

Optimization of the Robustness of Radiotherapy against Stochastic Uncertainties

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Benjamin Sobotta
aus Mühlhausen

Tübingen
2011

Tag der mündlichen Qualifikation:	25. 05. 2011
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Wilhelm Kley
2. Berichterstatter:	Prof. Dr. Dr. Fritz Schick

Zusammenfassung

In dieser Arbeit wird ein Ansatz zur Erstellung von Bestrahlungsplänen in der Krebstherapie präsentiert, welcher in neuartiger Weise den Fehlerquellen, die während einer Strahlentherapie auftreten können, Rechnung trägt. Ausgehend von einer probabilistischen Interpretation des Therapieablaufs, wird jeweils eine Methode zur Dosisoptimierung, Evaluierung und gezielten Individualisierung von Bestrahlungsplänen vorgestellt. Maßgebliche Motivation hierfür ist die, trotz fortschreitender Qualität der Bildgebung des Patienten während der Behandlung, immer noch unzureichende Kompensation von Lagerungsfehlern, Organbewegung und physiologischer Plastizität der Patienten sowie anderen statistischen Störungen. Mangelnde Berücksichtigung dieser Einflüsse führt zu einer signifikanten Abnahme der Planqualität und damit unter Umständen zur Zunahme der Komplikationen bei verringerter Tumorkontrolle.

Im Zentrum steht ein völlig neuartiger Ansatz zur Berücksichtigung von Unsicherheiten während der Dosisplanung. Es ist üblich, das Zielvolumen durch einen Saum zu vergrößern, um zu gewährleisten, dass auch unter geometrischen Abweichungen das gewünschte Ziel die vorgesehene Dosis erhält. Der hier vorgestellte Optimierungsansatz umgeht derlei Maßnahmen, indem den Auswirkungen unsicherer Einflüsse explizit Rechnung getragen wird. Um die Qualität einer Dosisverteilung hinsichtlich ihrer Wirksamkeit in verschiedenen Organen zu erfassen, ist es nötig, Qualitätsmetriken zu definieren. Die Güte einer Dosisverteilung wird dadurch von n Skalaren beschrieben. Die Schlüsselidee dieser Arbeit ist somit, sämtliche Eingangsunsicherheiten im Planungsprozess quantitativ bis in die Qualitätsmetriken zu propagieren. Das bedeutet, dass die Bewertung von Bestrahlungsplänen nicht mehr anhand von n Skalaren vorgenommen wird, sondern mittels n statistischer Verteilungen. Diese Verteilungen spiegeln wider, mit welchen Unsicherheiten das Ergebnis der Behandlung aufgrund von Eingangsunsicherheiten während der Planung und Behandlung behaftet ist.

Der in dieser Arbeit beschriebene Optimierungsansatz rückt folgerichtig die oben beschriebenen Verteilungen an Stelle der bisher optimierten skalaren Nominalwerte der Metriken in den Mittelpunkt der Betrachtung. Dadurch werden alle möglichen Szenarien zusammen mit deren jeweiligen Wahrscheinlichkeiten mit in die Optimierung einbezogen. Im Zuge der Umsetzung dieser

neuen Idee wurde als erstes das klassische Optimierungsproblem der Bestrahlungsplanung neu formuliert. Die Neuformulierung basiert auf Mittelwert und Varianz der Qualitätsmetriken des Planes und bezieht ihre Rechtfertigung aus der Tschebyschew-Ungleichung. Der konzeptionell einfachste Ansatz zur Errechnung von Mittelwert und Varianz komplexer Computermodelle wie des Patientenmodells ist die klassische Monte Carlo Methode. Allerdings muss dafür das Patientenmodell sehr häufig neu evaluiert werden. Für eine Reduzierung der Anzahl der Neuberechnungen konnte jedoch ausgenutzt werden, dass kleinen Änderungen der unsicherheitsbehafteten Parameter nur leichte Fluktuationen der Qualitätsmetriken folgen, d.h. der Ausgang der Behandlung hängt stetig von den unsicheren geometrischen Parametern ab. Durch diesen funktionalen Zusammenhang wird es möglich, das Patientenmodell durch ein Regressionsmodell auszutauschen. Dazu wurden Gauß-Prozesse, eine Methode zur nichtparametrischen Regression, verwendet, die mit Hilfe von wenigen Evaluierungen des Patientenmodells angelernt werden. Nach der Anlernphase nimmt der Gauß-Prozess die Stelle des Patientenmodells in den folgenden Berechnungen ein. Auf diesem Wege können sämtliche zur Optimierung relevanten Größen sehr effizient und häufig berechnet werden, was dem iterativen Optimierungsalgorithmus entgegen kommt. Die dadurch erreichte Beschleunigung des Optimierungsalgorithmus ermöglicht dessen Anwendung unter realen Bedingungen. Um die klinische Tauglichkeit des Algorithmus zu demonstrieren, wurde dieser in ein Bestrahlungsplanungssystem implementiert und an Beispielpatienten vorgeführt.

Zur Evaluierung optimierter Bestrahlungspläne wurden die klinisch etablierten Dosis-Volumen-Histogramme (DVHs) unter Einbeziehung der zusätzlich durch probabilistische Patientenmodelle bereitgestellten Informationen erweitert. Durch die Verwendung von Gauß-Prozessen kann die DVH-Unsicherheit unter verschiedenen Bedingungen auf aktueller Computerhardware in Echtzeit abgeschätzt werden. Der Planer erhält auf diese Weise wertvolle Hinweise bezüglich der Auswirkungen verschiedener Fehlerquellen als auch einen „Überblick“ über die Effektivität potentieller Gegenmaßnahmen. Um die Präzision der Vorhersagen auf Gauß-Prozessen zu verbessern wurde die Bayesian Monte Carlo Methode (BMC) maßgeblich erweitert mit dem Ziel, neben Mittelwert und Varianz auch die statistische Schiefe zu berechnen.

Der letzte Teil dieser Arbeit befasst sich mit der Verbesserung von Bestrahlungsplänen, die gegebenenfalls dann notwendig wird, wenn der vorliegende Plan noch nicht den klinischen Ansprüchen genügt. Oftmals, beispielsweise durch zu streng verschriebene Schranken für die Dosis im Normalgewebe, kommt es im Zielvolumen zu punktuellen Unterdosierungen. In der Praxis ist somit von erheblicher Bedeutung, das Zusammenspiel aller Verschreibungen

zu durchleuchten, um einen bestehenden Plan in möglichst wenigen Schritten den Vorgaben anzupassen. Mit anderen Worten: Der vorgestellte Ansatz hilft dem Planer, lokale Konflikte binnen kürzester Zeit zu lösen.

Es konnte gezeigt werden, dass die statistische Behandlung unvermeidlicher Störgrößen in der Strahlentherapie mehrere Vorteile birgt. Zum einen ermöglicht der gezeigte Algorithmus, Bestrahlungspläne zu erzeugen, die individuell auf den Patienten und dessen Muster geometrischer Unsicherheiten zugeschnitten sind. Dies war bislang nicht der möglich, da die momentan verwendeten Sicherheitssäume die Unsicherheiten nur summarisch und für das Zielvolumen behandeln. Zum anderen bietet die Evaluierung von Bestrahlungsplänen unter expliziter Berücksichtigung statischer Einflüsse dem Planer wertvolle Anhaltspunkte für das zu erwartende Ergebnis einer Behandlung.

Eines der größten Hindernisse für eine weitere Steigerung der Effizienz der Strahlentherapie stellt die Behandlung geometrischer Unsicherheiten dar, die, sofern sie nicht eliminiert werden können, durch die Bestrahlung eines deutlich vergrößerten Volumens kompensiert werden. Diese ineffiziente und wenig schonende Vorgehensweise könnte durch die hier vorgestellte Behandlung des Optimierungsproblems der Strahlentherapie als statistisches Problem abgelöst werden. Durch die Verwendung von Gauß-Prozessen als Substitute des Originalmodells konnte ein Algorithmus geschaffen werden, welcher in klinisch akzeptablen Zeiten im statistischen Sinne für alle Organe robuste Dosisverteilungen liefert.

Contents

1	Introduction	2
2	Robust Treatment Plan Optimization	5
2.1	Main Objective	7
2.1.1	Explicit Incorporation of the Number of Fractions	13
2.2	Efficient Treatment Outcome Parameter Estimation	14
2.2.1	Gaussian Process Regression	16
2.2.2	Uncertainty Analysis with Gaussian Processes	21
2.2.3	Bayesian Monte Carlo	24
2.2.4	Bayesian Monte Carlo vs. classic Monte Carlo	26
2.3	Accelerating the Optimization Algorithm	28
2.3.1	Constrained Optimization in Radiotherapy	29
2.3.2	Core Algorithm	29
2.3.3	Convergence Considerations	31
2.3.4	Efficient Computation of the Derivative	32
2.4	Evaluation	35
2.5	Discussion and Outlook	40
3	Uncertainties in Dose Volume Histograms	43
3.1	Discrimination of Dose Volume Histograms based on their respective Biological Effects	45
3.2	Quantitative Analysis	46
3.3	Accounting for Asymmetry	48
3.3.1	Skew Normal Distribution	49
3.3.2	Extension of Bayesian Monte Carlo to compute the Skewness in Addition to Mean and Variance	50
3.4	Determination of Significant Influences	53
3.5	Summary and Discussion	54

4	Spatially Resolved Sensitivity Analysis	56
4.1	The Impact of the Organs at Risk on specific Regions in the Target	57
4.1.1	Pointwise Sensitivity Analysis	58
4.1.2	Perturbation Analysis	60
4.2	Evaluation and Discussion	62
5	Summary & Conclusions	64
A	Calculations	76
A.1	Solutions for Bayesian Monte Carlo	76
A.1.1	$E_x[\mu(X)]$	77
A.1.2	$E_x[\mu^2(X)]$	77
A.1.3	$E_x[\mu^3(X)]$	78
A.2	Details of the Incorporation of Linear Offset into the Algorithm for the Variance Computation	78
A.3	Details of the Incorporation of Linear Offset into the Algorithm for the Skewness Computation	80
A.3.1	$E_x[\mu^2(X)m(X)]$	80
A.3.2	$E_x[\mu(X)m^2(X)]$	80
A.3.3	$E_x[m^3(X)]$	81
B	Details	82
B.1	Relating the Average Biological Effect to the Biological Effect of the Average	82
B.2	Mean, Variance and Skewness of the Skew Normal Distribution	83
C	Tools for the analysis of dose optimization: III. Pointwise sensitivity and perturbation analysis	84
D	Robust Optimization Based Upon Statistical Theory	92
E	On expedient properties of common biological score functions for multi-modality, adaptive and 4D dose optimization	103
F	Special report: Workshop on 4D-treatment planning in actively scanned particle therapy - Recommendations, technical challenges, and future research directions	111
G	Dosimetric treatment course simulation based on a patient-individual statistical model of deformable organ motion	119

List of Abbreviations

ARD	Automatic Relevance Determination
BFGS method	Broyden Fletcher Goldfarb Shanno method
BMC	Bayesian Monte Carlo
CT	Computed Tomography
CTV	Clinical Target Volume
DVH	Dose-Volume Histogram
EUD	Equivalent Uniform Dose
GP	Gaussian Process
GTV	Gross Tumor Volume
IMRT	Intensity Modulated Radiotherapy
MC	Monte Carlo
MRI	Magnetic Resonance Imaging
NTCP	Normal Tissue Complication Probability
OAR	Organ At Risk
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PRV	Planning Organ-at-Risk Volume
PTV	Planning Target Volume
R.O.B.U.S.T.	Robust Optimization Based Upon Statistical Theory

RT	Radiotherapy
TCP	Tumor Control Probability

Chapter 1

Introduction

Radiotherapy is the application of ionizing radiation to cure malignant disease or slow down its progress. The principal goal is to achieve local tumor control to suppress further growth and spread of metastases, while, at the same time, immediate and long term side effects to normal tissue should be minimized. This trade-off between normal tissue sparing and tumor control is the fundamental challenge of radiotherapy and continues to drive development in the field.

In the planning stage, prior to treatment, images from various modalities, e.g. CT, MRI and PET, are used to create a virtual model of the patient. Based on this model, the dose to any organ can be precisely predicted. From a medical viewpoint, the idealized dose distribution tightly conforms to the delineated target volume. The main task of treatment planning is to bridge the gap between the idealized dose distribution and a distribution that is physically attainable. To accomplish this, the method of constrained optimization has proven particularly useful. The idea is to maximize the dose in the target under the condition that the exposure of the organs at risk must not exceed a given threshold. In order to quantify the merit of a given dose distribution, the optimizer employs cost functions. In recent years, biological cost functions experienced increasing popularity. Biological cost functions, in contrast to physical cost functions, incorporate the specific dose response of a given organ into the optimization.

Typically, radiation is administered in several treatment sessions. Fractionated radiotherapy exploits the inferior repair capabilities of malignant tissue. Through fractionation, the damage to normal tissue is significantly reduced. This is the key to reach target doses as high as 80 Gy. In most cases, radiotherapy is concentrated on a region around the tumor. Depending on the particular scenario, the target region may be expanded to also include adjacent tissue such as lymph nodes. Apart from clinical reasons for

such an expansion, a margin around the tumor is typically added to compensate for movement that occurs during treatment. Such movement comprises motion that may occur during a treatment session as well as displacements that happen between treatment fractions.

Unfortunately, motion compensation via margins comes at the price of a severely increased dose to surrounding normal tissue. Furthermore, the margin concept is entirely target centered and completely neglects potentially hazardous effects of motion on the organs at risk. Several approaches have been put forth to alleviate this problem [BABN06, BYA⁺03, CZHS05, ÓW06, CBT06, SWA09, MKVF06, YMS⁺05, WvdGS⁺07, TRL⁺05, LLB98]. Perhaps the simplest extension to the margin concept is coverage probability [BABN06]. The basic idea is to superimpose several CT snapshots to generate an occupancy map of any organ of interest. The resulting per voxel occupancies are used as weights in the optimization. Through these occupancy maps, the specific motion of the organs in essence replaces the more unspecific margins. Even though the problem of overlapping volumes is alleviated, it is not solved. The probability clouds of the organs may still very well have common regions. In contrast, other authors [BYA⁺03, MKVF06, SWA09, TRL⁺05] use deformable patient models to track the irradiation of every single volume element. These methods evaluate dose in the reference geometry while the dose is calculated in the deformed model. Compared to coverage probability, this method is more demanding, as it requires deformation vector fields between the CT snapshots and the optimization on several geometry instances. Additionally, this technique introduces further sources of error.

The core of this thesis is an entirely new concept for uncertainty management. The key difference to the previously outlined methods is that it pays tribute to the inherent uncertainties in a purely statistical manner. All treatment planning techniques up to this point culminate in the optimization of nominal values to reflect the treatment plan merits. The approach in the present work handles all treatment outcome metrics as statistical quantities. That is, instead of optimizing a nominal value, *our approach directly optimizes the shape of dose metric distribution*. Because it is independent of geometric considerations, any treatment parameter, whose uncertainty is quantifiable, can be incorporated into our framework. Requiring no changes to the infrastructure of the optimizer, it seamlessly integrates into the existing framework of constrained optimization in conjunction with biological costfunctions. The new method will also reveal its true potential in conjunction with heavy particle therapy because many concepts that are currently in use will fail and other sources of error, such as particle range uncertainties, besides geometric deviations have to be considered.

Once a plan has been computed, it needs to be evaluated. Dose volume histograms (DVHs) are used to express the merit of a 3D dose distribution into an easy to grasp 2D display. Today, DVHs are ubiquitous in the field radiation oncology. Classic DVHs reflect the irradiation of an organ only for a static geometry, hence provide no indication how the irradiation pattern changes under the influence of uncertainties. Consequently, to assess the influence of organ movement and setup errors, DVHs should reflect the associated uncertainties. This work augments DVHs such that the assumed or measured error propagates into the DVH, adding error bars for each source of uncertainty individually, or any combination of them. Furthermore, it is possible to discriminate DVHs based on biological measures, such as the realized equivalent uniform dose (EUD).

