following treatment application. The number of fractions chosen for application is indicated by color in the legends. Line style refers to the fraction number used for probabilistic optimization: solid lines (— / — / —) for 1 fraction, dashed lines (--- / --- / ---) for 5 fractions and dotted lines (···· / ···· / ····) for 30 fractions. Similarly presented in Wahl et al. (2018a).

work by Park et al. (2013) and Perkó et al. (2016), behavior of dose uncertainty under fractionation could be confirmed with APM's computations. Statistical methods, which are able to compute (estimates of) the variance for at least two fractionation schemes (e.g. Kraan et al. 2013; Park et al. 2013; Unkelbach and Oelfke 2004), may rely on linear regression as indicated in Fig. V.1e and in Wahl et al. (2018a). The regression method is, however, subject to sampling uncertainty, which may lead to unphysical results like negative random or systematic error components if not handled carefully. Also, generalization to probabilistic optimization of the aforementioned methods is not in all cases trivial or feasible within reasonable computational complexity and accuracy.

While this study used an uncertainty model based on common assumptions in literature, it excludes a set of non-trivial uncertainties. Amongst these are, for example, deforming anatomies due to motion or tumor shrinking (Chan and Mišić 2013; Knopf et al. 2013; Kraan et al. 2013; Liu et al. 2012b; Lomax 2008a). Preparation of more sophisticated uncertainty models adapting to clinical workflows and protocols, e.g. image guidance, is yet to be addressed throughout academic methodologies. For the actual quantification of such non-linear effects the linear variance model would no longer be applicable. Some other protocols, however, like prescribed dose depending on the fractionation scheme, could be easily incorporated in APM. Other effects, like varying uncertainty mitigation efforts depending on the fraction number would require re-computations. In these scenarios an APM-based method can still provide fast decision guidance for practical uncertainty mitigation efforts under fractionation, potentially relying on the qualitative sub-sampling approximations discussed in Chapter IV.

For all generated treatment plans (conventional as well as probabilistic for different fraction numbers), the analytical computation with APM was complemented with an analysis based on random samples (compare Kraan et al. 2013; Park et al. 2013). The increasing complexity (for each treatment sample, 30 fractions were computed) enforced reduction of treatment sample number. Based on Fig. IV.7, 100 treatment samples were chosen, yielding reasonable accuracy for both expectation value and standard deviation. More extreme quantiles (e.g. the 5% and 95% percentiles) have yet to be interpreted carefully, since they are subject to only few samples and therefore inaccurate. Hence, the analysis in this study focused mainly on median DVHs.

Probabilistic optimization under fractionation based on APMs computations demonstrated fractionation-dependent trade-offs between nominal dosage and treatment plan robustness. This trade-off was realized through more pronounced modulation in the optimized beam fluences as the number of fractions increase. The reason for this increase in modulation relies in the decreased random variance influence component, which corresponds to a possible cancellation of error scenarios across fractions. APM's capability to model the full interplay between systematic and random components within optimization is at least challenging to reproduce in other frameworks. Within photon therapy, fractionation effects on probabilistic or robust optimization have already been evaluated (Chan, Bortfeld, and Tsitsiklis 2006; Unkelbach and Oelfke 2004). In proton therapy, however, the exact incorporation of the number of fractions usually requires further simplifications. Lowe et al. (2016) and Lowe et al. (2017), for example, completely separate random and systematic errors by dimension, i.e., range errors are purely systematic while set-up errors are purely random. Unkelbach, Chan, and Bortfeld (2007) only evaluate systematic errors. In other approaches, an infinite amount of fractions is assumed while additionally imposing uncertainty on the random errors' probability distribution itself, allowing computation in the infinite limit (i.e., no random dose variance component) while avoiding over-modulation of fluences to encourage random error cancellation which would lead to extremely heterogeneous dose distributions (Fredriksson 2012; Unkelbach and Oelfke 2004).

The random sampling studies performed have run-times of several days, compared to computation of a full fractionation spectrum including optimization with APM within minutes. This encourages clinical application. However, limitations of the underlying pencil-beam algorithm as well as a limited uncertainty model are drawbacks for certain treatment sites, for example in lung treatments (Taylor, Kry, and Followill 2017).

So far APM has only proven within expected value optimization of the penalized least-squares objective. In state-of-the-art treatment planning, however, more complex objective functions than penalized least-squares and constraints are used (compare Section II.2). Within statistical methods based on scenarios, statistical representations of those functions are easily obtained; within the APM framework, they require analytical probabilistic representations. Furthermore, optimization of the expectation value is a non-conservative method (Fredriksson 2012), i.e., while the result is more robust than the conventional plan, it is not particular robust against worst-cases. Continuously "sliding" between worst-case robustness and expectation improvement could be realized by conditional value-at-risk optimization (Fredriksson 2012; Rockafellar and Uryasev 1997; Romeijn et al. 2003), which would, however, require analytical representations of the re-

spective objectives within *APM*. Analytical probabilistic representations of constraint functions would even allow optimization subject to exact quantification of the desired robustness. The following Chapters *VI* and *VII* will address these issues; First, by defining probabilistic representations of a set of treatment plan metrics in Chapter *VI*, and second by conceptualizing probabilistic optimization using novel probabilistic objective functions and quantile constraints based on these plan metrics.



# Analytical Models for Probabilistic Treatment Plan Quality Metrics

This chapter approaches further propagation of uncertainty from the uncertain dose distribution to plan metrics used in the clinic. Similar to the originally proposed [APM](#) concept explained in [Chapter III](#), the goal is to model expectation value and variance, i. e., the first two moments, of the respective plan metrics in closed-form. Thereby the probability distribution over dose, which is assumed to be a multivariate normal distribution, is used as a clean start. This facilitates generalization of the models to other methods that are able to provide at least the first two moments of the probability distribution over dose.

The analytical computations are validated against sample statistics on patient data. Further, new visualization techniques are developed which may possibly aid treatment planners to assess uncertainties.

## VI.1 Overview

Clinical decisions are not taken by solely performing a visual inspection of the dose distribution, but also based on mainly dose-based treatment plan metrics for the [VOIs](#). So far, [APM](#) can only provide expected value and covariance in dose. While from these moments certain quality metrics for robustness can be derived, i. e., the [SDVHs](#) and [FSDVHs](#) used in [Chapters IV](#) and [V](#), uncertainty over dose-dependent treatment plan quality metrics is not directly available, but requires new explicit probabilistic models. Therefore, to strengthen the argument for using an analytical method like [APM](#) clinically, the framework needs to be extended to address uncertainty in the clinically relevant plan metrics.

As laid out in [Chapter III](#), [APM](#)'s analytical computations differ from other established methods. Where other methods provide statistical estimates of the moments of the probability distribution over dose, [APM](#) provides an approximation computed from mathematically exact expressions in closed-form. One advantage of scenario-based methods is, however, that computed dose scenarios also directly allow statistics on dose-dependent plan metrics, e. g. extreme doses or [DVH](#) statistics.

When using a rigorous analytical approach like [APM](#), which provides a parametrization of the

probability distribution over dose, one may, for example, sample from this distribution to deduct the plan metrics' statistics. Perkó et al. (2016), for example, use this technique, however based on a numerically constructed PCE. APM, providing expectation value and covariance of dose, may do the same based on marginal multivariate normal distributions within respective VOIs.

Alternatively, dose uncertainty may be propagated analytically with exact or approximate closed-form expressions. Based on the statistical moments of the probability distribution over dose, such analytical expressions would generalize further to scenario sampling using the sample mean and sample covariance, for example.

In radiation therapy, approaches for developing analytical representations of treatment plan quality metrics are, however, few and far between. Henríquez and Castrillón (2008a), for example, derive the expected value of DVHs using heuristic assumptions for the dose's probability distribution, by modeling the possible contribution to a histogram bin as Bernoulli experiments. Their heuristic assumption lacks, however, a correlation model and is therefore not suited for quantification of higher moments of a DVH. Analytical propagation of uncertainties through plan metrics like EUD or minimum and maximum dose has—to the best of the author's knowledge—not yet been attempted in literature in the area of radiation therapy. Yet, such problems are not unknown to mathematics, where, for example, extreme value statistics is an area of extensive research with applications in finance, meteorology, and other fields (Coles 2001).

Application of mathematical results on probabilistic modeling of treatment plan quality metrics is, however, problematic. Often, analytically proven results exist for assumptions not valid in the context of a dose distribution. For example, extreme value statistics is well studied for i. i. d. random variables, where the extreme value theorem shows that under certain conditions their maximum can be represented with the generalized extreme value distribution (Coles 2001; Mcfadden 1978). This is not applicable for the dose's probability distribution, which is physically bound to correlations. For the general case of correlated and dependent random variables, current mathematical research tackles the derivation of order statistics or generalized extreme value distributions (e. g. Bertin and Clusel 2006; Majumdar and Pal 2014). While some identities and simplifications about the distribution of order statistics are applicable under special conditions (Chen 2014; Kotz, Johnson, and Balakrishnan 2000; Tong 1990), a general exact theory is not yet available. A long known iterative method for approximating moments of the maximum of a multivariate normal is given by Clark (1961). Arellano-Valle and Genton (2008) derived exact expressions for the resulting distributions under assumption of multivariate normals or Student's  $t$ . However, they are difficult to handle with large numbers of variables due to the dependency on the full multivariate cumulative distribution.

For the generalized mean (representing EUD), a literature research revealed no general attempts of analytically deriving its probability distribution or its moments. Some engineering research, particularly in hydrology, evaluated statistical properties of the generalized mean when the averaged random samples are following a log-normal distribution (Jensen 1998). Also for special cases of the generalized mean, i. e., the harmonic mean where  $\kappa = -1$  (Limbrunner, Vogel, and Brown 2000) and the geometric mean for  $\kappa \rightarrow 0$  (Landwehr 1978), studies of their expectation and variance exist. These studies, however, are usually based on independent log-normal

samples, and are therefore only partly applicable to the correlated problem at hand.

The approaches shortly reviewed may work under conditions that are not given in the case of a dose distribution on a discretized voxel grid with a large number of voxels. Thus, in the following, mainly based on a multivariate normal model of the probability distribution over dose, analytical representations for the first two moments, i. e., expectation value and variance, are derived based on two approaches. At first, an approach similar to the existing *APM*-framework is applied. Second, if no solution can be found, a Taylor expansion of the respective function describing the plan metric is used to compute the first two moments. These derivations are evaluated based on the three patients already used in Chapters IV and V.

While concrete values of mean and standard deviation of such plan metrics might already be helpful for evaluation of plan robustness (McGowan et al. 2015), it stays difficult to understand implications on the dose distribution within the full probability space. The assessment of single scenarios, i. e., worst-case or random, gives a single snapshot which is difficult to bring into line with other, possible realizations, especially regarding their probability.

Here, the multivariate normal dose model can be exploited, since it can be used to regenerate scenario samples with desired properties. Based on a method proposed by Hennig (2013), marginal distributions, i. e., over a slice through the dose cube or the dose within a *VOI*, are deducted and sampled such that they form a series of equiprobable samples. This thesis qualitatively investigates and presents such visualizations created with the methods of Hennig (2013).

The remainder of this chapter begins with definition of the multivariate model and derivation of analytical uncertainty propagation expressions for certain plan metrics in Section VI.2. Within the same section, the sampling technique for generation of smooth animations is introduced. Section VI.3 evaluates the derived analytical expressions on their validity and compares to sample statistics, finalizing with exemplary sample series used for animation. The chapter concludes with a discussion of the results in Section VI.4 under critical evaluation of the underlying assumptions.

## VI.2 Methods

### VI.2.1 The multivariate dose model

As a function of spatial coordinates  $\mathbf{r}$ , the dose  $d(\mathbf{r})$  under uncertainty can be understood as a random field. Since *APM* enables the computation of the expectation value of dose at a coordinate  $\mathbf{r}$  as well as the covariance between dose values at two spatial coordinates, a Gaussian random field may be used as descriptor. As the computations are applied on the discretized patient data, this corresponds to assuming that the (random) dose vector  $\mathbf{d}$  describing the dose in all voxels is distributed as a multivariate normal distribution with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ , i. e.,

$$\mathbf{d} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) . \quad (\text{VI.1})$$

Note that, contrary to previous chapters, Eq. (VI.1) spares indexing of  $\boldsymbol{\mu}$  and  $\Sigma$  with  $\boldsymbol{d}$ , for better readability of the derivations to come.

Assumption Eq. (VI.1) is, of course, debatable for various reasons and requires short discussion: First, the dose physically can never be negative, which is inherently violated by assumption of a multivariate normal distribution which has infinite support. Second, the common assumptions on input uncertainty, i. e., spatial setup and calculational range uncertainties (compare Section II.3), impose strong correlations across pencil-beams, such that the central limit theorem does not deliver a basis for argumentation since the sources of uncertainty are not independent.

On the other hand, calculations with and on multivariate Gaussians are well understood, especially in the multivariate case. Therefore, it is well suited for a first model to approach empirical investigation. Use of other multivariate distributions comes with additional implications, i. e., they have limitations considering arbitrary linear correlations (compare e. g. Kotz, Johnson, and Balakrishnan 2000). As an exception, for example, the multivariate log-normal distribution, which is a logarithmic transformation of the multivariate normal distribution, may be named and also find its use in Section VI.2.3.3. Further, the application of treatment plans over multiple fractions may resolve correlation and enable argumentation with the central limit theorem. Hence, the following derivations will rely mainly on the assumption of Eq. (VI.1), and more discussion on this issue will be exerted in Section VI.4.

## VI.2.2 Analytical uncertainty propagation methods

Approaches for analytical uncertainty propagation through a function  $I(\boldsymbol{d})$  describing a dose metric may either attempt to compute or approximate a full probability distribution with density  $f_{I(\boldsymbol{d})}$ , or to find representations for lower-order moments like  $\mathbb{E}[I(\boldsymbol{d})]$  and  $\mathbb{E}[I(\boldsymbol{d})^2]$ . Considering that the underlying assumption (VI.1) of the probability distribution over dose is approximate, derivation of a full probability distribution over a dependent metric  $I(\boldsymbol{d})$  would comprise an accurate description where, in principle, there is none.

Instead, the initial motivation of APM, i. e., the computation of lower moments of  $I(\boldsymbol{d})$ , is followed. Besides the analytical computation of moments by integration used in the original works by Bangert, Hennig, and Oelfke (2013) (compare Section III.1, esp. Eq. (III.1)), uncertainties are commonly propagated through a function with the commonly known (and taught) first-order Taylor expansion (Farrance and Frenkel 2012; Ku 1966), often neglecting correlations. Both methods are summarized below, with the Taylor-based method being expanded to second order (for the expectation value) under presence of correlations.

### VI.2.2.1 Integration

To find the  $\nu$ -th moment of a treatment plan metric  $I(\boldsymbol{d})$  dependent on the dose distribution  $\boldsymbol{d}$  in  $n$  voxels by integration, one needs to solve

$$\mathbb{E}[I(\boldsymbol{d})^\nu] = \int_{\mathbb{R}^n} I(\tilde{\boldsymbol{d}})^\nu f_{\tilde{\boldsymbol{d}}}(\tilde{\boldsymbol{d}}; \boldsymbol{\mu}, \Sigma) \, \mathrm{d}\tilde{\boldsymbol{d}} \quad (\text{VI.2})$$

where  $f_{\mathbf{d}}$  is the probability density function (PDF) over dose  $\mathbf{d}$ , parametrized by its mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$ . When applying the initial assumption of Eq. (VI.1), one has  $f_{\mathbf{d}} = \mathcal{N}(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma)$ .

Since most of the plan metrics introduced in Section II.2.1 are, in some way, expressed as sums  $I(\mathbf{d}) \propto \sum_i g(d_i)$  over functions  $g(d_i)$  of voxel doses  $d_i$  within the respective structure, one can outline a general recipe for solving Eq. (VI.2) for  $\nu = \{1, 2\}$ .

For the first central moment, i. e.,  $\nu = 1$ , setting  $I(\mathbf{d}) \stackrel{!}{=} \sum_i g(d_i)$ , Eq. (VI.2) becomes

$$\mathbb{E} [I(\mathbf{d})^{\nu=1}] = \int_{\mathbb{R}^n} \left[ \sum_i g(d_i) \right]^{\nu=1} f_{\mathbf{d}}(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}} \quad (\text{VI.3a})$$

$$= \sum_i \int_{\mathbb{R}^n} g(d_i) f_{\mathbf{d}}(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}} \quad (\text{VI.3b})$$

$$= \sum_i \int_{\mathbb{R}} g(d_i) f_{d_i}(\tilde{d}_i; \mu_i, \sigma_i^2) \, d\tilde{d}_i. \quad (\text{VI.3c})$$

In step (VI.3a) the sum rule for integration was used to pull the sum out of the integral. Then, using the definition of marginal probability densities (Eq. (A.1)), one can “marginalize” the integrals to obtain Eq. (VI.3c).

For  $\nu = 2$ , analogous steps can be taken, using  $[\sum_i g(d_i)]^2 = \sum_{il} g(d_i)g(d_l)$ :

$$\mathbb{E} [I(\mathbf{d})^{\nu=2}] = \int_{\mathbb{R}^n} \left[ \sum_i g(d_i) \right]^2 f_{\mathbf{d}}(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}} \quad (\text{VI.4a})$$

$$= \sum_{il} \int_{\mathbb{R}^n} g(d_i)g(d_l) f_{\mathbf{d}}(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}} \quad (\text{VI.4b})$$

$$= \sum_{il} \iint_{\mathbb{R}^2} g(d_i)g(d_l) f_{d_{i;l}}(\tilde{d}_{i;l}; \mu_{i;l}, \Sigma_{i;l}) \, d\tilde{d}_i d\tilde{d}_l. \quad (\text{VI.4c})$$

Hence the computations simplify into a sum over integrals of mixed terms  $g(d_i)g(d_l)$  against bivariate marginal densities over components  $i$  and  $l$ . Note that this approach is on a higher level similar to the computations exercised for the general APM framework by Bangert, Hennig, and Oelfke (2013) presented in Chapter III. There, to evaluate the uncertainty of dose at a certain coordinate, sums over the pencil-beams with uncertain positioning need to be evaluated, also making use of the respective univariate and bivariate marginal distributions.

### VI.2.2.2 Taylor expansion

As alternative to finding an exact solution for integral (VI.2), uncertainty might be estimated expanding the function  $I(\mathbf{d})$  as a Taylor series around the mean dose vector  $\boldsymbol{\mu}$ . Up to second order, this results in

$$I(\mathbf{d}) \approx I(\boldsymbol{\mu}) + (\mathbf{d} - \boldsymbol{\mu})^T [\nabla I](\boldsymbol{\mu}) + \frac{1}{2} (\mathbf{d} - \boldsymbol{\mu})^T H_I(\boldsymbol{\mu}) (\mathbf{d} - \boldsymbol{\mu}), \quad (\text{VI.5})$$

where  $[\nabla I](\boldsymbol{\mu})$  denotes the gradient and  $H_I(\boldsymbol{\mu})$  the Hessian matrix evaluated at  $\boldsymbol{\mu}$ .

Using  $\mathbb{E}[\mathbf{d} - \boldsymbol{\mu}] = \mathbf{0}$ ,  $\mathbb{E}[(\mathbf{d} - \boldsymbol{\mu})^T (\mathbf{d} - \boldsymbol{\mu})] = \Sigma$  and  $\text{Var}[I(\mathbf{d})] = \mathbb{E}[I(\mathbf{d})^2] - \mathbb{E}[I(\mathbf{d})]^2$  one can derive the following approximations for expectation value and variance of  $I$  (Benaroya and Mi Han 2005; Gustafsson and Hendeby 2012):

$$\mathbb{E}[I(\mathbf{d})] = I(\boldsymbol{\mu}) + \frac{1}{2} \text{tr}(H_I(\boldsymbol{\mu}) \Sigma), \quad (\text{VI.6a})$$

$$\text{Var}[I(\mathbf{d})] = [\nabla I(\boldsymbol{\mu})]^T \Sigma \nabla I(\boldsymbol{\mu}). \quad (\text{VI.6b})$$

Note that in Eq. (VI.6a), the trace term  $\text{tr}(H_I(\boldsymbol{\mu}) \Sigma)$  can be efficiently evaluated using the element-wise (i. e., Hadamard) product using  $\text{tr}(H_I(\boldsymbol{\mu}) \Sigma) = \sum_{il} (H_I(\boldsymbol{\mu}) \circ \Sigma)_{il}$ . Note that while Eq. (VI.6a) includes the approximation up to second order, only the first order approximation was kept in Eq. (VI.6b). The reason for this is that a higher-order term in Eq. (VI.6b) would increase computational complexity beyond dimensionality of the covariance matrix, which was reasoned to be out of scope for this chapter, yet might be worth investigation in the future.

### VI.2.3 Analytical computations

In this section the derivations for the first non-central moment, i. e., the expectation value, and second central moment, i. e., the variance, of several clinical plan quality metrics will be described. In case of the DVH derivations, additionally the covariance between single dose-volumes is computed.

Since the presented clinical plan metrics make most sense when applied to a subset of dose voxels, i. e., a VOI, all derivations should be understood to be carried out for the marginal distribution in the voxel-space of the respective VOI. More precisely, the random dose vector is now marginalized to  $\mathbf{d} = (d_1, d_2, \dots, d_n)^T$  over the voxels  $i = 1, 2, \dots, n$  contained in an arbitrary VOI with  $n$  voxels, following the multivariate normal distribution  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ .

#### VI.2.3.1 Dose-volume (histograms)

Henríquez and Castrillón (2008a) already derived analytical expressions for the expectation value and an upper bound on the variance of a DVH. Their derivations are based on a heuristic assumption of the underlying probability distribution to be uniform with unknown correlation. Other types of distributions can be investigated (Henríquez and Castrillón 2008b). However, since they do not make assumptions on correlation of uncertainty between two dose voxels, these investigations are limited to expectation value of dose. In the following, their result for the expectation value will be confirmed by using explicit integration as described in Section VI.2.2, and then extended to an explicit quantification of the (co)variance of a DVH.

Henríquez and Castrillón (2010) also claimed to be able to derive  $\alpha$ -DVHs with their approach. Their model is, however, based on independent contributions of the voxel doses to the DVH accumulation and is therefore to be seen critically. A short recap of the method is given in at the end of this section.

**Expectation Value:** For the computation of the expectation value of a DVH, Henríquez and Castrillón (2008a) noticed that for a single point of the DVH, i. e., a dose-volume DVH( $\hat{d}$ ) at a given dose value  $\hat{d}$ , the evaluation of the step function term  $\Theta(d_i - \hat{d})$  within Eq. (II.9) corresponds to a Bernoulli experiment in case of uncertain  $d_i$ :

$$\mathbb{E} \left[ \Theta(d_i - \hat{d}) \right] = p \quad (\text{VI.7})$$

where the Bernoulli-parameter  $p$  is the probability obtained from  $d_i$ 's cumulative distribution function (CDF)  $F_{d_i}$  as  $p = 1 - F_{d_i}(\hat{d})$ .

Through linearity of the expectation value, Henríquez and Castrillón (2008a) deduced that the expectation of a dose-volume  $\mathbb{E}[\text{DVH}(\hat{d})]$  is then given by

$$\begin{aligned} \mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})] &= \mathbb{E} \left[ \frac{1}{n} \sum_i \Theta(d_i - \hat{d}) \right] = \frac{1}{n} \sum_i \mathbb{E} \left[ \Theta(d_i - \hat{d}) \right] \\ &= \frac{1}{n} \sum_i [1 - F_{d_i}(\hat{d})] . \end{aligned} \quad (\text{VI.8})$$

Equation (VI.8) holds true for any kind of distribution function  $F_{d_i}$ , and can therefore be seamlessly applied with the multivariate normal uncertainty model from Eq. (VI.1) with the respective marginal, i. e., a univariate normal CDF.

The result of Eq. (VI.8) may also be confirmed by explicit evaluation of the integral (VI.2). Since the respective computations are helpful to understand following derivations for the (co)variance of dose-volume histograms, the explicit steps to obtain the same result as in Eq. (VI.8) are given below, using the ‘‘marginalization’’ demonstrated in Eq. (VI.3):

$$\mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})] = \int_{\mathbb{R}^n} \text{DVH}(\hat{d}; \tilde{\mathbf{d}}) f_{\mathbf{d}}(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} \quad (\text{VI.9a})$$

$$= \int_{\mathbb{R}^n} \frac{1}{n} \sum_i \Theta(d_i - \hat{d}) f_{\mathbf{d}}(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} \quad (\text{VI.9b})$$

$$= \frac{1}{n} \sum_i \int_{-\infty}^{\infty} \Theta(d_i - \hat{d}) f_{d_i}(\tilde{d}_i) d\tilde{d}_i \quad (\text{VI.9c})$$

$$= \frac{1}{n} \sum_i \int_{\hat{d}}^{\infty} f_{d_i}(\tilde{d}_i) d\tilde{d}_i \quad (\text{VI.9d})$$

$$\mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})] = \frac{1}{n} \sum_i [1 - F_{d_i}(\hat{d})] \quad (\text{VI.9e})$$

Here, in step (VI.9d) the zero-valued part of the Heaviside function was canceled out by adaptation of the integration limits. Equation (VI.9e) then directly follows from Eq. (VI.9d) through the definition of the CDF itself.

Applying the initial assumption (VI.1) that dose follows a multivariate normal distribution,

Eq. (VI.9e) can be written as

$$\begin{aligned}
\mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})] &= \frac{1}{n} \sum_i \left[ 1 - \int_{-\infty}^{\hat{d}} \mathcal{N}(\delta; \mu_i, \sigma_i^2) d\delta \right] \\
&= \frac{1}{n} \sum_i \left[ 1 - \frac{1}{2} \left( 1 + \text{erf} \left( \frac{\hat{d} - \mu_i}{\sigma_i \sqrt{2}} \right) \right) \right] \\
&= \frac{1}{n} \sum_i \frac{1}{2} \text{erfc} \left( \frac{\hat{d} - \mu_i}{\sigma_i \sqrt{2}} \right)
\end{aligned} \tag{VI.10}$$

using the complementary error function  $\text{erfc}(x) = 1 - \text{erf}(x)$ . Note that the error function is not an elementary analytical function but merely a shorter representation, since the integral over the normal distribution in Eq. (VI.10) cannot be analytically solved.

**Variance:** For the variance of a dose-volume  $\text{Var}[\text{DVH}(\hat{d})]$ , Henríquez and Castrillón (2008a) only gave an upper bound

$$\text{Var} [\text{DVH}(\hat{d})] \leq \frac{1}{n} \sum_i [1 - F_{d_i}(\hat{d})] F_{d_i}(\hat{d}), \tag{VI.11}$$

presumably due to the lack for a sophisticated uncertainty model including correlations.

Yet, if the probability distribution over dose, including correlations, is available, the second non-central moment of a dose-volume point may be obtained by solving

$$\mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})^2] = \int_{\mathbb{R}^n} \text{DVH}(\hat{d}; \mathbf{d})^2 f_d(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}}. \tag{VI.12}$$

To solve Eq. (VI.12) a similar approach as exercised in Eq. (VI.9) can be taken, now using the bivariate “marginalization” from Eq. (VI.4):

$$\int_{\mathbb{R}^n} (\text{DVH}(\hat{d}; \mathbf{d}))^2 f_d(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} = \int_{\mathbb{R}^n} \left[ \frac{1}{n} \sum_i \Theta(\tilde{d}_i - \hat{d}) \right]^2 f_d(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} \tag{VI.13a}$$

$$= \frac{1}{n^2} \int_{\mathbb{R}^n} \sum_{il} [\Theta(d_i - \tilde{d}) \Theta(d_l - \tilde{d})] f_d(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} \tag{VI.13b}$$

$$= \frac{1}{n^2} \sum_{il} \int_{\hat{d}}^{\infty} \int_{\hat{d}}^{\infty} f_{d_{i;l}}(\tilde{\mathbf{d}}_{i;l}) d\tilde{d}_l d\tilde{d}_i \tag{VI.13c}$$

$$= \frac{1}{n^2} \sum_{il} [1 - F_{d_{i;l}}(\hat{d} \cdot \mathbf{1})] \tag{VI.13d}$$

This result is analogous to Eq. (VI.9e), now involving the double sum over all voxel combinations  $il$  evaluating the marginal bivariate CDF  $F_{d_{i;l}}(\hat{d} \cdot \mathbf{1})$ .

The variance of the dose-volume  $\text{Var}[\text{DVH}(\hat{d}; \mathbf{d})]$  is now given by

$$\begin{aligned} \text{Var} [\text{DVH}(\hat{d}; \mathbf{d})] &= \mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})^2] - \mathbb{E} [\text{DVH}(\hat{d}; \mathbf{d})]^2 \\ &= \frac{1}{n^2} \sum_{il} \left[ 1 - F_{d_{i;l}}(\hat{d} \cdot \mathbf{1}) \right] - \left\{ \frac{1}{n} \sum_i \left[ 1 - F_{d_i}(\hat{d}) \right] \right\}^2. \end{aligned} \quad (\text{VI.14})$$

Under assumption of a multivariate normal distribution, Eq. (VI.14) thus boils down to evaluations of univariate and bivariate normal probabilities. While both univariate and bivariate probabilities can not be analytically evaluated, at least for the univariate case expression with the error function is well established, and most programming languages deliver accurate implementations. For the bivariate case, which is less common, such numerically accurate evaluations can be performed with, for example, the works from Genz (2004).

Note that then, due to the symmetry of the covariance matrix  $\Sigma = \Sigma^T$  and the invariance of the Heaviside-step under powers  $\Theta(x)^a = \Theta(x)$ , the complexity of the evaluation of term (VI.13d) within Eq. (VI.14) may be reduced:

$$\text{Var} [\text{DVH}(\hat{d}; \mathbf{d})] = \frac{1}{n^2} \sum_i \frac{1}{2} \text{erfc} \left( \frac{\hat{d} - \mu_i}{\sigma_i \sqrt{2}} \right) + \frac{2}{n^2} \sum_{l>i} \int_{\hat{d}}^{\infty} \int_{\hat{d}}^{\infty} \mathcal{N}(\tilde{\mathbf{d}}_{i;l}; \boldsymbol{\mu}_{i;l}, \boldsymbol{\Sigma}_{i;l}) d\tilde{d}_l d\tilde{d}_i. \quad (\text{VI.15})$$

**Covariance:** With only slight changes in the above computations, also the covariance of two DVH-points, i. e.,  $\text{Cov}[\text{DVH}(\hat{d}_p; \mathbf{d}), \text{DVH}(\hat{d}_q; \mathbf{d})]$  with respective dose parameters  $\hat{d}_p$  and  $\hat{d}_q$ , can be computed by computation of  $\mathbb{E}[\text{DVH}(\hat{d}_p; \mathbf{d})\text{DVH}(\hat{d}_q; \mathbf{d})]$ :

$$\mathbb{E} [\text{DVH}(\hat{d}_p; \mathbf{d})\text{DVH}(\hat{d}_q; \mathbf{d})] = \int_{\mathbb{R}^n} \text{DVH}(\hat{d}_p; \tilde{\mathbf{d}})\text{DVH}(\hat{d}_q; \tilde{\mathbf{d}}) f_{\mathbf{d}}(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} \quad (\text{VI.16a})$$

$$= \frac{1}{n^2} \int_{\mathbb{R}^n} \left[ \sum_i \Theta(\tilde{d}_i - \hat{d}_p) \right] \left[ \sum_l \Theta(\tilde{d}_l - \hat{d}_q) \right] f_{\mathbf{d}}(\tilde{\mathbf{d}}) d\tilde{\mathbf{d}} \quad (\text{VI.16b})$$

$$= \frac{1}{n^2} \sum_{il} \int_{\hat{d}_p}^{\infty} \int_{\hat{d}_q}^{\infty} f_{\mathbf{d}_{i;l}}(\tilde{\mathbf{d}}_{i;l}) d\tilde{d}_l d\tilde{d}_i \quad (\text{VI.16c})$$

Analogous to Eq. (VI.14), the covariance is then given by

$$\begin{aligned} \text{Cov} [\text{DVH}(\hat{d}_p; \mathbf{d}), \text{DVH}(\hat{d}_q; \mathbf{d})] \\ = \mathbb{E} [\text{DVH}(\hat{d}_p; \mathbf{d})\text{DVH}(\hat{d}_q; \mathbf{d})] - \mathbb{E} [\text{DVH}(\hat{d}_p; \mathbf{d})] \mathbb{E} [\text{DVH}(\hat{d}_q; \mathbf{d})]. \end{aligned} \quad (\text{VI.17})$$

Under the assumption of a multivariate normal, the numerical evaluation of Eq. (VI.16) is, as for the variance, possible with methodology developed by Genz (2004).

**Computation of  $\alpha$ -DVHs and DVCMs:** Henríquez and Castrillón (2010), based on their aforementioned earlier work (Henríquez and Castrillón 2008a; Henríquez and Castrillón 2008b), claimed to be able to compute confidence levels  $\alpha$  for DVHs, i. e.,  $\alpha$ -DVHs. They model an  $\alpha$ -DVH based on the assumption that the respective  $\alpha$  value restricts the probability of dose voxels contributing to the DVH (compare their Bernoulli model from Eq. (VI.8)) independently to define

$$\alpha\text{-DVH}(\hat{d}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \stackrel{!}{=} \frac{1}{n} \sum_i \Theta \left[ 1 - \alpha - F_{d_i} \left( \hat{d}; \mu_i, \Sigma_{ii} \right) \right]. \quad (\text{VI.18})$$

Note that Eq. (VI.18) only depends on the respective marginal CDF  $F_{d_i}$ , which contradicts the observations made in Eqs. (VI.12) to (VI.16) that the variability of a DVH depends on the full covariance of dose  $\boldsymbol{\Sigma}$ . Consequently their approach must be seen critically and will therefore be explicitly evaluated in Section VI.3.