The very nature of the radiotherapy optimization problem, i.e. contradictory objectives, almost always demands further refinements to the initial plan. In the realm of constrained optimization, this translates into the need to predict the changes to the dose distribution if a constraint is altered. While this can be done by sensitivity analysis in terms of treatment metrics, the introduced method in this work introduces spatial resolution to trade-off analysis. It is especially suitable to eliminate so-called cold spots, i.e. regions in the target that experience underdosage. The algorithm screens the optimization setting and determines which constraints need to be loosened in order to raise the dose in the cold spot to an acceptable level. This is especially useful if a multitude of constraints can be the cause of the cold spot. This way, tedious trial and error iterations are reduced.

To sum up, this thesis introduces a novel framework for uncertainty management. In contrast to earlier publications in the field, it *explicitly* accounts for the fact that the inherent uncertainties in radiotherapy render all associated quantities random variables. During optimization and evaluation, all random variables are rigorously treated as such. The algorithm seamlessly integrates as cost function into existing optimization engines and can be paired with arbitrary dose scoring functions. To provide the physician with essential information needed to evaluate a treatment plan under the influence of uncertainty, DVHs, a cornerstone in every day clinical practice, were augmented to reflect the consequences of potential error sources. The expert is provided with valuable information to facilitate the inevitable trade-off between tumor control and normal tissue sparing. The last part of this thesis deals with a new tool for the *efficient* customization of a treatment plan to the needs of each individual patient.

Chapter 2

Robust Treatment Plan Optimization

In the presence of random deviations from the planning patient geometry, there will be a difference between applied and planned dose. Hence, the delivered dose and derived treatment quality metrics become random variables [MUI06, BABN04, LJ01, vHRL00].

Evidently, the statistical nature of the treatment also propagates into the treatment quality metrics, e.g. dose volume histogram (DVH) points or equivalent uniform dose (EUD) [Nie97]. Any one of these metrics becomes a random variable, having its own patient and treatment plan specific probability distribution. This distribution indicates the likelihood that a certain value of the metric will be realized during a treatment session. Because the outcome distribution is sampled only a few times, namely the number of treatment sessions, it becomes important to consider the width of the distribution in addition to the mean value, fig. 2.1(a) [UO04, UO05, BABN04]. The wider a distribution becomes, the more samples are necessary to obtain reliable estimates of the mean. In the treatment context, this translates into the need to minimize the distribution width of the quality metrics to arrive at the planned mean of the metric with a high probability. This is to assure that the actual prescription is met as closely as possible in every potentially realized course of treatment.

The standard approach to address this problem is to extend the clinical target volume (CTV) to obtain a planning target volume (PTV) [oRUM93]. Provided that the margin is properly chosen [vHRL02, SdBHV99], the dose to any point of the CTV remains almost constant across all fractions, leading to a narrow distribution of possible treatment outcomes, fig. 2.1(b). However, the PTV concept only addresses the target, not the organs at risk (OARs). The addition of margins for the OARs was suggested by [MvHM02], lead-

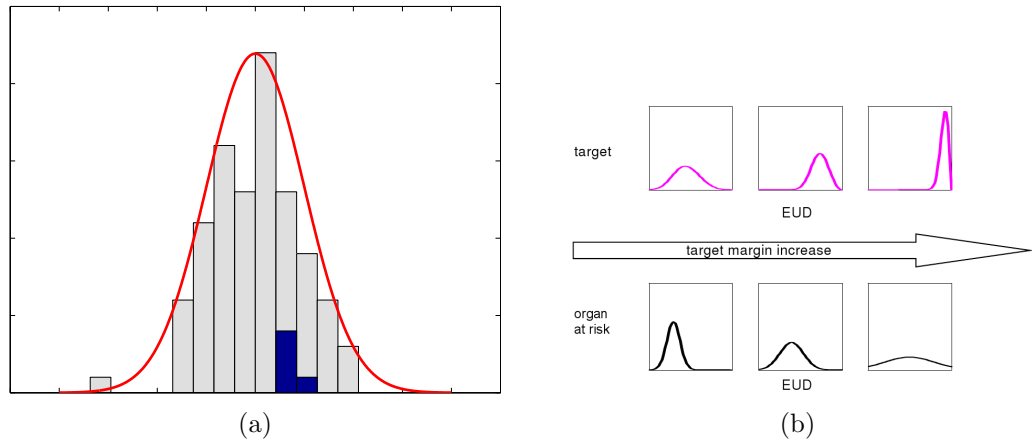


Figure 2.1: (a) The mean of only a few samples (blue) may not resemble the mean of the distribution. This is confounded if the width of the distribution increases. (b) The treatment outcome distribution (e.g. in terms of EUD) plotted against target margin width. If the target margin becomes larger, the target distribution becomes more and more located (top 3 panels). The contrary is the case regarding the organ at risk distribution. With increasing target margin, this distribution widens toward higher doses as the gradient moves into organ space.

ing to planning risk volumes (PRVs). This increases the overlap region of mutually contradictory plan objectives even more [SH06].

In order to overcome these problems, numerous approaches have been suggested [BABN06, BYA⁺03, CZHS05, ÓW06, CBT06, SWA09, MKVF06, YMS⁺05, WvdGS⁺07, TRL⁺05, LLB98] that explicitly incorporate the magnitude of possible geometric errors into the planning process. Depending on the authors' view on robustness, and thereby the quantities of interest, the proposed methods are diverse.

The method of coverage probabilities can be regarded as the most basic extension of the PTV concept in that it assigns probability values to a classic or individualized PTV/PRV to indicate the likelihood of finding the CTV/OAR there [SdBHV99]. The work of [BABN06] addresses the problem of overlapping volumes of interest by incorporating a coverage probability based on several CTs into the optimization process. Through these occupancy maps, the specific motion of the organs in essence replaces the more unspecific margins. Even though the problem of overlapping volumes is alleviated, it is not solved. The respective probability clouds of the organs may still very well have common regions, since there is no notion of correlated motion. A related concept, namely the optimization of the expectation of

tumor control probability (TCP) and normal tissue complication probability (NTCP) was investigated [WvdGS⁺07]. Notice that coverage probability is still based on a static patient model, i.e. the dose is evaluated in dose space.

In contrast, other authors use deformable patient models to track the irradiation of every single volume element (voxel). This means that while dose scoring and evaluation is done in a reference geometry of the patient, the dose is calculated on the deformed model. The authors of [BYA⁺03] brought forth a method of accounting for uncertainties by substituting the nominal dose in each voxel of a reference geometry with an effective dose. The effective dose is obtained by averaging the doses in the voxel over several geometry instances. This also allows for the use of biological cost functions. A similar approach with the addition of dose recomputation for each geometry instance has been presented [MKVF06, SWA09, TRL⁺05]. The authors of [CZHS05] and [ÓW06] propose a method that models organ movement as uncertainties in the dose matrix. In essence, their method addresses the robustness of the dose on a per voxel basis. Thus, their method is restricted to physical objective functions.

Naturally, all the above methods require precise information about the patient movement prior to treatment. Due to the limits of available information, these deformation models themselves provide for another source of errors. A method to take these errors of the model of motion into account has also been suggested [CBT06].

In consequence, even concepts that go beyond margins to alleviate the impact of geometric variation have to deal with uncertain outcome metrics.

In this chapter, a general probabilistic optimization framework, R.O.B.-U.S.T. - Robust Optimization Based Upon Statistical Theory, is proposed for any type of cost function, that pays tribute to the statistical nature of the problem by controlling the treatment outcome distribution of target and OARs alike. Any treatment parameter, for which the uncertainty is quantifiable, can be seamlessly incorporated into the R.O.B.U.S.T. framework. By actively controlling the width of the outcome distributions quantitatively during the optimization process, the finite amount of treatment fractions can also be considered.

2.1 Main Objective

As the exact patient geometry configuration during each treatment fraction is unknown, the delivered dose becomes a statistical quantity. Due to the statistical nature of the delivered dose, all associated quality metrics become random variables, in the following denoted by Y for objectives and Z for con-

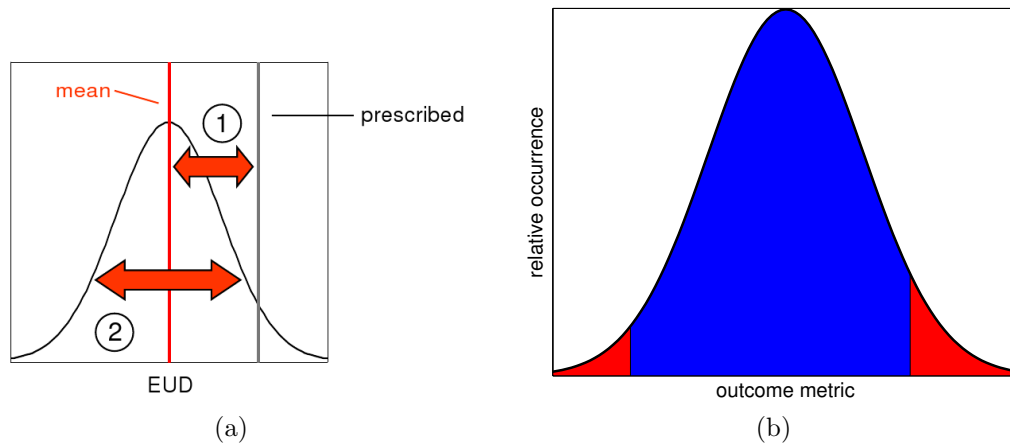


Figure 2.2: (a) The indispensable properties of a treatment outcome distribution: 1. the mean (red) should be close to the prescribed dose (black) 2. The distribution should be as narrow as possible, because patients are treated in only a few treatment sessions. (b) In case of the target, the red area, depicting undesirable outcomes, is minimized. This corresponds to maximizing the blue area. In other words, the optimization aims to squeeze the treatment outcome distribution into the set interval.

straints. Because the geometric uncertainties are specific for every individual patient, the probability distribution of the quality metrics for any given plan is also patient specific. Hence, it is essential to consider the shape of the outcome metric distribution $p(y)$ instead of just a nominal value y for the outcome.

Two basic properties are postulated, that any outcome distribution of a robust plan should possess, fig. 2.2(a). First of all, it is important that the planned value of the metric and the mean of the distribution coincide, i.e. the delivered dose should resemble the planned one. If the treatment had *very* many fractions, this condition would suffice to ensure that the proper dose is delivered. In this case, sessions with a less than ideal delivery would be compensated by many others. However, in a realistic scenario a failed treatment session cannot be compensated entirely. It is therefore also necessary to take the outcome distribution width into consideration. By demanding that the outcome distribution should be as narrow as possible, it is ensured that any sample taken from it, i.e. any treatment session, is within acceptable limits. This issue becomes more serious for fewer treatment sessions.

Consequently, any robust treatment optimization approach should keep control over the respective outcome distributions of the involved regions of

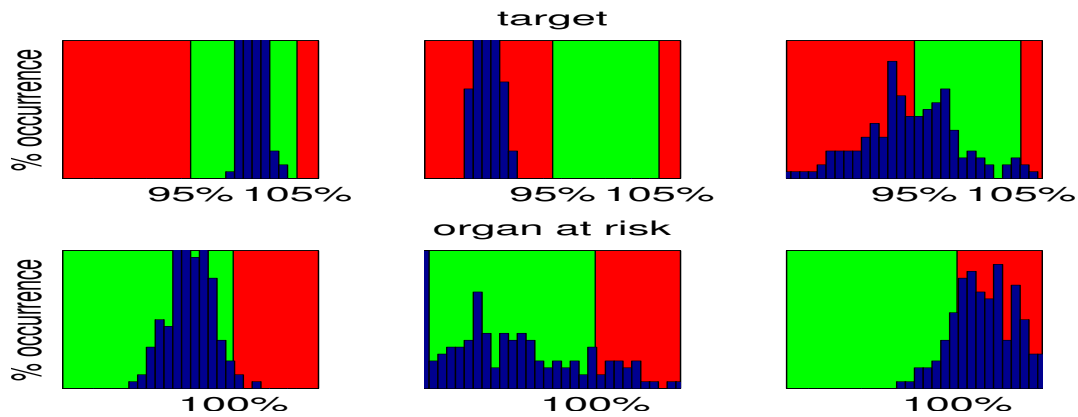


Figure 2.3: Exemplary outcome distributions in relation to the aspired treatment outcome region. The green area denotes acceptable courses of treatment whereas red indicates unfavorable treatment outcomes. The top row illustrates the two-sided target cost function (e.g. the green area could indicate 95%-105% of the prescribed dose). The distribution in the left panel is entirely located within the acceptable region, indicating that the patient would receive the proper dose during all courses of treatment. While the distribution width in the middle panel is sufficiently narrow, its mean is too low, pointing towards an underdosage of the target. The distribution on the right panel is too wide and its mean too low. In practice, this means that there is a significantly reduced chance that the patient will receive proper treatment. The bottom panels show the one-sided OAR cost function. From left to right the complication risk increases. Notice, that even though the mean in the middle panel is lower than on the leftmost panel, the overall distribution is less favorable because of a probability of unacceptable outcomes.

interest. In case of the target, it is desirable to keep the distribution narrow and peaked at the desired dose. The organ at risk distributions, however, should be skewed towards low doses and only a controlled small fraction should be close or above the threshold. This ensures that in the vast majority of all possible treatment courses, the irradiation of the organ at risk is within acceptable limits, fig. 2.3.

The course of treatment is subject to several sources of uncertainty (e.g. setup errors). Suppose each of these sources is quantified by a continuous geometry parameter \mathbf{x} (e.g. displacement distance), its value indicating a possible realization during a treatment session. The precise value of any of these geometry parameters is unknown at treatment time, making them statistical quantities. Thus as random variables, X , they have associated probability distributions, $p(\mathbf{x})$, indicating the probability that a certain value \mathbf{x} is realized. These input distributions need to be estimated either prior to planning based on patient specific or population data, or data acquired during the course of treatment.

The central goal is to obtain reliable estimates of the outcome distribution $p(y)$ given the input uncertainty $p(\mathbf{x})$. First a number of samples, $\mathbf{x}_{1\dots N}$, is drawn from the uncertainty distributions. Then, for all those geometry instances the outcome metric, $y_k = f(D, \mathbf{x}_k)$, $0 \leq k \leq N$, is computed. If this is repeated sufficiently often, i.e. for large N , the treatment outcome distribution, $p(y)$, can be obtained and alongside an estimate of the probability mass, M , present in the desired interval, \mathbf{I}_{tol} .

$$M[Y] = \int d\mathbf{x} p(\mathbf{x}) \begin{cases} 1 & \Leftrightarrow f(\mathbf{x}) \in \mathbf{I}_{\text{tol}} \\ 0 & \Leftrightarrow f(\mathbf{x}) \notin \mathbf{I}_{\text{tol}} \end{cases} \quad (2.1)$$

The interval \mathbf{I}_{tol} is the green area in fig. 2.3.

However, there are two major problems with the practical implementation of this approach. First and foremost, for dependable estimates, large numbers of samples are required, rendering this method becomes prohibitively expensive, especially with dose recalculations involved. Secondly, the quantity of interest, $M[Y]$ (2.1), cannot be computed in closed form, an essential requirement for efficient optimization schemes. The next paragraph will introduce a cost function with similar utility that does not suffer from these limitations.

Typical prescriptions in radiotherapy demand that the interval contains large parts of the probability mass, e.g. 95% of all outcomes must lie within. This means that the exact shape of the actual distribution loses importance while concentrating on the tails by virtue of the *Chebyshev inequality*. Loosely speaking, it states that if the variance of a random variable is small, then

the probability that it takes a value far from its mean is also small. If Y is a random variable with mean μ and variance σ^2 , then

$$\mathbf{P}(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad k \geq 1 \quad (2.2)$$

limits the probability of a sample being drawn k standard deviations away from the mean, *independently* of the density $p(y)$. For the problem at hand, k is calculated by determining how many standard deviations fit in the prescription interval \mathbf{I}_{tol} . If most of the probability mass is already concentrated in the desired area, its spread is relatively small compared to the extent of the acceptable region. This means that k is large, and by virtue of (2.2), the probability of a sample lying outside \mathbf{I}_{tol} is low. This holds *regardless* of what the real outcome distribution may look like.