With the computations from Eqs. (VI.10) to (VI.16), however, a different approach might be taken to compute  $\alpha$ -DVHs and DVCM. With the expectation value and (co)variance, available for the DVH-points, it is possible to parameterize a probability distribution over the DVH. With marginal probabilities, DVCMs (Gordon and Siebers 2009; Gordon et al. 2010) can be computed (compare Section II.3.2.2). DVCMs were originally based on sample statistics of all DVH-points. They give an approximation to the two-dimensional map of the probability for the desired volume fractions  $v$  and dose thresholds  $\hat{d}$  (see Eq. (II.25)).

Using the expected value and variance over all computed DVH-points, one may parametrize a probability distribution to substitute  $P(v \leq \text{DVH}(\hat{d})) = F(v; \mathbb{E}[\text{DVH}(\hat{d})], \text{Var}[\text{DVH}(\hat{d})])$  in Eq. (II.25) with the CDF  $F$ :

$$\text{DVCM}(v, \hat{d}) = F \left( v; \mathbb{E}[\text{DVH}(\hat{d})], \text{Var}[\text{DVH}(\hat{d})] \right) \quad (\text{VI.19})$$

While the DVH-points could be, again, assumed to be normally distributed, this violates the physical and mathematical constraint of  $v \in [0, 1]$ . A probability measure over such volume fractions might be better represented using a beta distribution (compare Appendix A.1.3). While the beta-distribution is usually parametrized by two shape parameters, parametrization with mean and variance is also possible using the transformation rules given in Eq. (A.11). As desired for DVH-point modeling, the beta distribution then has only support in the interval  $[0, 1]$ .

### VI.2.3.2 Mean Dose

Since the arithmetic mean  $\bar{d}$  of the dose  $d$  is a linear function of dose, its moments can be exactly derived with a first order Taylor expansion as given in Eq. (VI.6). Alternatively, also integration can be applied, where the solution can be directly achieved using Eqs. (VI.3) and (VI.4) with

$g(d_i) = d_i n^{-1}$ . Hence, the (trivial) result can be directly stated as

$$\mathbb{E} [\bar{\mathbf{d}}] = \frac{1}{n} \sum_i \mathbb{E} [d_i], \quad (\text{VI.20a})$$

$$\text{Var} [\bar{\mathbf{d}}] = \frac{1}{n^2} \sum_{il} \Sigma_{il} = \frac{1}{n^2} \left( \sum_i \sigma_i^2 + 2 \sum_{l>i} \Sigma_{il} \right). \quad (\text{VI.20b})$$

### VI.2.3.3 Equivalent uniform dose

**EUD**, as explained in Section II.2.1.3, uses a generalized mean (see Eq. (II.12)) to encode the sensitivity of structures to inhomogeneity in dose, e. g. hot-spots in OARs. At this point, the reader should be notified *ex ante* that analytical formulation of the expectation value and variance of **EUD** turned out to be of particular difficulty, and the derivations do not successfully lead to the exact desired solution. The derivations are included in this manuscript nevertheless, since they might be of interest for certain readers. Others might skip the following paragraphs and re-enter reading at the end of this section, when the Taylor series expansion is exerted.

**Analytical integration:** Attempting to compute the  $\nu$ -th moment of **EUD** by integration, i. e., Eq. (VI.2), would require solving

$$\begin{aligned} \mathbb{E} [\text{EUD}_\kappa(\mathbf{d})^\nu] &= \int_{\mathbb{R}^n} \left[ \left( \frac{1}{n} \sum_i \tilde{d}_i^\kappa \right)^{\frac{1}{\kappa}} \right]^\nu f_d(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}} \\ &= \int_{\mathbb{R}^n} \left( \frac{1}{n} \sum_i \tilde{d}_i^\kappa \right)^{\frac{\nu}{\kappa}} f_d(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}}. \end{aligned} \quad (\text{VI.21})$$

Integral (VI.21) is in several ways problematic. First, the “marginalization” explained in Section VI.2.2.1 for  $\nu = 1$  and  $\nu = 2$  is not applicable, due to the non-integer exponent  $\nu\kappa^{-1}$ ; instead, one would need to apply, for example, the generalized multinomial theorem for non-integer exponents, expanding into infinite series. Second, the assumption of  $f_d$  being a multivariate normal density raises problems for  $\kappa < 1$  if  $\kappa \in \mathbb{R}$ , since the multivariate normal is supported on the complete real space, opposed to the physical reality of the dose being always positive.

To deal with the first problem, i. e., the non-integer power of the sum within Eq. (VI.21), an ansatz for a step-wise solution can be tried, by first finding the expectation value of the sum of powers

$$\mathbb{E} \left[ \left( \sum_i d_i^\kappa \right)^\nu \right] = \int_{\mathbb{R}^n} \left[ \sum_i \tilde{d}_i^\kappa \right]^\nu f_d(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma) \, d\tilde{\mathbf{d}} \quad (\text{VI.22})$$

and afterwards propagating the result in a second step through the inverse exponent  $\kappa^{-1}$ . Equation (VI.22) can now be “marginalized” according to Eqs. (VI.3) and (VI.4). For  $\nu = 1$ , Eq. (VI.22)

reduces to the elementary problem of finding

$$\mathbb{E} [d_i^\kappa] = \int_{-\infty}^{\infty} \tilde{d}_i^\kappa f_{d_i}(\tilde{d}_i; \mu_i, \sigma_i^2) d\tilde{d}_i. \quad (\text{VI.23})$$

For  $\kappa \in \mathbb{N}$ , Eq. (VI.23) represents the respective raw moments of  $f_{d_i}$ . In the following, two attempts to solve Eq. (VI.22) via Eq. (VI.23) will be presented.

*An ansatz for  $\mathbf{d} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ :* In case of a normal distribution, the raw integer moments can be directly obtained from generalized Hermite polynomials (compare Eq. (A.14) and Roman and Rota 1978). For non-integer  $\kappa$ , solution of integral (VI.23) is not trivial. Alternatively, with the help of *fractional calculus* a *fractional moment* for  $\kappa > 0$  can be defined by evaluating the fractional derivative of the moment generating function of  $f_{d_i}$  at its origin (Cressie, Borkent, and Gupta 1986). In the case of  $f_{d_i}(\tilde{d}_i; \mu_i, \sigma_i^2) = \mathcal{N}(\tilde{d}_i; \mu_i, \sigma_i^2)$ , alternatively integral (A.13) from Gradshteyn and Ryzhik (2000) may be used to find the solution

$$\begin{aligned} \mathbb{E} [d_i^\kappa] &= \int_{-\infty}^{\infty} \tilde{d}_i^\kappa \mathcal{N}(\tilde{d}_i, \mu_i, \sigma_i^2) d\tilde{d}_i \\ &= \sigma_i^\kappa \cdot (-i\sqrt{2})^\kappa U\left(-\frac{\kappa}{2}, \frac{1}{2}, -\frac{\mu_i^2}{2\sigma_i^2}\right), \quad \kappa > -1, \end{aligned} \quad (\text{VI.24})$$

where  $U(\alpha, \beta, \zeta)$  denotes Tricomi's confluent hypergeometric function, also known as Kummer's confluent hypergeometric function of the second kind (Olver et al. 2010; Slater 1960). A similar solution is also given in a summary on moments of the normal distribution by Winkelbauer (2012).

To find a moment constrained to the positive dose domain—in the hope to find a simpler expression than Eq. (VI.24)—solutions for the absolute raw moment  $\mathbb{E}[|d_i|^\kappa]$  could be of value. Winkelbauer (2012) provide a solution as

$$\begin{aligned} \mathbb{E} [|d_i|^\kappa] &= \int_{-\infty}^{\infty} |\tilde{d}_i|^\kappa \mathcal{N}(\tilde{d}_i, \mu_i, \sigma_i^2) d\tilde{d}_i \\ &= \sigma_i^\kappa \cdot 2^{\frac{\kappa}{2}} \frac{\Gamma(\frac{1+\kappa}{2})}{\sqrt{\pi}} {}_1F_1\left(-\frac{\kappa}{2}, \frac{1}{2}, -\frac{\mu_i^2}{2\sigma_i^2}\right), \quad \kappa > -1, \end{aligned} \quad (\text{VI.25})$$

where  ${}_1F_1(\alpha, \beta, \zeta)$  is Kummer's confluent hypergeometric function of the first kind (Olver et al. 2010; Slater 1960). Hence, using the absolute value  $|d_i|$  does not facilitate computation of the expectation  $\mathbb{E}[|d_i|^\kappa]$ , since both  $U$  and  ${}_1F_1$  require evaluation of a hypergeometric series.

Solving Eq. (VI.22) for  $\nu = 2$  to find the second raw moment, even makes things more complicated; now, the bivariate integrals

$$\mathbb{E} \left[ \left( \sum_i d_i^\kappa \right)^2 \right] = \sum_{i'l} \iint_{\mathbb{R}^2} \tilde{d}_i^\kappa \tilde{d}_l^\kappa f_{d_{i;l}}(\tilde{d}_{i;l}; \boldsymbol{\mu}_{i;l}, \boldsymbol{\Sigma}_{i;l}) d\tilde{d}_{i;l} \quad (\text{VI.26})$$

are involved. Despite literature research and usage of symbolic programming languages, no successful ansatz could be defined to solve Eq. (VI.26). Hence, to stay within the scope of this

manuscript, and also because the solutions for  $\nu = 1$  are not able to capture  $-\infty < \kappa < \infty$ , this approach was disbanded for dose following a multivariate normal distribution.

The complexity of Eqs. (VI.22) to (VI.26) and their solutions was mainly attributed to the normal distribution being supported on the full real space  $\mathbb{R}^n$ , since exponentiation  $x^\kappa$  with  $\kappa \in \mathbb{R}$  is not consistently defined for negative base  $x$ . One may, however, chose a PDF which is supported only within the positive orthant  $\mathbb{R}_+^n$ . To keep the handiness of the multivariate normal with respect to the posed problem, usage of the log-normal distribution suggests itself, since mean and covariance exhibit straightforward transformation rules (compare Appendix A.1.2). In the following, the respective derivations are exercised.

*A workaround with the log-normal distribution:* Given the mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$  over dose, the parameters  $\tilde{\boldsymbol{\mu}}$  and  $\tilde{\Sigma}$  of the corresponding log-normal distribution  $\mathcal{LN}(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$  can be computed with the transformations (A.8). The parameters  $\tilde{\boldsymbol{\mu}}$  and  $\tilde{\Sigma}$  then correspond to the mean and covariance of the logarithm of dose supposedly following a multivariate normal distribution, i. e.,  $\ln \mathbf{d} \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$ .

The considerations of Jensen (1998) for independent log-normal samples already showed that setting  $f_d(\boldsymbol{\mu}, \Sigma) = \mathcal{LN}(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$  in Eqs. (VI.23) and (VI.26) leads to the solution

$$\begin{aligned} \mathbb{E}[d_i^\kappa] &= \int_0^\infty \tilde{d}_i^\kappa \mathcal{LN}(\tilde{d}_i; \tilde{\mu}_i, \tilde{\sigma}_i^2) d\tilde{d}_i \\ &= \exp\left(\kappa \tilde{\mu}_i + \frac{\kappa^2 \tilde{\sigma}_i^2}{2}\right) \end{aligned} \quad (\text{VI.27})$$

to compute the first moment. Note that this solution can be directly obtained from the moment-generating function  $M_{\ln d}(\mathbf{t}) = \mathbb{E}[\exp(\mathbf{t} \cdot \ln \mathbf{d})]$ , i. e., the moments of a log-normally distributed random variable are given from the moment-generating function of the normal distribution of its logarithm evaluated at  $\mathbf{t} = \mathbf{1}$  (compare Appendix A.1.2).

Similarly, integral (VI.26) can then be solved as

$$\begin{aligned} \mathbb{E}[d_i^\kappa d_l^\kappa] &= \iint_{\mathbb{R}_+^2} \tilde{d}_i^\kappa \tilde{d}_l^\kappa \mathcal{LN}(\tilde{d}_{i;l}; \tilde{\boldsymbol{\mu}}_{i;l}, \tilde{\Sigma}_{i;l}) d\tilde{d}_{i;l} \\ &= \exp\left[\kappa(\tilde{\mu}_i + \tilde{\mu}_l) + \kappa^2 \left(\frac{\tilde{\Sigma}_{ii} + \tilde{\Sigma}_{ll}}{2} + \tilde{\Sigma}_{il}\right)\right] \end{aligned} \quad (\text{VI.28})$$

to compute the second moment.

With the expectation value and the variance of the sum of powers available, the second step, i. e., the propagation through the  $\kappa$ -th root, can be performed with the result obtained from Eq. (VI.22). Note that this assumes a log-normal distribution for the sum of powers based on its obtained expected value and covariance. In reality, this distribution may, however, exhibit completely different shapes for large negative or positive  $\kappa$ , and cannot be deducted from the previous derivations (however, it is possible to compute higher moments when assuming dose following a log-normal distribution). Therefore results with this two-step approach, which again assumes a log-normal distribution at the second step, should be handled carefully.

**Using Taylor expansion:** Due to the complexity of the integration ansatz, uncertainty propagation using Taylor expansion is of interest for EUD, for which the gradient and the Hesse matrix of Eq. (II.12)

$$\nabla_{\mathbf{d}} \text{EUD}^{\kappa}(\mathbf{d}) = \frac{\text{EUD}^{\kappa}(\mathbf{d})}{\sum_i d_i^{\kappa}} \mathbf{d}^{\circ(\kappa-1)}, \quad (\text{VI.29})$$

$$H_{\text{EUD}^{\kappa}}(\mathbf{d}) = (\kappa - 1) \frac{\text{EUD}^{\kappa}(\mathbf{d})}{(\sum_i d_i^{\kappa})^2} \cdot \left[ \sum_i d_i^{\kappa} \cdot \left( \mathbf{d}^{\circ(\kappa-2)} \right)^T I - \mathbf{d}^{\circ(\kappa-1)} \left( \mathbf{d}^{\circ(\kappa-1)} \right)^T \right], \quad (\text{VI.30})$$

are required, where  $I$  is the unit matrix and the operation  $\mathbf{d}^{\circ\kappa}$  denotes element-wise exponentiation of  $\mathbf{d}$  with exponent  $\kappa$ .

Equations (VI.29) and (VI.30) may then be inserted in Eq. (VI.6) to obtain an approximation for  $\mathbb{E}[\text{EUD}^{\kappa}(\mathbf{d})]$  and  $\text{Var}[\text{EUD}^{\kappa}(\mathbf{d})]$ .

#### VI.2.3.4 Minimum / Maximum dose

The minimum and maximum of dose within a VOI is not a smooth function of the dose vector  $\mathbf{d}$ . This is not an issue in determination of its nominal value, but as soon as differentiation or integration is involved, i. e., for optimization or, in this case determining its moments with Eq. (VI.2) or Eq. (VI.6), other methods or workarounds have to be found.

As indicated in Section VI.1, mathematical theory exists within extreme value theory. For sequences of i. i. d. random variables, a family of the so called *generalized extreme value distributions* can be defined, to which the maximum of the sequence converges with increasing length of the sequence.

The random dose vector  $\mathbf{d}$  is, however, not an i. i. d. sequence of random numbers, but a discretization of a three-dimensional random field with complex correlation patterns, represented as multivariate normal distribution (see Eq. (VI.1)). Such generalized extreme value statistics to sets of correlated random variables are investigated in literature (Bertin and Clusel 2006; Majumdar and Pal 2014), and iterative approaches have been known since the work of Clark (1961) to approximate the first and second moment of the maximum or the full PDF of the maximum (Arellano-Valle and Genton 2008). Those approaches however, are often unsuited due to computational issues when the number of random variables is large, or work under restrictions that are not given directly in the problem of the random dose vector at hand.

In optimization, however, extrema are often handled by smooth approximations. `matRad`, for example, uses a smoothly approximated maximum using a variation of the ‘‘LogSumExp’’ function  $\max(x) \approx L(x) = \ln(\sum_i \exp x_i)$  (Wieser et al. 2017b), which can serve as smooth, convex maximum or minimum approximation.

More interestingly, since the generalized mean is the basis of EUD (compare Section II.2.1.3),

EUD can be used as a smooth approximation of the minimum and maximum itself:

$$\min(\mathbf{d}) = \lim_{\kappa \rightarrow -\infty} \text{EUD}^\kappa(\mathbf{d}), \quad (\text{VI.31a})$$

$$\max(\mathbf{d}) = \lim_{\kappa \rightarrow +\infty} \text{EUD}^\kappa(\mathbf{d}). \quad (\text{VI.31b})$$

Consequently, using the generalized mean/EUD as smooth maximum approximation, directly allows us to recycle the derivations from above (Section VI.2.3.3). When choosing a large absolute  $\kappa \gg 0$ , expectation and variance of the maximum of dose may then be approximated with

$$\mathbb{E}[\max(\mathbf{d})] \approx \mathbb{E}[\text{EUD}^\kappa(\mathbf{d})], \quad (\text{VI.32a})$$

$$\text{Var}[\max(\mathbf{d})] \approx \text{Var}[\text{EUD}^\kappa(\mathbf{d})]. \quad (\text{VI.32b})$$

For the minimum, Eq. (VI.32) holds analogously with  $\kappa \ll 0$ .

## VI.2.4 Smooth visualizations

Visualization of uncertainties is an important aspect to encourage patient-specific treatment plan uncertainty analysis—and eventually robust and probabilistic treatment planning—within clinical workflows. Sampling based uncertainty propagation has the advantage of being able to directly supply the treatment planner with explicit dose scenarios computed from a sample from the input uncertainty space. Such samples are not generated when using a forward probabilistic approach like APM. Since APM is able to provide mean and covariance of dose, however, the assumption of dose being normally distributed made in Eq. (VI.1) allows to generate dose vector samples  $\mathbf{d}_s$ .

The usual approach for the computation of samples from a multivariate normal distribution, in this case given by  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ , is to factorize the covariance matrix with a Cholesky-decomposition such that  $LL^T = \Sigma^d$ . With the upper triangular matrix  $L$ , it is then possible to generate a sample  $\mathbf{d}_s$  through linear transformation of a sample  $\mathbf{u}_s$  from the multivariate standard normal distribution  $\mathcal{N}(\mathbf{0}, I)$ :

$$\mathbf{d}_s = \boldsymbol{\mu} + L\mathbf{u}_s. \quad (\text{VI.33})$$

The i. i. d. samples  $\mathbf{u}_s$  can be easily generated with numerous pseudo-random number generators.

Alternatively, instead of using the Cholesky-decomposition, an eigenvalue or singular-value decomposition can be used.

### VI.2.4.1 The method by Hennig (2013)

While Eq. (VI.33) now allows generation of random scenarios, selection of samples can also be done on purpose. Hennig (2013) describes a method to draw a series of equiprobable samples that, displayed one after each other, result in a smooth looped animation of samples of a Gaussian process. This is achieved by choosing the standard normal samples  $\mathbf{u}_s$  such that they define a

circular path on the hyper-sphere in the space of  $\mathcal{N}(\mathbf{0}, I)$ . Linearly transformed with Eq. (VI.33), these samples are then located on an ellipse on an ellipsoid.

The workflow described by Hennig (2013) is summarized below:

1. Draw a random start sample  $\mathbf{u}_1$  from the multivariate standard normal.
2. Compute a direction  $\mathbf{t}$  that is a tangent on the hyper-sphere on which  $\mathbf{u}_1$  is located. This is done by drawing a second multivariate standard normal sample  $\mathbf{v}$ , and then use the Gram-Schmidt process for orthogonalization to obtain  $\mathbf{t} = \mathbf{v} - (\mathbf{v}^T \mathbf{u}_1) \mathbf{u}_1$ .
3. Choose a number of frames  $F$  (for the animation) and distribute  $F + 1$  grid points  $l_s$  equidistantly in  $[0, 2\pi]$  (the last point  $l_{F+1}$  is discarded, as in the grand circle  $l_1$  and  $l_{F+1}$  correspond to the same coordinate). Use the exponential map  $\mathbf{u}_s = \text{Exp}(\mathbf{u}_1, l_s \mathbf{t})$  to obtain samples  $\mathbf{u}_s$  on the grand circle on the hyper-sphere.

When all  $\mathbf{u}_s$  are available, one can apply transformation (VI.33) to obtain the corresponding sample animation for a general multivariate normal  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

The method of Hennig (2013) draws a random start coordinate  $\mathbf{u}_1$ . With high probability, this starting coordinate falls into a thin shell with radius  $\sqrt{n}$  where nearly all probability mass is located for large dimensionality  $n$  of the sample vectors (Mackay 2005). Therefore, in almost all cases, reasonably probable samples are produced.

## VI.3 Results

### VI.3.1 Technical evaluation of the analytical models

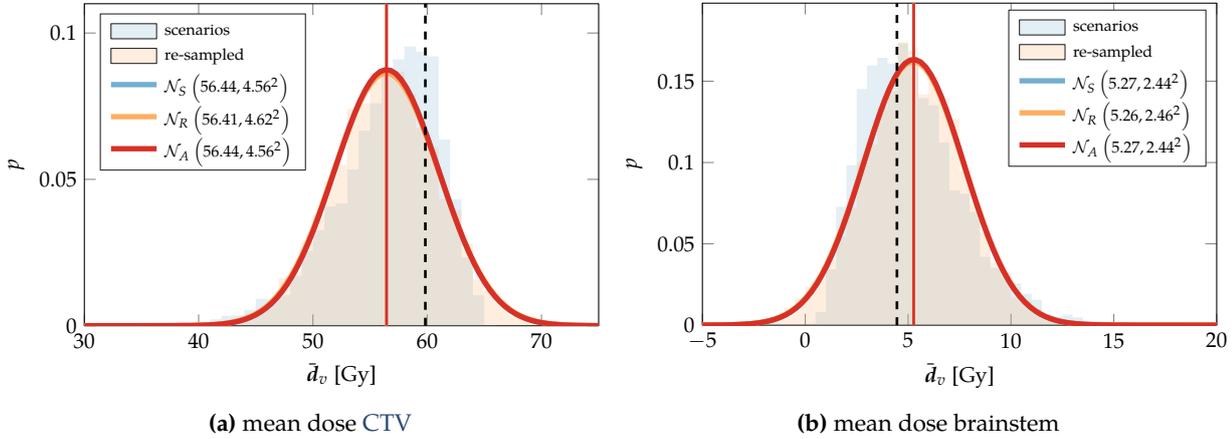
At first, a short technical evaluation of the analytical models derived in Section VI.2.3 was performed. As benchmark, the 5,000 samples used in Chapter IV for validation of the moment computations were recycled to produce marginal sample mean vectors  $\hat{\boldsymbol{\mu}}_v$  and covariance matrices  $\hat{\boldsymbol{\Sigma}}_v$  for respective VOIs  $v$ . For the latter, the CTV and the brainstem were evaluated.

First, the accuracy with respect to the underlying assumed probability distribution is evaluated. Hence, artificial multivariate random samples were generated to evaluate the accuracy of the analytical models under perfect conditions (i. e., the dose *actually* follows the assumed probability distribution). Since, in most cases, a multivariate normal distribution was assumed as the underlying probability distribution of dose, random samples could be easily generated using Eq. (VI.33).

#### VI.3.1.1 Mean dose

Figure VI.1 compares sample statistics of mean dose in the respective VOI to its analytical computation within APM with Eq. (VI.20) and  $\mathbf{d} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_v, \hat{\boldsymbol{\Sigma}}_v)$ . The normalized histograms present values from the 5,000 sampled dose scenarios and from the 5,000 re-samples based on Eqs. (VI.1) and (VI.33). Their difference in shape clearly shows that the multivariate model from Eq. (VI.1)

is not a perfectly accurate description of reality. The obtained expected value and variance of the mean dose  $\bar{d}$ , however, are accurate within sample accuracy. Thus their Gaussian representations are almost indistinguishable. This behavior is expected, as the first two moments of the mean dose are *exactly* described by a Taylor approximation up to first order, and thus independent of higher moments of the probability distribution underlying  $d$ .



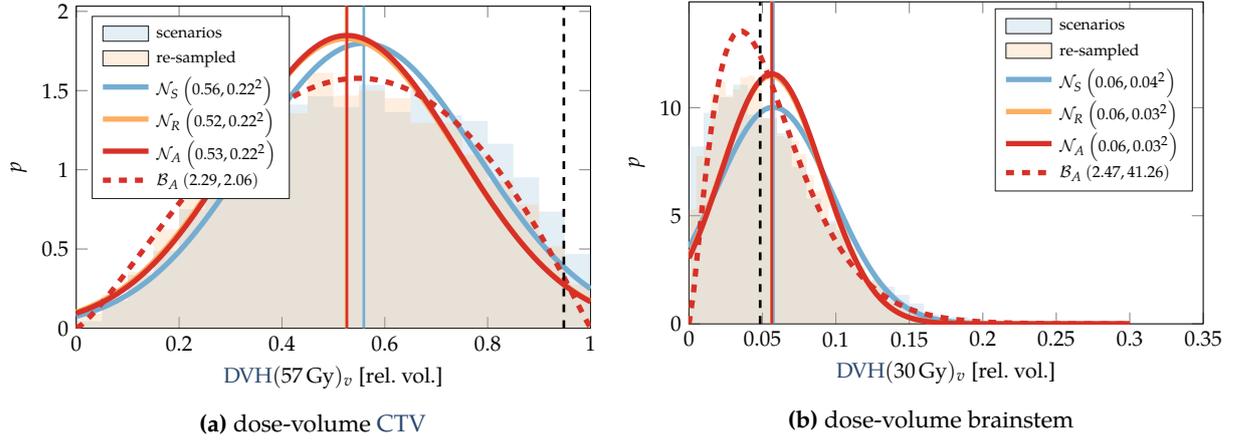
**Figure VI.1:** Technical evaluation of analytical probabilistic mean dose for the CTV (a) and the brainstem (b) of the intracranial patient. Sampled values from the original 5,000 dose scenarios as well as from 5,000 re-samples using the multivariate normal dose model are shown as histograms. Their sample mean (vertical solid lines) and sample standard deviation are compared to the analytically computed mean and standard deviation by displaying the respectively parametrized normal distributions  $\mathcal{N}_S$ ,  $\mathcal{N}_R$  and  $\mathcal{N}_A$ . Thereby  $\mathcal{N}_S$  is derived from the dose scenario samples and  $\mathcal{N}_R$  is derived from the re-sampled scenarios.  $\mathcal{N}_A$  gives the normal distribution obtained from APM's moment computations. The vertical dashed line indicates the mean dose value obtained from the nominal dose scenario.

### VI.3.1.2 Dose-volume histogram-points

A similar analysis was performed for DVH-points in Fig. VI.2. Here, mean and variance derived from the re-sampled data show small, yet significant differences to mean and variance. Since APM's novel analytical computations from Eqs. (VI.10) and (VI.14) are exact under assumption of a multivariate normal distribution, they match almost perfectly with the sample statistics based on the re-sampled data.

Figure VI.2 additionally proposes an alternative description of the underlying probability distribution of a DVH-point. Since DVHs represent relative volumes and are therefore bound to the interval, a beta distribution  $\mathcal{B}(\alpha, \beta)$  with shape parameters  $\alpha$  and  $\beta$  is assumed (compare Appendix A.1.3). Since the Beta-distribution itself has only support on the interval  $[0, 1]$ , it can serve as a more physically sensible representation of the probability distribution of volume fractions. Transforming the expected value and variance obtained by APM with Eq. (A.11), the respective PDFs could be computed.

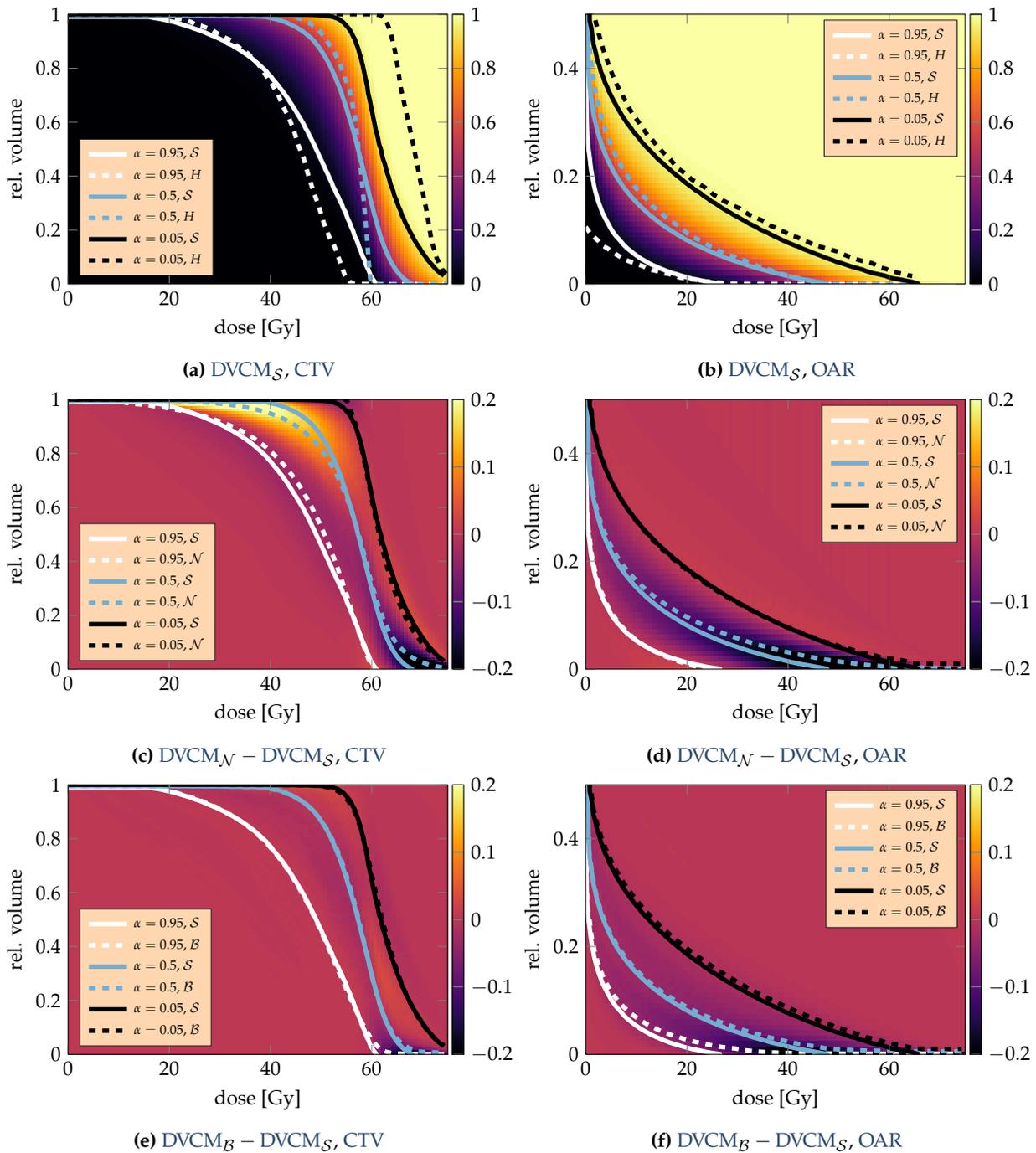
Based on Fig. VI.2 one can qualitatively argue that the beta-distribution better captures the skewness of the sampled distribution. To quantify this statement comparisons to cumulative probabilities obtained through DVCMs (compare Section II.3.2.3, Eq. (II.25)) were exerted in Fig. VI.3, based on the sample statistics. DVCMs display the probability  $P(v \leq \text{DVH}(\hat{d}))$  for all (desired) combinations of dose parameter  $\hat{d}$  and volumes  $v$ .



**Figure VI.2:** Technical evaluation of analytical probabilistic dose-volume for the CTV (a) and the brainstem (b) of the intracranial patient. Sampled values from the original 5,000 dose scenarios as well as from 5,000 re-samples using the multivariate normal dose model are shown as histograms. Their sample mean (vertical solid lines) and sample standard deviation are compared to the analytically computed mean and standard deviation by displaying the respectively parametrized normal distributions  $\mathcal{N}_S$ ,  $\mathcal{N}_R$  and  $\mathcal{N}_A$ . Thereby  $\mathcal{N}_S$  is derived from the dose scenario samples and  $\mathcal{N}_R$  is derived from the re-sampled scenarios.  $\mathcal{N}_A$  gives the normal distribution obtained from APM's moment computations. The vertical dashed line indicates the DVH-point obtained from the nominal dose scenario. Additionally, an exemplary beta-distribution  $\mathcal{B}(\alpha, \beta)$  was parametrized using the transformation rules from Eq. (A.11).

To compare to the assumption of a normal distribution and a beta distribution, the respective DVCMs were computed through their quantile functions. The difference of the DVCMs from sample statistics and the DVCMs deduced from parametrized normal distributions are shown in Figs. VI.3c and VI.3d. For the beta distribution, a similar analyses are given in Figs. VI.3e and VI.3f. Additionally,  $\alpha$ -DVHs (compare Eq. (II.26)) for three values of  $\alpha = \{0.05, 0.5, 0.95\}$  corresponding to the 5-th percentile, the median, and the 95-th percentile DVH. Additionally, the  $\alpha$ -DVHs are compared to evaluate the methodology proposed by Henríquez and Castrillón (2010).

Figure VI.3 shows that the assumption of beta distributed DVH-points is more accurate than assumption of a normal distribution, especially for the CTV. There, differences between the sampled DVCM and the DVCM reconstructed from normal probabilities frequently exceed 20% (of maximum probability 1), while comparison with the DVCM constructed from beta distributions exceeds 10% probability difference only in the maximum dose area. For the OAR, also the beta distribution induces large differences of about 20%, yet such differences are still more pronounced in the approximation with a normal distribution. These differences are partly mitigated when in  $\alpha$ -DVHs computation. There, assumptions of a normal distribution might be also acceptable; difference does not exceed 5% relative volume, yet the  $\alpha$ -DVHs deduced from the beta distribution are accurate within 2% relative volume. The method suggested by Henríquez and Castrillón (2010) leads to substantially deviating  $\alpha$ -DVHs, with differences in relative volume of up to 60% in the target.



**Figure VI.3:** Assessment of the cumulative distribution of a DVH displayed as DVCM. (a) and (b) show the cumulative distribution based on the 5,000 scenario samples, i. e.,  $P(v \leq DVH(\hat{d}))$  for the CTV and the brainstem, respectively. From the DVCMs  $\alpha$ -DVHs were derived for  $\alpha = 0.05, 0.5, 0.95$ , based on the sampled statistics ( $S$ ) and based on the method proposed by Henruez and Castrillon (2010) ( $H$ ). In (c,d) the difference between the sampled  $DVCM_S$  and the  $DVCM_N$  obtained from assumption of marginal normal distributions is displayed, together with the respective  $\alpha$ -DVHs. (e,f) display the analogous difference of sampled  $DVCM_S$  to  $DVCM_B$  derived by assumption of marginal beta distributions.

### VI.3.1.3 EUD approximations

Figure VI.4 evaluates the approaches used for computing expectation and variance of EUD. For the CTV,  $\kappa = -20.5$  was assumed, i. e., the CTV is sensitive to cold-spots in dose (compare Section II.2.1.3). For the OAR increased sensitivity to hot-spots was assumed with  $\kappa = 5.1$ . The values were chosen to demonstrate feasibility with fractional exponents  $\kappa \notin \mathbb{N}$  and were based on Niyazi et al. (2013) and Wu et al. (2002).

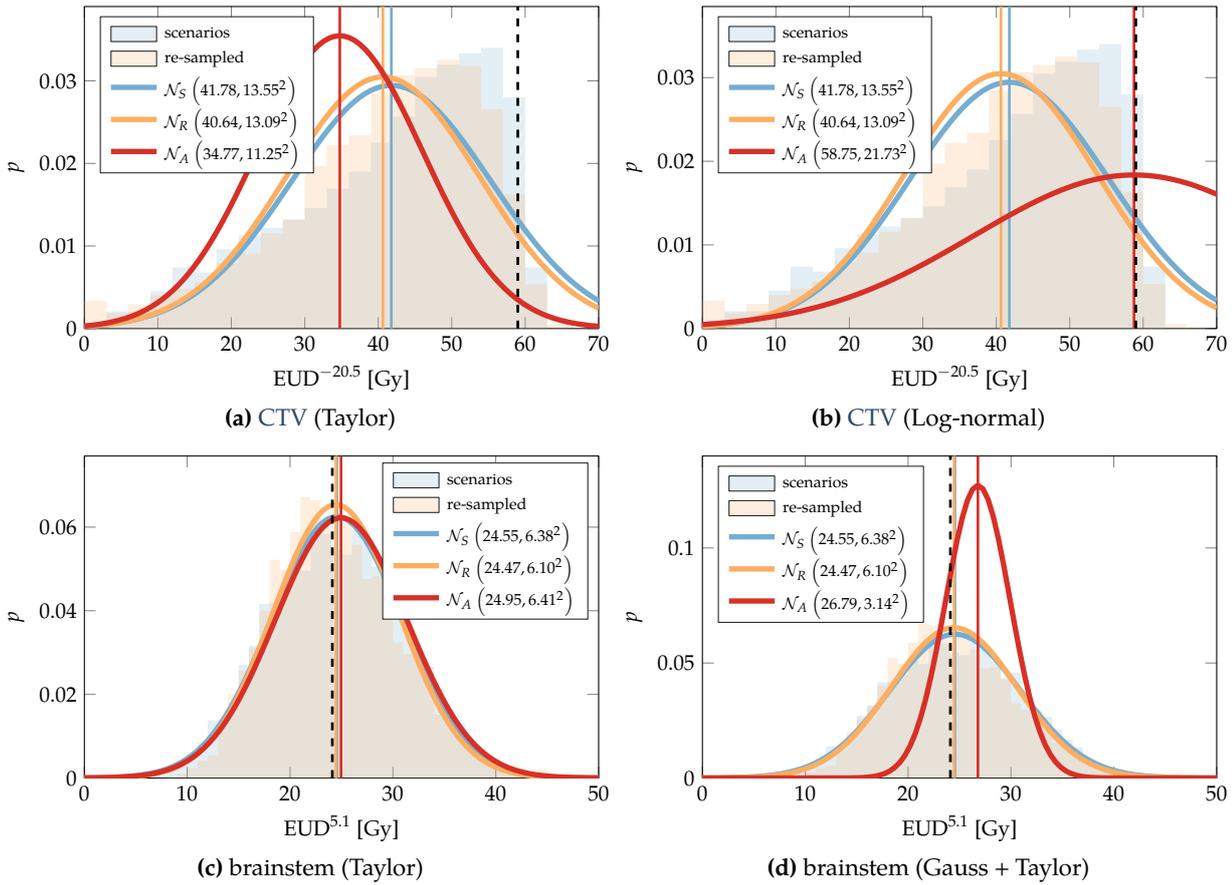
Considering the analytical computations, especially for the CTV with  $\kappa = -20.5$ , significant differences between the analytical computations and sample statistics can be observed. The expected value of EUD computed with Taylor expansion underestimates the sample mean, serving as a benchmark, substantially by 10%. The approach using the multivariate log-normal model in Fig. VI.4b for dose heavily overestimates the expected value by approximately 50%. Comparable deviations are present for the standard deviation. In the EUD approximation for the OAR, the differences are smaller for the Taylor expansion, which slightly overestimates the expected value and accurately models standard deviation. The log-normal model failed completely for numerical reasons, and is therefore not shown.

As a substitute for the failing log-normal approach, a different combined approach is shown in Fig. VI.4d. There, the expressions for the fractional moment of a normal distribution from Eq. (VI.24) were used to estimate the expected value of the sum of powers. Since no expressions could be derived for the variance of this sum, it was approximated with Taylor expansion. The moments of the EUD were then derived using—again—Eq. (VI.24) to propagate the moments of the sum of powers through the  $\kappa$ -th root. This approach yields a too compact Gaussian approximation over the resulting EUD while overestimating the expected value.

The failure of the log-normal approach can be mainly attributed to large variance in combination with the comparably heavy tail of the log-normal distribution. For the CTV, where the variance is usually smaller than the prescribed dose, this already yields substantial overestimation. For the OAR, in which dose gradients manifest and therefore the variance might be more than ten times higher than the actual value, the covariance term in Eq. (VI.26) for estimation of the sum of powers becomes numerically unstable, and even if it can be correctly evaluated, results in an unreasonably high variance estimate for the sum of powers compared to the expectation value. The large variance also leads to deviations in the Taylor-based approach, where the limitations of a first and second order approximation become more and more prominent with increasing  $|\kappa|$ .

Hence, also the maximum and minimum approximation with large absolute EUD exponents ( $|\kappa| = 100$ ) relying on the Taylor approximated moments show large deviations. At first sight, it might surprise that now the minimum is overestimated and maximum is underestimated, contrary to evaluation in Fig. VI.4. Reasons for this might be also found in the imperfect approximation with a non-infinite exponent, which underestimates the absolute extrema in general.

Across all evaluations, comparisons of the re-sampled statistics show slight, yet significant (w. r. t. sampling error) deviation from the EUD sample statistics obtained from the dose scenarios for both CTV and OAR. Yet qualitatively the differences between Gaussians parametrized from the original and re-sampled appear marginal, especially for the EUD computation of the



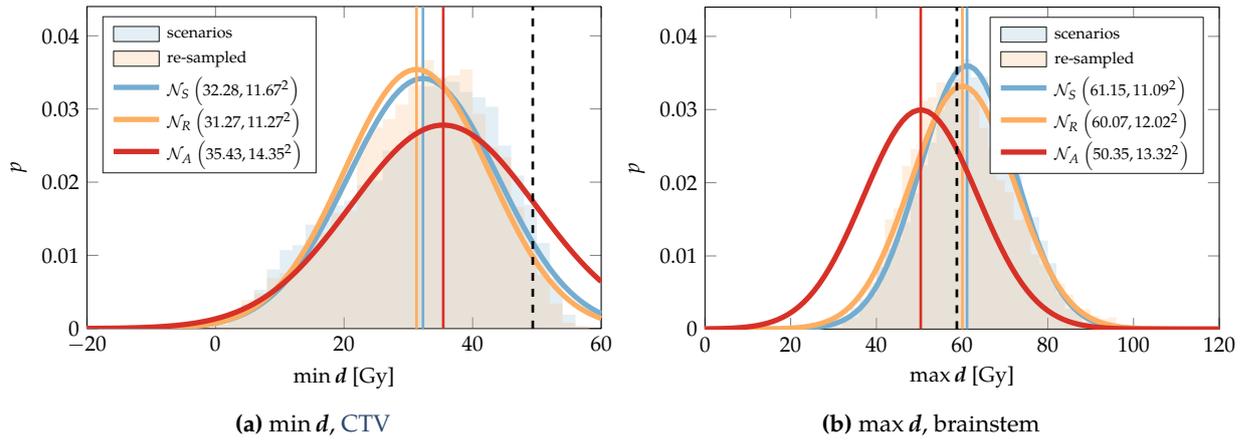
**Figure VI.4:** Technical evaluation of analytical probabilistic EUD for the CTV with  $\kappa = -20.5$  (a,b) and the brainstem with  $\kappa = 5.1$  (c,d) of the intracranial patient using Taylor expansion (a,c) and the Log-normal approximation (b). (d) shows an alternative approach that combines Taylor expansion for variance with the Gaussian approach. Sampled values from the original 5,000 dose scenarios as well as from 5,000 re-samples using the multivariate normal dose model are shown as histograms. Their sample mean (vertical solid lines) and sample standard deviation are compared to the analytically computed mean and standard deviation by displaying the respectively parametrized normal distributions  $\mathcal{N}_S$ ,  $\mathcal{N}_R$  and  $\mathcal{N}_A$ . Thereby  $\mathcal{N}_S$  is derived from the dose scenario samples and  $\mathcal{N}_R$  is derived from the re-sampled scenarios.  $\mathcal{N}_A$  gives the normal distribution obtained from the respective analytical computation with APM. The vertical dashed line indicates the EUD obtained from the nominal dose scenario.

OAR.

## VI.3.2 Evaluation of analytical models on all patients

To better pin down the results from above, analytical treatment plan metric calculation was performed on conventionally optimized treatment plans for all three patient datasets already used in Chapters IV and V. Sample statistics were estimated from the 100 treatment samples used in Chapter V for using a single fraction and 30 fractions. The analysis under more fractions was included to investigate behavior of the analytical models under decreasing variance.

Figure VI.6 illustrates means and their one standard deviation error bars for selected plan metrics. As expected from Section VI.3.1, the best agreement can be seen again for the mean dose  $\bar{d}$ . Largest differences occur for the expected minimum and maximum estimates based on Taylor expansion of EUD. In general, Fig. VI.6 confirms that probabilistic EUD based on Taylor expansion

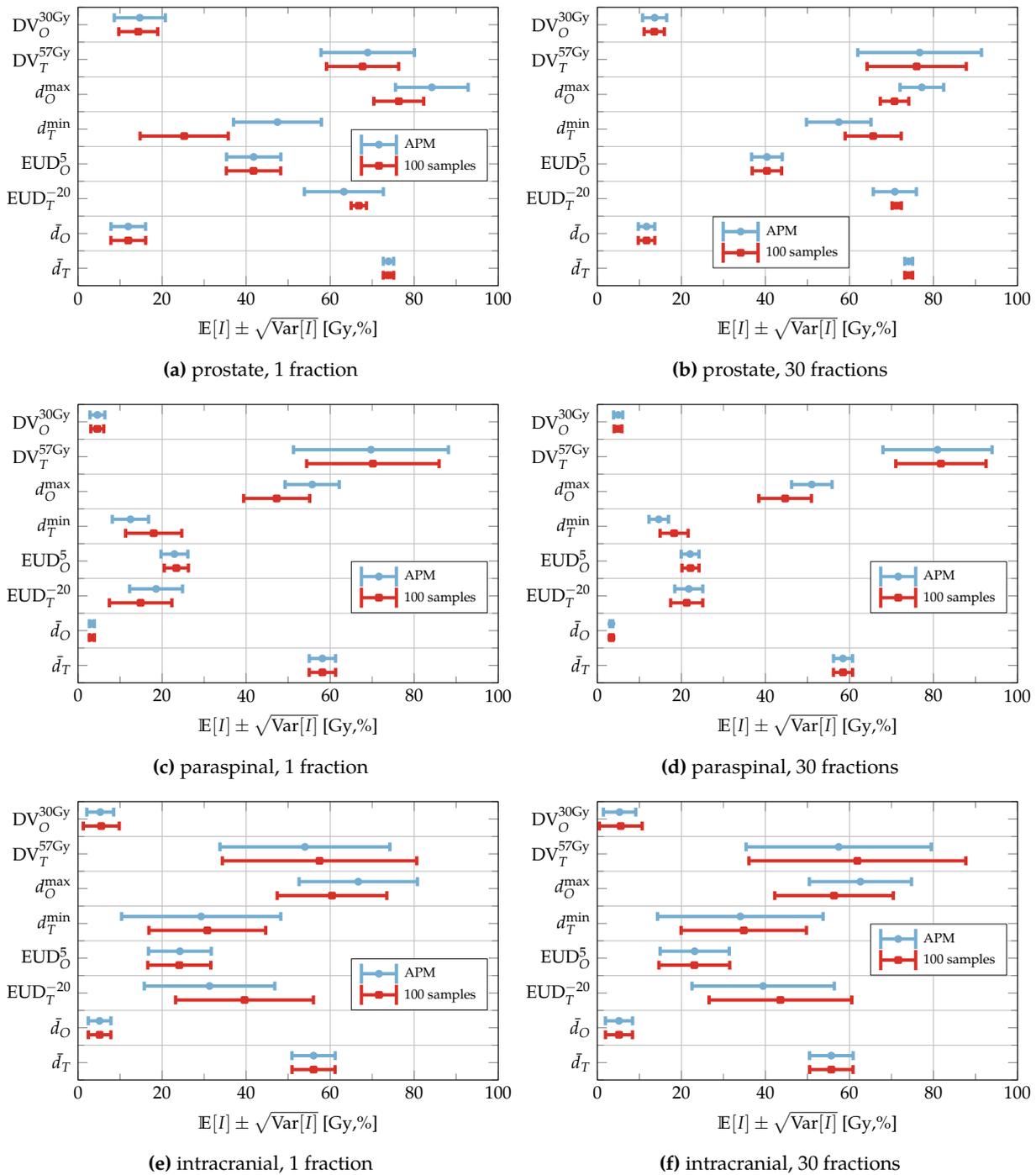


**Figure VI.5:** Technical evaluation of analytical probabilistic dose extrema using  $EUD^{\kappa=100}$  for the maximum dose approximation in the CTV (a) and  $EUD^{\kappa=-100}$  for the minimum approximation in the brainstem (b). Sampled values from the original dose scenarios as well as from the re-sampling using the multivariate normal dose model are shown as histograms. Their sample mean and sample standard deviation are compared to the analytical computed mean and standard deviation by displaying the respectively parametrized normal distribution.

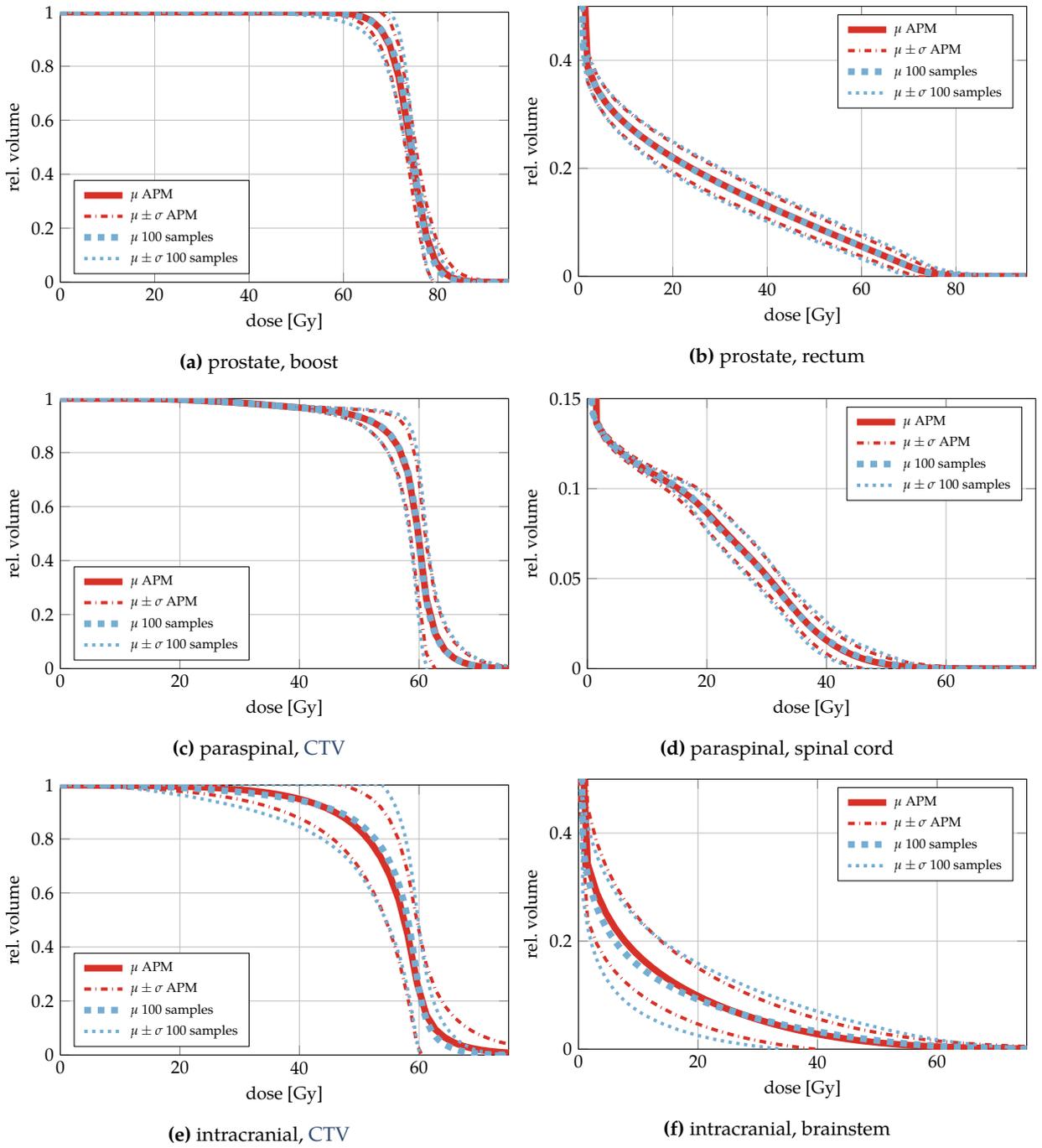
shows larger disagreement with increasing exponent  $|\kappa|$ . Extreme deviations occur especially in the prostate case, which also exhibits the largest relative dose variance because of larger range error due to longer WEPLs and the assumption of a bigger random spatial set-up uncertainty.

The two dose-volume points evaluated per case in Fig. VI.6 show maximum deviations of up to 9% in mean and up to 15% in standard deviation. These maximum values realize in the intracranial case. Hence, the deviations in the mean value correspond to deviations already observed in Fig. VI.2. The differences in standard deviations are larger. Note that the reference sampling calculations rely on 100 scenarios only, which according to Eq. (IV.6) translates into relative sampling errors of 7% for the standard deviation. The accuracy of probabilistic DVH-computation was further evaluated by comparing the analytical first and second moment computations against their sample statistics for full DVHs, as displayed in Fig. VI.7.

Figure VI.7 shows a good agreement between analytical and sampled computations for the prostate and the paraspinal case. There, differences between analytic and sampled expected DVH are within 1% to 3%, and differences in standard deviation reach 4% to 5%, only exceeding this values in the vicinity of the maximum dose where the DVH falls to zero. For the intracranial case, however, the larger deviations observed before remain. An explanation might be given by the smaller VOIs, increasing sensitivity to local shape of the real probability distribution. Furthermore, since for all DVHs a treatment with 30 fractions was assumed, the strong relative sensitivity of the intracranial case to systematic errors in brainstem and CTV leads to larger deviations from the true probability distribution over dose to a multivariate normal distribution. The reason for this is that with larger sensitivity to random errors, with increasing fraction number the shape of the probability distribution over dose will become more and more similar to a multivariate normal according to the central limit theorem. This effect can also be seen in the comparison of scalar plan metrics in Fig. VI.6.



**Figure VI.6:** Comparison of plan metrics computed from the sample covariance matrix using APM and their direct sample statistics for all three patient datasets. Results are presented for conventionally optimized treatments applied in a single (left column, a,c,e) and in 30 fractions (right column, b,d,f). EUDs, mean and minimum/maximum dose (approximated with  $EUD^{-100}/EUD^{100}$ , respectively) are given in Gy on the horizontal axis, dose-volumes are given in % of VOI-volume. The subscript  $T$  indicates computation for the respective target (prostate: boost, paraspinal: CTV, intracranial: CTV). On the other hand, the subscript  $O$  indicates computation for a OAR (prostate: rectum, paraspinal: spinal cord, intracranial: brainstem).



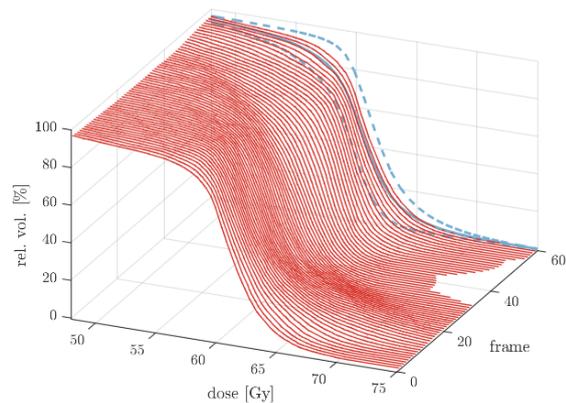
**Figure VI.7:** Analytically computed expected DVHs and their standard deviation compared to DVHs from sample statistics for all three patient cases. 30 fractions were assumed. For each patient case, the DVH for a target and an OAR is shown.

### VI.3.3 Visualization examples

The method described in Section VI.2.4 was applied to exemplarily produce (a) a smooth sample animation of a dose slice and (b) a smooth sample animation of a DVH for the paraspinal case in Fig. VI.8 for two treatment plans: one was generated using conventional, the other one using probabilistic optimization. To do so, the marginal covariances within the slice (for the slice animation) as well as the CTV (for the DVH animation) were computed with APM under the assumptions of 5 fractions. Computations of scenario samples subject to the input uncertainty model were thus not required. Comparing to the respective standard deviation bounds, the sample path is qualitatively reasonable for the DVH. Similar observations can be made when comparing the slice distributions to the respective standard deviation in Fig. V.2e. The visualizations provide an intuitive qualitative understanding where uncertainties might occur. As illustrated in Fig. VI.8 they can be used for a quick comparison of treatment plans with different robustness settings.

(a) dose samples [Gy], conv. opt., 5 fractions

(b) dose samples [Gy], prob. opt., 5 fractions



(c) prob. opt., CTV, 5 fractions

(d) conv. opt., CTV, 5 fractions, all DVHs

(This figure is animated in the electronic PDF version of this thesis.)

**Figure VI.8:** Looped smooth animations of a dose slice (a,b) and a DVH (c) based on the para-spinal case. (a) depicts samples drawn according to Section VI.2.4.1 from a conventionally optimized, non-probabilistic treatment plan. (b) shows similar samples from the probabilistically optimized plan, for which (c) adds the corresponding DVHs for the CTV, where the animated sample (—) is shown with the expected DVH (—) with the one standard deviation bound (- - -). (d) displays all samples from (c) in a three-dimensional static image, in case the animations are not playing or the print version is at hand. The number of frames equals 40 for (a,b) and 60 for (c,d), with the animation running for 4 s.

## VI.4 Discussion

This chapter introduced exact as well as approximative analytical expressions for the expected value and the variance of clinical plan metrics. The computations were based on the assumption of a multivariate normal probability distribution over the dose vector  $d$ .

As one of the more trivial treatment plan indicators, the mean dose could be computed at high accuracy due to its linear dependence on the dose distribution. Focus was laid especially on the exact derivation of expressions for expected value and variance DVH-points, since they are an omnipresent metric in treatment plan evaluation. For the expected DVH, derivations previously exercised by Henríquez and Castrillón (2008a) could be confirmed, and extended by analytically exact computations of (co)variance of DVHs, which generalize to other multivariate distributions over dose as long as the CDF exists and can be parametrized with mean and covariance. Note that these results contradict later work of Henríquez and Castrillón (2010), who claim to have derived  $\alpha$ -DVHs without dependence on the covariance of dose, which would inherently yield variance without regard for correlations between dose voxels. That this is not the case is clearly proven by the requirement of the marginal bivariate CDFs. The results from Henríquez and Castrillón (2010) are further undermined in Fig. VI.3, and therefore their methodology should not be applied when accurate  $\alpha$ -DVHs estimates are desired.

Evaluations of the probabilistic DVH computation on patient data showed reasonable accuracy, with larger deviations arising in the intracranial case. The reason for that is the dominance of systematic error in the evaluated VOIs, which is expected to be reflected in a more non-Gaussian probability distribution over dose. For the probability distribution over single DVH-points, the use of a beta distribution was qualitatively suggested, since the beta distribution is only defined on the interval  $[0, 1]$  and therefore able to provide a physically sound probabilistic model over a volume fraction.

Furthermore, analytical derivations for other treatment plan metrics were attempted; for EUD it was shown that exact analytical derivations are particularly difficult due to the fractional exponents that may be used. Under the assumption of normally distributed dose, this requires computation of fractional moments of dose, which was possible for the expected value but could not be solved for the second fractional moment. The computations might, however, become analytically tractable under assumption of different probability distributions that are not supported on the negative orthant  $\mathbb{R}_+^n$ , which was demonstrated by assuming a multivariate log-normal distributions. However, in particular this assumptions yields problems when the variance is large due to the heavy tail of the log-normal distribution. Hence, the approach of exact analytical computation was not further pursued for the EUD. However, also the negative results in this approach might be worthy for future attempts or when new mathematical theory regarding multivariate fractional moments of normal distributions is developed, or when other multivariate probability distributions are pulled up. For example, the recently introduced “generalized multivariate log-gamma distribution” (Demirhan and Hamurkaroglu 2011) with better control over its shape might be a candidate.

As an alternative for exact analytical probabilistic computations of EUD an approximation us-

ing Taylor expansion was suggested. The Taylor-approximation yielded significantly deviating results, which may be, however, qualitatively useful, especially in cases where the absolute exponent  $|\kappa|$  does not take large values. It was also tried to re-use the Taylor expansion on minimum or maximum approximations using large  $|\kappa|$ , where especially the expectation value exhibits unacceptable deviations. The variance, however, had better approximative quality throughout most probabilistic EUD computations, with the exception of the prostate case.

An alternative for extreme value expectations and variance might be the derivation of those quantities from the inverse probabilistic DVHs and DVCMs. Gordon and Siebers (2009) and Gordon et al. (2010) derived a dose coverage probability from DVCM by taking iso-volume slices of the DVCM. Such iso-slices might be used as a workaround to obtain extreme value statistics as well as order statistics for the dose coverage of a specific volume, that is finding approximate probability distributions over inverse DVH-points (compare Section II.2.1.2). An analysis of the quality of such approximations has, however, been omitted, since this chapter focused on providing purely analytical expressions and approximations. The numeric inversion of the probabilistic DVH dose not fall into this concept.

Furthermore, a new visualization technique was suggested and demonstrated. This technique bases on the considerations of Hennig (2013) for animated sampling of Gaussian processes such that the samples are smoothly connected. Since in this chapter the dose is modeled as a discrete representation of a Gaussian random field, i. e., a three-dimensional Gaussian process, the method is directly applicable.

Such animations, as exemplary presented in Fig. VI.8, may provide added value to the assessment of uncertainties, also in clinical environments. First, they are easy to compute, when a marginal mean vector and covariance matrix are available. These can be not only obtained by analytical methods like APM, but also computed up to respective accuracy (compare Chapter IV) using sampled scenarios, and are therefore comparably versatile. Relying on APM for computation of expected value and covariance of dose, however, opens up new possibilities. Using the separated covariance matrices introduced in Chapter VI, influence of random and systematic components may be directly investigated intuitively, without tedious computation of fractionated treatment samples. Since only the marginal Gaussian approximation is required, they are lightweight to compute for sub-slices or within VOIs. Again, however, the biggest disadvantage is the multivariate Gaussian model, which does not give a completely accurate representation of the “true” probability distribution over dose.



# Constrained Probabilistic Optimization

This chapter proposes a framework for probabilistic optimization with non-linear constraints on explicit quantiles of treatment plan quality metrics. Contrary to stochastic optimization approaches relying on sample statistics, a full analytical pipeline is established relating probabilistic objectives and constraints to the optimization variable, i. e., the pencil-beam weights. Thereby *APM* enables direct control over probabilities in inverse treatment planning.

## VII.1 Overview

Robust and probabilistic treatment planning is usually approached with worst-case and stochastic optimization techniques, respectively, particularly scenario-based expected value approximation (e. g. Unkelbach et al. 2009, using stochastic gradient descent) and minimax worst case optimization (Fredriksson 2012).

This restriction to scenario-based and stochastic optimization techniques goes hand in hand with the limited pool of likewise scenario-based uncertainty propagation methods. Alternatively, for photon therapy, suggestions exist to base optimization on probability distributions to allow probabilistic optimization with closed-form objective functions (e. g. Sobotta, Söhn, and Alber 2010; Unkelbach and Oelfke 2004).

Probabilistic optimization using expected objective functions, as exercised in Chapters IV and V, may be classified as a non-conservative method (Fredriksson 2012); while the probability mass is shifted to improve expectation of the desired objective, robustness against worst cases is not guaranteed. Methods optimizing for the worst case, on the other hand, squeeze the probability mass into a confidence interval without explicit optimization of expectation.

To find a balance between both methods, Fredriksson (2012), for example, discussed conditional value at risk optimization. Sobotta, Söhn, and Alber (2010) defined closed-form objectives on confidence intervals of the approximate normal distribution of volume-based treatment plan quality indicators, to minimize their variation.

*APM*, in its interpretation as a probabilistic dose engine provides the probabilistic analogues to the dose influence matrix, i. e., the expected dose influence matrix and the covariance influence tensor (compare Chapter III). This parallelism in formulation for mapping pencil-beam weights to the moments of the probability distribution over dose can be continued to optimization. With the

derivations from Chapter VI, objectives or constraints can be defined based on an approximative probability distribution over treatment plan metrics. Instead of the nominal dose gradient, these then require the dose's expected value and (co)variance gradient. With probabilistic optimization of Eq. (III.22) based on APM, this approach is already partly implemented. There, only the variance gradient is needed for minimization of the integral variance. This is, however, directly bypassed by usage of the helper matrix  $\Omega$  (compare Eqs. (III.23) and (III.24)).

To extend probabilistic optimization beyond the closed-form probabilistic penalized least-squares from Eq. (III.22), this chapter attempts to generalize the probabilistic optimization problem to, in principle, arbitrary objective and constraint functions depending on expected dose and dose covariance. Based on Chapter VI, new objectives are proposed, including a generalization of the probabilistic penalized least-squares to *piece-wise* penalized least-squares (Eq. (II.17)). Furthermore, constraints based on quantiles of an approximative Gaussian distribution are introduced.

The methodology including the analytical formulations of the objective and constraint functions for optimization are conceptualized in Section VII.2. Section VII.3 then provides a proof-of-concept by exemplary optimization of treatment plans on a one-dimensional prototype anatomy. All concepts are made available online as open-source MATLAB code on GitHub within the "APMToolbox".<sup>1</sup> The chapter is concluded by a critical discussion considering computational applicability of the proposed methodology on full-grown patient datasets.

## VII.2 Methods

The following methodological principles will introduce constrained probabilistic optimization, based on the models from Chapter VI. Thereby, an approach parallel to nominal dose concepts is followed. That is, optimization of nominal dose (compare Section II.2.2) is translated to optimization of mean and covariance over dose, nominal objectives are substituted by its expectation values, and constraints shall work similar to nominal dose constraints, yet with an additional desired confidence parameter.