Hence, instead of trying to obtain the entire outcome distribution, it is *assumed* that its tails can be approximated by a Gaussian to compute all quantities necessary for the optimizer. This is justified by the *Chebyshev inequality* (2.2). The idea is shown in fig. 2.2(b). Given this assumption, it is sufficient to compute mean, $\mathbf{E}_{\mathbf{x}}[Y]$, and variance, $\text{var}_{\mathbf{x}}[Y]$. It is important to note that even if $p(y)$ does not resemble a normal distribution, the above approximation is still reasonable in the context of the proposed prescription scheme. It is sufficient to assume that the tails of $p(y)$ (two sided for target volumes, one sided for OARs) can be described by the tails of the Gaussian $\mathcal{N}(y|\mathbf{E}_{\mathbf{x}}[Y], \text{var}_{\mathbf{x}}[Y])$. This approximation becomes more powerful for fractionated treatments because of the *central limit theorem*[BT02]. It states that the sum of n random variables is normally distributed if n is sufficiently large.

Incorporating the assumption above, the probability mass enclosed in the interval of interest can be approximated analytically,

$$M[Y] \approx \int_{\min \mathbf{I}_{\text{tol}}}^{\max \mathbf{I}_{\text{tol}}} dy \mathcal{N}(y|\mathbf{E}_{\mathbf{x}}[Y], \text{var}_{\mathbf{x}}[Y]) \quad (2.3)$$

which is crucial for the formulation of an objective function for the optimizer. The entire optimization problem reads

$$\text{maximize } M[Y] \quad (2.4a)$$

$$\text{subject to } 1 - M[Z_k] \leq c_k \quad (2.4b)$$

$$\text{and } \phi_j \geq 0 \quad \forall j. \quad (2.4c)$$

Loosely speaking, this amounts to maximizing the probability mass of treatment outcomes for the tumor inside the prescription interval via (2.4a), while

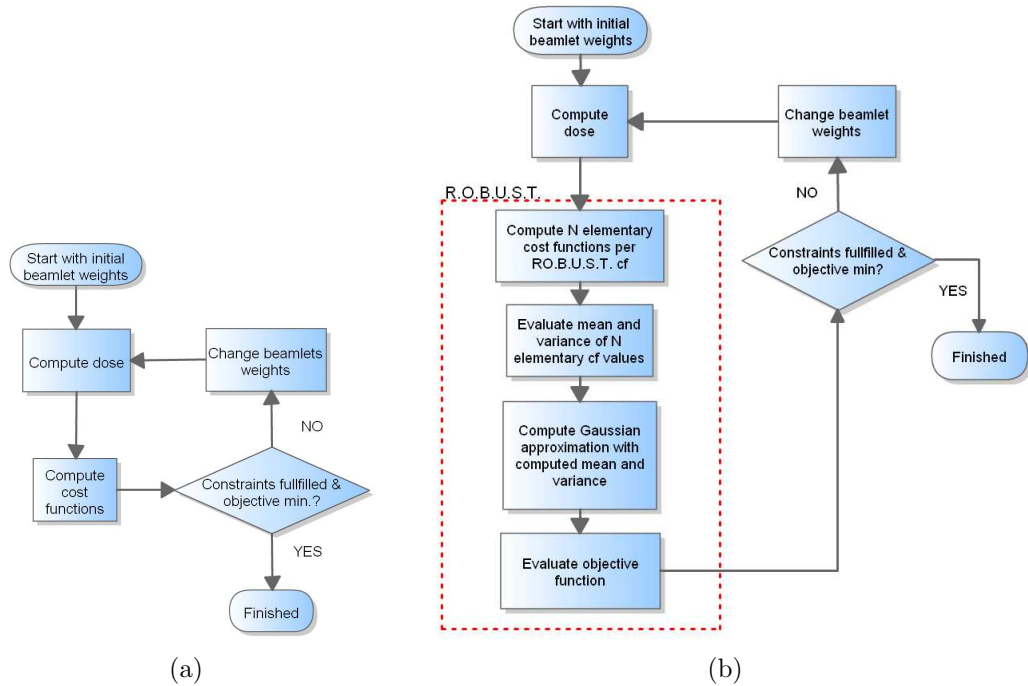


Figure 2.4: (a) Constrained optimization in conjunction with classic cost functions. (b) The R.O.B.U.S.T. probabilistic approach. Notice that the algorithm is entirely concealed from the optimizer. Hence, no alterations are necessary.

at the same time allowing only a controlled portion of the organ at risk distribution over a set threshold in (2.4b).

Since the Gaussian approximation tries to squeeze the probability mass into the interval, it points the optimizer into the right direction, *independently* of the actual shape of $p(y)$. This is guaranteed by the *Chebyshev inequality*. Further refinements regarding the error bounds are possible, but beyond the scope of this work. For the optimization of (2.4) to work, it is important that the *Chebyshev inequality* exists, but not necessarily how accurate it is. The Chebyshev bounds can be tightened by re-adjusting the imposed constraints c_k in (2.4b). Notice that the actual distribution depends on the intermediate treatment plans of the iterative optimization, and hence constantly changes its shape. The more the dose distribution approaches the optimization goal, the less probability mass stays outside the prescription interval and the better the Gaussian approximation and the tighter the Chebyshev bound become.

The algorithm for the cost function computation works as follows, c.f. fig. 2.4:

1. For each of the N geometry instances the cost function, f , is computed, $y_k = f(D, \mathbf{x}_k)$, $1 \leq k \leq N$
2. Mean, $E_{\mathbf{x}}[Y]$, and variance, $\text{var}_{\mathbf{x}}[Y]$, of those N samples is determined.
3. Use mean and variance to approximate a Gaussian $\mathcal{N}(y|E_{\mathbf{x}}[Y], \text{var}_{\mathbf{x}}[Y])$, fig. 2.2(b).
4. Compute the objective function, (2.3), and return the value to the optimizer.

Results of the algorithm, as well as a side by side comparison against different robust optimization methods can be found in [SMA10].

2.1.1 Explicit Incorporation of the Number of Fractions

It was mentioned earlier that the spread of the outcome distribution becomes increasingly important as one moves towards smaller fraction numbers. This is substantiated by the *law of large numbers* in probability theory [BT02]. In this section, the previous observation is quantified and the cost function augmented accordingly.

In the present context, the final outcome A_n can be regarded as the sample mean of all n sessions Y_i

$$A_n = \frac{\sum_{i=1}^n Y_i}{n}. \quad (2.5)$$

However, notice that the quantity of interest is typically the EUD of the cumulative dose. Taking into account the curvature properties of the EUD functions, one can establish a relation between the two quantities to prove that it is sufficient to consider the average EUD instead of the EUD of the average dose. Given a convex function f and a random variable D , *Jensen's inequality* [Jen06] establishes a link between the function value at the average of the random variable and the average of the function values of the random variable, namely

$$f(E[D]) \leq E[f(D)]. \quad (2.6)$$

The converse holds true if f is concave. With the general form of the EUD

$$\text{EUD}(D) = g^{-1} \left(\frac{1}{V} \sum_{v \in V} g(D_v) \right) \quad (2.7)$$

it can be shown that (2.5) provides a lower bound for the tumor control and an upper bound for normal tissue damage. For details regarding this argument, please refer to appendix B.1 and [SSSA11].

The random variable A_n , (2.5) has mean $E_{\mathbf{x}}[A_n] = E_{\mathbf{x}}[Y] \equiv \mu$ and variance $\text{var}_{\mathbf{x}}[A_n] = \frac{\text{var}_{\mathbf{x}}[Y]}{n} \equiv \frac{\sigma^2}{n}$. Again, the Chebyshev inequality, (2.2), is employed to obtain an approximation for the probability mass inside the prescription interval. With $k \equiv \frac{\varepsilon}{\sigma}$ and the mean and variance of A_n ,

$$\mathbf{P}(|A_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad (2.8)$$

is obtained, where ε is the maximum tolerated error. The probability of a treatment violating the requirements, i.e. the realized outcome is more than ε away from the prescription, is proportional to the variance of a single fraction, σ^2 , and goes to zero as the number of treatment fractions, n , increases. Further refinements can be made by reintroducing the Gaussian approximation. The accuracy of the approximation increases with the number of fractions, n , as indicated by the *central limit theorem* [BT02]. It states that the sum of a large number of independent random variables is approximately normal. The actual convergence rate depends on the shape of Y_i . For example, if Y_i was uniformly distributed, $n \sim 8$ would suffice [BT02]. The ensuing probability mass is computed analogously to (2.3),

$$\begin{aligned} M[A_n] &\approx \int_{\min \mathbf{I}_{\text{tol}}}^{\max \mathbf{I}_{\text{tol}}} dy \mathcal{N}(y|E_{\mathbf{x}}[A_n], \text{var}_{\mathbf{x}}[A_n]) \\ &= \int_{\min \mathbf{I}_{\text{tol}}}^{\max \mathbf{I}_{\text{tol}}} dy \mathcal{N}(y|E_{\mathbf{x}}[Y], \frac{\text{var}_{\mathbf{x}}[Y]}{n}). \end{aligned} \quad (2.9)$$

By explicitly including fractionation effects in the cost function, (2.9), the influence of the variance is regulated. That is, one *quantitatively* accounts for the fact that a given number of fractions offers a certain potential for treatment errors to even out. Notice that in the limit of *very* many fractions, the initial statement that the mean value alone suffices to ensure proper treatment is recovered.

2.2 Efficient Treatment Outcome Parameter Estimation

To estimate the robustness of a treatment plan, it is necessary to capture the response of the treatment outcome metric if the uncertain inputs are varied according to their respective distributions. Only if the response is sufficiently

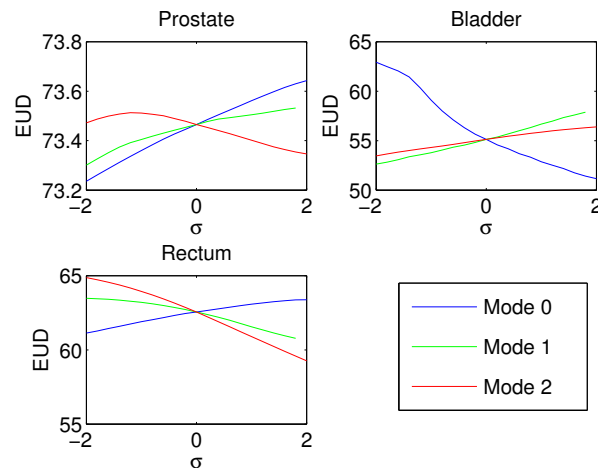


Figure 2.5: Sensitivity of the EUDs of different organs to gradual shape changes for a prostate cancer patient. The latter were modelled with principal component analysis (PCA) of a set of CT scans. The plots show the effects of the first three eigenmodes. For example, the first mode roughly correlates to A-P motion induced by the varying filling of rectum and bladder. Notice that the prostate EUD is fairly robust toward the uncertainties in question.

well behaved, i.e. does not depart from its tolerance limits, the treatment plan can be considered robust.

This section deals with the estimation of the treatment outcome distribution. This is commonly referred to as *uncertainty analysis*. The most direct way to do this is the Monte Carlo (MC) method. Samples are drawn from the input uncertainty distributions and the model is rerun for each setting. While MC is conceptionally very simple, an excessive number of runs may be necessary. For computer codes that have a non-negligible run time, the computational cost can quickly become prohibitive.

To alleviate this problem, an emulator for the real patient model is introduced. The underlying assumption is that the patient model output is a function $f(\cdot)$ that maps inputs, i.e. uncertain parameters X , into an output $y = f(\mathbf{x})$, i.e. the dose metric. The behavior of the EUDs of four exemplary prostate cancer plans is shown in fig. 2.5.¹ An emulation $\mu(\cdot)$ in place of the patient model $f(\cdot)$ is employed. If the emulation is good enough, then the results produced by the analysis will be sufficiently close to those that would have been obtained using the original patient model. Because the

¹Notice that due to the continuous nature of physical dose distributions, and their finite gradients, the smoothness assumption holds true for all physical and biological metrics commonly employed in dose optimization.

computation of the approximate patient model response using the emulator is *much* faster than the direct evaluation of the patient model, comprehensive analyses become feasible.

To replace the patient model, Gaussian Processes (GPs) [RW06, Mac03] are used. In the present context, they are used to learn the response of the patient model and subsequently mimic its behavior. GPs have been used to emulate expensive computer codes in a number of contexts / applications [SWMW89, KO00, OO02, OO04, PHR06, SWN03]. Notice that GP regression is also commonly referred to as “kriging” in literature.

This section is divided as follows. First, a brief introduction in Gaussian process regression is provided. Afterwards, it is presented how GPs can be used to greatly accelerate treatment outcome Monte Carlo analyses by substituting the patient model. It turns out that under certain conditions the introduced computations can be further accelerated while increasing precision. This technique, as well as benchmarks, is subject of the remainder of the section.

2.2.1 Gaussian Process Regression

In the following, the sources of uncertainty during treatment time are quantified by the vector-valued random variable X . The associated probability distribution, $p(\mathbf{x})$, indicates the likelihood that a certain vector \mathbf{x} is realized. The dimensionality of \mathbf{x} is denoted by S , i.e. $\mathbf{x} \in \mathbb{R}^S$. For a given dose distribution and any given vector \mathbf{x} , the patient model computes the outcome metric $y \in \mathbb{R}$. The input uncertainty inevitably renders the outcome metric a random variable Y .

Some fundamentals of Gaussian processes are only briefly recaptured, as a comprehensive introduction is beyond the scope of this work. For a detailed introduction, consult [RW06], [Mac03, ch. 45] or [SS02, ch. 16].

To avoid confusion, it is stressed that the method of Gaussian processes is used to provide a non-parametric, assumption-free, smooth interpretation of $f(\mathbf{x})$, for which otherwise only a few discrete observations are available. The GP does not make any assumptions about the nature of the probability distribution $p(\mathbf{x})$.