### VII.2.1 The constrained probabilistic optimization problem

First, the optimization problem from Eq. (II.13) needs to be adapted to mean and covariance of dose. Instead of optimizing  $\mathcal{F}$ , its expectation  $\mathbb{E}[\mathcal{F}]$  or similar probabilistic objectives are optimized. These, as well as confidence constraints, will not depend on the nominal dose distribution, but on its expectation value  $\mu$  and covariance  $\Sigma$ .

As for nominal dose  $d$ ,  $\mu$  and  $\Sigma$  depend on the fluence vector  $w$ , with the relationship expressed with the help of the APM framework (Eqs. (III.12) and (III.19)). With objectives and constraints depending on  $\mu$  and  $\Sigma$ , this allows a separation of the problem analogous to the case of nominal dose. The optimization problem can then be expressed in terms of a composite expected objective

<sup>1</sup><https://github.com/e0404/APMtoolbox>

$\mathbb{E}[\mathcal{F}]$  and likewise composite  $m$  inequality and  $p$  equality constraints  $c_i$  and  $h_j$ :

$$\begin{aligned} \mathbf{w}^* = \arg \min_{\mathbf{w}} \quad & \mathbb{E}[\mathcal{F}(\mathbf{d}(\mathbf{w}))] = \mathbb{E}[\mathcal{F}(\boldsymbol{\mu}(\mathbf{w}), \boldsymbol{\Sigma}(\mathbf{w}))] \\ \text{subject to} \quad & \mathbf{w} \geq \mathbf{0} \\ & c_i(\mathbf{w}) = c_i(\boldsymbol{\mu}(\mathbf{w}), \boldsymbol{\Sigma}(\mathbf{w})) \leq 0 \quad i = 1, \dots, m \\ & h_j(\mathbf{w}) = h_j(\boldsymbol{\mu}(\mathbf{w}), \boldsymbol{\Sigma}(\mathbf{w})) = 0 \quad j = 1, \dots, p \end{aligned} \quad (\text{VII.1})$$

Optimization of problem (VII.1) to find an optimal fluence  $\mathbf{w}$  can then be approached with a quasi-Newton interior-point solver<sup>2</sup> like problems (II.13) & (III.21) before.

The required first-order derivatives can then, analogous to Eq. (II.14), separated into (a) the derivatives of objective and constraints w. r. t.  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ , and (b) derivatives of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  w. r. t. the fluence vector  $\mathbf{w}$ . For a better understanding, the derivatives w. r. t. to  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  will now be discussed separately.

First, consider a scalar function  $f(\boldsymbol{\mu}(\mathbf{w}))$ . The derivative of  $f$  w. r. t.  $\mathbf{w}$  can be found analogously to Eq. (II.14) and is given by

$$\frac{\partial f}{\partial \mathbf{w}} = \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{w}} \right]^T \cdot \frac{\partial f}{\partial \boldsymbol{\mu}} = \mathcal{D}^T \frac{\partial f}{\partial \boldsymbol{\mu}}, \quad (\text{VII.2})$$

with the expected dose influence matrix  $\mathcal{D}$ .

For a second scalar function  $g(\boldsymbol{\Sigma}(\mathbf{w}))$  depending on the covariance, the situation is a little more complicated. Petersen and Pedersen (2012) give a general expression for derivatives of such multivariable function compositions as

$$\frac{\partial g(\boldsymbol{\Sigma})}{\partial w_j} = \sum_{il} \frac{\partial g(\boldsymbol{\Sigma})}{\partial \Sigma_{il}} \frac{\partial \Sigma_{il}}{\partial w_j} = \text{tr} \left[ \left( \frac{\partial g(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} \right)^T \frac{\partial \boldsymbol{\Sigma}}{\partial w_j} \right]. \quad (\text{VII.3})$$

Hence, each element of  $\boldsymbol{\Sigma}$  needs to be derived by each element of  $\mathbf{w}$ . Since the elements of  $\boldsymbol{\Sigma}$  are computed as quadratic forms over sub-matrices of  $\mathcal{C}$  (compare Eq. (III.19)), the derivative can then be expressed as a dot product of the fourth dimension of  $\mathcal{C}$  with  $\mathbf{w}$ :

$$\left( \frac{\partial \boldsymbol{\Sigma}}{\partial \mathbf{w}} \right)_{ijl} = 2 \sum_m \mathcal{C}_{ijlm} w_m \Leftrightarrow \frac{\partial \boldsymbol{\Sigma}}{\partial \mathbf{w}} = 2 \mathcal{C} \cdot_4 \mathbf{w}, \quad (\text{VII.4})$$

where a more compact notation  $\cdot_4$  for the dot product along the 4-th dimension  $\mathcal{C}$  of was introduced.

Consequently, one can express Eq. (VII.3) as

$$\frac{\partial g(\boldsymbol{\Sigma})}{\partial \mathbf{w}} = 2 (\mathcal{C} \cdot_4 \mathbf{w}) :_{1,3} \frac{\partial g}{\partial \boldsymbol{\Sigma}} \quad (\text{VII.5})$$

<sup>2</sup>Within the `APMToolbox`, `fmincon` from `MATLAB`'s optimization toolbox was used with an L-BFGS approximation of the Hessian matrix.

introducing the double-dot product  $\cdot_{1,3}$  along the first and third dimension as

$$\left[ (\mathcal{C} \cdot_4 \mathbf{w}) \cdot_{1,3} \frac{\partial g}{\partial \Sigma} \right]_j = \sum_{il} (\mathcal{C} \cdot_4 \mathbf{w})_{ijl} \left( \frac{\partial g}{\partial \Sigma} \right)_{il} = \sum_{ilm} \mathcal{C}_{ijlm} w_m \left( \frac{\partial g}{\partial \Sigma} \right)_{il}. \quad (\text{VII.6})$$

For the example of an expected objective  $\mathbb{E}[\mathcal{F}](\boldsymbol{\mu}(\mathbf{w}), \Sigma(\mathbf{w}))$  depending on  $\boldsymbol{\mu}$  and  $\Sigma$ , combination of the separated problems outlined in Eqs. (VII.2) to (VII.6) allows expression of the full gradient w. r. t.  $\boldsymbol{\mu}$  as

$$\frac{\partial \mathbb{E}[\mathcal{F}]}{\partial \mathbf{w}} = \mathcal{D}^T \frac{\partial \mathbb{E}[\mathcal{F}]}{\partial \boldsymbol{\mu}} + 2 (\mathcal{C} \cdot_4 \mathbf{w}) \cdot_{1,3} \frac{\partial \mathbb{E}[\mathcal{F}]}{\partial \Sigma}. \quad (\text{VII.7})$$

Depending on the form of the mean and covariance derivatives, this product may be expressed more compactly on-the-fly. On this level of abstraction, however, a general procedure for definition of probabilistic objective functions can be outlined: For each probabilistic objective function depending on (a marginal of) the multivariate dose, i. e., on  $\boldsymbol{\mu}$  and  $\Sigma$ , its gradient w. r. t.  $\boldsymbol{\mu}$  and matrix derivative w. r. t.  $\Sigma$  have to be found. The same holds true for probabilistic constraints, since Eqs. (VII.2) to (VII.6) also generalize to constraint functions depending on  $\boldsymbol{\mu}$  and  $\Sigma$ .

## VII.2.2 Probabilistic objectives & constraints

In the following, probabilistic objectives and inequality constraints for use in optimization problem Eq. (VII.1) will be proposed. For the objectives, this section will limit on the extension of probabilistic least-squares minimization to a piece-wise least-squares approach, while the analytical models from Chapter VII will be picked up to enable quantile constraints.

### VII.2.2.1 Expected piece-wise squared deviation

In Chapters III to V, probabilistic treatment planning based on APM was limited to optimization of an expected least-squares objective, stated in Eq. (III.22) as introduced by Bangert, Hennig, and Oelfke (2013). A least-squares objective cannot implement tolerance doses in OARs and is therefore limited in customizability. Hence, it became standard to instead use a piece-wise version of the objective—at least in the OARs—where the contribution of the objective function is limited to positive (or negative) deviations of a defined maximum (or minimum) tolerance dose (compare Section II.2.2.2 and e. g. Wu and Mohan 2000).

**Heuristic approach:** It is, however, common to extend the least-squares objective function  $\mathcal{F}^{\text{LS}}$  with tolerances, e. g.  $d^{\text{max}}$  as introduced in Eq. (II.17) and denoted as  $\mathcal{F}^{\text{LS}+}$ . One may think of a heuristic adaptation of the probabilistic least squares approach by applying the Heaviside step to the expected value deviation:

$$\mathbb{E}[\mathcal{F}^{\text{LS}}(\mathbf{d})]_+ \stackrel{!}{=} \frac{1}{n} \sum_i \left[ \Sigma_{ii} + \Theta(\mu_i - d_i^{\text{max}}) (\mu_i - d_i^{\text{max}})^2 \right]. \quad (\text{VII.8})$$

Equation (VII.8), however, has its limitations. A voxel with expected value  $\mu_i$  slightly below the threshold  $d_i^{\max}$  will now only contribute with its variance. This strongly simplifies the situation; there might, depending on the variance, still be a high probability that this voxel's nominal dose realization exceeds this threshold, in which case the expected value should be further reduced. Furthermore, a voxel dose that is far below the threshold, may be allowed to exhibit larger variance, which is also not considered.

**Exact approach:** To incorporate this effect, the following derivations target a closed-form solution for the expected value of Eq. (II.17):

$$\begin{aligned}\mathbb{E} \left[ \mathcal{F}^{\text{LS}+}(\mathbf{d}) \right] &= \mathbb{E} \left[ \frac{1}{n} \sum_i \Theta(d_i - d_i^{\max}) (d_i - d_i^{\max})^2 \right] \\ &= \int_{\mathbb{R}^n} \left[ \frac{1}{n} \sum_i \Theta(\tilde{d}_i - d_i^{\max}) (\tilde{d}_i - d_i^{\max})^2 \right] f_d(\tilde{\mathbf{d}}) \, \mathrm{d}\tilde{\mathbf{d}}.\end{aligned}\tag{VII.9}$$

Here, the first moment is again computed via integration, following the general concept of APM as in Chapters III and VI.

From the lessons learned in Chapter VI, Eq. (VII.9) can be solved by ‘‘marginalization’’, i. e., pulling the sum out of the integral and then integrate against the marginal probability density  $f_{d_i}$ :

$$\mathbb{E} \left[ \mathcal{F}^{\text{LS}+}(\mathbf{d}) \right] = \frac{1}{n} \sum_i \int_{-\infty}^{\infty} \Theta(\tilde{d}_i - d_i^{\max}) (\tilde{d}_i - d_i^{\max})^2 f_{d_i}(\tilde{d}_i) \, \mathrm{d}\tilde{d}_i.\tag{VII.10}$$

Assuming  $f_d(\tilde{\mathbf{d}}) = \mathcal{N}(\tilde{\mathbf{d}}; \boldsymbol{\mu}, \Sigma)$ . Through the linear transformation/substitution  $x = d_i - d_i^{\max}$  and  $\tilde{\mu} = \mu_i - d_i^{\max}$  and dropping the subscript  $\sigma = \sigma_i$  for clarity, the integral in Eq. (VII.10) reduces to the elementary problem

$$\begin{aligned}\mathbb{E} \left[ \Theta(x) x^2 \right] &= \int_{-\infty}^{\infty} \Theta(\tilde{x}) \tilde{x}^2 \mathcal{N}(\tilde{x}; \tilde{\mu}, \sigma^2) \, \mathrm{d}\tilde{x} \\ &= \int_0^{\infty} \tilde{x}^2 \mathcal{N}(\tilde{x}; \tilde{\mu}, \sigma^2) \, \mathrm{d}\tilde{x}.\end{aligned}\tag{VII.11}$$

Note that without the Heaviside step  $\Theta(x)$ , Eq. (VII.11) would resolve to the second raw moment of  $\mathcal{N}(x, \tilde{\mu}, \sigma^2)$ , i. e.,  $\sigma^2 + \tilde{\mu}^2$ , leading to Eq. (III.22). The step, however, changed the integration bounds in Eq. (VII.11). Hence the integration has to be exercised explicitly, e. g. through partial integration. A solution can be found by rearranging the integral (A.15) from Owen (1980) as shown in Eq. (A.16) to obtain:

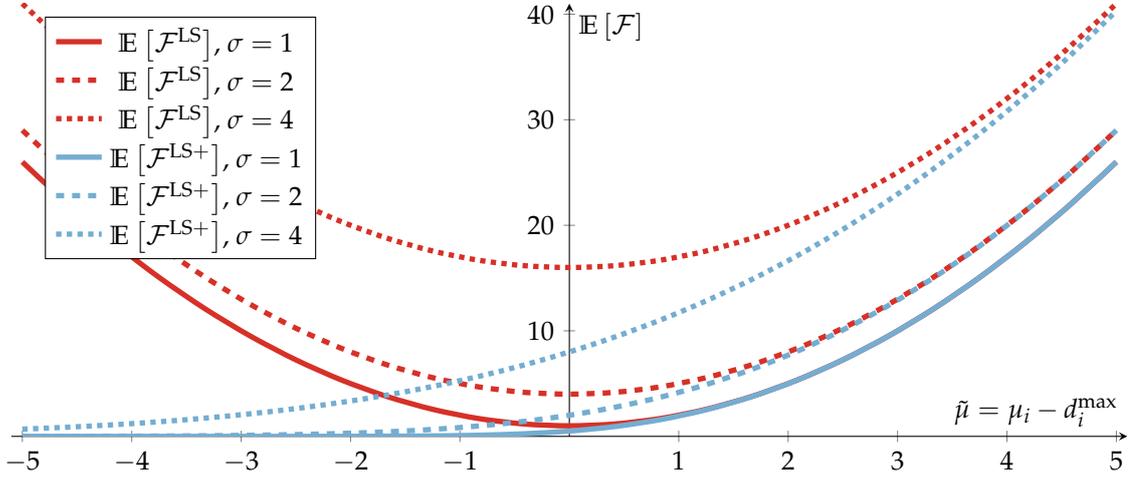
$$\begin{aligned}\mathbb{E} \left[ \Theta(x) x^2 \right] &= \left[ -(\tilde{\mu}^2 + \sigma^2) \Phi\left(\frac{\tilde{\mu} - x}{\sigma}\right) - \sigma^2 (x + \tilde{\mu}) \mathcal{N}(x; \tilde{\mu}, \sigma^2) \right]_{x=0}^{x=\infty} \\ &= (\tilde{\mu}^2 + \sigma^2) \Phi\left(\frac{\tilde{\mu}}{\sigma}\right) + \tilde{\mu} \sigma^2 \mathcal{N}(0; \tilde{\mu}, \sigma^2),\end{aligned}\tag{VII.12}$$

where  $\Phi(x)$  is the CDF of the standard normal distribution.

Now  $\tilde{\mu}$  just needs to be re-substituted in Eq. (VII.12) to obtain

$$\mathbb{E}[\mathcal{F}^{\text{LS}+}] = \frac{1}{n} \sum_i \left[ (\mu_i - d_i^{\text{max}})^2 + \sigma_i^2 \right] \Phi \left( \frac{\mu_i - d_i^{\text{max}}}{\sigma_i} \right) + (\mu_i - d_i^{\text{max}}) \sigma_i^2 \mathcal{N}(0; \mu_i, \sigma_i^2). \quad (\text{VII.13})$$

Figure VII.1 illustrates the difference in expectation value  $\mathbb{E}[\mathcal{F}^{\text{LS}+}]$  imposed by the positivity operator compared to  $\mathbb{E}[\mathcal{F}^{\text{LS}}]$ . When the mean deviation falls below zero, i. e.,  $\tilde{\mu} = \mu_i - d_i^{\text{max}} < 0$ ,



**Figure VII.1:** Comparison of the expected value of the least-squares objective  $\mathbb{E}[\mathcal{F}^{\text{LS}}]$  and the least-squares objective  $\mathbb{E}[\mathcal{F}^{\text{LS}+}]$  with positivity operator for a single voxel  $i$ . Dependence of the mean deviation  $\tilde{\mu} = \mu_i - d_i^{\text{max}}$  is shown under assumption of three different variances.

the expectation value gradually decreases. Note that at  $\tilde{\mu} = 0$ , the expected values are exactly  $2\mathbb{E}[\mathcal{F}^{\text{LS}+}] = \mathbb{E}[\mathcal{F}^{\text{LS}}]$ . Further, Fig. VII.1 shows that simple heuristics, i. e., applying the positivity operator directly on the expected deviation in  $\mathbb{E}[\mathcal{F}^{\text{LS}}]$ , would overestimate the importance of the variance contribution.

All derivations up to here can be analogously carried out for a minimum tolerance value to obtain  $\mathbb{E}[\mathcal{F}^{\text{LS}-}]$ .

**Derivatives:** For efficient optimization of both the heuristic objective  $\mathbb{E}[\mathcal{F}^{\text{LS}}(d)]_+$  from Eq. (VII.8) and the exact objective  $\mathbb{E}[\mathcal{F}^{\text{LS}+}]$  from Eq. (VII.12) with a quasi-Newton method, their first order derivatives w. r. t.  $w$  are required. With Eq. (VII.7), the derivatives w. r. t.  $w$  are computed from the derivatives w. r. t.  $\mu$  and  $\Sigma$ . Since both expected objective functions do not depend on the off-diagonal elements of  $\Sigma$ , only the derivatives w. r. t. variance have to be derived.

For the heuristic approach from Eq. (VII.8), these derivatives can be trivially obtained from Eq. (III.24). For Eq. (VII.12), they are more tedious but also straightforward, resulting in

$$\frac{\partial \mathbb{E}[\mathcal{F}^{\text{LS}+}]}{\partial \mu_i} = \frac{2}{n} \cdot \left[ \sigma_i^2 \mathcal{N}(0; \tilde{\mu}_i, \sigma_i^2) + \tilde{\mu}_i \Phi \left( \frac{\tilde{\mu}_i}{\sigma_i} \right) \right], \quad (\text{VII.14})$$

$$\frac{\partial \mathbb{E}[\mathcal{F}^{\text{LS}+}]}{\partial \sigma_i^2} = \frac{1}{n} \cdot \Phi \left( \frac{\tilde{\mu}_i}{\sigma_i} \right). \quad (\text{VII.15})$$

where for clarity the substitution  $\tilde{\mu}_i = \mu_i - d_i^{\text{max}}$  was used as before.

### VII.2.2.2 Constraints on the $\alpha$ -DVH

In Section VI.2.3.1 closed-form solutions for the expected value and the (co)variance of a DVH were presented, based on mean and covariance of a multivariate normal probability distribution over dose. Further, it was shown in Section VI.3.1 and illustrated in Fig. VI.3 that using the computed expected value and the variance, marginal probabilities over the DVH-points can be parametrized to obtain DVCMs and therefore also  $\alpha$ -DVHs.  $\alpha$ -DVHs correspond to the  $\alpha$ -quantiles of the underlying probability distribution, and can be represented with the quantile function  $Q(\alpha; \mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}])$  of the probability distribution over a DVH-point. Then, minimum and maximum  $\alpha$ -DVH-constraints can be defined as

$$c^{\min \alpha\text{-DVH}} = v^{\min} - Q(\alpha; \mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}]) \stackrel{!}{\leq} 0, \quad (\text{VII.16a})$$

$$c^{\max \alpha\text{-DVH}} = Q(1 - \alpha; \mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}]) - v^{\max} \stackrel{!}{\leq} 0. \quad (\text{VII.16b})$$

Equation (VII.16a) ensures that a DVH-point obeys the inequality  $P(\text{DVH} \leq v^{\min}) \leq \alpha$ , i. e., the DVH-point may fall below  $v^{\min}$  in  $\alpha \cdot 100$  % of all scenarios. To have a similar definition of the maximum constraint in Eq. (VII.16b), the probability argument for  $Q$  is changed to  $1 - \alpha$ ; Equation (VII.16b) then only allows  $\alpha \cdot 100$  % of scenarios to exceed  $v^{\max}$ .

In principle, the choice of  $Q$  is not restricted to a single specific probability distribution over the respective DVH-point. For example, Fig. VI.3 showed that, compared to a normal distribution, choosing a beta distribution yielded a more accurate description of the probability over the DVH-points. However, in the following a normal distribution  $\mathcal{N}(\mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}])$  will be assumed for the sake of clarity, since the beta distribution requires transformations of expected value and variance of the DVH-point, and, to make matters worse, its quantile function is given by the inverse regularized incomplete beta function.

Then the quantile function  $Q(\alpha; \mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}])$ , i. e., the inverse of the CDF, is given by

$$Q(\alpha; \mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}]) = \mathbb{E}[\text{DVH}] + \sqrt{2 \text{Var}[\text{DVH}]} \text{erf}^{-1}(2\alpha - 1). \quad (\text{VII.17})$$

For efficient optimization, the derivatives of  $Q$  w. r. t. to the mean  $\mu$  and covariance  $\Sigma$  of dose are required. Since the first summand of  $Q$  in Eq. (VII.17) is just given by  $\mathbb{E}[\text{DVH}]$ , which is computed as a sum over marginal CDFs (compare Eq. (VI.10)), its first order derivatives w. r. t.  $\mu$  and  $\Sigma$  can be obtained through the definition of the CDF:

$$\frac{\partial \mathbb{E}[\text{DVH}]}{\partial \mu_i} = \frac{1}{n} \cdot \mathcal{N}(\hat{d}; \mu_i, \Sigma_{ii}), \quad (\text{VII.18a})$$

$$\frac{\partial \mathbb{E}[\text{DVH}]}{\partial \Sigma_{ii}} = \frac{1}{n} \cdot \frac{(\hat{d} - \mu_i) \mathcal{N}(\hat{d}; \mu_i, \Sigma_{ii})}{2\Sigma_{ii}}. \quad (\text{VII.18b})$$

In Eq. (VII.18b), only the diagonal elements, i. e., the variance derivatives, are required because  $\mathbb{E}[\text{DVH}]$  does not depend on covariance elements of  $\Sigma$ .

The derivatives in Eq. (VII.18) would already suffice to construct a median (in the case of

$\mathcal{N}(\mathbb{E}[\text{DVH}], \text{Var}[\text{DVH}])$  corresponding to an expected DVH constraint, since at  $\alpha = 0.5$  the second summand in Eq. (VII.17) vanishes. This may be convenient, because then the more costly evaluation of the covariance of dose is not required.

Moreover, the second summand in Eq. (VII.18) w. r. t.  $\boldsymbol{\mu}$  and  $\Sigma$  requires differentiation of  $\sqrt{\text{Var}[\text{DVH}]}$ . With the differentials

$$d \left( \sqrt{2 \text{Var}[\text{DVH}] \text{erf}^{-1}(2\alpha - 1)} \right) = \frac{\text{erf}^{-1}(2\alpha - 1)}{\sqrt{2 \text{Var}[\text{DVH}]}} d \text{Var}[\text{DVH}] \quad (\text{VII.19a})$$

$$d(\mathbb{E}[\text{DVH}^2] - \mathbb{E}[\text{DVH}]^2) = d\mathbb{E}[\text{DVH}^2] - 2\mathbb{E}[\text{DVH}] d\mathbb{E}[\text{DVH}] \quad (\text{VII.19b})$$

the problem reduces to finding the derivatives of  $\mathbb{E}[\text{DVH}^2]$  w. r. t.  $\boldsymbol{\mu}$  and  $\Sigma$ , as the derivatives of  $\mathbb{E}[\text{DVH}]$  have already been obtained in Eq. (VII.18).

The raw second moment  $\mathbb{E}[\text{DVH}^2]$  is, according to Eq. (VI.14), given by a sum over bivariate normal probabilities. While the derivatives w. r. t.  $\boldsymbol{\mu}$  and  $\Sigma$  of the PDF are well known (e. g. for likelihood methods), the derivatives of the CDF are more cumbersome than in Eq. (VII.18).

If one denotes  $F_2(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \rho)$  as the bivariate normal CDF of a bivariate normal distribution with mean  $\boldsymbol{\mu}$ , variances  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)^T$  and correlation coefficient  $\rho$ , the partial derivatives along the first dimension  $x_1$  are given as (Kotz, Johnson, and Balakrishnan 2000)

$$\frac{\partial F_2(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \rho)}{\partial x_1} = \mathcal{N}(x_1; \mu_1, \sigma_1) F_1 \left( \frac{x_2 - \mu_2}{\sigma_2}; \rho \frac{x_1 - \mu_1}{\sigma_1}; 1 - \rho^2 \right), \quad (\text{VII.20})$$

where  $F_1$  is then the CDF of the respective univariate normal distribution.

Kotz, Johnson, and Balakrishnan (2000) also state the partial derivatives w. r. t.  $\rho$  of a standardized bivariate normal CDF:

$$\frac{\partial F_2(\mathbf{x}; \mathbf{0}_2, \mathbf{1}_2, \rho)}{\partial \rho} = \mathcal{N}(\mathbf{x}; \mathbf{0}_2, \mathbf{1}_2, \rho). \quad (\text{VII.21})$$

From Eqs. (VII.20) and (VII.21), using substitution/change of variables, the partial derivatives of  $\mathbb{E}[\text{DVH}^2] = n^{-2} \sum_{il} [1 - F_2(\hat{d} \cdot \mathbf{1}; \boldsymbol{\mu}_{i;l}, \Sigma_{i;l})]$  w. r. t.  $\boldsymbol{\mu}$  and  $\Sigma$  become

$$\frac{\partial \mathbb{E}[\text{DVH}^2]}{\partial \mu_i} = \frac{1}{n^2} \cdot \mathcal{N}(\hat{d}; \mu_i, \Sigma_{ii}) \sum_l F_1 \left( \frac{\hat{d} - \mu_l}{\sqrt{\Sigma_{ll}}}; \frac{\rho_{il} (\hat{d} - \mu_i)}{\sqrt{\Sigma_{ii}}}, 1 - \rho_{il}^2 \right), \quad (\text{VII.22a})$$

$$\frac{\partial \mathbb{E}[\text{DVH}^2]}{\partial \Sigma_{ii}} = \frac{1}{n^2} \cdot \frac{(\hat{d} - \mu_i) \mathcal{N}(\hat{d}; \mu_i, \Sigma_{ii})}{2\Sigma_{ii}} \sum_l F_1 \left( \frac{\hat{d} - \mu_l}{\sqrt{\Sigma_{ll}}}; \frac{\rho_{il} (\hat{d} - \mu_i)}{\sqrt{\Sigma_{ii}}}, 1 - \rho_{il}^2 \right), \quad (\text{VII.22b})$$

$$\frac{\partial \mathbb{E}[\text{DVH}^2]}{\partial \Sigma_{i \neq l}} = -\frac{1}{n^2} \cdot \mathcal{N}(\hat{d} \cdot \mathbf{1}; \boldsymbol{\mu}_{i;l}, \Sigma_{i;l}). \quad (\text{VII.22c})$$

Together with Eq. (VII.19), this allows computation of the covariance derivative of  $\sqrt{\text{Var}[\text{DVH}]}$ . Hence, all derivatives required for optimization with the constraint functions from Eq. (VII.16) could be obtained analytically.

### VII.2.2.3 Mean dose

In Section II.2.2.2, the mean dose objective function  $\mathcal{F}^{\vec{d}}$  was introduced in Eq. (II.19) as a bare linear minimizer of the mean dose  $\vec{d}$ . The probabilistic analog to such an objective, i. e., the expected objective  $\mathbb{E}[\mathcal{F}^{\vec{d}}]$  is straight forward due to linearity of the expectation value as the arithmetic mean over  $\mu$  (compare Eq. (VI.20)):

$$\mathbb{E}[\mathcal{F}^{\vec{d}}] = \mathbb{E}[\vec{d}] = \bar{\mu}. \quad (\text{VII.23})$$

Its gradient is given by

$$\nabla_{\mu} \bar{\mu} = \frac{1}{n} \mathbf{1}, \quad (\text{VII.24})$$

with the vector  $\mathbf{1}$  comprised of unit elements.

With Eq. (VI.20) it is possible to add a variance minimization objective

$$\text{Var}[\mathcal{F}^{\vec{d}}] = \text{Var}[\vec{d}] = \frac{1}{n^2} \sum_{il} \Sigma_{il} \quad (\text{VII.25})$$

with covariance derivative

$$\frac{\partial \text{Var}[\mathcal{F}^{\vec{d}}]}{\partial \Sigma} = \frac{1}{n^2} \mathbf{1} \mathbf{1}^T, \quad (\text{VII.26})$$

where  $\mathbf{1} \mathbf{1}^T$  yields the  $n \times n$  matrix of all ones.

Combining Eqs. (VII.25) and (VII.26) with Eq. (VII.7), objective and gradient may also be computed directly with a pre-computed helper matrix similar to  $\Omega$  introduced in Eq. (III.23). To do so, a contraction-like operation on  $\mathcal{C}$  along the voxel dimensions must be applied:

$$\check{\Omega}_{jm} = \sum_{il} \mathcal{C}_{ijlm}. \quad (\text{VII.27})$$

Similar to Eqs. (III.23) and (III.24) objective and gradient w. r. t.  $w$  are then simplified to

$$\text{Var}[\mathcal{F}^{\vec{d}}] = w^T \check{\Omega} w, \quad (\text{VII.28a})$$

$$\nabla_w \text{Var}[\mathcal{F}^{\vec{d}}] = 2\check{\Omega} w. \quad (\text{VII.28b})$$

A mean dose variance objective  $\text{Var}[\mathcal{F}^{\vec{d}}]$  could thus be efficiently optimized with low memory requirements.

Note that the derivations from Eqs. (VII.23) to (VII.28) also directly translate to the implementation of probabilistic constraints on expected mean dose and variance. Furthermore, when assuming  $\vec{d} \sim \mathcal{N}(\mathbb{E}[\vec{d}], \text{Var}[\vec{d}])$ , constraints on the  $\alpha$ -quantiles of  $\vec{d}$  similar to the  $\alpha$ -DVH constraints from Eq. (VII.16) can be implemented, using the respective quantile function  $Q(\alpha; \mathbb{E}[\vec{d}], \text{Var}[\vec{d}])$ . By substitution of  $\text{Var}[\text{DVH}]$  with  $\text{Var}[\vec{d}]$  in the differential Eq. (VII.19a), also the required derivatives are directly available.

### VII.2.2.4 Equivalent uniform dose

In Section II.2.2.2 objectives for minimizing nominal EUD were introduced as used in `matRad` (Wieser et al. 2017b). Using the Taylor-approximated moments from Section VI.2.3.3, analogous to the previously established mean dose objectives, the expected value and variance of the nominal objective from Eq. (II.18) can be directly based on the considerations from Section VI.2.3.3.

Objective and gradient w. r. t.  $\boldsymbol{\mu}$  of expected EUD are then given by

$$\mathbb{E} [\mathcal{F}^{\text{EUD}^\kappa}(\boldsymbol{d})] \approx \text{EUD}^\kappa(\boldsymbol{\mu}) + \frac{1}{2} \text{tr} \left( H_{\text{EUD}^\kappa(\boldsymbol{\mu})} \boldsymbol{\Sigma} \right), \quad (\text{VII.29})$$

$$\frac{\partial \mathbb{E}[\mathcal{F}^{\text{EUD}^\kappa}]}{\partial \boldsymbol{\mu}} \approx \frac{\text{EUD}^\kappa(\boldsymbol{\mu})}{\sum_i \mu_i^\alpha} \boldsymbol{\mu}^{\circ(\alpha-1)}, \quad (\text{VII.30})$$

Note that in Eq. (VII.30) the gradient of the second order term w. r. t.  $\boldsymbol{\mu}$  was omitted to facilitate efficient computation, which at the same time also leads to a vanishing covariance gradient.

The covariance gradient of EUD variance can be similarly approximated by

$$\frac{\partial \text{Var}[\mathcal{F}^{\text{EUD}^\kappa}]}{\partial \boldsymbol{\Sigma}} \approx \nabla_{\boldsymbol{\mu}} \text{EUD}^\kappa(\boldsymbol{\mu}) [\nabla_{\boldsymbol{\mu}} \text{EUD}^\kappa(\boldsymbol{\mu})]^T. \quad (\text{VII.31})$$

As the case for mean dose above, Eqs. (VII.29) to (VII.31) apply for constraints on expected EUD or its variance, and can also be used for the definition of  $\alpha$ -quantile constraints.