Assume a data set $\mathcal{S} = \mathbf{x}_n, y_n$ with N samples is given, stemming from a patient model. It is the goal to utilize this data to predict the model response $f(\mathbf{x}^*)$ at a previously unseen location, e.g. geometry configuration, \mathbf{x}^* . The key idea is to incorporate that for any reasonable quality metric f , $f(\mathbf{x}') \approx f(\mathbf{x}^*)$ holds if \mathbf{x}' is sufficiently close to \mathbf{x}^* , i.e. f is a smooth, continuous function. For example, if \mathbf{x} denotes the displacement of the patient from the planning position, minor geometry deviations should entail only a small

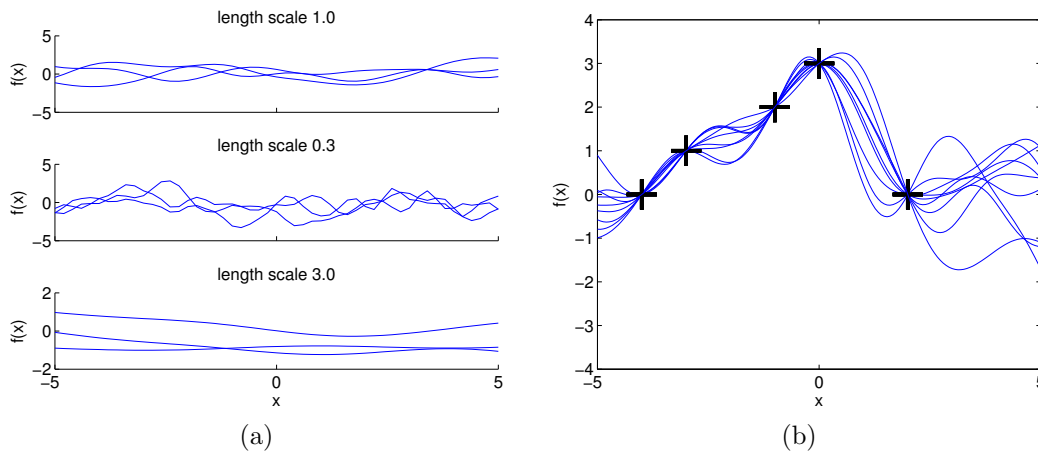


Figure 2.6: (a) Random functions drawn from a Gaussian process prior. Notice the influence of the scaling *hyperparameter* on the function. (b) Random functions drawn from a Gaussian process posterior where sample positions are marked with crosses. Notice that all functions pass through the observed points.

change in the dose metric y .¹

GPs can be thought of as a generalization of linear parametric regression. In linear parametric regression, a set of J basis functions $\phi_j(\mathbf{x})$ is chosen and is linearly superimposed:

$$\hat{f}(\mathbf{x}) = \sum_{j=1}^J \alpha_j \phi_j(\mathbf{x}) = \alpha^\top \cdot \phi(\mathbf{x}) \quad (2.10)$$

Bayesian inference of the weighting parameters α follows two steps. First, a *prior* distribution is placed on α , reflecting the beliefs of the modeller *before* having seen the data. The prior, $p(\alpha)$, is then updated in light of the data \mathcal{S} via Bayes' rule leading to the *posterior* distribution, $p(\alpha|\mathcal{S}) \propto p(\mathcal{S}|\alpha) \cdot p(\alpha)$. The posterior reflects the *updated knowledge* of the model.

The properties of the model in (2.10) depend crucially on the choice of basis functions $\phi(\mathbf{x})$. Gaussian process modelling circumvents this problem by placing the prior directly on the space of functions, fig. 2.6(a), eliminating the need to explicitly specify basis functions, thus allowing for much greater flexibility. Following the scheme of Bayesian inference, the GP prior is updated using Bayes' rule yielding the posterior, when observations become available, fig. 2.6(b).

By definition, a Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution [RW06]. That

is, the joint distribution of the observed variables, $[f(\mathbf{x}_1) \dots f(\mathbf{x}_N)]^\top$, is a multivariate Gaussian.

$$(f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)) \sim \mathcal{N}(f|\boldsymbol{\mu}, \Sigma) \quad (2.11)$$

with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ . A Gaussian process can be thought of a generalization of a Gaussian distribution over a finite dimensional vector space to an infinite dimensional function space [Mac03]. This extends the concept of random variables to random functions.

A Gaussian process is completely specified by its mean and covariance function:

$$\mu(\mathbf{x}) = \mathbb{E}_f[f(\mathbf{x})] \quad (2.12a)$$

$$\begin{aligned} \text{cov}(f(\mathbf{x}), f(\mathbf{x}')) &= \mathbb{E}_f[(f(\mathbf{x}) - \mu(\mathbf{x})) \\ &\quad (f(\mathbf{x}') - \mu(\mathbf{x}'))] \end{aligned} \quad (2.12b)$$

In the context of Gaussian processes, the covariance function is also called kernel function,

$$\text{cov}(f(\mathbf{x}), f(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}'). \quad (2.13)$$

It establishes the coupling between any two dose metric values $f(\mathbf{x})$, $f(\mathbf{x}')$. A common choice for $k(\cdot, \cdot)$ is the so-called squared exponential covariance function

$$k_{\text{se}}(\mathbf{x}, \mathbf{x}') = \omega_0^2 \exp\left(-\frac{1}{2}(\mathbf{x}' - \mathbf{x})^\top \Omega^{-1}(\mathbf{x}' - \mathbf{x})\right), \quad (2.14)$$

where Ω is a diagonal matrix with $\Omega = \text{diag}\{\omega_1^2 \dots \omega_S^2\}$. If the locations \mathbf{x} , \mathbf{x}' are close to each other, the covariance in (2.14) will be almost one. That in turn implies that the respective function values, $f(\mathbf{x})$, $f(\mathbf{x}')$ are assumed to be almost equal. As the locations move apart, the covariance decays rapidly along with correlation of the function values. Given some covariance function, the covariance matrix of the Gaussian in (2.11) is computed with $\Sigma_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1 \dots N$.

Because it governs the coupling of the observations, the class of functions which is embodied by the Gaussian process, is encoded in the covariance function. In (2.14) the behavior is steered by $\boldsymbol{\omega}$, the so-called *hyperparameters*. While ω_0 is the signal variance, $\omega_{1 \dots S}$ are scaling parameters. The scaling parameters reflect the degree of smoothness of the GP. Their influence on the prior is shown in fig. 2.6(a). In (2.14) a separate scaling parameter is used for each input dimension. This allows for the model to adjust to the influence of the respective input dimensions. This is known as *Automatic Relevance Determination* [Mac95, Nea96]. If the influence of an input dimension is

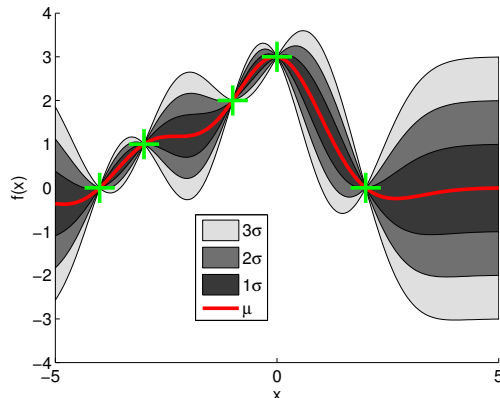


Figure 2.7: The mean function as well as the predictive uncertainty of a GP.

rather small, the length scale in that dimension will be large, indicating that function values are relatively constant (or vice versa).

Given this prior on the function f , $p(f|\boldsymbol{\omega})$, and the data $\mathcal{S} = \{\mathbf{x}_n, y_n\}$, the aim is to get the predictive distribution of the function value $f(\mathbf{x}^*)$. Because the patient model typically exists on a voxel grid, its discretization may also have an influence on y , especially since the dependency of y on some kind of continuous movement is modelled. Consequently, it must be taken into account that the data may be contaminated by noise, i.e.

$$y = f(\mathbf{x}) + \varepsilon. \quad (2.15)$$

Standard practice is to assume Gaussian noise, $\varepsilon \sim \mathcal{N}(\varepsilon|0, \omega_n^2)$. Hence, an extra contribution, ω_n^2 , is added to the variance of each data point. The final kernel matrix now reads

$$K_{ij} = \Sigma_{ij} + \omega_n^2 \delta_{ij} \quad (2.16)$$

where δ is the Kronecker delta. By additionally accounting for possible noise in the data, the constraint that the posterior has to go exactly through the data points is relaxed. This is a very powerful assumption, because it allows for the data explaining model to be simpler, i.e. the function is less modulated. Thus, the hyperparameters of the GP are larger. This, in turn, entails that fewer sample points are needed to make predictions with high accuracy. The trade-off between model simplicity and data explanation within the GP framework follows *Occam's razor* [RG01]. In other words, the model is *automatically* chosen as simple as possible while being at the same time as complex as necessary. Patient model uncertainties that arise due to imprecise image segmentation, discretization errors, etc. are effectively handled this way.

Within the GP framework, the predictive distribution, $p(f_*|\mathbf{x}_*, \mathcal{S}, \boldsymbol{\omega})$, for a function value $f_* \equiv f(\mathbf{x}_*)$ is obtained by conditioning the joint distribution of unknown point and data set, $p(f_*, \mathcal{S}|\boldsymbol{\omega})$, on the data set [Mac03]. Since all involved distributions are Gaussians, the posterior predictive distribution is also a Gaussian,

$$p(f_*|\mathbf{x}_*, \mathcal{S}, \boldsymbol{\omega}) = \mathcal{N}(f_*|\mu(\mathbf{x}_*), \sigma^2(\mathbf{x}_*)). \quad (2.17a)$$

$$\text{with mean } \mu(\mathbf{x}_*) = \mathbf{k}(\mathbf{x}_*)^\top K^{-1}\mathbf{y} \quad (2.17b)$$

$$\text{and variance } \sigma^2(\mathbf{x}_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*)^\top K^{-1}\mathbf{k}(\mathbf{x}_*) \quad (2.17c)$$

where $\mathbf{k}(\mathbf{x}_*) \in \mathbb{R}^{N \times 1}$ is the vector of covariances between the unknown point and data set, i.e. $[\mathbf{k}(\mathbf{x}_*)]_j \equiv k(\mathbf{x}_j, \mathbf{x}_*)$. Fig. 2.7 shows the mean function (2.17b) as well as the confidence interval (2.17c) for an exemplary dataset. For a detailed derivation of (2.17), please refer to [RW06, chap. 2]. Notice that the mean prediction (2.17b) is a linear combination of kernel functions centered on a training point. With $\boldsymbol{\beta} \equiv K^{-1}\mathbf{y}$ (2.17b) can be rewritten as

$$\mu(\mathbf{x}_*) = \mathbf{k}(\mathbf{x}_*)^\top \boldsymbol{\beta}. \quad (2.18)$$

Loosely speaking, the function value $f(\mathbf{x}_*)$ at an unknown point \mathbf{x}_* is inferred from the function values of the data set \mathcal{S} weighted by their relative locations to \mathbf{x}_* .

It is important to notice that Gaussian process prediction does *not* return a single value. Instead, it provides an entire probability distribution over possible values $f(\mathbf{x}_*)$, (2.17). However, the mean (2.17b) is often regarded as approximation, but the GP also provides a distribution around that mean, (2.17c), which describes the likelihood of the real value being close by.

At last it is briefly covered how the *hyperparameters* are computed. Again, this is presented in much greater detail in [RW06, ch. 5], [Mac03, ch. 45], [SS02, ch. 16]. Up to this point the hyperparameters $\boldsymbol{\omega}$ were assumed as fixed. However, it was shown that their values critically affect the behavior of the GP, fig. 2.6(a). In a sound Bayesian setting, one would assign a prior over the hyperparameters and average over their posterior, i.e.

$$p(f_*|\mathbf{x}_*, \mathcal{S}) = \int d\boldsymbol{\omega} p(f_*|\mathbf{x}_*, \mathcal{S}, \boldsymbol{\omega}) p(\boldsymbol{\omega}|\mathcal{S}). \quad (2.19)$$

However, the integral in (2.19) cannot be computed analytically. A common choice in this case is *not* to evaluate the posterior of the parameters $\boldsymbol{\omega}$, but to approximate it with a delta distribution, $\delta(\boldsymbol{\omega} - \boldsymbol{\omega}^*)$. Thus, (2.19) simplifies to:

$$p(f_*|\mathbf{x}_*, \mathcal{S}) \approx p(f_*|\mathbf{x}_*, \mathcal{S}, \boldsymbol{\omega}^*) \quad (2.20)$$

The preferred set of hyperparameters maximizes the marginal likelihood

$$\boldsymbol{\omega}^* = \underset{\boldsymbol{\omega}}{\operatorname{argmax}} p(\mathcal{S}|\boldsymbol{\omega}) \quad (2.21)$$

It can be computed in closed form and optimized with a gradient based scheme. In the implementation the negative log likelihood is optimized

$$-\log[p(\mathcal{S}|\boldsymbol{\omega})] = -\frac{1}{2} \log \det K - \frac{1}{2} \mathbf{y}^\top K^{-1} \mathbf{y} - \frac{N}{2} \log(2\pi) \quad (2.22)$$

with the BFGS method. Since the kernel matrix K is symmetric and positive-definite, it can be inverted using Cholesky decomposition. Generally, the selection of the hyperparameters is referred to as *training* of the Gaussian process.

2.2.2 Uncertainty Analysis with Gaussian Processes

The central goal of this work is to efficiently compute the treatment outcome distribution under the influence of uncertain inputs. The benefits of GP regression to accelerate classical Monte Carlo computations are investigated in the following. The most straightforward way to use the GP emulator is to simply run an excessive number of evaluations based on the GP mean, (2.17b). This approach implicitly exploits the substantial information gain in the *proximity* of each sample point. This proximity information is lost in classical Monte Carlo methods.

The application of Gaussian processes as a model substitution is inherently a two stage approach. First, the Gaussian process emulator is built and subsequently used in place of the patient model to run any desired analysis. Only one set of runs of the patient model is needed, to compute the sample points necessary to build the GP. Once the GP emulator is built, there is no more need for any additional patient model evaluations, no matter how many analyses are required of the simulator's behavior. This contrasts with conventional Monte Carlo based methods. For instance, to compute the treatment outcome for a slightly altered set of inputs, fresh runs of the patient model are required.

GPs are used to learn the behavior of a head and neck patient model under rigid movement. The primary structures of interest, Chiasm and GTV, are directly adjacent to each other as depicted in fig. 2.9. This is the reason why a slight geometry change leads not only to a target EUD decline, but also to a rapid dose increase in the chiasm.

For the experiment, 1000 sample points, i.e. displacement vectors, were drawn at random. For all runs, the same dose distribution was used. To

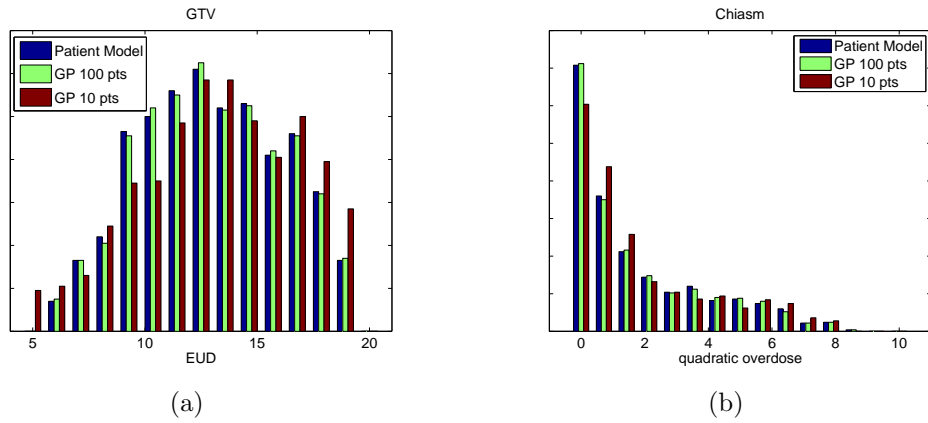


Figure 2.8: Treatment outcome histograms obtained with different methods for (a) the GTV and (b) the Chiasm. The blue bar stands for 1000 evaluations of the patient model. The green and red bars stand for 1000 evaluations of a Gaussian process trained on 100 or 10 points, respectively.