## VII.3 Results

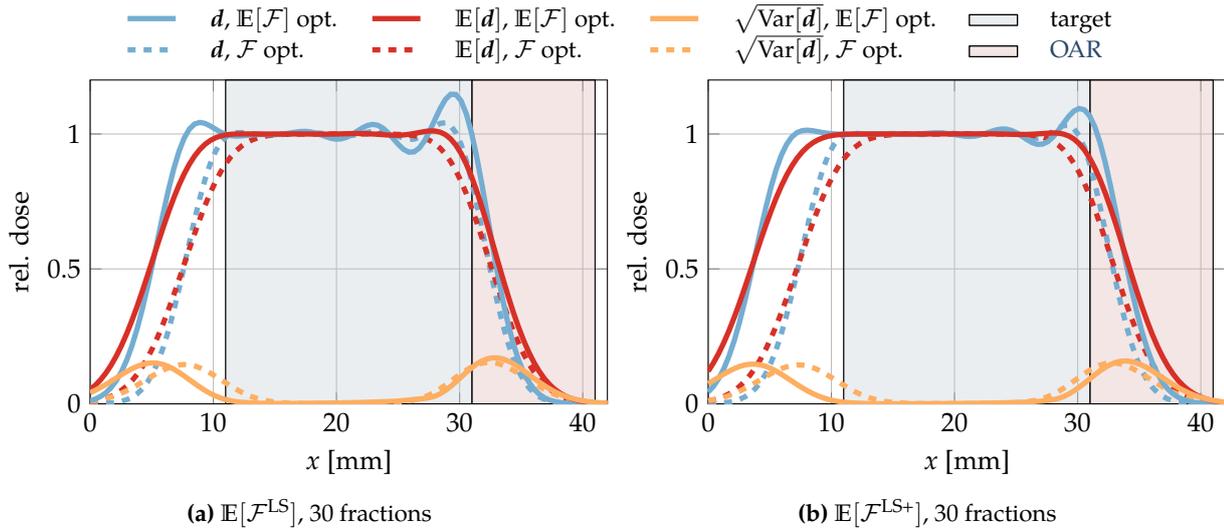
In the following, the in Section VII.2.2 introduced probabilistic objective and constraint functions are evaluated on a one-dimensional prototype anatomy from the `APMToolbox`, where the full tensor  $\mathcal{C}$  can be stored and thus the expressions from Section VII.2 can be evaluated without approximations.

This “1D prototype” represents a target structure of 25 mm width with directly adjacent OAR of half the target’s size. The anatomy is discretized into 100 “voxels” (which are actually one-dimensional line elements in this case), as shown in Fig. VII.2a, and is irradiated with a lateral profile of pencil-beams. The prototype features 20 artificial Gaussian pencil-beams with a dense spacing of 2 mm and a width of  $\lambda_x = 2.5$  mm. Lateral set-up uncertainty over the pencil-beam positions of  $\pm(1 \text{ mm}^{\text{sys}} + 2 \text{ mm}^{\text{rand}})$  (one standard deviation) was assumed for the probabilistic computations, with all pencil-beams exhibiting perfect correlation. A dose of  $d^* = 1$  was prescribed to the target. For optimization, penalties on the objectives of  $p_T = 5,000$  for the target,  $p_{\text{OAR}} = 200$  for the OAR and  $p_{\text{NT}} = 1$  for the unclassified normal tissue were chosen. All analytical derivatives from the previous section were validated with finite difference checks.

### VII.3.1 Evaluation of probabilistic objectives

Figure VII.2 shows nominal and probabilistic lateral dose profiles on the 1D prototype after nominal and expected optimization with the common least-squares and the novel piece-wise least-

squares objective. The corresponding nominal and probabilistic DVHs are supplied in Fig. VII.3.

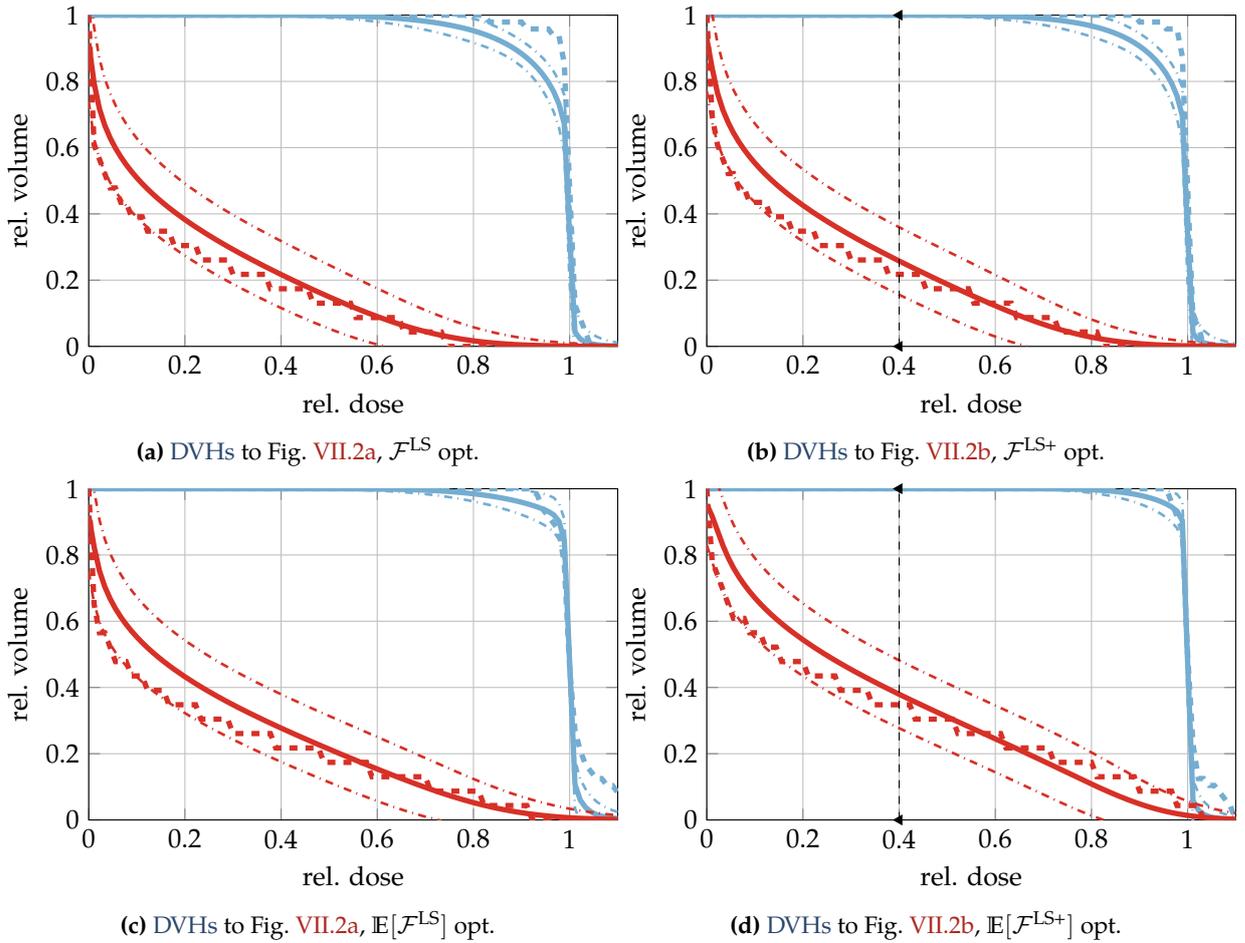


**Figure VII.2:** Comparison of the expected least squares objective and piece-wise least squares objective. Nominal dose as well as its expected value and standard deviation are shown. Plans optimizing the nominal objectives are shown as dashed lines, while plans optimizing the expected objectives are shown as solid lines. (a) shows the lateral profile after nominal and probabilistic optimization with a common least squares objective  $\mathcal{F}^{\text{LS}}$ . (b) shows optimization with the piece-wise least squares objective  $\mathcal{F}^{\text{LS}+}$  and its novel expectation model as derived in Eq. (VII.12). Thereby tolerance doses of  $d^{\text{max OAR}} = 0.4$  and  $d^{\text{max NT}} = 0.6$  were used on the OAR and the unclassified normal tissue, respectively.

As already shown in Chapters IV and V in the full-fledged patient analyses, the minimization of an expected least-squares objective improves conformity of the expected dose compared to the conventionally optimized plan. This is achieved by automatically generating a margin on the one side, and induction of increased dose modulation of the nominal dose at the interface of target to OAR. These results line up with findings from Chapter V and from literature (Fredriksson 2012; Unkelbach and Oelfke 2004). These state that increasing the number of fractions leads to increased modulation of the nominal dose profile, due to the random fraction realizations blurring the heterogeneous nominal dose into a smooth expectation value of dose over all fractions.

In the conventionally optimized plan, introduction of a piece-wise squared objective relaxes the dose penalization within the OAR coverage (compare, in particular, the DVHs in Fig. VII.3a and Fig. VII.3b). The newly introduced *expected* piece-wise squared objective  $\mathbb{E}[\mathcal{F}^{\text{LS}+}]$  carries this relaxation to probabilistic optimization, and mitigates the strong modulation effect of nominal dose to some extent, visible in the profile in Fig. VII.2b and the DVHs in Fig. VII.3d. This can be explained with the reduced “height” of the step from OAR to target, due to the non-zero maximum dose tolerance value.

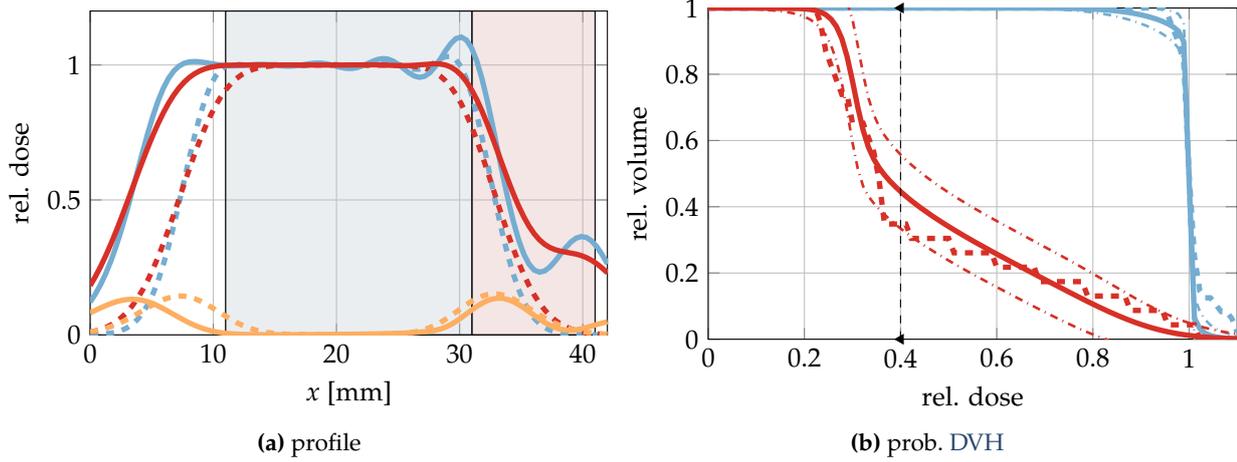
In general, the introduction of the relaxation with the objective function  $\mathcal{F}^{\text{LS}+}$  has more impact in probabilistic optimization. For example, when comparing Figs. VII.3b and VII.3d, the  $\mathbb{E}[\text{DVH}(\hat{d} = 0.4)]$  increases from a volume fraction of about 0.25 in Fig. VII.3b to 0.38 in Fig. VII.3d in the OAR. With the common least-squares objective, the corresponding volume fraction equals 0.22 and increases only to 0.28. On the other hand, the piece-wise least-squares objective results in more robust target coverage. Hence, the expected piece-wise least-squares objective trades expected OAR sparing against improved expected target coverage.



**Figure VII.3:** DVHs corresponding to the dose profiles in Fig. VII.2, i. e., the lateral profiles optimized with nominal (a,b) and expected (c,d) least squares (a,c) and piece-wise least squares (b,d) objective. Solid lines refer to  $\mathbb{E}[\text{DVH}]$  in the target (—) and the OAR (—). The dashdotted lines (— · — / — · —) correspond to the  $\pm 1$  standard deviation bounds. The respective nominal DVHs are complementary given as dashed lines (— · — / — · —). The maximum tolerance  $d^{\text{max}}$  for the piece-wise objective in the OAR is indicated by the vertical dashed line (— · —) in the DVHs.

Substituting the exact expected piece-wise squared objective with the heuristic adaptation from Eq. (VII.8) leads to substantial differences in optimization outcome as shown in Fig. VII.4. Below the tolerance value  $d^{\text{max}}$ , no more reduction of the expected dose is intended by the objective in the OAR, which leads to unnecessary increase of the expected as well as the nominal dose. This results in a flattening of the expected dose profile below  $d^{\text{max}}$  in the OAR. The reason for this flattening lies in remaining variance term in Eq. (VII.8). When the expected dose falls below  $d^{\text{max}}$ , further minimization of the variance goes hand in hand with a reduction of the expected dose gradient, i. e., a flattening of the profile.

Reducing the number of fractions, as exemplary shown for an optimization of  $\mathbb{E}[\mathcal{F}^{\text{LS}+}]$  under assumption of a single fraction in Fig. VII.5, forces the optimizer to create a flat dose profile in the target. Due to the larger uncertainty in the single fraction case, a bigger margin has to be created than in the 30 fraction plan from Figs. VII.2b and VII.3d, substantially compromising OAR sparing. Similar effects were observed with  $\mathbb{E}[\mathcal{F}^{\text{LS}}]$  in Chapter V and by Unkelbach and Oelfke (2004).



**Figure VII.4:** Probabilistic dose profile (a) and DVH (b) after optimization with the “fake” piece-wise expected value from Eq. (VII.8) under assumption of 30 fractions.

Legends in (a) are similar to Fig. VII.2, i. e., the dotted lines represent nominal dose (— · — · —), its expected value (— · — · —) and standard deviation (— · — · —) after optimization with the nominal objective  $\mathcal{F}^{LS+}$ , whereas the respective solid lines (— / — / —) represent the profiles after optimization with the “fake”  $\mathbb{E}[\mathcal{F}^{LS+}]_+$ .

In (b), legends are similar to Fig. VII.3; the  $\mathbb{E}[\text{DVH}]$  within the target (—) and the OAR (—) is shown with the respective standard deviation bounds (— · — / — · —) and the corresponding nominal DVH (— · — / — · —).

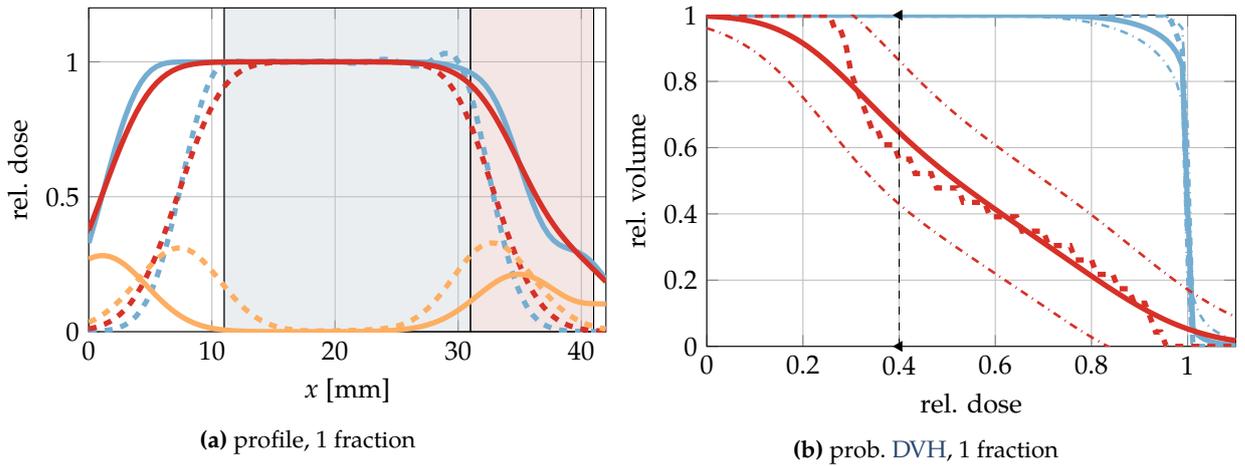
## VII.3.2 Optimization with $\alpha$ -DVH constraints

### VII.3.2.1 Minimum $\alpha$ -DVH constraint on the target

Based on the ratio behind classical margins for photon therapy (Herk et al. 2000), the treatment planner may want to constrain the optimization to ensure that 95 % of the target are covered with 95 % of the prescribed dose. In the DVH of the conventionally optimized lateral profile (Fig. VII.2b), this minimum DVH constraint is already met by the nominal DVH (Fig. VII.3b). Considering uncertainty, however, this constraint is hardly met, which is also reflected by the expected DVH and the associated standard deviation bands following below this point. After probabilistic optimization the situation improves, i. e., the probability mass is shifted producing more conformal expected dose and smaller standard deviation bands (compare also the results from Chapter V). However, the inverse planning process does not provide explicit control over the confidence that this constrained is met.

To enable this control, a minimum  $\alpha$ -DVH constraint as introduced in Section VII.2.2.2 may be used. In Fig. VII.6, the constraint function from Eq. (VII.16a) with  $\alpha = \Phi(-1) \approx 0.159$  was added to optimization, enforcing the  $\alpha$ -DVH corresponding to the lower one standard deviation bound to stay above the original nominal constraint. That is, in approximately 84 % of all cases, 95 % of the target must be covered in at least 95 % of the prescribed dose.

Fulfillment of the minimum  $\alpha$ -DVH constraint to the target requires more dose to the OAR and remaining healthy tissue, i. e., target robustness is traded against OAR sparing. Compared to using only the objective, the constraint yields control over the respective  $\alpha$ -DVH. In Fig. VII.6b, the inequality constraint corresponds to the lower one standard deviation bound, and is thus precisely fulfilled after optimization.



**Figure VII.5:** Probabilistic dose profile (a) and DVH (b) after optimization with expected piece-wise least-squares objective  $\mathbb{E}[\mathcal{F}^{LS+}]$  under assumption of a single fraction.

Legends in (a) are similar to Fig. VII.2, i. e., the dotted lines represent nominal dose (---), its expected value (-.-.-) and standard deviation (-.-.-) after optimization with the nominal objective  $\mathcal{F}^{LS+}$ , whereas the respective solid lines (— / — / —) represent the profiles after optimization with  $\mathbb{E}[\mathcal{F}^{LS+}]$ .

In (b), legends are similar to Fig. VII.3; the  $\mathbb{E}[\text{DVH}]$  within the target (—) and the OAR (—) is shown with the respective standard deviation bounds (--- / ---) and the corresponding nominal DVH (--- / ---).

### VII.3.2.2 Maximum $\alpha$ -DVH constraint on the OAR

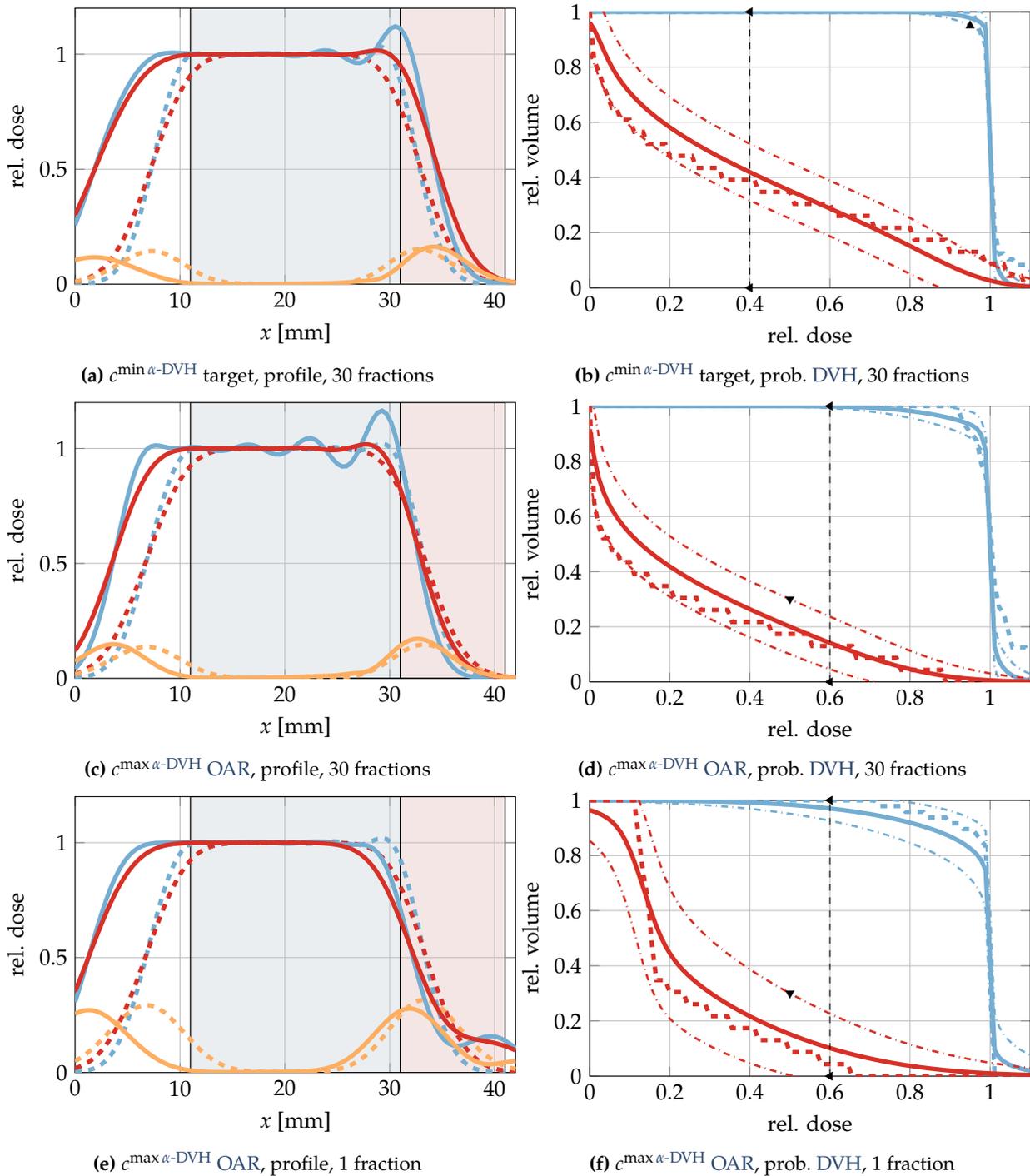
Similarly, a maximum  $\alpha$ -DVH constraint can be applied to the OAR. Figures VII.6c and VII.6d shows the results for a treatment plan optimized with a relaxed piece-wise objective ( $d^{\max} = 0.6$  in the OAR), but a maximum  $\alpha$ -DVH constraint ensuring that the upper standard deviation bound does not exceed a volume fraction of 30% at  $\hat{d} = 0.5$ . Contrary to before, now target coverage is traded against OAR robustness. However, please bear in mind that in the conventionally optimized plan without DVH constraint (Fig. VII.3b), the constraint was already fulfilled (nominally as well as probabilistically), however at substantially worse expected target coverage than after probabilistic optimization with the enabled  $\alpha$ -DVH constraint.

Additionally, Fig. VII.6e and Section VII.3.2.1 exemplary show the situation with a maximum  $\alpha$ -DVH constraint in the case of a single fraction. As discussed before, the typical oscillations under multiple fractions disappear. Furthermore, the increased uncertainty makes it more difficult to enforce the  $\alpha$ -DVH constraint, leading to an even larger decrease of the expected target coverage, which cannot be sustained under the increased plan variability.

### VII.3.2.3 Dependence on the uncertainty model

Note that behavior of the optimization changes under assumption of different correlation models over the pencil-beam uncertainties, especially for the constraints. So far, under assumption of perfect correlation, the uncertainty could be mitigated by creating a margin and the oscillating modulations near the interface of target and OAR, because the perfect correlation induces near-zero variance when the dose is homogeneous.

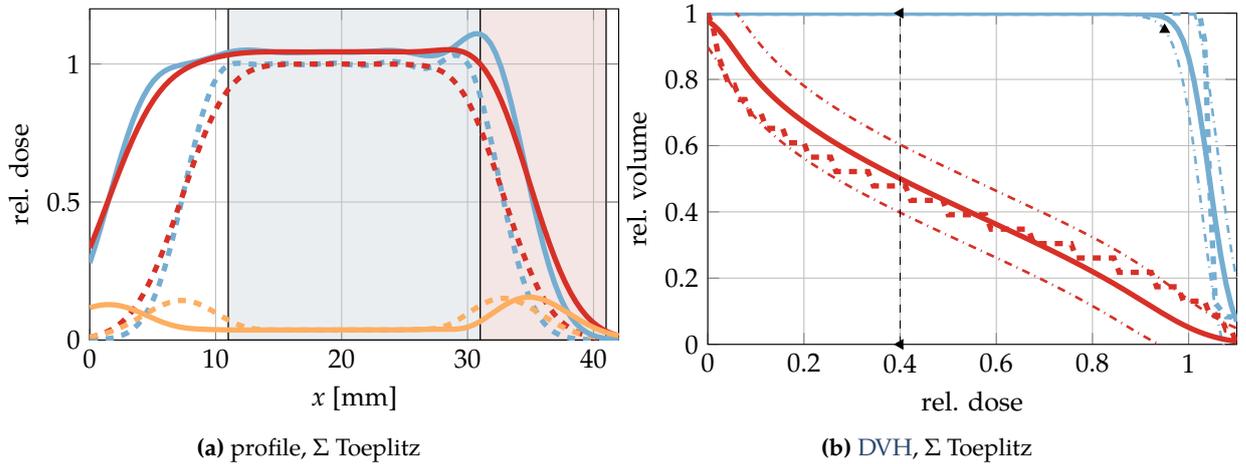
Figure VII.7 shows how the optimization changes under a different uncertainty model; there, the correlation matrix  $C^p$  exhibits Toeplitz structure, i. e., with increasing distance of the pencil-



**Figure VII.6:** Lateral profiles (a,c,d) and probabilistic DVHs (b,d,e) after optimization with  $\alpha$ -DVH constraints. (a,b) show a treatment applied in 30 fractions enforcing a minimum  $\alpha$ -DVH constraint ( $\blacktriangle$ ) on the target at  $\text{DVH}(0.95) \geq 0.95$  on the lower standard deviation bound, i. e.,  $\alpha = \Phi(-1)$ . In (c,d) the minimum  $\alpha$ -DVH constraint was substituted by a maximum  $\alpha$ -DVH constraint on the OAR (with similar probability). (e,f) shows a plan with similar criteria as in (c,d) under assumption of only a single fraction.

Remaining figure legends in (a,c,e) are similar to Fig. VII.2, i. e., the dotted lines represent nominal dose (---), its expected value (-.-.-) and standard deviation (---) after optimization with the nominal objectives and constraints, whereas the respective solid lines (— / — / —) represent the profiles after probabilistic optimization.

In (b,d,f) legends are similar to Fig. VII.3; the  $\mathbb{E}[\text{DVH}]$  within the target (—) and the OAR (—) is shown with the respective standard deviation bounds (--- / ---) and the corresponding nominal DVH (--- / ---).



**Figure VII.7:** Lateral dose profile optimized with a minimum  $\alpha$ -DVH-constraint on the target. The correlation matrix over pencil-beam uncertainties exhibits Toeplitz structure with a slight exponential decrease of  $\exp[-\kappa|\Delta n|]$  with  $\alpha = \ln(10) \cdot 10^{-2}$ , where  $\Delta n$  is difference in pencil-beam index.

Figure legends in (a) are similar to Fig. VII.2, i. e., the dotted lines represent nominal dose (---), its expected value (---) and standard deviation (-.-.-) after optimization with the nominal objectives and constraints, whereas the respective solid lines (— / — / —) represent the profiles after probabilistic optimization.

In (b) legends are similar to Fig. VII.3; the  $\mathbb{E}[\text{DVH}]$  within the target (—) and the OAR (—) is shown with the respective standard deviation bounds (--- / ---) and the corresponding nominal DVH (-.-.- / -.-.-).

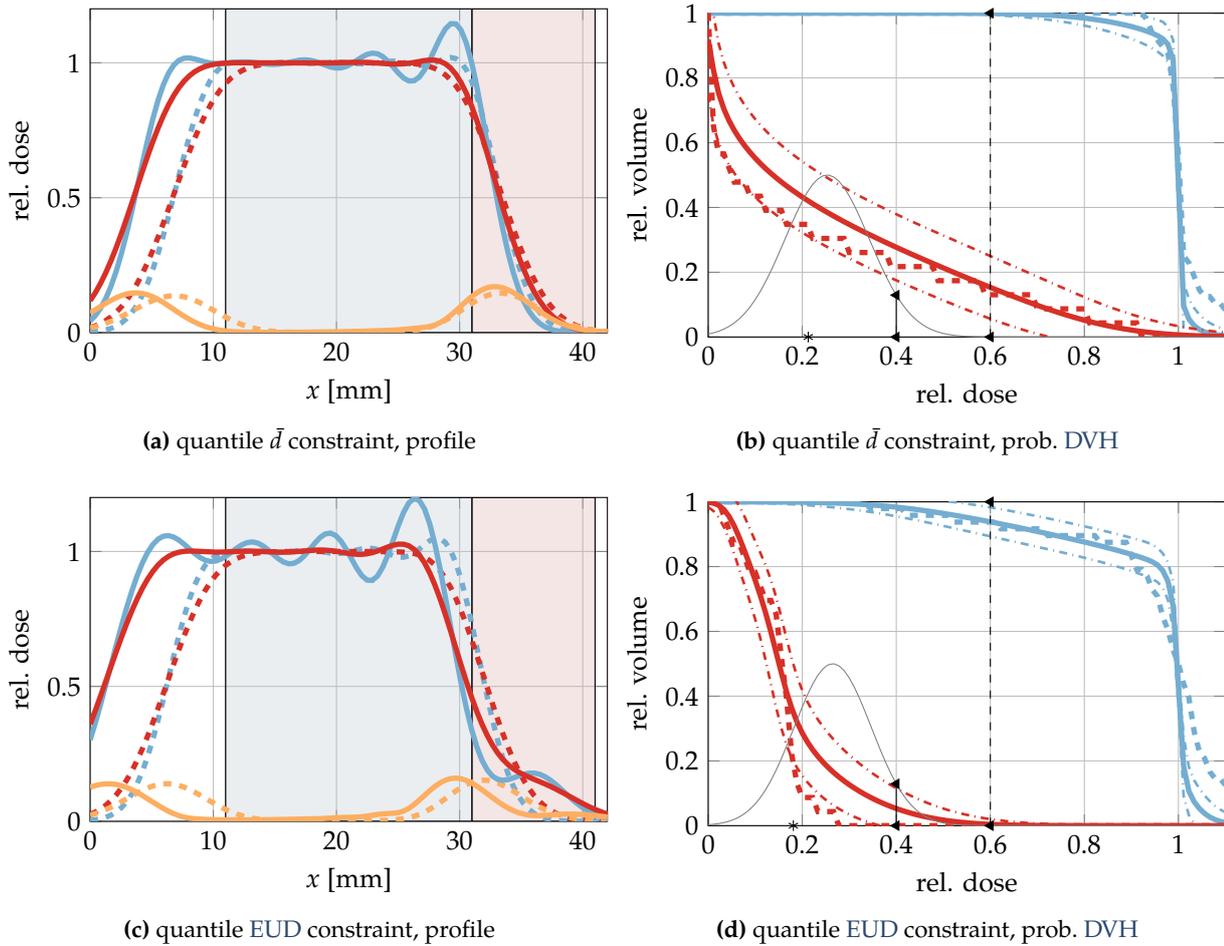
beams their displacement correlation slightly decreases. This leads to a non-zero variance even if the dose is perfectly homogeneous, which requires the expected dose in the target to be increased to exceed the prescribed dose by approximately one standard deviation (i. e., the constraint requirement).

### VII.3.3 Optimization with probabilistic EUD & mean dose constraints

Chapter VI also derived approximate results for EUD, since the attempted analytically exact approaches turned out to be unsuitable. These approximate results, however, did not yield stable results, especially with large (negative) exponents as used for the target or when used as minimum / maximum approximation. For OARs however, the approximations lead to reasonable results, and therefore this section will briefly present results comparing a probabilistic mean dose constraint with a probabilistic EUD constraint.

Figure VII.8 displays the results after using quantile mean and EUD (with  $\kappa = 5.5$ ) constraints on the OAR combined with probabilistic (piece-wise) least squares objectives. A strict probability constraint of  $\alpha = 0.95$  was selected, i. e., only 5% of scenarios are allowed to violate the mean / EUD value of  $0.4d^*$ .

Since the EUD constraint aggravates the sensitivity to high doses compared to the mean dose, the DVHs get pushed to the left, with increased low-dose coverage. Both EUD and mean dose quantile constraints. Again, since multiple fractions were assumed, the constraints are often mitigated with the help of “oscillating” nominal dose in the target. Both mean dose and EUD constraints are not violated significantly.



**Figure VII.8:** Comparison of mean dose and EUD quantile constraints. (a) shows the lateral profile after optimization with a mean dose quantile constraint  $P(\bar{d} > 0.4) \leq 0.05$ , which was substituted in (c) with an EUD quantile constraint  $P(\text{EUD}^* > 0.4) \leq 0.05$ . (b,d) show corresponding nominal and probabilistic DVHs, where the approximated Gaussian of analytically computed mean/EUD is shown (—). The constraints are visualized as solid black lines ( $\blackleftarrow\blackrightarrow$ ), and the nominal value indicated (\*).

The remaining figure legends in (a,c) are similar to Fig. VII.2, i. e., the dotted lines represent nominal dose (---), its expected value (-.-.-) and standard deviation (-.-.-) after optimization with the nominal objectives and constraints, whereas the respective solid lines (— / — / —) represent the profiles after probabilistic optimization.

In (b,d) legends are similar to Fig. VII.3; the E[DVH] within the target (—) and the OAR (—) is shown with the respective standard deviation bounds (--- / ---) and the corresponding nominal DVH (-.-.- / -.-.-).

## VII.4 Discussion

This chapter demonstrated a novel method enabling probabilistic optimization of expected objectives and quantile constraints. These objective and constraint functions used the findings of Chapter VI, and therefore base on a completely analytical propagation from the input space, i. e., pencil-beam weights and the respective uncertainty model. This facilitated implementation of the optimization as the “probabilistic analog” to nominal dose optimization relying on a quasi-Newton interior-point algorithm provided with analytical objective and constraint first-order derivatives and Hessian approximation via L-BFGS.

Similar to nominal dose optimization as, for example, used in `matRad` (Wieser et al. 2017b), the computation of the objective and constraint derivatives factorizes; they can be computed w. r. t.

the respective marginal expected value and covariance over dose (compare Chapter VI), and afterwards multiplied with the expected value and covariance derivatives of the probabilistic dose influence provided by the original APM framework.

This makes the novel concept versatile and difficult to handle at the same time. Versatile, because it also generalizes to other methods that can provide the expected value and covariance gradient of probabilistic dose influence. Difficult to handle, because especially the covariance influence  $\mathcal{C}$ , despite being sparse, is voluminous with its dimensionality both squared in number of voxels and pencil-beams. For a rule-of-thumb estimate of the complexity, note that the dense covariance influence for the evaluated prototype stores  $100^2 \times 20^2$  elements and thus already requires 30 MB of RAM, compared to the dense expected dose influence only storing  $100 \times 20$  elements, i. e., 16 kB, in RAM. These numbers show that for actual application on large-scale 3D patient cases, where a sparse expected dose influence matrix easily has several million non-zero entries, it is infeasible to store the full  $\mathcal{C}$ , even in a sparse format. Hence, its values require, for example, on-the-fly computation whenever the pencil-beam weights change, which may be time consuming when using either common sampling based methods or the APM framework. While the latter exhibited the possibility to precompute certain calculations (i. e., the helper matrix  $\Omega$ ), this does not generalize to the more complex objective and constraint functions required for, most importantly,  $\alpha$ -DVHs.

The ability of APM to efficiently evaluate single elements  $\mathcal{C}_{ijlm}$  on request may encourage the use of decomposing approximation techniques on  $\mathcal{C}$  (Grasedyck, Kressner, and Tobler 2013). Such approximations of large tensors are of big interest in academic as well as commercial applications. For example, on the tested 1D lateral profile prototype used within this chapter, heuristic attempts of computing CANDECAMP/PARAFAC (Acar, Dunlavy, and Kolda 2011; Acar et al. 2011) or Tucker decompositions (Kolda and Sun 2008) of  $\mathcal{C}$  using the MATLAB Tensor Toolbox (Bader and Kolda 2015) showed promising results in terms of size reduction and speed-up of the evaluation of Eqs. (VII.4) and (VII.7), i. e., the gradient computations, and Eq. (III.19), i. e., the covariance evaluation. Additionally, tensor train decompositions (Oseledets 2011a; Oseledets 2011b; Oseledets, Tyrtshnikov, and Zamarashkin 2011) may be an alternative. In the future, such approximations may be explicitly implemented and evaluated w. r. t. computational feasibility and approximative quality on the full-fledged problem on patient cases. Also, considering that the dose covariance matrix exhibits strong spatial structure depending on the input correlation model, functional representations of the spatial component (i. e., voxel indices  $i$  and  $l$ ), might facilitate a representation of  $\mathcal{C}$  that allows usage on larger, real world patient datasets.

Evaluation on the prototype, nonetheless, successfully demonstrated the theoretical applicability of the developed concepts. The derived objectives and constraints could be easily used with standard interior-point optimization methods using analytical gradients, without implementing custom adaptations to the optimization routine.

Regarding probabilistic quantile constraints, the implementation and evaluation of  $\alpha$ -DVH constraints was highlighted. These allowed definition of concrete probabilities, under which the respective DVH-point inequality should be enforced. For both such minimum and maximum quantiles, the interior-point optimization was able to follow those constraints, by creating dose

distributions that intrinsically define a margin around the patient, but also in exploiting fractionation assumptions. The automatic generation of margins is a desired effect, which is underlined by other works trying to adapt conventional margin adaptations schemes from the photon world (Herk 2004; Herk et al. 2000) to robust proton therapy optimization (e. g. Fredriksson and Bokrantz 2016; Knopf et al. 2013). The exploitation of the fractionation scheme, i. e., inducing heterogeneity to the nominal dose since the profile “flattens” when multiple scenarios cancel each other, is often classified as an undesired effect, and may be mitigated by additionally imposing uncertainty on the random error assumption (Fredriksson 2012; Unkelbach and Oelfke 2004). One should consider that these “oscillations” of the nominal dose profile are only that severe because of the ideal anatomy in the evaluated prototype. When multiple beams imping on the target subject to multiple uncertainty sources, these effects are not as strong, which was already demonstrated within this thesis in Chapter V.

In addition to the  $\alpha$ -DVH constraints, also EUD and mean dose quantile constraints were evaluated, where optimization also provided sensible results. Such constraints might be useful since often clinical objectives are based on those indicators (Marks et al. 2010). In this context, the mean dose quantile constraint—which on top could be efficiently implemented—may be used without any problems, since its analytically exact quantification is accurate according to the results from Chapter VI. Caution should be used, however, with EUD, since the approximations derived in Chapter VI are not particularly stable, especially when the absolute exponents are large. If larger accuracy is required for the  $\alpha$ -DVH-constraints, one could rely on the quantile function of a beta distribution. This would require some additional calculations including handling of the regularized incomplete beta function, but is in general, within the realms of possibility.

Not implemented and evaluated were expected objectives or constraints using minimum dose and maximum dose, despite the evaluation of approximations with EUD discussed in Chapter VI. The reason for this is the insufficient accuracy/stability that was concluded from the respective results in Chapter VI. For the experimental reader, however, it would be straightforward to continue on one’s own behalf using the EUD quantile constraint.

Focus regarding new probabilistic objectives was laid on extension of the expected least-squares objective to the expected value of a piece-wise extension which is more commonly used in conventional treatment planning. An exact analytical solution could be found that can overcome limitations of heuristic adaptations to simulate such an objective. Its evaluation is, however, more expensive than the expected least squares objective, which allowed pre-computations of the helper matrix  $\Omega$ . Note that the derivations made for probabilistic  $\alpha$ -DVHs, EUD and mean dose constraints may also be used to implement objectives to optimize their expectation value or variance, comparable to the works by Sobotta, Söhn, and Alber (2010).

Hence, concluding this chapter, one has to highlight the versatility of this concept once again. First, it offers a clear, closed-form description between the probabilistic optimization objective and the dose uncertainty. The parametrization of the plan metrics with a probability distribution allows for a continuous transition between robustness against exactly defined worst-cases (i. e., high/low probability quantile constraints) or probabilistic expected value optimization (i. e., probabilistic objectives or median probability quantile constraints). Constraints behave like their

nominal counterpart, and require just a single additional parameter, i. e., the desired confidence. This gives freedom to the treatment planner, who can “play around” with different settings, while at the same time lightens the workload with respect to ensuring treatment plan robustness.

It is, however, to evaluate how the concept competes with other established approaches, when the aforementioned computational hurdles are overcome. Especially these computational requirements, are a risk factor in proving the concepts applicability. Even if future studies render the concept disadvantageous, it might, however, still be of educational and conceptual value to understand the relation between input uncertainties and output plan metrics. Furthermore, the analytical models within *APM* as well as the optimization are separated into two problems, namely (a) modeling of the expected dose distribution  $\mu$  and its covariance  $\Sigma$  depending on fluence  $w$ , and (b) modeling of the probability over plan metrics based on  $\mu$  and its covariance  $\Sigma$ . Hence, the methodology from this chapter could be also used in other frameworks able to approximate  $\mu$  and  $\Sigma$  and their derivatives. Then, the closed-form expressions of the here presented objectives and constraints could be realized and also be helpful to implement an efficient optimization routine within these frameworks.

## Conclusion

### VIII.1 Summary

Based on the *Analytical Probabilistic Modeling* framework introduced by Bangert, Hennig, and Oelfke (2013), this thesis developed probabilistic analogs to conventional proton treatment planning concepts. This includes implementation of *APM* as a probabilistic dose calculation engine as well as the first application to patient data, mainly covered by Chapters IV and V. In the remaining chapters, novel models were developed to enable a closed-form probabilistic description of treatment plan metrics that can be used as objectives or constraints in probabilistic treatment planning workflows.

*APM* was implemented into the image processing platform *DIPP/MITK* as a plugin (*MITKrad*), which was developed completely from scratch. With careful software design it was attempted to build a platform that encourages extension and sustainable use, while also archiving the developed algorithms relevant for this thesis. To showcase the developed concepts and enable reproducibility, especially of the newly developed concepts from Chapters VI and VII, simplified *MATLAB* code has been provided online within the *APMToolbox*.

Chapter IV described the implementation of *MITKrad*, which enables basic treatment planning workflows including the definition of beam geometry, nominal dose calculation and treatment plan optimization (using *IPOPT*) including display of the final result. *APM* was incorporated into *MITKrad* as a probabilistic dose engine, i. e., with an extended interface *APM*'s nominal dose can be computed and optimized, as well as the first and second moment of dose. The implementations were then used to create exemplary non-probabilistic as well as probabilistic treatment plans on three patient datasets, which were comparable with results achieved through other academic methods. Further, the probabilistic dose calculation, i. e., the computed expectation value and standard deviation of dose, were validated against multivariate random sampling, showing nearly perfect agreement. While a run-time measurement showed competitive computation times for *APM*, particularly when it comes to optimization, approximation techniques were proposed and evaluated. These enable qualitative approximations for probabilistic analysis as well as treatment plan optimization.

Thereafter, in Chapter V, the by Bangert, Hennig, and Oelfke (2013) proposed method to handle fractionation within *APM* was overhauled and evaluated. The original formulation by Bangert,

Hennig, and Oelfke (2013) was already of constant time complexity in the number of fractions, but required re-computations when the number of fractions changed. The new formulation, however, allowed the expression of covariance influence as linear combination of a systematic and random term. Hence, only a single computation is required for probabilistic evaluation and optimization under varying fractionation schemes. This was exploited in the creation of treatment plans under different fractionation assumptions, where the assumptions of more fractions allowed the optimization to reduce dose to OARs while maintaining relative robustness of target coverage. Plans that are probabilistically optimized under assumption of a single fraction, but then applied in multiple fractions, did not share this reduction of dose to the OARs.

With Chapters IV and V working in the domain of propagation of parametric uncertainties to dose, Chapter VI attempted to carry the propagation to dose-dependent clinical plan metrics. Exact analytical expressions were found for the first two moments of DVHs and mean VOI dose, which yielded accurate results under the assumption of a multivariate normal distribution over dose. For EUD, analytically exact and applicable representations could not be derived. Approximations were attempted with a Taylor-expansion on EUD, which were also used to approximate minimum and maximum. However, evaluation on the three patient datasets shows that these were less accurate than the exact derivations for DVH and mean dose.

Last but not least, Chapter VII used the closed-form expressions from Chapter VI to conceptualize constrained probabilistic optimization on treatment plan metrics with an interior-point method. To do so, the constrained probabilistic optimization problem was formulated analogously to the nominal constrained optimization problem, with expected dose and covariance over dose replacing nominal dose. Assumption of normal distributions over the respective treatment plan metrics enabled the definition of quantile constraints, which extend the nominal constraint function with a confidence probability. Furthermore, the previously used probabilistic least-squares objective was extended to allow positivity operators/tolerance values. On a one-dimensional test anatomy, lateral pencil-beam profiles were successfully optimized under various parametrizations of the different constraint functions, proving the concept.

## VIII.2 Innovation

This thesis provided the first implementation of a probabilistic dose calculation engine, based on the APM framework introduced by Bangert, Hennig, and Oelfke (2013). Compared to other concepts, for example the PCE approach by Perkó et al. (2016) or the Gaussian process model from Sobotta, Söhn, and Alber (2012), probabilistic analogs to the dose *influence* are provided, enabling direct usage for probabilistic treatment plan analysis *and* optimization. Furthermore, it is the first implementation that does not require multiple computations of dose scenarios. Instead, analogous to the nominal pencil-beam algorithm, distinct elements of expected dose influence and covariance influence can be independently computed, allowing novel approximation techniques and re-formulations.

The introduced re-formulation for fractionated treatments further distinguish the scientific work herein from previous attempts. That is the analysis of treatment plans under optimiza-

tion with different fractionation assumptions, where *APM* is capable of computing the complete fractionation spectrum of the covariance influence. Compared to other works attempting the modeling of random and systematic uncertainties (Lowe et al. 2016; Lowe et al. 2017; Perkó et al. 2016) in proton therapy, the methodology described herein is, to the best of the author's knowledge, the first able to model the full-fledged interplay of systematic and random errors without simplifications at constant time complexity *while also* generalizing to optimization.

Closed-form uncertainty propagation from dose uncertainty to uncertainty in treatment plan metrics, and its subsequent use for prototyping constrained probabilistic optimization comprises the main conceptual scientific contribution. The closed-form derivations of probabilistic DVHs surpass the sole found previous approach (Henríquez and Castrillón 2008a; Henríquez and Castrillón 2008b), which was limited to exact expectation value and heuristic, yet inexact modeling of  $\alpha$ -DVHs. The transition to probabilistic constraint functions gives the first method that enables a scenario-free incorporation of such confidence constraints, which allow to *prescribe* a desired probability to the respective plan metric that is enforced during optimization, if possible. This exceeds the capabilities of other methods relying on computed scenarios (Mescher, Ulrich, and Bangert 2017, for photon therapy; Fredriksson 2012, using conditional value-at-risk optimization).

### VIII.3 Limitations & Outlook

This thesis proved the applicability of an analytical method to describe random and systematic uncertainties over the course of a radiation treatment. The model for random and systematic uncertainties so far does, however, neglect numerous other sources of uncertainty. Amongst these are intra- and inter-fractional anatomical motion and deformation (Knopf et al. 2013; Liu et al. 2012a; Lomax 2008a; Ulrich et al. 2017). Also, biological effects and uncertainties of particle irradiation are not considered in this work, but already part of ongoing research with *APM* (Wieser et al. 2017a). Future work needs to address these issues; either by introduction of voxel-based uncertainty assumptions or through an approximate projection into the BEV uncertainty model. *APM* is especially suited for such incorporation of high-rank correlation models. The closed-form mapping may, in this context, also prove valuable to estimate this uncertainty model.

In Chapter IV, it was shown that *APM* provides accurate estimates of expected dose and standard deviation of dose subject to common uncertainty assumptions over the input parameters. This validation, however, compared to the nominal formulation of the analytical pencil-beam dose calculation within *APM*. It has been shown, however, that nominal analytical pencil-beam dose calculation algorithms themselves have their limitations with extremely heterogeneous anatomies as given, for example, in the case of lung treatments (Taylor, Kry, and Followill 2017). While the nominal inaccuracy might be mitigated by relying on Monte Carlo dose calculation, it has to be quantified if the superior accuracy of Monte Carlo dose calculation is also advantageous in the presence of uncertainties, or if the probability space neutralizes this effect and a probabilistic dose calculation engine like *APM* might be applicable nevertheless. Furthermore, accuracy of *APM*'s underlying pencil-beam algorithm might be improved by using multiple lateral Gaussian components (Bellinzona et al. 2015) or pencil-beam fine-sampling (Soukup, Fippel, and Alber

2005). Also, the extension to biologically effective dose—which is part of ongoing research for carbon-ions (Wieser et al. 2017a)—deserve an investigation for protons since recent studies challenge the assumption of constant radio-biological effectiveness of protons compared to photons (Jones 2016; Mohan et al. 2017).

Regarding the analytical models developed in Chapter VI, further work could focus on finding more accurate probabilistic models for EUD and extreme values. For the latter, academic developments regarding extreme value/order statistics should be kept track of. Suggested methods to derive probabilistic estimates for extreme values or order statistic for correlated random variables do, however, mostly rely on rely in some way on iterative evaluations, and it is unclear if exact theory that is computationally feasible in the radiotherapy context with thousands of correlated random dose voxels will be available in the near future. If only used for probabilistic analysis, it might be more reasonable to re-sample from the approximate multivariate normal distribution to estimate such extreme values. Additionally,  $\alpha$ -DVHs constraints, which could be successfully established within this thesis, can be used as surrogates to evaluate and control uncertainty over extreme dose values in treatment planning.

The true potential of the derived analytical models is revealed when used in optimization. However, if the concepts from Chapter VII shall find their way into clinical application, it is crucial to find a way for efficient handling or approximation of the covariance influence information. There, collaboration with the machine learning community could be of value, who often investigate problems in which the complete computation or storage of the model is infeasible due to dimensionality. For example, tensor decompositions could be evaluated (e. g. Grasedyck, Kressner, and Tobler 2013). In particular, careful analysis and exploitation of structure of the covariance influence tensor could be of value, since patient and irradiation geometry impose characteristic spatial structure. Alternatively, massive parallelism on distributed computer systems would be an option. The growing availability of commercial cloud systems could also make such solutions feasible and affordable in clinical environments without sophisticated computing infrastructure.

## VIII.4 Closing words

In conclusion, the findings of this thesis show that APM in its original form might be valuable for fractionated treatments in comparably static indications, as uncertainty estimates can be efficiently computed and mitigated with probabilistic optimization for the full fractionation spectrum. Extension of the uncertainty model and the underlying pencil-beam algorithm could enable use in more complex indications. Further, a merge with ongoing research would generalize to biological effective dose and thus also heavier particles. With the extensions of APM to model uncertainty in clinical treatment plan metrics, probabilistic analogs to accepted nominal dose concepts in inverse treatment planning are now available. If future works can overcome computational hurdles, the novel presented models can serve as powerful tools adding the desired confidence as additional parameter to treatment plan optimization.

## Mathematical Appendix

### A.1 Probability distributions

#### A.1.1 Marginal distributions

**Marginal cumulative distribution function:** If the random vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  has the multivariate cumulative distribution function  $F_{\mathbf{X}} : \mathbb{R}^n \Rightarrow [0, 1]$ , the marginal cumulative distribution function over  $\mathbf{X}_{i;j;k;\dots}$  is given by

$$F_{\mathbf{X}_{i;j;k;\dots}}(\mathbf{x}_{i;j;k;\dots}) = F_{\mathbf{X}}(\infty, \dots, \infty, x_i, x_j, x_k, \dots, \infty). \quad (\text{A.1})$$

**Marginal probability density function:** Analogous to Eq. (A.1), if  $\mathbf{X}$  has the probability density  $f_{\mathbf{X}}$ , one can define marginal probability densities for  $\mathbf{X}_{i;j;k;\dots}$  as

$$f_{\mathbf{X}_{i;j;k;\dots}}(\mathbf{x}_{i;j;k;\dots}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(\infty, \dots, x_i, x_j, x_k, \dots, \infty) dx_1 \cdots dx_{p \neq i,j,k,\dots} \cdots dx_n, \quad (\text{A.2})$$

meaning that integration will be performed along all dimensions not indexed by  $i, j, k, \dots$

#### A.1.2 Log-normal distribution

**Univariate case:** A random variable  $X$  follows a log-normal distribution when its natural logarithm  $Y = \ln X$  follows a normal distribution  $\ln X \sim \mathcal{N}(\mu, \sigma)$  (Johnson, Kotz, and Balakrishnan 1994a). Its probability density is given by

$$f_X(x) = \frac{1}{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \quad (\text{A.3})$$

The  $t$ -th moments of the Log-normal distribution can be directly calculated from the moment-generating function of the normal distribution over  $Y$  through the logarithmic relationship be-

tween  $Y$  and  $X$ :

$$M_Y(t) = \mathbb{E} \left[ e^{tY} \right] = \mathbb{E} \left[ X^t \right] = \exp \left( \mu t + \frac{1}{2} \sigma^2 t \right). \quad (\text{A.4})$$

**Multivariate case:** In the multivariate case, if  $Y$  is now an  $n$ -dimensional random vector following a multivariate normal distribution  $Y \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , then  $X = \exp Y$  follows a multivariate log-normal distribution (Kotz, Johnson, and Balakrishnan 2000). Its density is given by

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\boldsymbol{\Sigma}|} \prod_i^n x_i} \exp \left[ -\frac{1}{2} (\ln \mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\ln \mathbf{x} - \boldsymbol{\mu}) \right] \quad (\text{A.5})$$

with the expectation value

$$\mathbb{E} [X_i] = \exp \left( \mu_i + \frac{1}{2} \Sigma_{ii} \right) \quad (\text{A.6})$$

and covariance

$$\text{Cov} [X_i, X_j] = \exp \left[ \mu_i + \mu_j + \frac{1}{2} (\Sigma_{ii} + \Sigma_{jj}) \right] \cdot \left( e^{\Sigma_{ij}} - 1 \right). \quad (\text{A.7})$$

Note that the inverse transformations of Eqs. (A.6) and (A.7) can be stated as

$$\mu_i = \ln \left( \mathbb{E} [X_i] - \frac{\sigma_i^2}{2} \right), \quad (\text{A.8a})$$

$$\Sigma_{ij} = \ln \left( \frac{\text{Cov}[X_i, X_j]}{\mathbb{E}[X_i] \mathbb{E}[X_j]} + 1 \right). \quad (\text{A.8b})$$

### A.1.3 Beta distribution

The beta distribution  $\mathcal{B}(\alpha, \beta)$  with shape parameters  $\alpha$  and  $\beta$  is defined through its density

$$f(x) = \mathcal{B}(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (\text{A.9})$$

where  $B(\alpha, \beta)$  is the Beta-function. The beta distribution is supported on the interval  $x \in [0, 1]$  (Johnson, Kotz, and Balakrishnan 1994b).

Expectation value and variance of a Beta-distributed random variable  $X$  are given by

$$\mathbb{E} [X] = \frac{\alpha}{\alpha + \beta}, \quad (\text{A.10a})$$

$$\text{Var} [X] = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}. \quad (\text{A.10b})$$

If (estimates of) expectation value and variance are given, e. g. through sample statistics, shape

$\alpha, \beta$  can be computed using the transformations

$$\alpha = \mathbb{E}[X] \frac{\mathbb{E}[X](1 - \mathbb{E}[X])}{\text{Var}[X]} - 1, \quad (\text{A.11a})$$

$$\beta = (1 - \mathbb{E}[X]) \left( \frac{\mathbb{E}[X](1 - \mathbb{E}[X])}{\text{Var}[X]} - 1 \right), \quad (\text{A.11b})$$

valid for  $\text{Var}[X] \leq \mathbb{E}[X](1 - \mathbb{E}[X])$ .

### A.1.4 Gaussian algebra & identities

**Products of multivariate normal distributions:** The product of two multivariate normal distributions can be again expressed with a multivariate normal distribution multiplied by a Gaussian normalization:

$$\begin{aligned} \mathcal{N}(x; \mu_1, \Sigma_1) \mathcal{N}(x; \mu_2, \Sigma_2) &= \mathcal{N}(\mu_1; \mu_2, \Sigma_1 + \Sigma_2) \mathcal{N}(x; \mu, \Sigma) \\ &\text{with } \Sigma^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1} \text{ and } \mu = \Sigma \left( \Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 \right). \end{aligned} \quad (\text{A.12})$$

## A.2 Used Integrals

### Fractional moment integral

Equation (3.462) from Gradshteyn and Ryzhik (2000) was used in Section VI.2.3.3 to derive the  $k$ -th fractional moment of a Normal distribution. It is given by

$$\int_{-\infty}^{\infty} (ix)^k \exp(-\beta^2 x^2 - iqx) dx = 2^{-\frac{k}{2}} \sqrt{\pi} \beta^{-k-1} \exp\left(-\frac{q^2}{8\beta^2}\right) \mathcal{D}_k\left(\frac{q}{\beta\sqrt{2}}\right) \quad (\text{A.13})$$

valid for  $\Re[\beta] > 0$ ,  $\Re[k] > -1$  and  $\arg ix = \frac{\pi}{2} \text{sign } x$  using the parabolic cylinder function  $\mathcal{D}_k$ .

Equation (3.463) from Gradshteyn and Ryzhik (2000) can be used to derive the  $n$ -th integer moments

$$\int_{-\infty}^{\infty} x^n \exp[-(x - \beta)^2] dx = (2i)^{-n} \sqrt{\pi} \text{He}_n(i\beta) \quad (\text{A.14})$$

using the  $n$ -th Hermite polynomials  $\text{He}_n(x)$ .

### Indefinite integral for the second moment

Owen (1980) states the indefinite integral

$$\int x^2 \phi(a + bx) dx = \frac{a^2 + 1}{b^3} \Phi(a + bx) - \left(\frac{bx - a}{b^3}\right) \phi(a + bx) \quad (\text{A.15})$$

with the probability density function  $\phi(x)$  and the cumulative distribution function  $\Phi(x)$  of the standard normal distribution.

Equation (A.15) can be reformulated to give an anti-derivative for the integral required to obtain the second raw moment of a normal distribution:

$$\int \frac{x^2}{\sigma} \phi\left(\frac{\mu-x}{\sigma}\right) dx = \frac{1}{\sigma} \frac{\frac{\mu^2}{\sigma^2} + 1}{-\frac{1}{\sigma^3}} \Phi\left(\frac{\mu-x}{\sigma}\right) - \frac{1}{\sigma} \frac{-\frac{x}{\sigma} - \frac{\mu}{\sigma}}{-\frac{1}{\sigma^3}} \phi\left(\frac{\mu-x}{\sigma}\right) \quad (\text{A.16a})$$

$$\Rightarrow \int x^2 \mathcal{N}(x, \mu, \sigma^2) dx = -(\mu^2 + \sigma^2) \Phi\left(\frac{\mu-x}{\sigma}\right) - \sigma^2 (x + \mu) \mathcal{N}(x, \mu, \sigma^2) . \quad (\text{A.16b})$$

## Code

```

__global__ void voxelKernel_uniformBlockCorrelations( double* p_covVal, float* p_omega,
BaseDataLibrary baseData, UniformBlockCovariance_3D err_R, UniformBlockCovariance_3D err_S,
unsigned int numSpots, const float* p_Weights, unsigned int numCurrentSpots_i, unsigned int
numCurrentSpots_l, const VoxelSpotInfo* p_voxelSpotInfo_i, const VoxelSpotInfo*
p_voxelSpotInfo_l, uint16_t nrOfFractions)
{
    unsigned int currentSpot_j = blockIdx.x * blockDim.x + threadIdx.x;
    unsigned int currentSpot_m = blockIdx.y * blockDim.y + threadIdx.y;

    cudaReal tensor_element = 0.0;

    if (currentSpot_j < numCurrentSpots_i && currentSpot_m < numCurrentSpots_l)
    {
        const VoxelSpotInfo* voxelSpotInfo_j = &p_voxelSpotInfo_i[currentSpot_j];
        const VoxelSpotInfo* voxelSpotInfo_m = &p_voxelSpotInfo_l[currentSpot_m];

        //Simple Naming, the compiler should not use extra registers for that
        const int si_j = voxelSpotInfo_j->spotIndex;
        const auto dX_j = voxelSpotInfo_j->distanceX;
        const auto dY_j = voxelSpotInfo_j->distanceY;
        const auto dE_j = voxelSpotInfo_j->edij;
        const auto rg_j = voxelSpotInfo_j->range;
        const auto rD_j = voxelSpotInfo_j->radDepth;
        const auto bw_j = p_Weights[si_j];

        const unsigned int rgIx_j = unsigned int(rg_j);
        const BaseDataEntry* bdEntry_j = &(baseData.entries[rgIx_j]);
        const auto widthSq_j = addSquared(bdEntry_j->initialSigma, bdEntry_j->lateralSigmas[
            unsigned int(rD_j)]);

        const int si_m = voxelSpotInfo_m->spotIndex;
        const auto dX_m = voxelSpotInfo_m->distanceX;
        const auto dY_m = voxelSpotInfo_m->distanceY;
        const auto dE_m = voxelSpotInfo_m->edij;
        const auto rg_m = voxelSpotInfo_m->range;
        const auto rD_m = voxelSpotInfo_m->radDepth;
        const auto bw_m = p_Weights[si_m];

        const unsigned int rgIx_m = unsigned int(rg_m);
        const BaseDataEntry* bdEntry_m = &(baseData.entries[rgIx_m]);
        const cudaReal widthSq_m = addSquared(bdEntry_m->initialSigma, bdEntry_m->lateralSigmas[
            unsigned int(rD_m)]);

        CovarianceInfo covR_tmp, covS_tmp;
        //Calculate the Elements
        //X
        GetCovariance(covR_tmp, si_j, abs(dX_j), si_m, abs(dX_m), err_R.errX);
        GetCovariance(covS_tmp, si_j, abs(dX_j), si_m, abs(dX_m), err_S.errX);
        //rand
        cudaReal tensor_element_corr = gauss2D<cudaReal>(dX_j, dX_m,
            widthSq_j + covR_tmp.jj + covS_tmp.jj,

```

```

        covR_tmp.jm + covS_tmp.jm,
        widthSq_m + covR_tmp.mm + covS_tmp.mm);
//sys
cudaReal tensor_element_uncorr = 1.0;
if (nrOfFractions > 1)
{
    tensor_element_uncorr *= gauss2D<cudaReal>(dX_j, dX_m,
        widthSq_j + covR_tmp.jj + covS_tmp.jj,
        covS_tmp.jm,
        widthSq_m + covR_tmp.mm + covS_tmp.mm);
}

//Calculate the Elements
//Y
GetCovariance(covR_tmp, si_j, abs(dY_j), si_m, abs(dY_m), err_R.errY);
GetCovariance(covS_tmp, si_j, abs(dY_j), si_m, abs(dY_m), err_S.errY);
//rand
tensor_element_corr *= gauss2D<cudaReal>(dY_j, dY_m,
    widthSq_j + covR_tmp.jj + covS_tmp.jj,
    covR_tmp.jm + covS_tmp.jm,
    widthSq_m + covR_tmp.mm + covS_tmp.mm);
//sys
if (nrOfFractions > 1)
{
    tensor_element_uncorr *= gauss2D<cudaReal>(dY_j, dY_m,
        widthSq_j + covR_tmp.jj + covS_tmp.jj,
        covS_tmp.jm,
        widthSq_m + covR_tmp.mm + covS_tmp.mm);
}

//Z Factor
GetCovariance(covR_tmp, si_j, rD_j, si_m, rD_m, err_R.errZ);
GetCovariance(covS_tmp, si_j, rD_j, si_m, rD_m, err_S.errZ);

//Pre-store covariance
const auto tmp_covZ_jj = covR_tmp.jj + covS_tmp.jj;
const auto tmp_covZ_mm = covR_tmp.mm + covS_tmp.mm;
const auto tmp_covZ_jm_uncorr = covS_tmp.jm;
const auto tmp_covZ_jm_corr = covR_tmp.jm + covS_tmp.jm;

const APM_Gaussian* gauss_j = bdEntry_j->GaussianDD;
const APM_Gaussian* gauss_m = bdEntry_m->GaussianDD;

cudaReal zElement_uncorr = 0.0;
cudaReal zElement_corr = 0.0;

for (int k = 0; k < bdEntry_j->nDepthGaussians; k++)
{
    for (int n = 0; n < bdEntry_m->nDepthGaussians; n++)
    {
        //rand
        zElement_corr += gauss_j[k].A * gauss_m[n].A *
            gauss2D<cudaReal>(rD_j - gauss_j[k].mu, rD_m - gauss_m[n].mu,
                gauss_j[k].sigma * gauss_j[k].sigma + tmp_covZ_jj,
                tmp_covZ_jm_corr,
                gauss_m[n].sigma * gauss_m[n].sigma + tmp_covZ_mm);

        //sys
        if (nrOfFractions > 1)
        {
            zElement_uncorr += gauss_j[k].A * gauss_m[n].A *
                gauss2D<cudaReal>(rD_j - gauss_j[k].mu, rD_m - gauss_m[n].mu,
                    gauss_j[k].sigma * gauss_j[k].sigma + tmp_covZ_jj,
                    tmp_covZ_jm_uncorr,
                    gauss_m[n].sigma * gauss_m[n].sigma + tmp_covZ_mm);
        }
    }
}

tensor_element_corr *= zElement_corr;

```

```

tensor_element_uncorr *= zElement_uncorr;

tensor_element = (tensor_element_corr + cudaReal(nrOfFractions - 1)*
    tensor_element_uncorr) / cudaReal(nrOfFractions);

//Subtract the square of the expectation value
tensor_element -= dE_j*dE_m;

//Store to omega matrix
p_omega[si_j*numSpots + si_m] += tensor_element;

//Multiply with beamlet weights for variance values
tensor_element *= bw_j*bw_m;

    customAtomicAdd<double>(p_covVal, tensor_element);
}
}

```

**Listing B.1:** CUDA kernel for evaluating (co)variance in a voxel (combination). Note that certain required functions as well as required data structures are not explicitly defined.