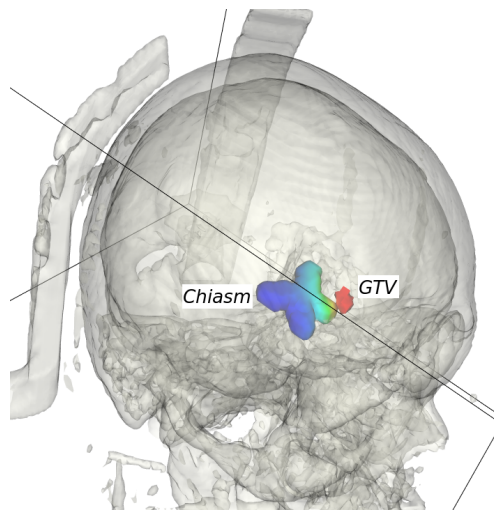


Figure 2.9: The GTV and Chiasm in a head and neck case. Both structures are colored according to the delivered dose (dose increasing from blue to red). Notice the extremely steep dose gradient between GTV and Chiasma.

no. of pat. model runs	10	100
MC on pat. model.	$12.3 \pm 31\%$	$13.36 \pm 10.0\%$
MC on GP	$13.1 \pm 0.3\%$	$12.9 \pm 0.3\%$
no. of pat. model runs	1000	100000
MC on pat. model.	$12.96 \pm 3.1\%$	$12.89 \pm 0.3\%$
MC on GP	$12.89 \pm 0.3\%$	–

Table 2.1: Computation of the average treatment outcome (EUD) for the GTV under random rigid movement. The second row shows the result for the standard approach of simply averaging over the computed outcomes. Accordingly, the error bound becomes tighter as more samples are involved. The third row shows the results of computing the average via MC on a Gaussian Process that mimics the patient model. The MC error is always at a minimum because in every run 10^5 samples were computed on the GP. Merely the number of original patient model evaluations to build the GP is varied.

establish a ground truth, the patient model was evaluated for each displacement vector and included in the analysis. To evaluate the Gaussian Process performance, only a small subset of points was chosen to build the emulator. Once built, the rest of the 1000 points were evaluated with the emulator. Fig. 2.8 shows the outcome histograms of GTV and chiasm respectively if the displacement in x-y-direction is varied by max. ± 1 cm. The plots show that as few as 10 points suffice to capture the qualitative behavior of the model. With only 100 points, i.e. 10% of the MC model runs, the GP histogram is almost indistinguishable from the patient model histogram obtained through classical MC.

For instance, one could use the emulator runs $\mu_1 = \mu(\mathbf{x}_1), \mu_2 = \mu(\mathbf{x}_2), \dots$ to compute the mean treatment outcome. It is important to note that with increasing sample size, the value for the mean does *not* converge to the true mean of the patient model, but the mean of the GP. However, given a reasonable number of sample points to build the GP, it is often the case that the difference between GP and actual patient model is small compared to the error arising from an insufficient number of runs when computing the mean based on the actual patient model.

To clarify this, consider the head and neck case illustrated in fig. 2.9. For all practical purposes, the true average treatment outcome of the GTV is 12.89 Gy based on 10^5 patient model runs. In table 2.1, the results of

both methods for a varying number of patient model runs are collected. The theoretical error margin of the Monte Carlo method, $1/\sqrt{N}$, is also included. The means based on the GP were computed using 10^5 points resulting in an error of 0.3%. Consequently, the residual deviations stem from errors in the Gaussian Process fit.

A similar argument can be made for the variance of the treatment outcomes.

2.2.3 Bayesian Monte Carlo

It was previously shown that the use of GP emulators to bypass the original patient model greatly accelerates the treatment outcome analysis. Because all results in the previous section were essentially obtained through Monte Carlo on the GP mean, this approach still neglects that the surrogate model itself is subject to errors. In the following, a Bayesian approach is presented that does not suffer from these limitations.

The R.O.B.U.S.T. costfunction (2.3) is based entirely on mean and variance of the outcome metric, Y . Consequently, it is the objective to obtain both quantities as efficiently as possible. The underlying technique to solve integrals (such as mean and variance) over Gaussian Processes has been introduced as Bayes Hermite Quadrature [O'H91] or *Bayesian Monte Carlo* (BMC) [RG03].

It was established that in case of Gaussian input uncertainty, $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, B)$, in conjunction with the squared exponential kernel, (2.14), mean and variance can be computed analytically [GRCMS03]. The remainder of this subsection deals with the details of computing these quantities.

Given the GP predictive distribution, (2.17), the additional uncertainty is incorporated by averaging over the inputs. This is done by integrating over \mathbf{x}_* :

$$p(f^*|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}, \mathcal{S}, \boldsymbol{\omega}) = \int d\mathbf{x}_* p(f^*|\mathbf{x}_*, \mathcal{S}, \boldsymbol{\omega}) p(\mathbf{x}_*|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \quad (2.23)$$

Notice that $p(f^*|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}, \mathcal{S}, \boldsymbol{\omega})$ can be seen as marginal predictive distribution, as the original predictive distributions has been marginalized with respect to \mathbf{x}_* . It only depends on the mean and variance of the inputs.

For notational simplicity, set $p(\mathbf{x}) \equiv p(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$. The quantities of interest are the expectation of the mean, $\mu_{f|\mathcal{S}}$, and variance, $\sigma_{f|\mathcal{S}}^2$ [OO04, PHR06],

$$\begin{aligned} \mu_{f|\mathcal{S}} &\equiv \mathbb{E}_{f|\mathcal{S}} [\mathbb{E}_{\mathbf{x}}[f(X)]] = \mathbb{E}_{\mathbf{x}} [\mathbb{E}_{f|\mathcal{S}}[f(X)]] \\ &= \mathbb{E}_{\mathbf{x}} [\mu(X)] \\ &= \int d\mathbf{x} p(\mathbf{x}) \mathbf{k}(\mathbf{x})^\top K^{-1} \mathbf{y} = \mathbf{z}^\top K^{-1} \mathbf{y}, \end{aligned} \quad (2.24)$$

where μ is the GP mean, (2.17b). The estimated variance of the function over $p(\mathbf{x}_*|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$ decomposes to

$$\begin{aligned} \sigma_{f|S}^2 \equiv E_{f|S} [\text{var}_{\mathbf{x}}[f(X)]] &= \underbrace{\text{var}_{\mathbf{x}} [E_{f|S}[f(X)]]}_{\text{Term 1}} \\ &+ \underbrace{E_{\mathbf{x}} [\text{var}_{f|S}[f(X)]]}_{\text{Term 2}} - \underbrace{\text{var}_{f|S} [E_{\mathbf{x}}[f(X)]]}_{\text{Term 3}} \end{aligned} \quad (2.25)$$

If the GP had no predictive uncertainty, i.e. $\text{var}_{f|S}[\cdot] = 0$, the expectation of the input variance with respect to the posterior would be solely determined by the variance of the posterior mean, i.e. term 1,

$$\begin{aligned} \text{var}_{\mathbf{x}} [E_{f|S}[f(X)]] &= \text{var}_{\mathbf{x}} [\mu(X)] \\ &= \int d\mathbf{x} p(\mathbf{x}) (\mathbf{k}(\mathbf{x})^\top K^{-1} \mathbf{y})^2 - E_{\mathbf{x}}^2 [\mu(X)] \quad (2.26a) \\ &= (K^{-1} \mathbf{y})^\top L (K^{-1} \mathbf{y}) - E_{\mathbf{x}}^2 [\mu(X)]. \end{aligned}$$

The remaining terms 2 & 3 account for the GP uncertainty, (2.17c),

$$\begin{aligned} E_{\mathbf{x}} [\text{var}_{f|S}[f(X)]] &= E_{\mathbf{x}} [\sigma^2(X)] \\ &= \int d\mathbf{x} p(\mathbf{x}) [k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x})^\top K^{-1} \mathbf{k}(\mathbf{x})] \quad (2.26b) \\ &= k_0 - \text{tr} [K^{-1} L] \end{aligned}$$

and

$$\begin{aligned} \text{var}_{f|S} [E_{\mathbf{x}}[f(X)]] &= E_{f|S} [(E_{\mathbf{x}}[f(X)] - E_{f|S}[E_{\mathbf{x}}[f(X)]])^2] \\ &= \int d\mathbf{x} p(\mathbf{x}) \int d\mathbf{x}' p(\mathbf{x}') \text{cov}_{f|S} [f(\mathbf{x}), f(\mathbf{x}')] \\ &= \int d\mathbf{x} p(\mathbf{x}) \int d\mathbf{x}' p(\mathbf{x}') k(\mathbf{x}, \mathbf{x}') \\ &\quad - \mathbf{k}(\mathbf{x})^\top K^{-1} \mathbf{k}(\mathbf{x}') \\ &= k_c - \mathbf{z}^\top K^{-1} \mathbf{z}. \end{aligned} \quad (2.26c)$$

Some basic, re-occurring quantities can be identified:

$$k_c = \int d\mathbf{x} p(\mathbf{x}) \int d\mathbf{x}' p(\mathbf{x}') k(\mathbf{x}, \mathbf{x}') \quad (2.27a)$$

$$k_0 = \int d\mathbf{x} p(\mathbf{x}) k(\mathbf{x}, \mathbf{x}) \quad (2.27b)$$

$$z_l = \int d\mathbf{x} p(\mathbf{x}) k(\mathbf{x}, \mathbf{x}_l) \quad (2.27c)$$

$$L_{jl} = \int d\mathbf{x} p(\mathbf{x}) k(\mathbf{x}, \mathbf{x}_l) k(\mathbf{x}, \mathbf{x}_j) \quad (2.27d)$$

The l -th sample point is denoted by \mathbf{x}_l . It becomes apparent from (2.27) that the problem can be reduced to an integral over some product of the input distribution and the kernel. Normally, high dimensional integrals as in (2.27) are hard to solve and analytically intractable. With $\Sigma_{\mathbf{x}} = \text{diag}\{\sigma_{\mathbf{x}_1}^2 \dots \sigma_{\mathbf{x}_S}^2\}$, the input distribution factorizes, $p(\mathbf{x}) = \prod_i p_i(\mathbf{x}_i) = \prod_i \mathcal{N}(x_i | \mu_{\mathbf{x}_i}, \sigma_{\mathbf{x}_i}^2)$ and one S -dimensional integral can be decomposed into S 1-dimensional integrals. Because the solutions for k_c, k_0, \mathbf{z} and L tend to be lengthy, they were moved to appendix A.1.

Summarizing the above, closed form expressions for mean and variance of the outcome Y were derived. The calculation consists of the following steps:

1. Generate a set of samples, $\mathcal{S} = \mathbf{x}_n, y_n$, from the original patient model
2. Utilize this sample set to train a Gaussian process, (2.17), that subsequently replaces the patient model
3. Compute mean, (2.24), and variance, (2.25), of outcome Y based on the GP

Aside from taking the GP uncertainty into account, BMC also eliminates MC errors due to a finite sample size. Because the computations are cheap, even compared to brute force MC on the GP, BMC is more expedient.

2.2.4 Bayesian Monte Carlo vs. classic Monte Carlo

To make R.O.B.U.S.T. computationally efficient, the number of recomputations of potential courses of treatment has to be minimal. Hence, it is of great interest to evaluate the predictive performance of Bayesian Monte Carlo opposed to classical Monte Carlo. To establish a ground truth, the EUDs for a fixed dose distribution for 10^6 randomly drawn PCA geometries were evaluated. For each setting, 100 runs with randomly drawn samples were computed. The trials include the first three most significant PCA modes.

Fig. 2.10 show the convergence properties of mean and variance as a function of the number of points used. For every organ, BMC consistently outperformed MC by several orders of magnitude. Compared to MC, BMC uses the information from each sample point more efficiently, which is the key to make R.O.B.U.S.T. feasible under realistic conditions. Exploiting the fact that minor perturbations in the inputs induce only slight changes in the outcome metrics ultimately leads to a major reduction of the number of points necessary to obtain reliable results for mean and variance.

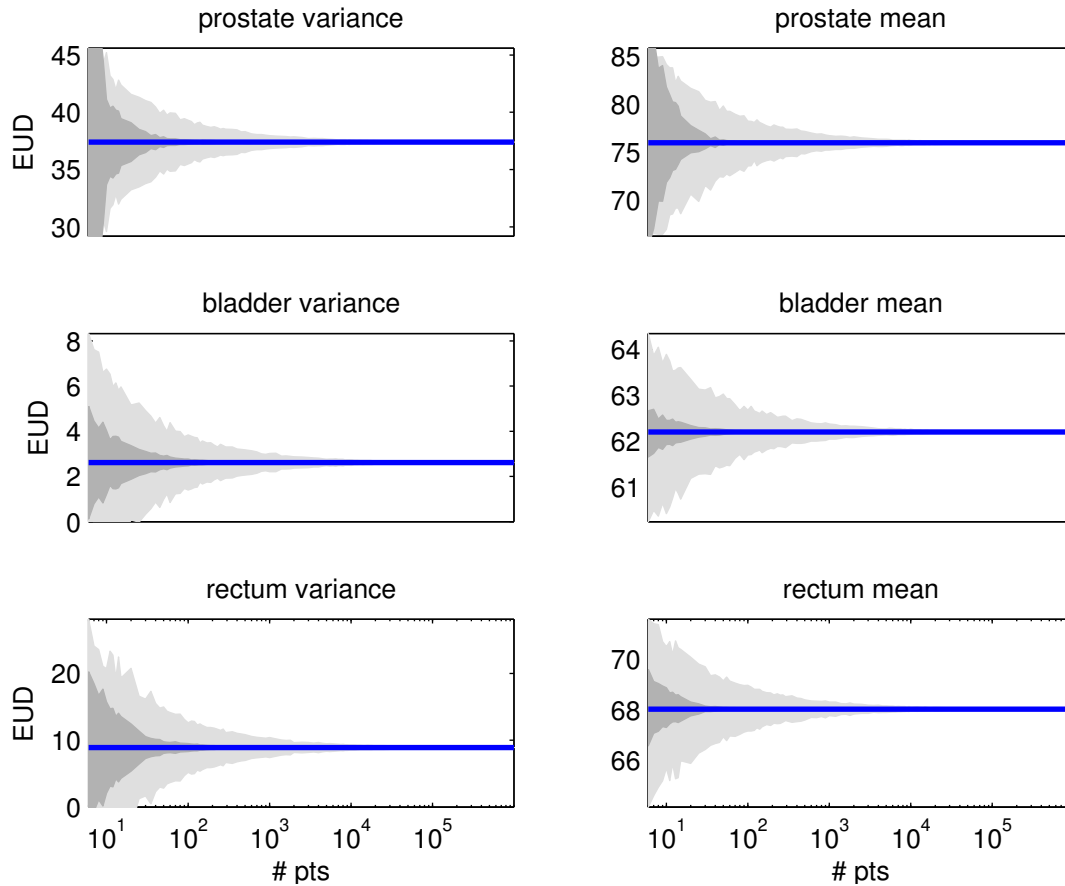


Figure 2.10: Convergence speed of BMC compared to classical MC for the predictive mean and predictive variance. The blue line denotes the ground truth. Light gray is the standard deviation (3σ) from MC and dark gray from BMC. For all organs in question, BMC reaches reliable results with a significantly reduced number of points, thus clearly outperforming classical MC.

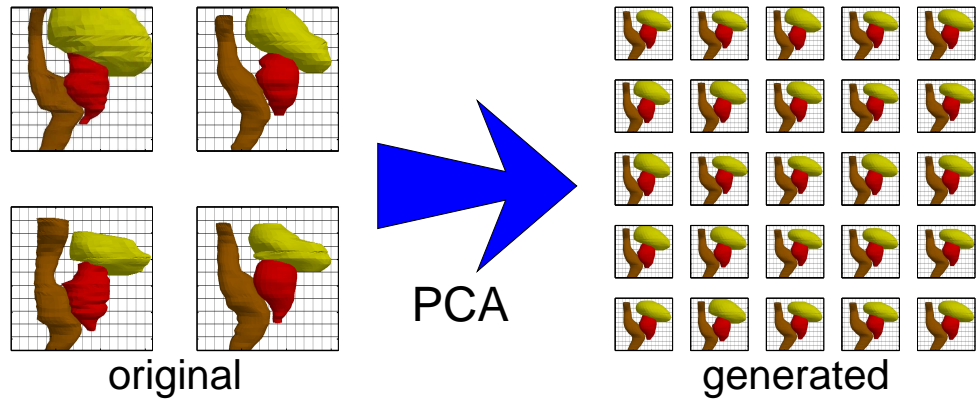


Figure 2.11: Based on only a few real CTs, Principal Component Analysis (PCA), can be used to generate an arbitrary amount of realistic geometry configurations[SBYA05].

2.3 Accelerating the Optimization Algorithm

In the proof of principle optimizer [SMA10], the main ingredients of the R.O.B.U.S.T. costfunction (2.3), mean and variance of the treatment outcome, were obtained via a crude Monte Carlo approach. However, classical Monte Carlo has many limitations and drawbacks. Firstly, in a realistic scenario, model evaluation runs are at a premium. Obtaining a large sample size, required for accurate computation of mean and variance, is prohibitive. Hence, further measures should be taken to utilize the available samples as efficiently as possible. Secondly, as [O’H87] argues, MC uses only the observations of the output response quantity $[y_1, \dots, y_N]$, not the respective positions $[\mathbf{x}_1, \dots, \mathbf{x}_N]$. Hence, MC estimates are error prone to an uneven distribution of samples. Quasi Monte Carlo techniques, such as Latin Hypercube [MBC79], have been introduced to alleviate this problem. However, they are still based on point estimates, hence make no full use of the information at hand.

It was shown that by introducing GPs as fast surrogates, these problems can be addressed. Additionally, in case of Gaussian input uncertainty, BMC can be used instead of classical Monte Carlo. All operations after the GP emulator is built can be carried out in closed form, thus very efficiently and with no precision issues (2.27). Furthermore, the uncertainty in the GP itself is accounted for (2.25).

Now the details of the integration of augmented R.O.B.U.S.T. into an existing IMRT optimizer are elaborated. It was integrated into the HYPER-

ION treatment planning system, which has been developed at the University of Tübingen. It was fused into its constrained optimizer, allowing for a combination of classical and R.O.B.U.S.T. cost functions. The underlying uncertainty model was based on the work of [SBYA05, SSA09]. They model organ deformation based on the dominant eigenmodes of geometric variability obtained via principal component analysis (PCA). This method takes into account the correlated movement between organs, hence there are no more overlap regions. Since it manages to capture the significant aspects of a patient's movement only in terms of a few parameters, namely the dominant eigenmodes, and, alongside, delivers (Gaussian) probabilities for any value of a mode being realized, it is very well suited as a basis for the algorithm. By varying the model parameters along the eigenmodes, an arbitrary number of realistic geometry configurations can be generated, fig. 2.11. In essence, it serves as a statistical biomechanical model.

2.3.1 Constrained Optimization in Radiotherapy

In this section a generic approach to IMRT optimization is reviewed [AR07]. The method of constrained optimization to maximize dose to the target while maintaining reasonable organ at risk dose levels is employed. In this context, the k th constraint, which is a function of the dose distribution D , is denoted by $g_k(D)$. For the sake of simplicity, assume that there is only one objective, $f(D)$. All cost functions can be physical (maxDose, DVH) or biological (EUD, gEUD). The incident beams are discretized into so-called beamlets or bixels. Each one of those beamlets is assigned a weight ϕ_j to model the fluence modulation across the beam. The dose distribution is the product of the dose operator T and the beamlet weights ϕ . The element T_{ij} relates the dose deposited in voxel i by beamlet j , thus the dose in voxel i is $D_i = \sum_j T_{ij}\phi_j$. The entire optimization problem can now be stated as

$$\text{find } \phi \text{ such that } f(T\phi) \mapsto \min \quad (2.28a)$$

$$\text{while } g_k(T\phi) \leq c_k \quad (2.28b)$$

$$\text{and } \phi_j \geq 0 \quad \forall j, \quad (2.28c)$$

where c_k is the upper limit of constraint k . It is evident that the optimum depends upon \mathbf{c} . Notice that in many cases the number of beamlets may exceed 10^4 , resulting in a very high dimensional optimization problem.

2.3.2 Core Algorithm

In this section, R.O.B.U.S.T. [SMA10] is augmented with the previously introduced techniques. Since the cost function, (2.3), is based on mean and

variance of the outcome metric, it is crucial to obtain both quantities as efficiently as possible. That is, to make the approach computationally feasible, the number of recomputations of potential courses of treatment, N , has to be as low as possible. This is done by making further assumptions about the behavior of the patient model. Namely, the functional dependency between treatment parameter and outcome is exploited.

While a complete patient model substitution by a GP is appealing because it would allow instant reoptimizations, say, under altered conditions, it is infeasible due to the extremely high dimensional nature of the IMRT problem. Instead, a hybrid approach is proposed, where in each optimization step only the behavior of the model with respect to the uncertain parameters is learned. This way, the number of beamlets is decoupled from the uncertainty handling.

The initial step is to sample several values $[\mathbf{x}_1, \dots, \mathbf{x}_N]$ from the support of X . Notice that the actual shape of the distribution $p(\mathbf{x})$ enters the analysis analytically in (2.24) and (2.25), hence it is sufficient to cover the support of $p(\mathbf{x})$ uniformly. In order to improve the filling of the input space, the implementation uses a Latin Hypercube scheme [MBC79].

In each iteration, that is, for a given set of fluence weights, ϕ , a dose distribution D is calculated. Now that D is available, the dose metric is evaluated in each training scenario, $y_i = f(\mathbf{x}_i)$. Subsequently, the GP is trained on the dataset $\mathcal{S} = \{\mathbf{x}_i, y_i\}$ that was generated in the previous step. Mean and variance of Y are calculated using (2.24) and (2.25) respectively. Equation (2.3) can now readily be calculated.

The fit can be improved by introducing a classical regression model prior to the GP fit. In practice this means that a regression model is fitted to the data and then a GP is used to explain the residuals. By using the GP only to represent the variation of the residuals the GP is allowed to have increased smoothness, so that even fewer training runs are required [O’H06]. In many cases, a simple linear offset already suffices. Notice, in fig. 2.5 the EUD exhibits linear tendencies in many cases. In the following, the non-linear GP, $\mu(\mathbf{x})$, is combined with a linear part, $m(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$. Mean and variance are extended to

$$\mathbb{E}_{\mathbf{x}}[\mu(X) + m(X)] = \mathbb{E}_{\mathbf{x}}[\mu(X)] + \mathbb{E}_{\mathbf{x}}[m(X)] \quad (2.29)$$

$$\begin{aligned} \text{var}_{\mathbf{x}}[\mu(X) + m(X)] &= \text{var}_{\mathbf{x}}[\mu(X)] + \text{var}_{\mathbf{x}}[m(X)] \\ &\quad + 2 \text{cov}_{\mathbf{x}}[\mu(X), m(X)]. \end{aligned} \quad (2.30)$$

For details regarding the computation of (2.29) and (2.30), please refer to the appendix A.2.

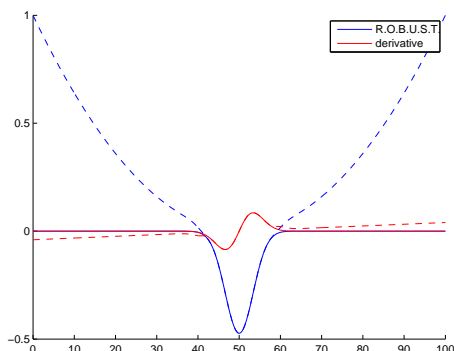


Figure 2.12: Unaltered (solid line) and augmented (dashed line) R.O.B.U.S.T. costfunctions along with their respective derivatives. The plots shows the value of (2.3) as a function of μ with $\sigma^2 = 10$ and $\mathbf{I}_{\text{tol}} = [48, 52]$. Notice that the unaltered R.O.B.U.S.T. gradient is practically zero as μ moves away from \mathbf{I}_{tol} , rendering such points infeasible as starting points for the IMRT optimization.

2.3.3 Convergence Considerations

Regarding the technical aspects of IMRT optimization, there is one major problem associated with the R.O.B.U.S.T. cost function, (2.3). If the outcome distribution is far away from the prescription interval, the cost function and its gradient. If such a point is chosen as a starting point, the optimization will not succeed, fig. 2.12.

There are two principal ways to address this problem. Firstly, is the use of a classic cost function, e.g. EUD, to optimize the dose distribution to a point where the probability mass inside the interval, (2.3) is non-zero and then launch R.O.B.U.S.T. from there. The second possibility is to augment (2.3) to have a non-zero gradient everywhere. In this work, the latter possibility was chosen.

The idea is to introduce an extra term to (2.3) that vanishes as soon as the enclosed probability mass turns non-zero. The heuristic used during this work is

$$\exp(-\eta \cdot M[Y]) \frac{(E_x[Y] - \mu_{\mathbf{I}})^2}{\mu_{\mathbf{I}}^2} - M[Y] \quad (2.31)$$

where $\mu_{\mathbf{I}}$ is the mean of the interval \mathbf{I} . The first term, $\exp(-\eta \cdot M[Y])$, is designed to vanish if the actual cost function $M[Y]$ becomes non-zero. The fading speed is determined by parameter η . For values far from the desired interval \mathbf{I} , the augmented cost function effectively minimizes the square dis-

tance from the expected outcome to the interval mean, $(\mathbb{E}_{\mathbf{x}}[Y] - \mu_{\mathbf{I}})^2$.

2.3.4 Efficient Computation of the Derivative

The efficiency of the optimization can be greatly improved if gradients are available. Again, since the cost function (2.3) consists of mean and variance, the gradients of the latter quantities are of interest.

$$\frac{\partial \mathbb{E}_{\mathbf{x}}[f(\phi, X)]}{\partial \phi} = \mathbb{E}_{\mathbf{x}} \left[\frac{\partial f(\phi, X)}{\partial \phi} \right] \quad (2.32a)$$

$$\frac{\partial \text{var}_{\mathbf{x}}[f(\phi, X)]}{\partial \phi} = \frac{\partial \mathbb{E}_{\mathbf{x}}[f^2(\phi, X)]}{\partial \phi} \quad (2.32b)$$

$$\begin{aligned} & - \frac{\partial \mathbb{E}_{\mathbf{x}}^2[f(\phi, X)]}{\partial \phi} \\ & = 2\mathbb{E}_{\mathbf{x}} \left[f(\phi, X) \frac{\partial f(\phi, X)}{\partial \phi} \right] \\ & \quad - 2\mathbb{E}_{\mathbf{x}}[f(\phi, X)] \mathbb{E}_{\mathbf{x}} \left[\frac{\partial f(\phi, X)}{\partial \phi} \right] \end{aligned} \quad (2.32c)$$

Because the gradients can be expressed in terms of expectation values, they can also be computed with BMC. Mind that each component of the gradient constitutes a dataset of its own that needs to be learned to conduct BMC. Notice also that optimizing the hyperparameters, i.e (2.22), involves the inversion of K , a $N \times N$ matrix, which is a $\mathcal{O}(N^3)$ process. Clearly, with the amount of beamlets for complicated cases being in the order of 10^4 , this becomes prohibitively expensive.

Nonetheless, the problem can still be handled by taking into account that neighboring beamlets in a beam will ‘see’ similar influences of fluctuations of uncertain parameters on the cost function gradient. In other words, adjacent beamlets will very likely have similar hyperparameters, i.e. with varying \mathbf{x} , $\frac{\partial f(\phi_i, \mathbf{x})}{\partial \phi_i}$ changes on similar length scales to $\frac{\partial f(\phi_j, \mathbf{x})}{\partial \phi_j}$ if beamlets i and j lie sufficiently close. In consequence, the *joint* likelihood of the datasets of several beamlets, $\prod_k p(\mathcal{S}_k, \boldsymbol{\omega})$ is optimized. In practice, this amounts to summing over the respective log marginal likelihoods, (2.22), for each beamlet.

$$\log \left[\prod_k p(\mathcal{S}_k, \boldsymbol{\omega}) \right] = \sum_k \log [p(\mathcal{S}_k, \boldsymbol{\omega})] \quad (2.33)$$

This is also known as *multi-task learning*.

The implementation currently groups the beamlets by beam. That is, for each beam a set of hyperparameters was learned during gradient computation. The rationale is that the impact of a given source of uncertainty has on

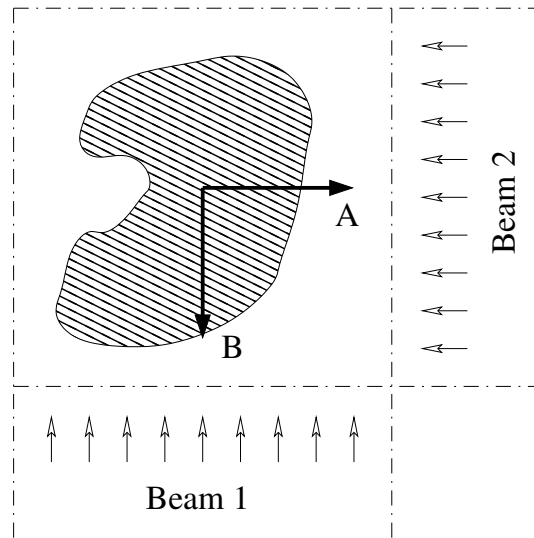


Figure 2.13: Suppose there are two principal directions of movement (A,B). While the impact of movement direction B is negligible for beam 1, it is quite significant for the second beam. The opposite is the case for direction A.

a beam may vary drastically with beam direction. The idea is illustrated in fig. 2.13. However, depending on the nature of the problem at hand, different partitioning schemes are conceivable.

Implementational Details of Multi-task Learning Gradient

The hyperparameters of a number of beamlet datasets are computed via gradient descent of (2.33). A naive implementation of the gradient of (2.33) via summation of the constituent derivatives is inefficient and slow. Because the optimization of (2.33) is performed frequently, an improved computation scheme for the gradient of (2.33) is presented. Starting from the gradient of log likelihood (2.22), [RW06, ch. 5],

$$\frac{\partial \log [p(\mathcal{S}, \boldsymbol{\omega})]}{\partial \omega_j} = \frac{1}{2} \text{tr} \left((\beta \beta^\top - K^{-1}) \frac{\partial K}{\partial \omega_j} \right) \quad (2.34)$$

the gradient of the multitask log likelihood (2.33) can be written as

$$\frac{\partial}{\partial \omega_j} \sum_k \log [p(\mathcal{S}_k, \boldsymbol{\omega})] = \sum_k \frac{1}{2} \text{tr} \left((\beta_k \beta_k^\top - K^{-1}) \frac{\partial K}{\partial \omega_j} \right) \quad (2.35)$$

$$= \frac{1}{2} \text{tr} \left(\left(\sum_k \beta_k \beta_k^\top - K^{-1} \right) \frac{\partial K}{\partial \omega_j} \right) \quad (2.36)$$

$$= \frac{1}{2} \text{tr} \left(\underbrace{(B_k B_k^\top - K^{-1})}_{\text{rank update}} \frac{\partial K}{\partial \omega_j} \right) \quad (2.37)$$

where B is a column matrix consisting of the individual β -vector for each of the beamlets. The term $(B_k B_k^\top - K^{-1})$ is identified with the symmetric rank update of matrix K^{-1} , an operation for which high-performance implementations, e.g. ATLAS [WPD01], exist.

Implementational Details of IMRT gradient computation

In the following, the attention is directed to the computation of $\frac{\partial f(\phi, X)}{\partial \phi}$, the derivative of the IMRT cost function.

Typically, IMRT cost functions define a mathematical measure of the dose. The measure of a set of voxels is the sum of its constituent subsets. In that sense, these subsets, labelled $\hat{f}(D_i)$ and $\hat{g}(D_i)$ respectively, are entirely local quantities, only depending upon the dose in a given voxel i . Hence, cost functions structurally resemble volume integrals. Notice that objectives as well as constraints can be arbitrary functions of the local dose [Alb08, RDL04].

In order to solve the IMRT optimization problem, (2.28), efficiently, the gradients of the objectives are needed. Applying the chainrule yields

$$\frac{\partial f}{\partial \phi_j} = \sum_i \frac{\partial \hat{f}}{\partial D_i} \frac{\partial D_i}{\partial \phi_j} \quad (2.38)$$

$$= \sum_i \frac{\partial \hat{f}}{\partial D_i} T_{ij}. \quad (2.39)$$

Notice that the summation in (2.39) is fairly expensive to compute because the index i runs over the entire voxel grid and needs to be repeated by the number of beamlets. Additionally, due to its sheer size, the dose operator remains in a compressed format in memory, requiring in-flight decompression. Since R.O.B.U.S.T. is carried out on a multitude of geometry instances, another penalty factor is added, making $\frac{\partial f(\phi, X)}{\partial \phi}$ very expensive to obtain. Again,

the problem is re-stated so it can be handled with existing high-performance linear algebra libraries, e.g. ATLAS [WPD01].