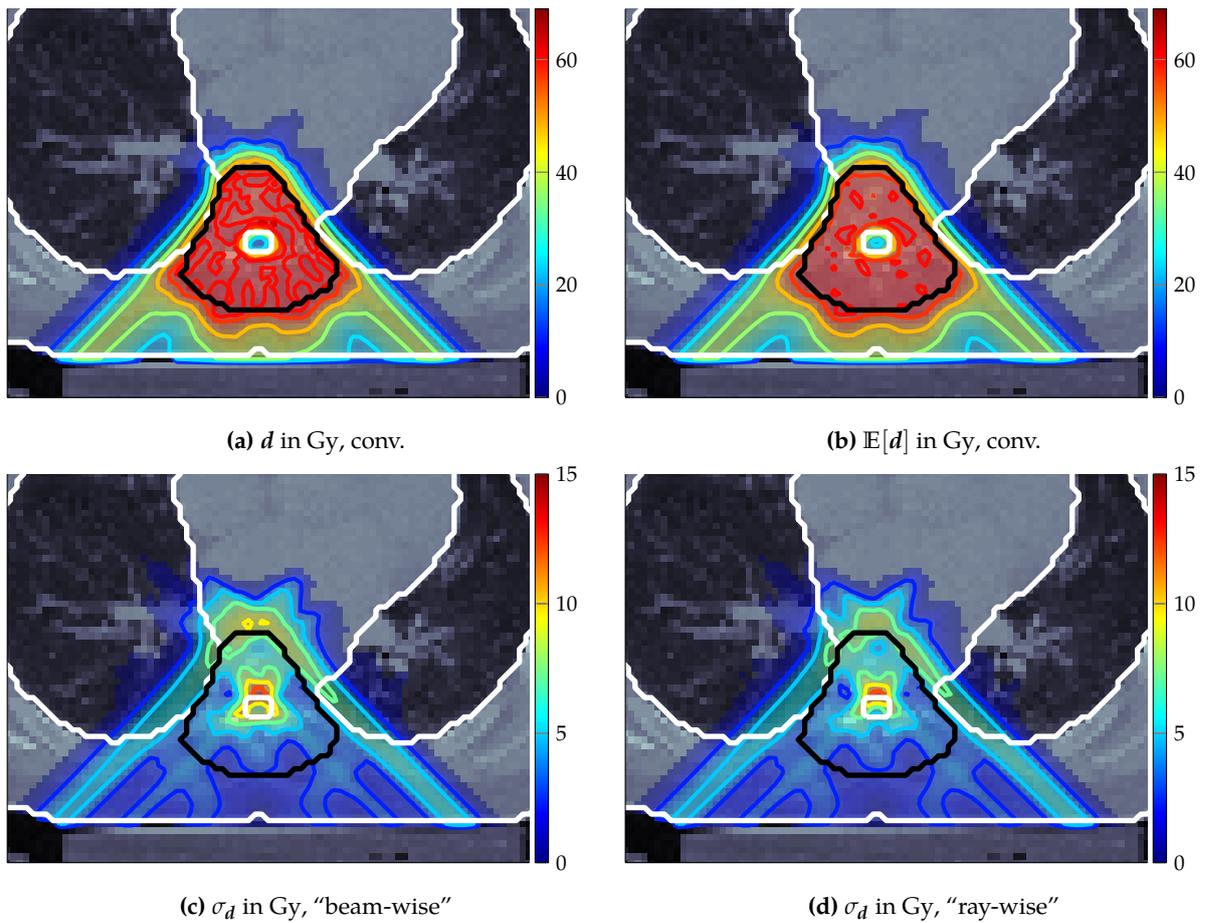
## Additional Data & Results

### C.1 Information on Patient Data

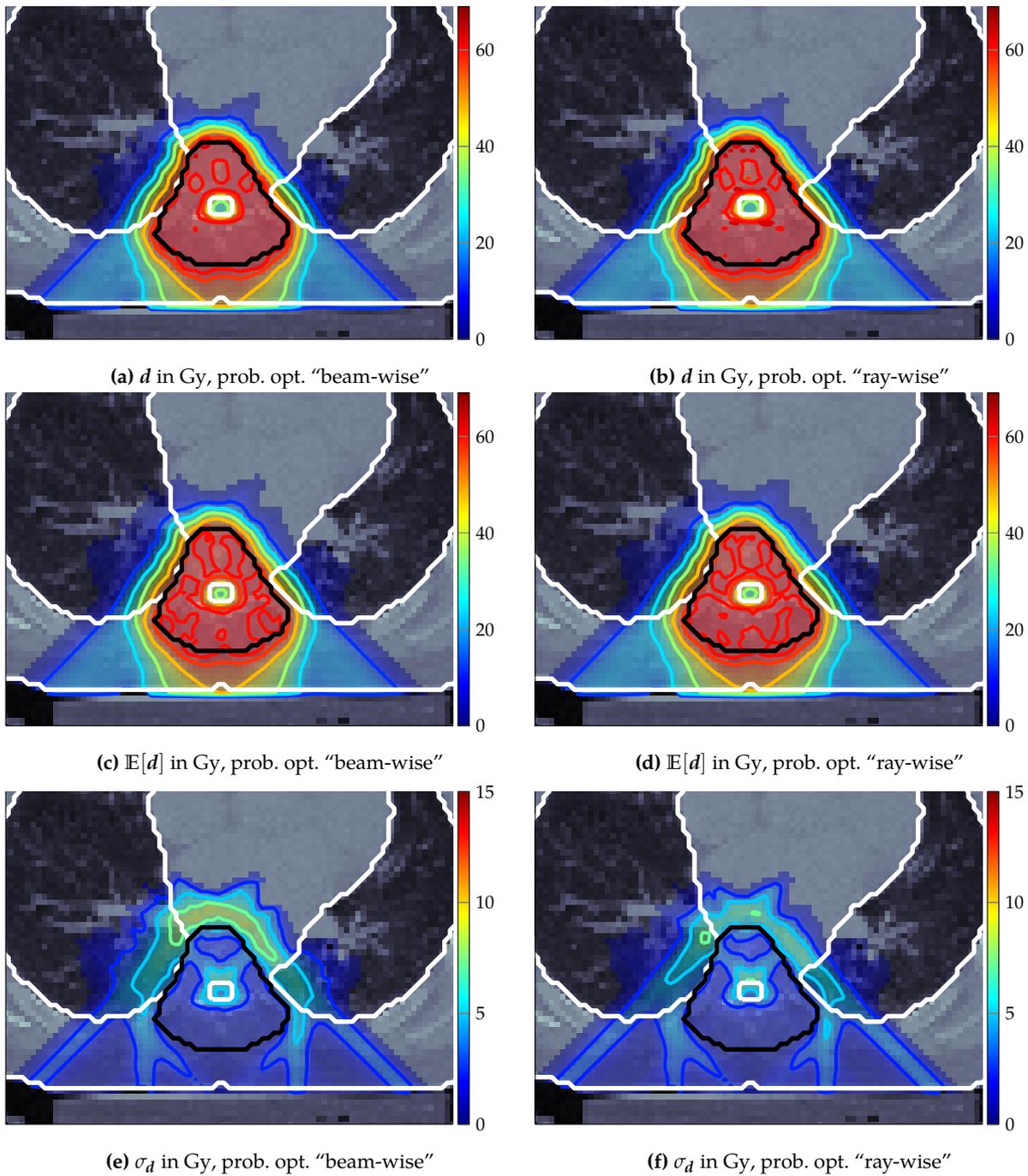
**Table C.1:** Information on the three patient datasets used for evaluation.

patient	intracranial	paraspinal	prostate
beam angles	60°, 120°	135°, 180°, 225°	90°, 270°
$d_v^*$ : CTV (Boost)	60 Gy	60 Gy	70 Gy (76 Gy)
$d_v^*$ : OARs	0 Gy	0 Gy	0 Gy
$p_v$	CTV: 1500 brainstem: 130	CTV: 3000 spinal cord: 500	PTV: 3000 boost: 3500 rectum: 200 bladder: 200
scanning grid	3 mm	4 mm	5 mm
grid margin	5 mm	5 mm	10 mm
#pencil-beams	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	$(1 \text{ mm})^{\text{sys}} + (2 \text{ mm})^{\text{rand}}$	$(1 \text{ mm})^{\text{sys}} + (2 \text{ mm})^{\text{rand}}$	$(1 \text{ mm})^{\text{sys}} + (3 \text{ mm})^{\text{rand}}$
range error	$(3.5 \%)^{\text{sys}} + (1 \text{ mm})^{\text{rand}}$	$(3.5 \%)^{\text{sys}} + (1 \text{ mm})^{\text{rand}}$	$(3.5 \%)^{\text{sys}} + (1 \text{ mm})^{\text{rand}}$

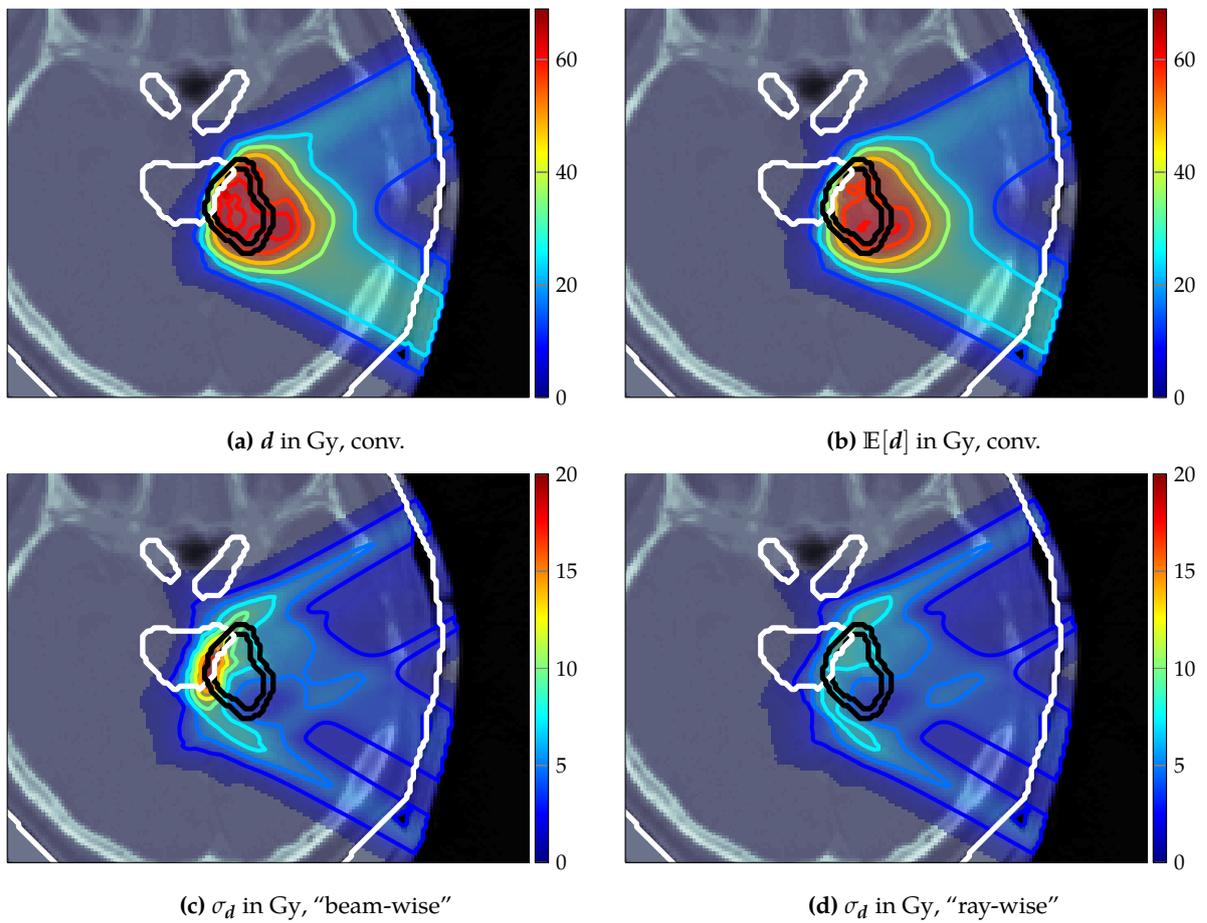
## C.2 Supplementary to Chapter IV



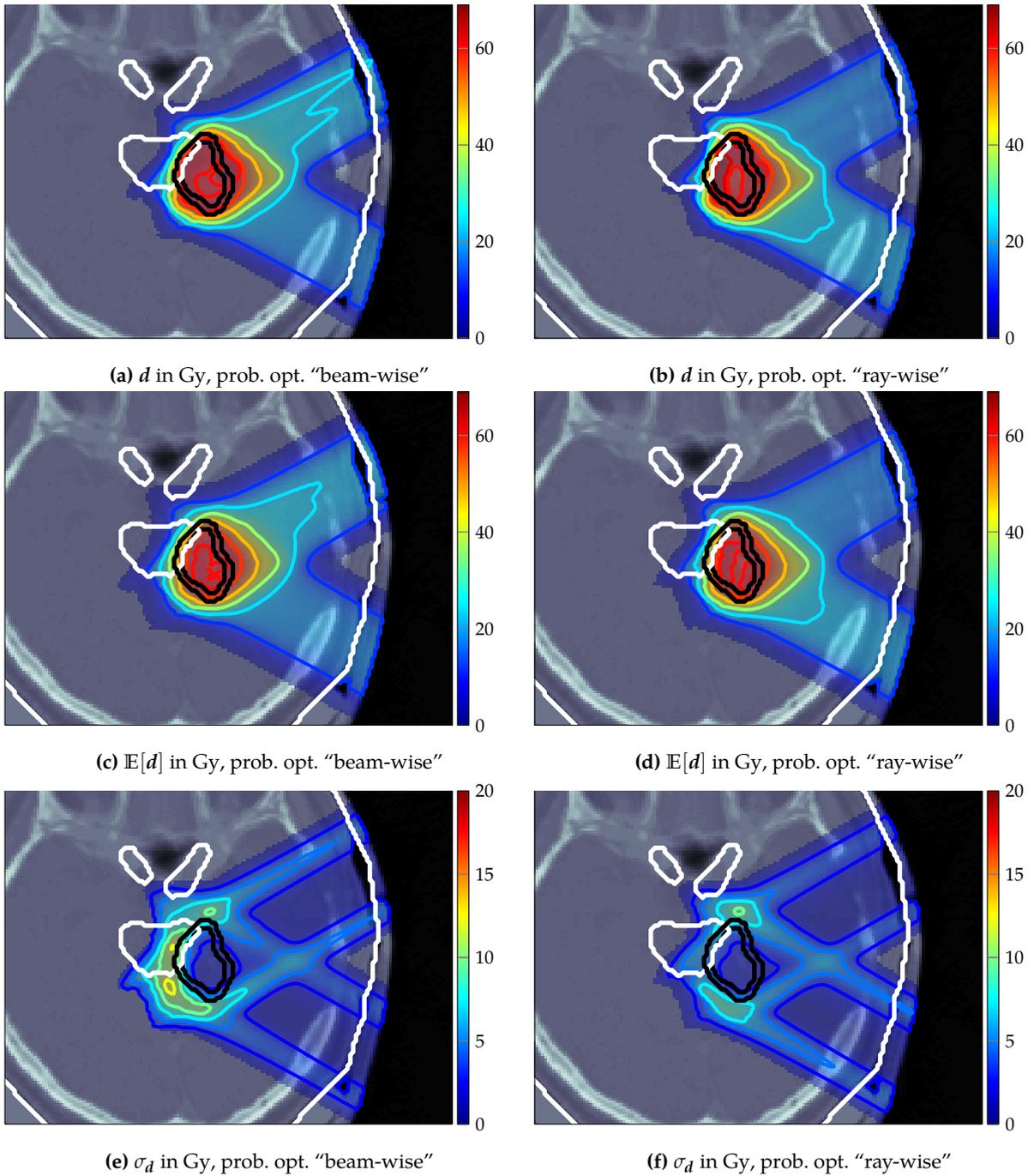
**Figure C.1:** Exemplary axial slice of (a) a conventionally optimized paraspinal treatment plan using a dose influence matrix computed with APM's dose approximation. Further, the expected dose could be computed (b) which is independent of the correlation model. Finally, the standard deviation of dose was computed, once assuming "beam-wise" correlations (c) and once assuming "ray-wise" correlations (d) for the range errors for a single fraction.



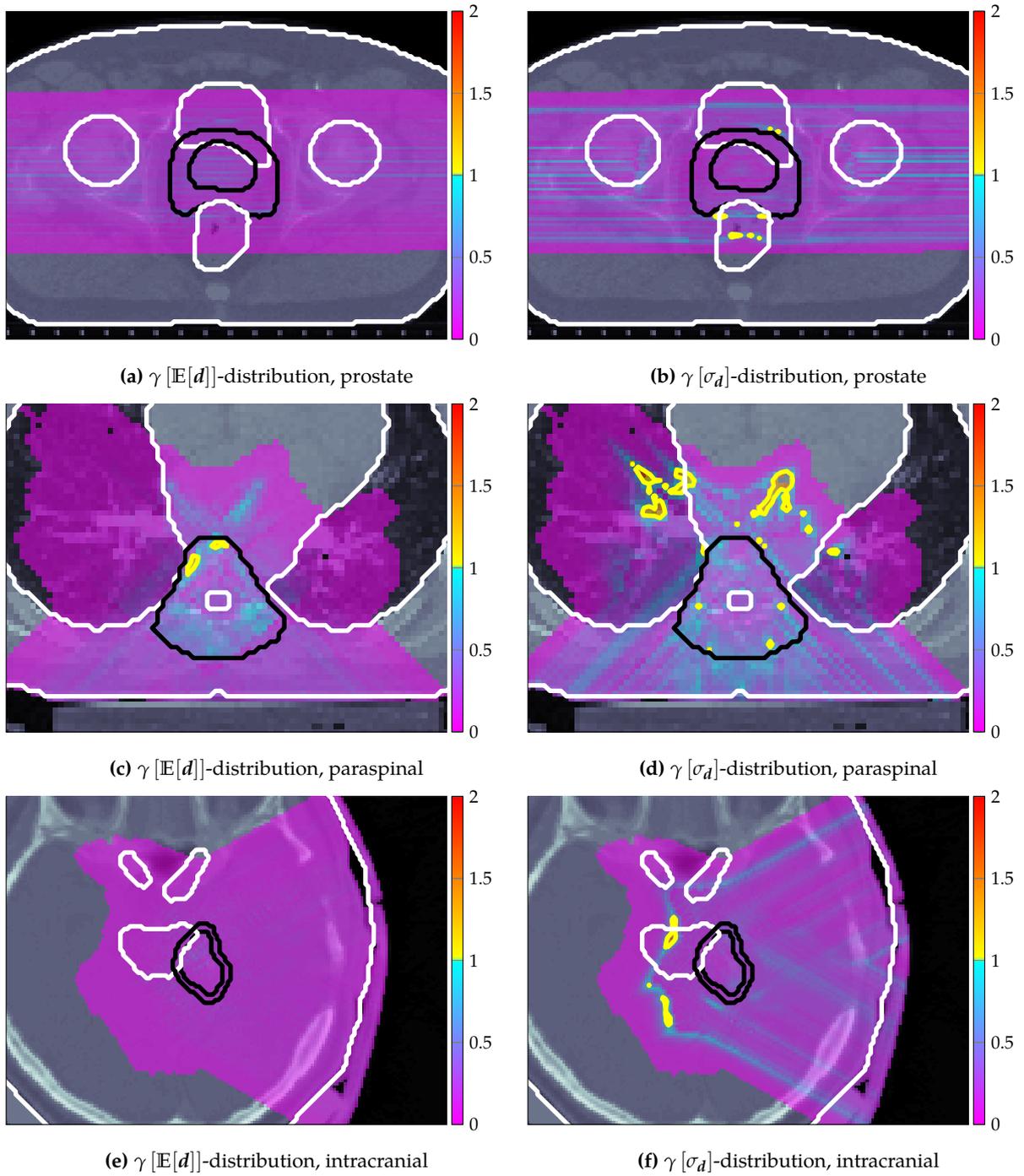
**Figure C.2:** Exemplary axial slices of distributions of nominal (a-b), expected (c-d) and standard deviation (e-f) of dose resulting from probabilistic optimization on the paraspinal case. For the plan shown in the left column (a,c,e), the "beam-wise" correlation model was assumed for probabilistic optimization. The right column (b,d,f) illustrates a probabilistically optimized plan under assumption of the "ray-wise" correlation model. Both treatment plans were optimized under 1 fraction.



**Figure C.3:** Exemplary axial slice of (a) a conventionally optimized intracranial treatment plan using a dose influence matrix computed with APM's dose approximation. Further, the expected dose could be computed (b) which is independent of the correlation model. Finally, the standard deviation of dose was computed, once assuming "beam-wise" correlations (c) and once assuming "ray-wise" correlations (d) for the range errors for a single fraction.

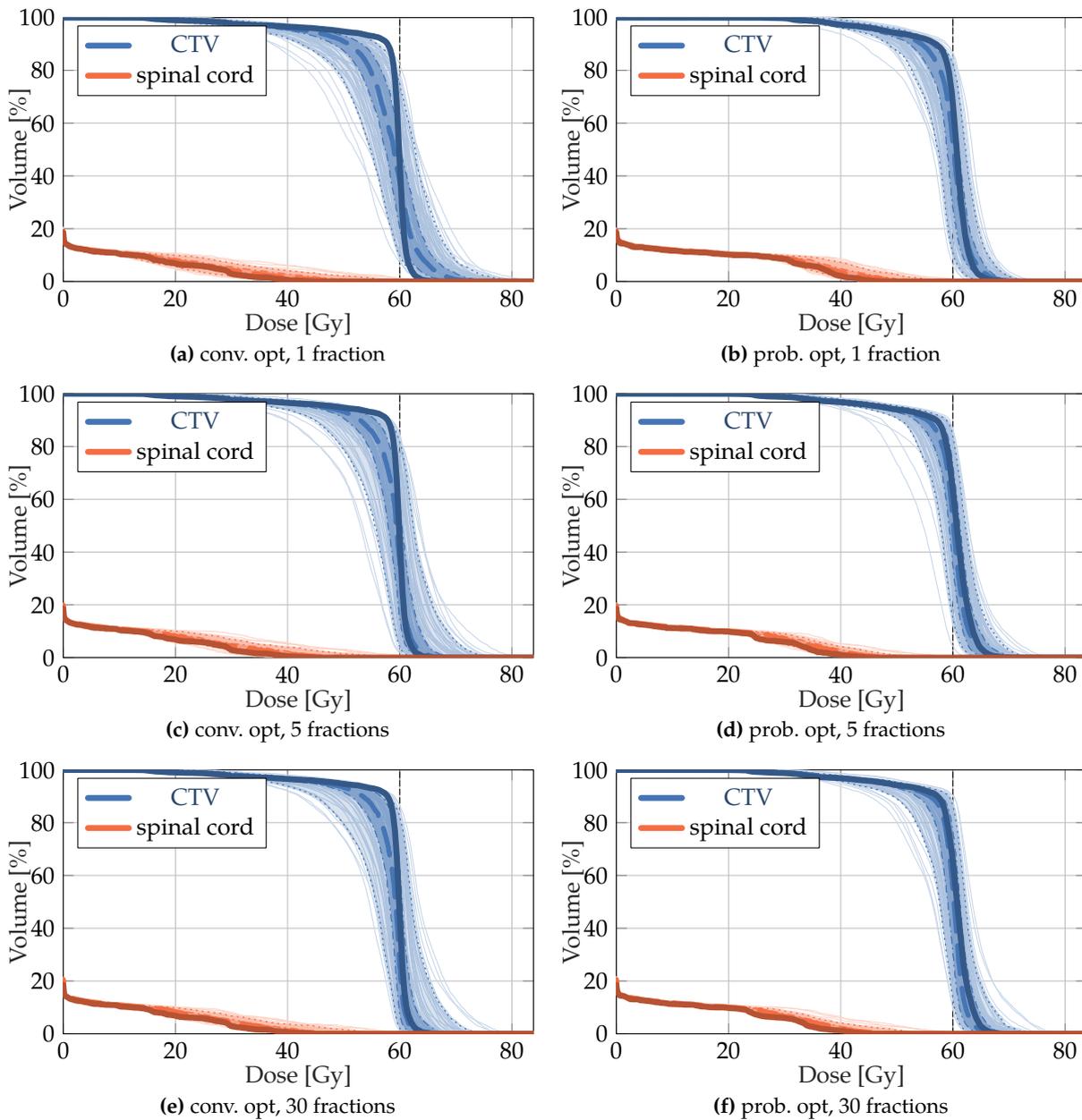


**Figure C.4:** Exemplary axial slices of distributions of nominal (a-b), expected (c-d) and standard deviation (e-f) of dose resulting from probabilistic optimization on the intracranial case. For the plan shown in the left column (a,c,e), the "beam-wise" correlation model was assumed for probabilistic optimization. The right column (b,d,f) illustrates a probabilistically optimized plan under assumption of the "ray-wise" correlation model. Both treatment plans were optimized under 1 fraction.

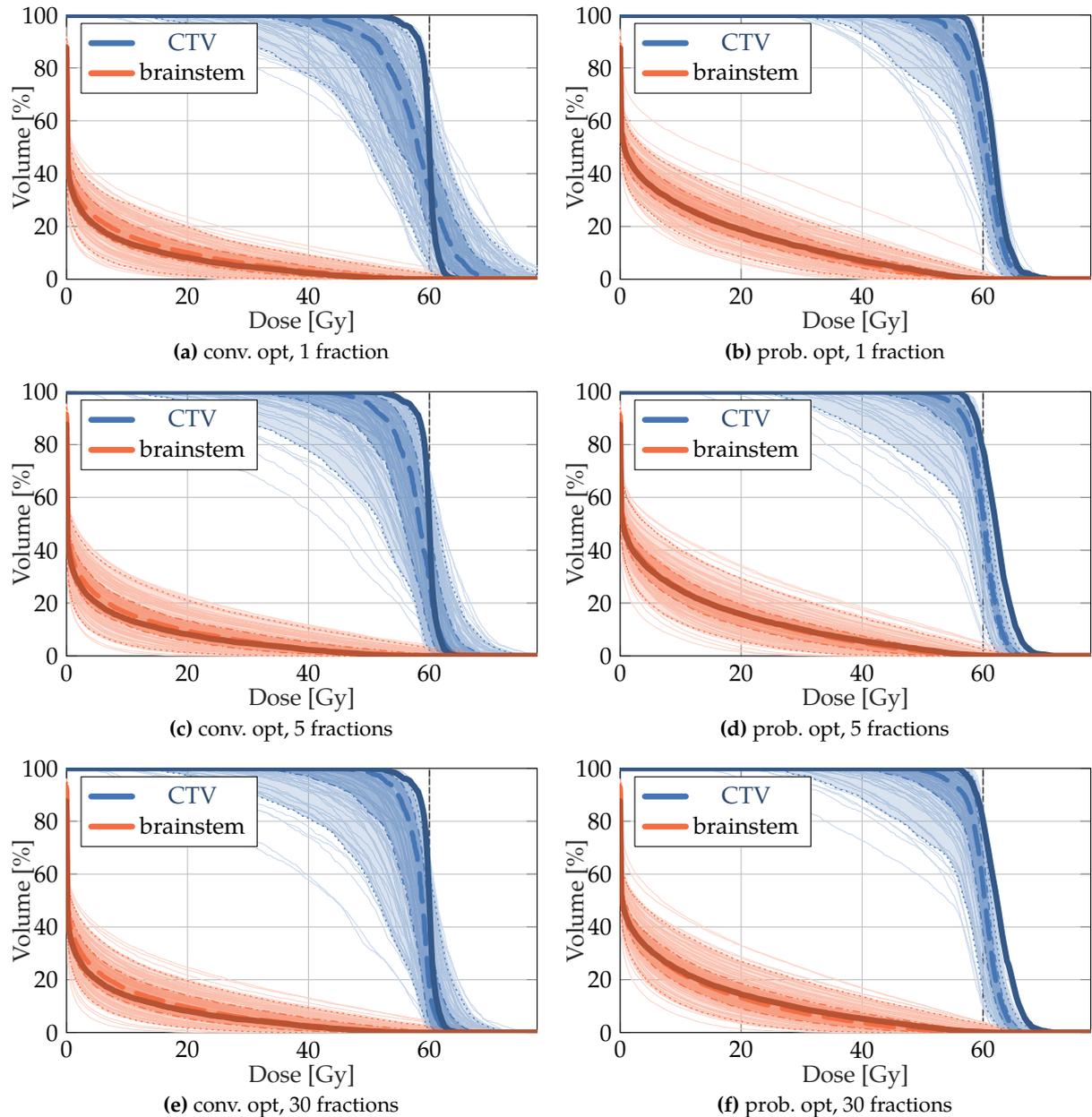


**Figure C.5:**  $\gamma_{2\%/2\text{mm}}$ -distributions validating APM's computations of expectation value (left column) and standard deviation (right column) against the respective sampled moment from the random sampling benchmark for exemplary axial slices of all three patient cases. The acceptance threshold of  $\gamma = 1$  is indicated as a color jump in the used colormap and through display of isolines at  $\gamma = 1$ .

## C.3 Supplementary to Chapter V



**Figure C.6:** Sampling based DVH-statistics for fractionated conventional and probabilistic treatment plans for the paraspinal case. (a,c,e) show the statistics for the conventionally optimized treatment planned applied in 1, 5 and 30 fractions, while (b,d,f) show the corresponding probabilistically optimized treatment plans. Nominal DVHs are represented by solid lines (—/—) while dashed lines (---/---) represent median DVHs (i. e., 50%  $\alpha$ -DVHs). The thin dotted (...../.....) and dash-dotted lines (-.-.-/-.-.-) enclosing the shaded areas refer to the 5% to 95%  $\alpha$ -DVHs and 25% to 75%  $\alpha$ -DVHs quantiles, respectively. The underlying DVHs for the treatment samples are indicated by the thin grayed out solid lines (—/—) in the background. The vertical dashed black line (-.-) indicates the prescribed dose to the respective target.



**Figure C.7:** Sampling based DVH-statistics for fractionated conventional and probabilistic treatment plans for the intracranial case. (a,c,e) show the statistics for the conventionally optimized treatment planned applied in 1, 5 and 30 fractions, while (b,d,f) show the corresponding probabilistically optimized treatment plans. Nominal DVHs are represented by solid lines (— / —) while dashed lines (--- / ---) represent median DVHs (i. e., 50 %  $\alpha$ -DVHs). The thin dotted (..... / ..... ) and dash-dotted lines (-.- / -.-) enclosing the shaded areas refer to the 5 % to 95 %  $\alpha$ -DVHs and 25 % to 75 %  $\alpha$ -DVHs quantiles, respectively. The underlying DVHs for the treatment samples are indicated by the thin grayed out solid lines (— / —) in the background. The vertical dashed black line (- -) indicates the prescribed dose to the respective target.

---

## Disclosure of Publications

This is a list of publications and conference contributions from the years 2015 to 2018 during my doctorate studies, disclosing possible relation to this thesis:

### As main author:

Wahl et al. (2015)	conference contribution (poster) of thesis findings
Kommer, Tugendhat, and Wahl (2015)	textbook, not related to doctoral studies
Wahl et al. (2016)	journal article, not related to doctoral studies
Wahl, Hennig, and Bangert (2016)	conference contribution (oral presentation) of thesis findings
Wahl et al. (2017)	journal article, contains thesis findings
Wahl et al. (2018a)	journal article, contains thesis findings
Wahl et al. (2018b)	conference contribution (poster) of thesis findings
Wahl et al. (2018c)	conference contribution (poster) of thesis findings

### As co-author:

Wieser et al. (2016)	conference contribution (oral presentation), related to the thesis' project, but not included or discussed in detail
Wieser et al. (2017b)	the software presented in this publication was used in the thesis, however the contributions to it by the author are unrelated to the thesis content
Wieser et al. (2017a)	related to the findings of this thesis, related to the thesis' project, but not included or discussed in detail
Wieser et al. (2018)	conference contribution (oral presentation), related to the thesis' project, but not included or discussed in detail



---

## Bibliography

- Acar, E., D. M. Dunlavy, and T. G. Kolda (2011). "A scalable optimization approach for fitting canonical tensor decompositions". In: *Journal of Chemometrics* 25.2, pp. 67–86.
- Acar, E. et al. (2011). "Scalable tensor factorizations for incomplete data". In: *Chemometrics and Intelligent Laboratory Systems* 106.1, pp. 41–56.
- Ahn, S. and J. A. Fessler (2003). *Standard Errors of Mean, Variance, and Standard Deviation Estimators*. Tech. rep.
- Albertini, F., E. B. Hug, and A. J. Lomax (2010). "The influence of the optimization starting conditions on the robustness of intensity-modulated proton therapy plans." In: *Physics in Medicine & Biology* 55.10, pp. 2863–78.
- (2011). "Is it necessary to plan with safety margins for actively scanned proton therapy?" In: *Physics in Medicine & Biology* 56.14, pp. 4399–4413.
- Arellano-Valle, R. B. and M. G. Genton (2008). "On the exact distribution of the maximum of absolutely continuous dependent random variables". In: *Statistics & Probability Letters* 78.1, pp. 27–35.
- Bader, B. W. and T. G. Kolda (2015). *MATLAB Tensor Toolbox, Version 2.6*. Available Online at <http://www.tensor toolbox.org/>.
- Bangert, M., P. Hennig, and U. Oelfke (2014). "Analytical probabilistic proton dose calculation and range uncertainties". In: *Journal of Physics: Conference Series*. Vol. 489. 1. IOP Publishing, p. 012002.
- Bangert, M., P. Hennig, and U. Oelfke (2013). "Analytical probabilistic modeling for radiation therapy treatment planning." In: *Physics in Medicine & Biology* 58.16, pp. 5401–19.
- Bednyakov, A. A. (2014). "On the Molière theory of multiple scattering of charged particles (1947–1948) and its critique in subsequent years". In: *Physics of Particles and Nuclei* 45.5, pp. 991–999.
- Bellinzona, V. et al. (2015). "On the parametrization of lateral dose profiles in proton radiation therapy". In: *Physica Medica* 31.5, pp. 484–492.
- Bellman, R. E. (1961). *Adaptive control processes: a guided tour*. 1st ed. Princeton University Press.
- Benaroya, H. and S. Mi Han (2005). "Probability Models in Engineering and Science". In: *Mechanical Engineering* 48.4, p. 740.
- Bertin, E. and M. Clusel (2006). "Generalised extreme value statistics and sum of correlated variables". In: *Journal of Physics A: Mathematical and Theoretical* 39, p. 7607.
- Bethe, H. (1930). "Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie". In: *Annalen der Physik* 397.3, pp. 325–400.
- Bloch, F. (1933). "Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie". In: *Annalen der Physik* 408.3, pp. 285–320.
- Bohoslavsky, R. et al. (2013). "Probabilistic objective functions for margin-less IMRT planning". In: *Physics in Medicine & Biology* 58.11, pp. 3563–3580.
- Bortfeld, T. et al. (2008). "Robust Management of Motion Uncertainty in Intensity-Modulated Radiation Therapy". In: *Operations Research* 56.6, pp. 1461–1473.
- Bortfeld, T. (1997). "An analytical approximation of the Bragg curve for therapeutic proton beams". In: *Medical Physics* 24.12, pp. 2024–2033.