The key is to extract the voxels that are occupied by the hull of the single instances of an organ. This submatrix is denoted by \bar{T}_{ij} . Because typically the volume relevant for R.O.B.U.S.T., e.g. the target, is small compared to the dose cube, \bar{T}_{ij} is stored uncompressed. Notice that this needs to be done only once prior to optimization. The matrix product of the row matrix of variation grids and \bar{T}_{ij} , yields a column matrix containing the derivatives of cost function f with respect to ϕ_j .

2.4 Evaluation

The results were obtained using the first 3 eigenmodes, i.e. $p(\mathbf{x}) = \prod_i^3 \mathcal{N}(x_i|0, \sigma_{\mathbf{x}_i}^2)$ of the PCA model. Notice that the Gaussians are zero-mean because PCA returns the deviation from a mean geometry. The likelihood of a certain configuration is indicated by its standard deviation $\sigma_{\mathbf{x}}$.

The basis for this analysis were the CT datasets of four prostate cancer patients, consisting of ~ 15 images each. Included into the optimization were prostate, rectum and bladder and biological optimization [AN99] was employed. The tolerance interval, \mathbf{I}_{tol} , for the prostate was chosen to range from 74 Gy EUD to 76 Gy EUD based on the Poisson cell kill model. Bladder as well as rectum were assumed to be governed by a serial complication model and the limits were set such that no more than 10% may exceed 57.5 Gy EUD and 65.5 Gy EUD respectively. The optimizer used 3mm voxels and, depending on patient, had about 5000-8000 2×4 mm beamlets. These were arranged in eight beams. The basis for the GP training were 70 geometries sampled prior to the optimization. In the present study the dose distribution is assumed to be invariant under organ deformations [BYA⁺03, WvdGS⁺07]. In practice, an optimization run took between and 20 and 30 minutes on a quad core workstation.

The investigation was twofold. Firstly, for each patient, N CTs were drawn at random from the image pool and the (PCA) patient model based on these CTs was used for optimization. The procedure was run 20 times. Secondly, it was investigated whether the treatment outcome distributions, as predicted by the optimizer, coincide with the outcome distributions that are obtained if the model of motion is built using all available imagery, which served as a gold standard. Both procedures were repeated for varying N .

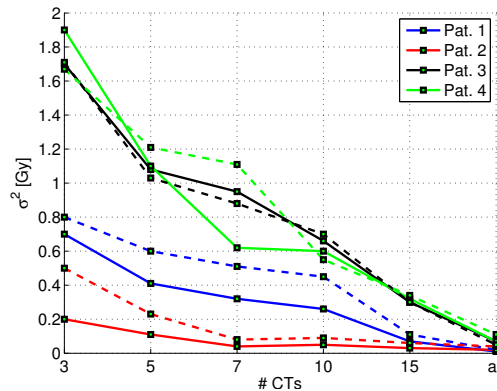


Figure 2.14: Std. dev. of predictive mean (solid) and predictive std. dev. (dashed) using different CTs each trial. (Each point represents 20 samples.)

Dependence of outcome variability on CT choice

In this section it is investigated how the projected treatment outcome is affected by the choice of CTs initially included into the optimization. To ascertain the credibility of the predictions, the spread of the relevant quantities, (2.24), (2.25) was investigated. In fig. 2.14 the standard deviation of both quantities with respect to the number of CTs are shown. In both cases, the standard deviation rapidly declines as more CTs are included. Starting from about 5 CTs, both standard deviations are around or below one gray.

Comparison with gold standard

A similar picture becomes apparent if the deviation from the gold standard is considered. The average difference from the gold standard is plotted in fig. 2.15(a). Again mean and standard deviation show similar tendencies. A closer look at fig. 2.15(a) as well as fig. 2.14 reveals that it is always the same patients who require more imagery than others. Naturally, the question arises for the possibility to identify these *before* a large amount of images is available. The key insight here is that the need for additional information comes from the increased movement of individual patients, i.e. large movement patients require more imagery. For similar prescriptions, the predictive standard deviation serves as a measure for a patient's geometric variability, i.e. how much the projected outcome is spread by the patient's movement. Fig. 2.15(b) shows the predictive standard deviations. It is possible to identify large movement patients starting from very few CTs. This way, patients which require more images can be identified from a very early stage.

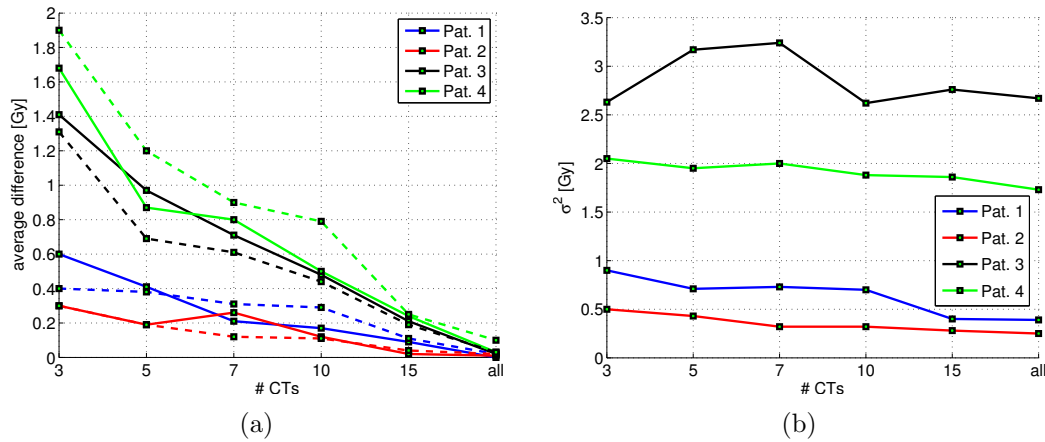


Figure 2.15: (a) Average differences of predictive mean (solid) and predictive std. dev. (dashed lines) to the all image gold standard model. (Each point represents 20 samples.) (b) Predictive standard deviations. Notice the elevated standard deviations for the large movement patients. (green, black)

Interpretation of the results

Exemplary optimization results of two patients are shown in fig. 2.16. The target EUD range for the prostate was set to 74 Gy - 76 Gy. The prescriptions for the organs at risk, i.e. bladder and rectum, were that no more than 10% of the outcome distribution may exceed 57.5 Gy and 65.5 Gy, respectively.

The treatment outcome distribution for the prostate of a large amplitude movement patient, fig. 2.16(a), is not entirely situated in the desirable dose range. This is because at least one, in case of fig. 2.16(a) both constraints are exhausted. The reason is a fairly large overlap region of the occupancy clouds of prostate and rectum in fig. 2.17(a). More specifically, this forces the optimizer to contain high doses to regions of very high probability of prostate occupancy in order to keep the radiation exposure of the rectum within predefined limits. In other words, the resulting dose distribution in fig. 2.17(a) pays tribute to the serial nature of the rectum by ensuring that the probability for high dose exposure is sufficiently low, while at the same time allowing lower doses to extend into regions of higher probability of rectum occupancy. In contrast, the high dose region in fig. 2.17(b) even encompasses regions of a low probability of prostate occupancy, especially towards the rectum. The resulting treatment outcome distribution of the prostate is fairly narrow and sharply peaked, fig. 2.16(b). Notice that the optimizer irradiates regions with very low prostate occupancy with high doses because the rectum constraint is still slack in fig. 2.16(b).

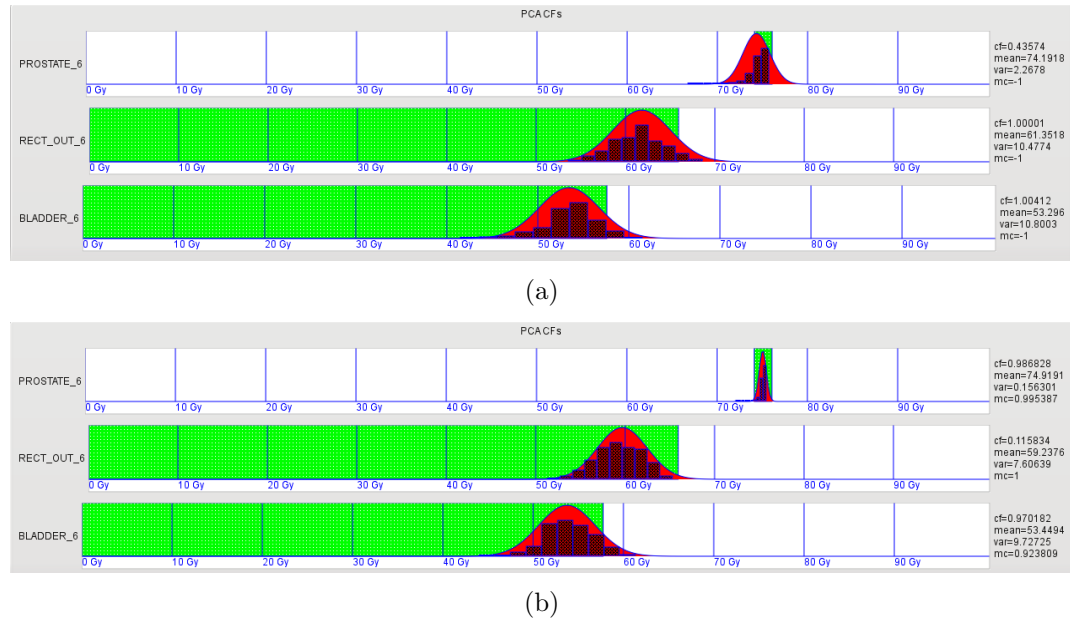
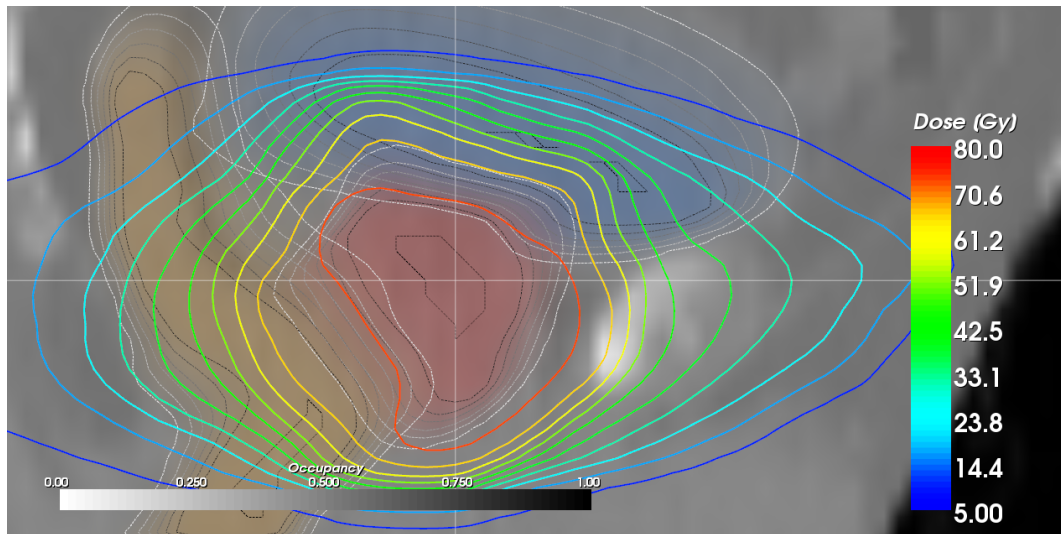
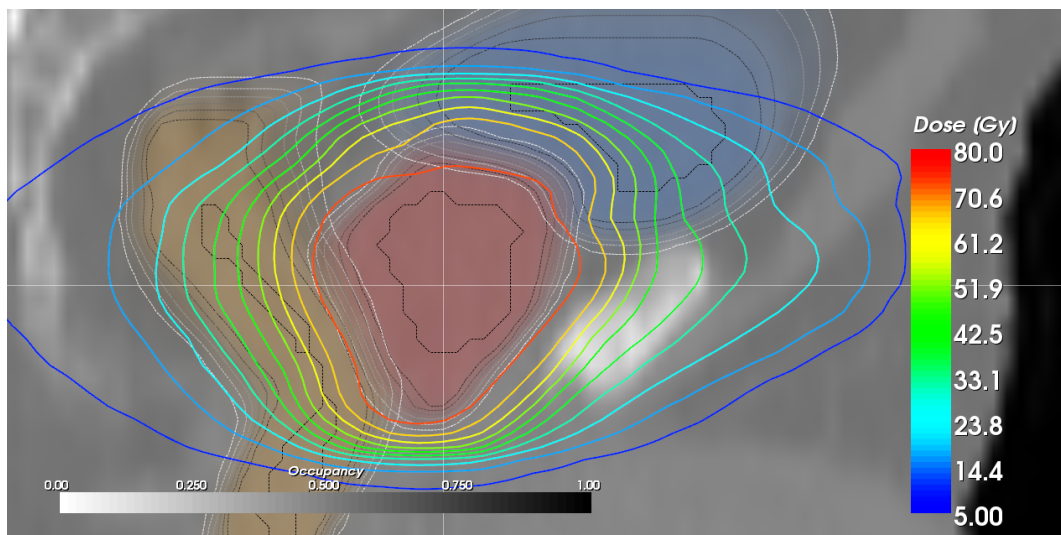


Figure 2.16: Partial screenshots of the R.O.B.U.S.T. front-end. There are three costfunctions that were handled with R.O.B.U.S.T. - tumor control in prostate, and serial complication in rectum and bladder, respectively. Tolerable EUD ranges are marked green. The Gaussian approximations, as seen by the optimizer, are illustrated in red. The histograms in the foreground are the treatment outcomes of the single geometry instances handled by the R.O.B.U.S.T. engine. Mind that these histograms are only for illustrational purposes, the optimizer is solely based on the Gaussian approximations. Shown are the final optimization results for (a) a large amplitude movement patient and (b) a patient with small amplitude movement. Notice that the rectum constraint in (b) is not reached.



(a)



(b)

Figure 2.17: Sagittal images of prostate (red), bladder (blue), and rectum (yellow) of a (a) large (b) small amplitude movement patient. The stippled lines indicate boundaries of organ occupancy levels. Darker tones indicate a higher probability of the volume of interest occupying a certain region. Notice that the overlap of prostate and rectum is significantly smaller in (b). Consequently, as indicated by the isodose lines, the optimizer is able to irradiate regions with a lower likelihood of prostate occupancy with full dose. The corresponding treatment outcome distributions are in fig. 2.16.

This case illustrates the difference between coverage probability and R.O.-B.U.S.T. While coverage probability introduces spatial weighting solely based on organ occupancy, R.O.B.U.S.T. maintains the spatial correlation between volume elements. In other words, contrary to coverage probability, R.O.B.-U.S.T. assesses the individual outcomes under the influence of uncertainty along with their respective probabilities. Hence, it can detect extraordinarily hazardous scenarios (e.g. a sizeable, compact, part of an organ moves into a high dose region) and adapt accordingly or, conversely, escalate the dose to the target as long as the effect to the organs at risk in conjunction with the probability of being realized is justifiable.

2.5 Discussion and Outlook

In the first part of this chapter, a new approach towards robust treatment plan optimization was presented. Accepting the fact that all outcome metrics are random variables in the presence of geometry uncertainties, it addresses robustness in the sense that under all possible uncertainties, the treatment outcome quality metrics should remain within a specified range across all treatment fractions. It focuses on controlling the position and width of the outcome distribution instead of a nominal value of the metric. By quantifying the uncertainty of the treatment outcome for each metric, it provides a measure for the robustness of the plan.

If the optimization problem is stated like (2.4), precise knowledge of the outcome distribution is not necessary and the Gaussian approximation is possible by virtue of the Chebyshev inequality. This greatly reduces the effort of uncertainty sampling.