- Bortfeld, T., S. B. Jiang, and E. Rietzel (2004). "Effects of motion on the total dose distribution". In: *Seminars in Radiation Oncology* 14.1, pp. 41–51.
- Casiraghi, M., F. Albertini, and A. J. Lomax (2013). "Advantages and limitations of the 'worst case scenario' approach in IMPT treatment planning." In: *Physics in Medicine & Biology* 58.5, pp. 1323–1339.
- Chan, T. C. Y., T. Bortfeld, and J. N. Tsitsiklis (2006). "A robust approach to IMRT optimization." In: *Physics in Medicine & Biology* 51.10, pp. 2567–2583.
- Chan, T. C. Y. and V. V. Mišić (2013). "Adaptive and robust radiation therapy optimization for lung cancer". In: *European Journal of Operational Research* 231.3, pp. 745–756.
- Chan, T. F., S. Osher, and J. Shen (2001). "The Digital TV Filter and Nonlinear Denoising". In: *IEEE Transactions on Image Processing* 10.2, pp. 231–241.
- Chen, E. J. (2014). "Selection and order statistics from correlated normal random variables". In: *Discrete Event Dynamic Systems* 24.4, pp. 659–668.
- Chen, W. et al. (2012). "Including robustness in multi-criteria optimization for intensity-modulated proton therapy". In: *Physics in Medicine & Biology* 57, pp. 591–608. arXiv: arXiv:1112.5362v1.
- Choi, B. and J. O. Deasy (2002). "The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning". In: *Physics in Medicine & Biology* 47.20, pp. 3579–3589.
- Chu, M. et al. (2005). "Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty". In: *Physics in Medicine & Biology* 50.23, pp. 5463–77.
- Clark, C. E. (1961). "The Greatest of a Finite Set of Random Variables". In: *Operations Research* 9.2, pp. 145–162. arXiv: 1602.01379.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer London, p. 208.
- Cook, S. (2013). *CUDA programming: a developer's guide to parallel computing with GPUs*. Waltham, MA: Morgan Kaufmann.
- Cressie, N., M. Borkent, and R. P. Gupta (1986). "The moment generating function has its moments". In: *Journal of Statistical Planning and Inference* 13, pp. 337–344.
- Demirhan, H. and C. Hamurkaroglu (2011). "On a multivariate log-gamma distribution and the use of the distribution in the Bayesian analysis". In: *Journal of Statistical Planning and Inference* 141.3, pp. 1141–1152.
- Desplanques, M. (2015). "An Open Source Software for Proton Treatment Planning". PhD Thesis. Politecnico di Milano.
- Farrance, I. and R. Frenkel (2012). "Uncertainty of Measurement: A Review of the Rules for Calculating Uncertainty Components through Functional Relationships." In: *The Clinical biochemist. Reviews* 33.2, pp. 49–75.
- Fredriksson, A. (2012). "A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization". In: *Medical Physics* 39.8, p. 5169.
- Fredriksson, A. and R. Bokrantz (2016). "The scenario-based generalization of radiation therapy margins". In: *Physics in Medicine & Biology* 61.5, pp. 2067–2082. arXiv: 1510.03300.
- Fredriksson, A., A. Forsgren, and B. Hårdemark (2011). "Minimax optimization for handling range and setup uncertainties in proton therapy". In: *Medical Physics* 38.3, p. 1672.
- Galassi, M. et al. (2016). *GNU Scientific Library Reference Manual*.
- Genz, A. (2004). "Numerical computation of rectangular bivariate and trivariate normal and t probabilities". In: *Statistics and Computing* 14.3, pp. 251–260.
- Goitein, M. (1985). "Calculation of the uncertainty in the dose delivered during radiation therapy". In: *Medical Physics* 12.5, pp. 608–612.
- Gordon, J. J. and J. V. Siebers (2009). "Coverage-based treatment planning: optimizing the IMRT PTV to meet a CTV coverage criterion." In: *Medical Physics* 36.3, pp. 961–973.

- Gordon, J. J. et al. (2010). "Coverage optimized planning: probabilistic treatment planning based on dose coverage histogram criteria." In: *Medical Physics* 37.2, pp. 550–563.
- Gottschalk, B. et al. (1993). "Multiple Coulomb scattering of 160 MeV protons". In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 74.4, pp. 467–490.
- Gradshteĭn, I. S. and I. M. Ryzhik (2000). "Definite Integrals of Elementary Functions". In: *Table of Integrals, Series, and Products*. Ed. by A. Jeffrey and D. Zwillinger. 6th ed. San Diego: Academic Press. Chap. 3-4, pp. 243–614.
- Grasedyck, L., D. Kressner, and C. Tobler (2013). "A literature survey of low-rank tensor approximation techniques". In: arXiv: 1302.7121.
- Gustafsson, F. and G. Hendeby (2012). "Some Relations Between Extended and Unscented Kalman Filters". In: *IEEE Transactions on Signal Processing* 60.2, pp. 545–555.
- Heinrich, H. et al. (2014). "GPU-accelerated ray-tracing for real-time treatment planning". In: *Journal of Physics: Conference Series* 489, p. 012050.
- Hennig, P. (2013). *Animating Samples from Gaussian Distributions*. Tech. rep. 8. Tübingen: Max Planck Institute for Intelligent Systems.
- Henríquez, F. C. and S. V. Castrillón (2008a). "A Novel Method for the Evaluation of Uncertainty in Dose-Volume Histogram Computation". In: *International Journal of Radiation Oncology\*Biophysics* 70.4, pp. 1263–1271.
- (2008b). "The effect of the different uncertainty models in dose expected volume histogram computation". In: *Australasian Physical and Engineering Sciences in Medicine* 31.3.
- (2010). "Confidence intervals in dose volume histogram computation". In: *Medical Physics* 37.4, pp. 1545–1553.
- Herk, M. van (2004). "Errors and margins in radiotherapy". In: *Seminars in Radiation Oncology* 14.1, pp. 52–64.
- Herk, M. van et al. (2000). "The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy". In: *International Journal of Radiation Oncology\*Biophysics* 47.4, pp. 1121–1135.
- Herlihy, M. and N. Shavit (2012). *The art of multiprocessor programming*. Rev. 1. Waltham, MA: Morgan Kaufmann, p. 508.
- Higham, N. J. (1988). "Computing a nearest symmetric positive semidefinite matrix". In: *Linear Algebra and its Applications* 103, pp. 103–118.
- Hong, L. et al. (1996). "A pencil beam algorithm for proton dose calculations". In: *Physics in Medicine & Biology* 41.8, pp. 1305–1330.
- Jensen, J. L. (1998). "Some statistical properties of power averages for lognormal samples". In: *Water Resources Research* 34.9, pp. 2415–2418.
- Johnson, N. L., S. Kotz, and N. Balakrishnan (1994a). *Continuous univariate distributions, Volume 1*. 2nd ed. Wiley.
- (1994b). *Continuous univariate distributions, Volume 2*. 2nd ed. Wiley.
- Jones, B. (2016). "Why RBE must be a variable and not a constant in proton therapy". In: *The British Journal of Radiology* 89.1063, p. 20160116.
- Knopf, A.-C. et al. (2013). "Adequate margin definition for scanned particle therapy in the incidence of intrafractional motion". In: *Physics in Medicine & Biology* 58.17, pp. 6079–6094.
- Kolda, T. G. and J. Sun (2008). "Scalable Tensor Decompositions for Multi-aspect Data Mining". In: *2008 Eighth IEEE International Conference on Data Mining*. IEEE, pp. 363–372.
- Kommer, C., T. Tugendhat, and N. Wahl (2015). *Tutorium Physik fürs Nebenfach*. 1st ed. Berlin, Heidelberg: Springer Spektrum.
- Kotz, S., N. L. Johnson, and N. Balakrishnan (2000). *Continuous multivariate distributions, Volume 1: Models and Applications*. 2nd ed. Wiley.

- Kraan, A. C. et al. (2013). "Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors". In: *International Journal of Radiation Oncology\*Biophysics* 87.5, pp. 888–896.
- Ku, H. (1966). "Notes on the use of propagation of error formulas". In: *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation* 70C.4, p. 263.
- Landwehr, J. M. (1978). "Some properties of the geometric mean and its use in water quality standards". In: *Water Resources Research* 14.3, pp. 467–473.
- Limbrunner, J. F., R. M. Vogel, and L. C. Brown (2000). "Estimation of Harmonic Mean of a Log-normal Variable". In: *Journal of Hydrologic Engineering* 5.1, pp. 59–66.
- Liu, W. et al. (2012a). "Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques." In: *Medical Physics* 39.6, pp. 3089–101.
- Liu, W. et al. (2012b). "Robust optimization of intensity modulated proton therapy". In: *Medical Physics* 39.2, p. 1079.
- Lomax, A. J. (2008a). "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions". In: *Physics in Medicine & Biology* 53.4, pp. 1043–1056.
- Lomax, A. (1999). "Intensity modulation methods for proton radiotherapy". In: *Physics in Medicine & Biology* 44.1, pp. 185–205.
- Lomax, A. J. (2008b). "Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties." In: *Physics in Medicine & Biology* 53.4, pp. 1027–1042.
- Lomax, A. J. et al. (2004). "The clinical potential of intensity modulated proton therapy." In: *Zeitschrift fur medizinische Physik* 14.3, pp. 147–52.
- Low, D. A. et al. (1998). "A technique for the quantitative evaluation of dose distributions." In: *Medical Physics* 25.5, pp. 656–61.
- Lowe, M. et al. (2016). "Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors". In: *Physics in Medicine & Biology* 61.1, pp. 413–429.
- Lowe, M. et al. (2017). "A robust optimisation approach accounting for the effect of fractionation on setup uncertainties". In: *Physics in Medicine & Biology* 62.20, pp. 8178–8196.
- Mackay, D. J. C. (2005). *Information Theory, Inference, and Learning Algorithms*. 4th ed. Cambridge University Press.
- Majumdar, S. N. and A. Pal (2014). "Extreme value statistics of correlated random variables". In: *Arxiv*, pp. 1–14. arXiv: 1406.6768.
- Marks, L. B. et al. (2010). "Use of normal tissue complication probability models in the clinic." In: *International Journal of Radiation Oncology\*Biophysics* 76.3 Suppl, S10–S19.
- MathWorks (2017). *MATLAB: C/C++, Fortran, Java, and Python API Reference*.
- Mcfadden, D. (1978). "Modeling the Choice of Residential Location". In: *transportation Research Record* 673, pp. 72–77.
- McGowan, S. E. et al. (2015). "Defining robustness protocols: a method to include and evaluate robustness in clinical plans." en. In: *Physics in Medicine & Biology* 60.7, pp. 2671–84.
- Mescher, H., S. Ulrich, and M. Bangert (2017). "Coverage-based constraints for IMRT optimization". In: *Physics in Medicine & Biology* 62.18, N460–N473.
- Mohan, R. et al. (1992). "Clinically relevant optimization of 3-D conformal treatments". In: *Medical Physics* 19.4, pp. 933–944.
- Mohan, R. et al. (2017). "Radiobiological issues in proton therapy". In: *Acta Oncologica* 56.11, pp. 1367–1373.
- Moore, J. A. et al. (2009). "Comparisons of treatment optimization directly incorporating random patient setup uncertainty with a margin-based approach." In: *Medical Physics* 36.9, pp. 3880–90.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc.

- Newhauser, W. D. and R. Zhang (2015). "The physics of proton therapy". In: *Physics in Medicine & Biology* 60.8, R155–R209.
- Niemierko, A. (1997). "Reporting and analyzing dose distributions: A concept of equivalent uniform dose". In: *Medical Physics* 24.1, pp. 103–110.
- (1999). "A generalized concept of equivalent uniform dose (EUD)". In: *Medical Physics* 26, p. 1100.
- Niyazi, M. et al. (2013). "Analysis of equivalent uniform dose (EUD) and conventional radiation treatment parameters after primary and re-irradiation of malignant glioma". In: *Radiation Oncology* 8.1, p. 287.
- Nocedal, J. and S. J. Wright (2006). *Numerical Optimization*. eng. 2nd ed. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer-Verlag New York.
- Nolden, M. et al. (2013). "The Medical Imaging Interaction Toolkit: challenges and advances 10 years of open-source development". In: *International Journal of Computer Assisted Radiology and Surgery* 8, pp. 607–620.
- Oelfke, U. and T. Bortfeld (2001). "Inverse planning for photon and proton beams." In: *Medical Dosimetry* 26.2, pp. 113–124.
- Olver, F. W. J. et al., eds. (2010). *NIST handbook of mathematical functions*. 1st ed. New York, NY: Cambridge University Press, p. 951.
- Oseledets, I. V. (2011a). "Tensor-Train Decomposition". In: *SIAM Journal on Scientific Computing* 33.5, pp. 2295–2317.
- Oseledets, I. (2011b). "DMRG Approach to Fast Linear Algebra in the TT-Format". In: *Computational Methods in Applied Mathematics* 11.3.
- Oseledets, I. V., E. E. Tyrtyshnikov, and N. L. Zamarashkin (2011). "Tensor-Train ranks for matrices and their inverses". In: *Comput. Meth. Appl. Math*, pp. 1–13.
- Owen, D. B. (1980). "A table of normal integrals". In: *Communications in Statistics - Simulation and Computation* 9.4, pp. 389–419.
- Paganetti, H., ed. (2012a). *Proton therapy physics*. 1st ed. CRC Press/Taylor & Francis.
- (2012b). "Range uncertainties in proton therapy and the role of Monte Carlo simulations". In: *Physics in Medicine & Biology* 57.11, R99–R117.
- Park, P. C. et al. (2013). "Statistical assessment of proton treatment plans under setup and range uncertainties". In: *International Journal of Radiation Oncology\*Biophysics* 86.5, pp. 1007–1013.
- Park, P. C. et al. (2012). "A Beam-Specific Planning Target Volume (PTV) Design for Proton Therapy to Account for Setup and Range Uncertainties". In: *International Journal of Radiation Oncology\*Biophysics* 82.2, e329–e336.
- Perkó, Z. et al. (2016). "Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion." In: *Physics in Medicine & Biology* 61.12, pp. 4646–64.
- Petersen, K. B. and M. S. Pedersen (2012). *The Matrix Cookbook*.
- Pflugfelder, D., J. J. Wilkens, and U. Oelfke (2008). "Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy." In: *Physics in Medicine & Biology* 53.6, pp. 1689–700.
- Pinter, C. et al. (2012). "SlicerRT: Radiation therapy research toolkit for 3D Slicer". In: *Medical Physics* 39.10, pp. 6332–6338.
- Rockafellar, R. T. and S. Uryasev (1997). "Optimization of conditional value-at-risk". In: *Journal of Risk* 2, pp. 21–41. arXiv: 1011.1669v3.
- Roman, S. M. and G.-C. Rota (1978). "The umbral calculus". In: *Advances in Mathematics* 27.2, pp. 95–188.
- Romeijn, H. E. et al. (2003). "A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning." In: *Physics in Medicine & Biology* 48.21, pp. 3521–3542.

- Schaffner, B., E. Pedroni, and A. Lomax (1999). "Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation". In: *Physics in Medicine & Biology* 44.1, pp. 27–41.
- Siddon, R. L. (1985). "Fast calculation of the exact radiological path for a three-dimensional CT array". In: *Medical Physics* 12.2, p. 252.
- Siggel, M. et al. (2012). "Boosting runtime-performance of photon pencil beam algorithms for radiotherapy treatment planning". In: *Physica Medica* 28.4, pp. 273–280.
- Slater, L. J. (1960). *Confluent Hypergeometric Functions*. 1st ed. Cambridge University Press, p. 247.
- Sobotta, B., M. Söhn, and M. Alber (2010). "Robust optimization based upon statistical theory". In: *Medical Physics* 37.8, pp. 4019–4028.
- Sobotta, B., M. Söhn, and M. Alber (2012). "Accelerated evaluation of the robustness of treatment plans against geometric uncertainties by Gaussian processes". In: *Physics in Medicine & Biology* 57.23, pp. 8023–8039.
- Soukup, M., M. Fippel, and M. Alber (2005). "A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations". In: *Physics in Medicine & Biology* 50.21, pp. 5089–5104.
- Steitz, J. et al. (2016). "Worst case optimization for interfractional motion mitigation in carbon ion therapy of pancreatic cancer". In: *Radiation Oncology* 11.1, p. 134.
- Taylor, P. A., S. F. Kry, and D. S. Followill (2017). "Pencil Beam Algorithms Are Unsuitable for Proton Dose Calculations in Lung". In: *International Journal of Radiation Oncology\*Biophysics* 99.3, pp. 750–756.
- Thörnqvist, S. et al. (2013). "Degradation of target coverage due to inter-fraction motion during intensity-modulated proton therapy of prostate and elective targets". In: *Acta Oncologica* 52.3, pp. 521–527.
- Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer Series in Statistics. New York, NY: Springer New York.
- Ulrich, S. et al. (2017). "Impact of respiratory motion on variable relative biological effectiveness in 4D-dose distributions of proton therapy". In: *Acta Oncologica* 56.11, pp. 1420–1427.
- Unkelbach, J. and U. Oelfke (2004). "Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions." In: *Physics in Medicine & Biology* 49.17, pp. 4005–4029.
- Unkelbach, J., T. C. Y. Chan, and T. Bortfeld (2007). "Accounting for range uncertainties in the optimization of intensity modulated proton therapy." In: *Physics in Medicine & Biology* 52.10, pp. 2755–73.
- Unkelbach, J. et al. (2009). "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning." In: *Medical Physics* 36.2009, pp. 149–163.
- Veiga, C. et al. (2016). "First Clinical Investigation of Cone Beam Computed Tomography and Deformable Registration for Adaptive Proton Therapy for Lung Cancer". In: *International Journal of Radiation Oncology\*Biophysics* 95.1, pp. 549–559.
- Wächter, A. and L. T. Biegler (2006). "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". In: *Mathematical Programming* 106.1, pp. 25–57.
- Wahl, N., P. Hennig, and M. Bangert (2016). "Probabilistic proton treatment planning using accelerated analytical probabilistic modelling". In: *Proceedings of the 18th International Conference on the Use of Computers in Radiation Therapy*. London.
- Wahl, N. et al. (2015). "Robust Planning for Intensity-modulated Proton Therapy using Analytical Probabilistic Modeling". In: *Proceedings to the 54th Annual Meeting for the Particle Therapy Cooperative Group (PTCOG) and the 2nd Annual Meeting of PTCOG – North America. International Journal of Particle Therapy* 2015. Vol. 2. 1, 314f.

- Wahl, N. et al. (2016). "Physically constrained voxel-based penalty adaptation for ultra-fast IMRT planning". In: *Journal of Applied Clinical Medical Physics* 17.4, pp. 172–189.
- Wahl, N. et al. (2017). "Efficiency of analytical and sampling-based uncertainty propagation in intensity-modulated proton therapy". In: *Physics in Medicine & Biology* 62.14, pp. 5790–5807.
- (2018a). "Analytical incorporation of fractionation effects in probabilistic treatment planning for intensity-modulated proton therapy". In: *Medical Physics* 45.4, pp. 1317–1328.
- (2018b). "Analytical probabilistic models for dose quality metrics and optimization objectives". In: *ESTRO 37 Abstract Book*. Barcelona: ELSEVIER, S486–S487.
- (2018c). "Smooth animations of the probabilistic analog to worst-case dose distributions". In: *ESTRO 37 Abstract Book*. Barcelona: ELSEVIER, S1030.
- Wieser, H.-P. et al. (2016). "Analytical probabilistic modeling of range and setup uncertainties in carbon ion therapy planning". In: *Proceedings of the 18th International Conference on the Use of Computers in Radiation Therapy*. London.
- Wieser, H.-P. et al. (2017a). "Analytical probabilistic modeling of RBE-weighted dose for ion therapy". In: *Physics in Medicine & Biology* 62.23, pp. 8959–8982.
- Wieser, H.-P. et al. (2017b). "Development of the open-source dose calculation and optimization toolkit matRad". In: *Medical Physics* 44.6, pp. 2556–2568.
- Wieser, H.-P. et al. (2018). "Simultaneous consideration of biological and physical uncertainties in robust ion therapy planning". In: *ESTRO 37 Abstract Book*. Barcelona: ELSEVIER, S44–S45.
- Winkelbauer, A. (2012). "Moments and Absolute Moments of the Normal Distribution". In: *arXiv* 2, pp. 1–3. eprint: 1209.4340.
- Wolf, I. et al. (2005). "The medical imaging interaction toolkit". In: *Medical Image Analysis* 9.6, pp. 594–604.
- Wu, Q. and R. Mohan (2000). "Algorithms and functionality of an intensity modulated radiotherapy optimization system." In: *Medical Physics* 27.4, pp. 701–711.
- Wu, Q. et al. (2002). "Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose". In: *International Journal of Radiation Oncology\*Biophysics* 52.1, pp. 224–235.
- Zhu, X. R. et al. (2014). "A single-field integrated boost treatment planning technique for spot scanning proton therapy." In: *Radiation Oncology* 9, p. 202.
- Ziegenhein, P. et al. (2008). "Speed optimized influence matrix processing in inverse treatment planning tools." In: *Physics in Medicine & Biology* 53, N157–N164.



---

# List of Used Software

## MITKrad

A research treatment planning plugin for use within *DIPP*/*MITK*. Developed by the author of this thesis, as well as Hans-Peter Wieser and Mark Bangert for the purpose of efficient computations that are part of this thesis' work. Described in more detail in Section [IV.2.1](#). [iv](#)

## matRad

An "open-source dose calculation and optimization toolkit" (Wieser et al. [2017b](#)), written in *MATLAB*, with contributions from the author of this thesis. Accessible under <http://www.matrad.org> [accessed 29.04.2018]. [5](#)

## APMToolbox

An open-source collection of *MATLAB* code to compute prototype problems with *APM* methodology. Developed by the author of this thesis, as well as Hans-Peter Wieser and Mark Bangert for the purpose of providing research results to encourage reproducibility. Contains novel models developed for the scope of this thesis, and also extended models for biological optimization subject to uncertainty with carbon ions (e. g. Wieser et al. [2017b](#)). Available on <https://github.com/e0404/APMToolbox> [accessed 29.04.2018]. [xv](#)

## DIPP

A DKFZ-internal branch of *MITK* (Nolden et al. [2013](#); Wolf et al. [2005](#)) bundling internal developments regarding *RT* and image processing. Accesible under <https://phabricator.mitk.org/project/profile/26/> [accessed 29.04.2018]. [32](#)

## Eigen

"Eigen is a C++ template library for linear algebra: matrices, vectors, numerical solvers, and related algorithms." (<http://eigen.tuxfamily.org/> [accessed 29.04.2018]). [34](#)

## IPOPT

A software library for large-scale nonlinear optimization, originally developed by Wächter and Biegler ([2006](#)) and maintained by <https://projects.coin-or.org/> [accessed 29.04.2018]. [10](#)

## ITK

An open-source collection of algorithms for image processing, especially registration and segmentation. Accessible under <https://itk.org/> [accessed 29.04.2018]. [34](#)

## MATLAB

A numerical computing environment based on the eponymous proprietary scripting language, developed by MathWorks. <https://www.mathworks.com> [accessed 29.04.2018]. [xv](#)

### **MATLAB Tensor Toolbox**

A toolbox for **MATLAB** facilitating computations with various tensor formats and decomposition methods (Bader and Kolda 2015). <http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.6.html> [accessed 29.04.2018]. 124

### **MITK**

An open-source software development toolkit for medical interactive image processing, based on **ITK** and **VTK**. Accesible under <https://mitk.org/> [accessed 29.04.2018]. iv

### **MPI**

The Message Passing Interface is a standard for communication protocols in distributed computing systems. Information on <http://www.mcs.anl.gov/research/projects/mpi/> [accessed 29.04.2018]. 38

### **OpenMP**

An interface for writing parallel programs in C++, C and Fortran for shared-memory multi-processor systems. Information on <http://www.openmp.org/> [accessed 29.04.2018]. 38

### **Qt**

A cross-platform toolkit for development of **GUIs**, available as proprietary or open-source software. <https://www.qt.io/> [accessed 29.04.2018]. 34

### **VTK**

An open-source toolkit for 3D visualization, image processing, and computer graphics. Accesible under <https://www.vtk.org/> [accessed 29.04.2018]. 158

---

## List of Figures

II.1	Bragg-curve . . . . .	5
II.2	Schematic illustration of the pencil-beam scanning grid . . . . .	6
II.3	Example of error scenarios for a prostate case irradiated with a single proton beam. . . . .	14
III.1	Visualization of <i>APM</i> pencil-beam uncertainties . . . . .	26
III.2	Exemplary correlation matrices of proposed correlation models. . . . .	29
IV.1	<i>MITKrad</i> within <i>MITK</i> / <i>DIPP</i> . . . . .	34
IV.2	Class structure of <i>MITKrad</i> . . . . .	36
IV.3	Probabilistic dose distributions for a prostate case computed with <i>APM</i> . . . . .	46
IV.4	Exemplary distributions after probabilistic optimization on a prostate case . . . . .	48
IV.5	<i>SDVH</i> comparison of conventional and probabilistic treatment plans using different correlation models. . . . .	49
IV.6	$\gamma_{3\%/3\text{mm}}$ -distributions . . . . .	50
IV.7	Accuracy of uncertainty propagation methods vs. their runtime . . . . .	54
IV.8	Approximated standard deviation distributions . . . . .	56
IV.9	Comparison of probabilistic treatment plans optimized with estimated $\hat{\Omega}$ vs. exact $\Omega$ . . . . .	57
V.1	Exemplary slice for a prostate treatment plan separating into random and systematic dose uncertainty. . . . .	67
V.2	Exemplary slices of separated random and systematic standard deviation of dose for the paraspinal and the intracranial case. . . . .	68
V.3	<i>FSDVHs</i> displaying the fractionated dose uncertainty spectrum of treatment plans. . . . .	69
V.4	Comparison of prostate beam modulations achieved by conventional and probabilistic optimization under fractionation. . . . .	70
V.5	Sampling based <i>DVH</i> -statistics for fractionated conventional and probabilistic treatment plans. . . . .	71
V.6	Median <i>DVHs</i> from sampling statistics for conventional and probabilistic treatment plans under fractionation. . . . .	72
V.7	Statistical analysis of probabilistic treatment plan quality of conventional and probabilistic treatment plans. . . . .	74
V.8	<i>FSDVHs</i> for treatment plans analyzed under several combinations of fraction number used for probabilistic optimization and following treatment application. . . . .	75
VI.1	Technical evaluation of analytical mean dose. . . . .	95
VI.2	Technical evaluation of analytical dose-volumes. . . . .	96
VI.3	Technical evaluation of analytical <i>DVCMs</i> and $\alpha$ - <i>DVHs</i> . . . . .	97
VI.4	Technical evaluation of probabilistic <i>EUD</i> . . . . .	99
VI.5	Technical evaluation of probabilistic dose extrema. . . . .	100
VI.6	Comparison of plan metrics computed from the sample covariance matrix using <i>APM</i> and their direct sample statistics. . . . .	101

VI.7	Analytically computed expected DVHs and their standard deviation compared to DVHs from sample statistics. . . . .	102
VI.8	Looped smooth animations of a dose slice and a DVH for a conventional and probabilistic treatment plan. . . . .	103
VII.1	Comparison of the expected value of the least-squares objective with and without positivity operator. . . . .	112
VII.2	Profile comparison of plans optimized with expected least squares and piece-wise least squares objective. . . . .	117
VII.3	DVHs comparison of plans optimized with expected least squares and piece-wise least squares objective. . . . .	118
VII.4	Probabilistic plan optimized with a “fake” piece-wise least squares objective. . . .	119
VII.5	Probabilistic plan optimized with a piece-wise least squares objective assuming a single fraction. . . . .	120
VII.6	Lateral profiles and probabilistic DVHs after optimization with $\alpha$ -DVH constraints. . . .	121
VII.7	Lateral dose profile optimized with a minimum $\alpha$ -DVH-constraint with Toeplitz-correlated pencil-beams. . . . .	122
VII.8	Comparison of mean dose and EUD quantile constraints. . . . .	123
C.1	Probabilistic dose distributions for a paraspinal case computed with APM . . . . .	140
C.2	Exemplary distributions after probabilistic optimization on a paraspinal case . . . .	141
C.3	Probabilistic dose distributions for an intracranial case computed with APM . . . .	142
C.4	Exemplary distributions after probabilistic optimization on a intracranial case . . . .	143
C.5	$\gamma_{2\%/2\text{mm}}$ -distributions . . . . .	144
C.6	Sampling based DVH-statistics for fractionated conventional and probabilistic treatment plans for the paraspinal case. . . . .	145
C.7	Sampling based DVH-statistics for fractionated conventional and probabilistic treatment plans for the intracranial case. . . . .	146

---

## List of Tables

IV.1 $\gamma$ -validation of APM and comparison to grid-based sampling. . . . .	52
IV.2 Detailed VOI-based analysis of APM's $\gamma$ -validation. . . . .	52
IV.4 Runtime analysis of APM's computations. . . . .	53
IV.5 RMSE and $\gamma$ -pass rates for approximated standard deviation distributions . . . . .	55
C.1 Information on the three patient datasets used for evaluation. . . . .	139



---

# Acknowledgments

While only one name marks the cover of this thesis, it couldn't have been realized without the help and support of many others.

First I would like to express my sincere gratitude to both of my referees, Prof. Jürgen Hesser and Prof. Oliver Jäkel.

Further thanks go to the members of my TAC, Prof. Florian Sterzing and Prof. Jan Unkelbach, for their valuable advice and nice atmosphere during the meetings. A special thanks goes to Prof. Wolfgang Schlegel for his advice in the early stages of this project.

As “fathers of APM”, I owe a great debt of gratitude to my group leader Mark Bangert and collaborator Philipp Hennig. Thank you, Mark, for all your efforts, the constant motivational boosts, as well as the atmosphere you create in your group that has driven me forward while at the same time allowing me to pursue my own ideas. Philipp, thank you for the awesome stays in your group in Tübingen and your valuable advice that broadened my perspective on this project immensely.

Not only the group leaders are accountable for the nice atmosphere; I would like to thank the whole group E0404 “Radiotherapy Optimization” of Mark Bangert at DKFZ, first and foremost my project buddy Hans-Peter: I will miss the research and conference stays, but may there be more Destille visits to come! Also Hubert, Silke, Lucas, Johann, Fabian, Rosario, Amit, and the others who crossed paths with this group; the time with you is greatly acknowledged. I also would like to thank the rest of the “Department of Medical Physics in Radiation Oncology” at DKFZ for all the friendly conversations and seminar sessions that pushed my research forward.

Thanks also to Philipp Hennig's group of “Probabilistic Numerics” at the MPI for Intelligent Systems in Tübingen. I still indulge in memories of the intense foosball battles with Hans, Simon, Michi, Edgar, Lukas, Filip, Alex, and Maren. Thank you all for letting me feel at home in Tübingen.

Support from beyond the office is not to underestimate. I would like to thank Christoph for providing valuable input<sup>1</sup> not only in academic questions, but also in questions of life, by supporting my caffeine addiction in presumably several thousands of Botanik and Marstall sessions since the beginning of my studies. Further thanks go to the friends who repeatedly enjoyed a beer, or two, with me in Destille, or conquered virtual lands in late evening sessions. Vic, you can be sure that my time with you is more than acknowledged; thank you for enduring me in hard times and enjoying life (which is not always easy given my constant ranting about its stupidities) in others.

I also like to thank my family for their continuous support and endurance. The last three years were not particularly easy, but we managed all of it together—I couldn't have done it all alone. And hence, last but certainly not least, I would like to send a gigantic “Thank You” to my mother. Despite everything, you were there in the beginning, and I wish you could have seen all of this happen; Thank you for all of it.

---

<sup>1</sup>“Seriously...Guys?” (Stirling Archer)