Because the method maintains different organ geometry setups, it is able to preserve correlations in geometric shifts. This means that there are no more ambiguous prescriptions. However, due to the fact that the method is offline, there may still be ambiguous dose volumes. Since it also explicitly takes the shape of the uncertainty distributions of the treatment parameters into account, it can detect a potentially hazardous scenario and assess its likelihood. This way, the plan will neither be overly conservative, nor lenient towards dangerous organ doses.

The robustness regarding the treatment outcome was investigated in terms of EUD. The method relies on sampling the distribution of each source of uncertainty. R.O.B.U.S.T. is very versatile because it can handle all sources of error and is only limited by the available patient uncertainty model. This is currently a very active field of research. In a number of works, e.g. [vSSLC07], new models of motion only depending upon very few

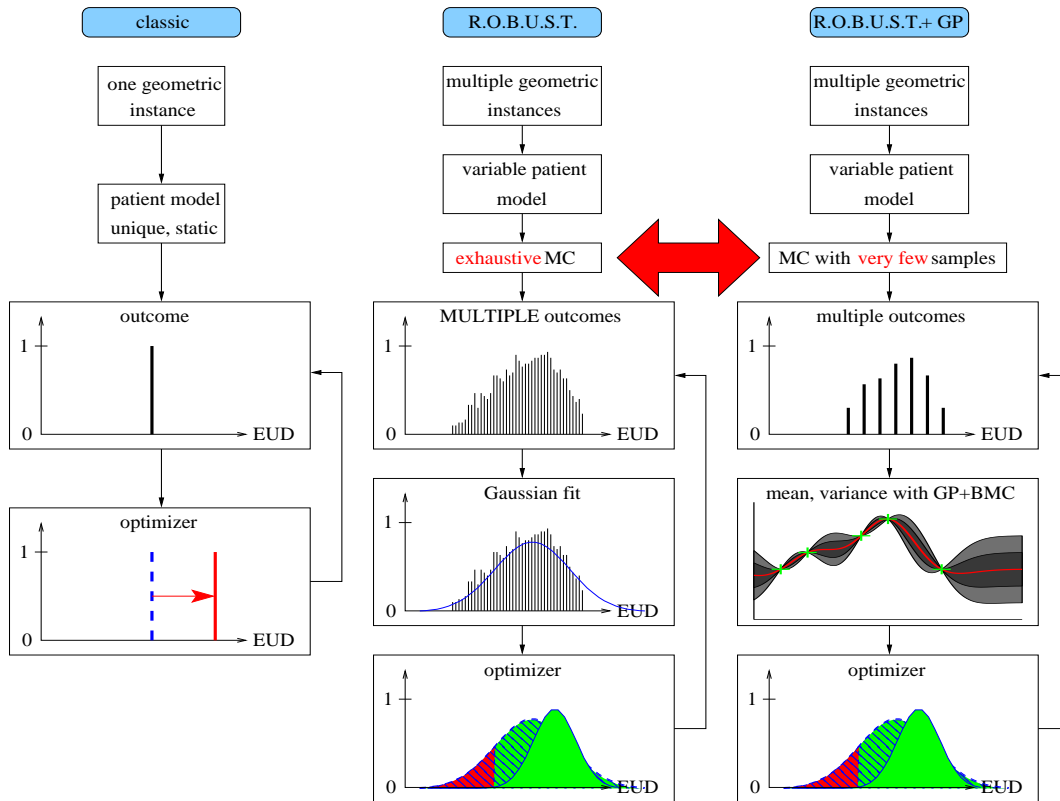


Figure 2.18: Side-by-side comparison of classical IMRT optimization, R.O.B.U.S.T. and R.O.B.U.S.T. with Gaussian Processes. The key to R.O.B.U.S.T. is the optimization of the outcome distribution in place of the nominal value. Gaussian Processes are introduced to obtain a precise estimate of mean and variance with only few samples. This leads to a greatly reduced computational effort.

parameters have been proposed. Each of these may leverage R.O.B.U.S.T.

In the second part of the chapter, an algorithm to facilitate the computation of mean and variance for any dose metric under arbitrary influence of uncertainty was introduced. It was described how to augment the previously presented probabilistic optimizer.

The method exploits the fact that practically all dose quality metrics undergo only minor changes if the fluctuations in the uncertain inputs are sufficiently small, i.e. the metrics are smooth, continuous functions of the inputs. Gaussian processes are used to mimic those functions and finally compute the essential quantities, i.e. mean and variance, analytically. Errors introduced by the Gaussian process fit are accounted for. Any source of uncertainty quantifiable in terms of probability distributions can be seamlessly included.

Side by side comparisons of the classic and augmented version showed that Gaussian process regression allows for a significant reduction of samples to be considered during treatment plan optimization, fig. 2.10.

The augmented R.O.B.U.S.T. algorithm in conjunction with PCA was implemented into the in-house planning system HYPERION. Investigations based on four prostate cancer patients showed that R.O.B.U.S.T. delivers dependable results with realistic computational effort. Large movement patients, requiring more imagery, are identified at a very early stage during treatment. This way the method leaves enough room for further measures to be taken, if necessary. The predicted outcome distributions can be updated in light of additional data acquired during the course of treatment. This allows for adaptive radiotherapy schemes.

Because R.O.B.U.S.T. is applicable to any kind of uncertainty, future work will include the evaluation of its performance in other fields and applications, e.g. the realm of particle therapy.

Chapter 3

Uncertainties in Dose Volume Histograms

The previous chapter dealt with uncertainty management at the planning stage, i.e. how to take uncertainties into account in the IMRT optimization process. Given a treatment plan, the complementary need arises to quantitatively evaluate its merits in the presence of uncertainties.

In that context, it is proposed to augment the well established Dose Volume Histograms (DVHs) with additional information about the impact of potential error sources.

Approaches pursuing similar goals have been presented [CvHMB02, MUO06, HC08, NHBT09]. The authors of [HC08] address DVH uncertainty on a per voxel basis. They propose to compute the DVH based on the average dose received by a voxel in the volume of interest. This entails that the uncertainty distribution for each voxel needs to be known. DVHs based on population data, (DVPH) have been introduced by [NHBT09]. They describe a method to incorporate the distribution of systematic and random errors into the DVH where the magnitude of the geometrical errors is calculated based on the population distribution. The basis of their algorithm is to simulate numerous treatment outcomes. Despite measures to re-use computed samples, their algorithm remains computationally expensive.

The approach presented in this work uses Gaussian process regression to detach the uncertainty analysis from the patient model. The two advantages that come with this approach are speed and the ability to interpolate DVHs *directly*, i.e. without the need to re-compute patient model parameters, e.g. rigid shifts or deformations for each previously unseen parameter set. The concept of interpolation of DVHs has been validated in [CvHMB02]. Their method of choice is B-spline interpolation. However, as the authors point out, while B-spline interpolation is applicable in higher dimensions, it scales

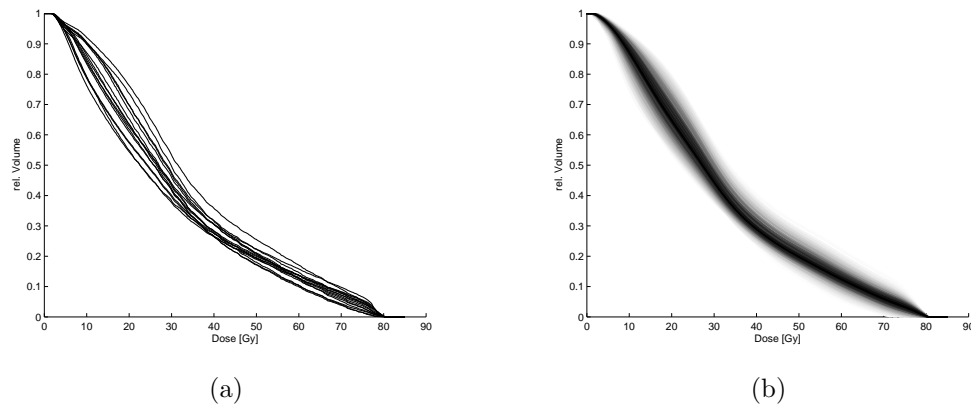


Figure 3.1: Rectum DVHs for a single prostate cancer patient under the influence of uncertainty. (a) Each DVH corresponds to one of 15CTs taken during the course of treatment. (b) Interpolated DVHs computed with a Gaussian process based on the 15 CTs obtained during treatment. Because the GP generates a DVH for any uncertainty realization, a continuous spectrum of DVH may be computed. A darker tone indicates a more likely scenario. Notice that, especially for doses around 40 Gy, the most likely scenario does *not* coincide with the mean.

very poorly in terms of memory consumption. Furthermore, it yields no statistical information.

In the present work, DVHs are extended in three ways. Firstly, the DVHs can be annotated by additional measures, such as biological effect (EUD). For example, the DVHs of an organ at risk that correspond to a certain EUD in the target may be extracted. Secondly, by propagating the assumed or measured error distribution into the DVH, error bars may be supplied. Because the algorithm establishes a mapping from input uncertainty to DVH error bars, it is also possible to screen the influence of each input separately. This substantial information gain facilitates the inevitable trade-off between organ at risk sparing and tumor control. And thirdly, it is also shown that a sensitivity analysis can be implemented very efficiently with the tools developed in the previous two points.

3.1 Discrimination of Dose Volume Histograms based on their respective Biological Effects

Since each uncertain parameter may take a different value at treatment time, a multitude of scenarios may be realized. Fig. 3.1(a) shows several DVHs that arise for different treatment scenarios. It is evident from the plot that single DVHs suffer from limited expressiveness. In order to break the one image (CT) \Leftrightarrow one DVH constraint, i.e. move from a few to a host of DVHs regardless of the available imagery, it is necessary to interpolate DVHs across the uncertainty space. The goal of this section is to derive a method that provides an overview over the treatment plan merits as a whole. That is, instead of only emphasizing the *single* patient geometry instances, *all* possible DVHs along with their likelihood and biological effect are presented.

Mathematically, DVHs can be understood as a function of dose, $h : \mathbb{R}^+ \mapsto [0, 1]$, i.e. for any given dose d , $h(d)$ returns the percentage of the organ receiving d Gy or less. Suppose the treatment is subject to random influences reflected by the S dimensional random variable X , e.g. displacements from a mean position. Since the DVH reflects the irradiation of a volume of interest, it is also influenced by the random variable, $h(d, X)$. This is shown in fig. 3.1(a) for 15 distinct samples (CTs) of X . To overcome the limit of only being able to display DVHs of previously recorded CTs, Gaussian process regression is employed to interpolate between the patient geometry samples.

Thus, the goal is to learn the behavior of the DVH as a function of dose and the uncertain parameters, $h(d, X)$, i.e. emulate a $\mathbb{R}^{S+1} \mapsto [0, 1]$ function with a Gaussian process. The distinct benefit is that this enables interpolation in uncertainty space. The GP training data are randomly chosen points from the DVHs depicted in fig. 3.1(a). The result is shown in fig. 3.1(b), where GP interpolated DVHs are plotted for several realizations of the uncertain parameters X . Notice the transition from discretely spaced DVHs in fig. 3.1(a) to a continuous representation in fig. 3.1(b). The DVHs are shaded based on their likelihood of being realized, i.e. $p(\mathbf{x})$.

In order to assess the merits of a treatment plan, it is useful to consider biological measures. That means, in addition to purely physical doses, the expected biological effect is taken into account. Typically, the function that maps a physical dose to a biological effect is highly organ specific because organs react differently depending on the irradiation pattern. Using GPs, a multitude of DVHs can be generated and with added post-processing, additional information can be incorporated, fig. 3.2. One could also color the

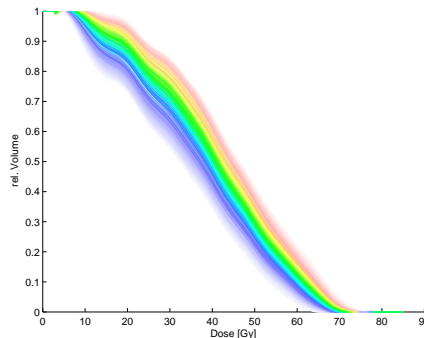


Figure 3.2: Bladder DVHs of a prostate cancer treatment plan. The hue encodes the biological effect (EUD). The likelihood is reflected by the saturation. Notice the broad green region of equal toxicity. The toxicity information in conjunction with the likelihood of a given scenario, provides an overview of the treatment plan merits.

DVHs of the organs at risk based on the EUD of the target to reveal cross-correlations.

GP regression of DVHs provides a qualitative overview of the risk for a certain organ that comes with a given treatment plan. A quantitative analysis of the risk is the subject of the second part of this chapter.

3.2 Quantitative Analysis

Recall that, given an S -dimensional uncertain input, the DVH is a function of dose d and S -dimensional random variable X , $h(d, X)$. In the previous section, a method to compute a host of DVHs, independent of the number of available imagery, was introduced. This was made possible by the use of Gaussian process regression to interpolate the DVHs at values \mathbf{x} , e.g. displacements, where no CT is available. If the probability distribution of the treatment uncertainties X is Gaussian, $p(\mathbf{x}) = \mathcal{N}(\mathbf{x})$, error bars indicating the probability of a deviation from the nominal value can be efficiently computed. Next, the mathematical foundations for the DVH error bar computation are laid out.

In essence, the DVH error bars are independently computed at each dose d_0 . DVHs at a given dose, $h(d = d_0)$, can be interpreted as distribution of a treatment metric, i.e. the statistical methods that were introduced in the previous chapter can be re-applied. Notice, however, that in the previous chapter the integral in (2.24) and (2.25) ran over all inputs of the

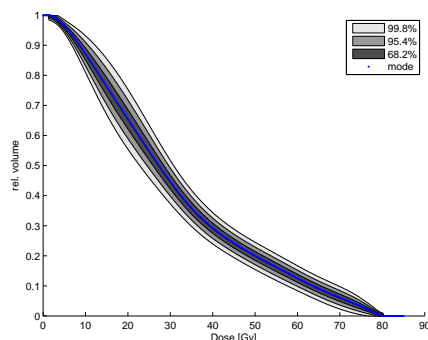


Figure 3.3: Rectum DVH with confidence levels. At every dose, a Gaussian is computed via BMC. The quantiles are marked in the plot.

GP, whereas now the dose is *not* integrated out, but held constant at $d = d_0$. Within the framework that was presented in the previous chapter, this can be conveniently done by setting

$$p(d) = \delta(d - d_0) = \mathcal{N}(d|d_0, a) \quad \text{with } a \rightarrow 0. \quad (3.1)$$

Thus, the mean and variance of the DVHs at dose d_0 are given by

$$\mathbb{E}_{\mathbf{x}}[h(d_0, X)] = \int d\mathbf{x} p(\mathbf{x}) h(d_0, \mathbf{x}) \quad \text{and} \quad (3.2)$$

$$\text{var}_{\mathbf{x}}[h(d_0, X)] = \int d\mathbf{x} p(\mathbf{x}) h^2(d_0, \mathbf{x}) - \mathbb{E}_{\mathbf{x}}^2[h(d_0, X)] \quad (3.3)$$

respectively. Together they provide an indicator of the expected DVH of the treatment along with an estimate of possible deviations. Again, (3.2) and (3.3) can be solved analytically if $p(\mathbf{x}) = \mathcal{N}(\mathbf{x})$.

With the methods at hand, the first step to compute DVH errors bars is to learn the behavior of the DVH as a function of dose and the uncertain parameters, $h(d, X)$. The basis of the GP training remains identical to the previous section, i.e. several samples are taken at random from the DVHs of each CT, fig. 3.1(a). The resulting GP is shown in fig. 3.1(b). Once the GP is trained, mean (3.2) and variance (3.3) are computed using *Bayesian Monte Carlo* at every dose. With mean and variance at hand, the maximum entropy principle suggests to assume a Gaussian at each dose. Given this normality assumption, quantiles can readily be computed. Using the GP shown in fig. 3.1(b) to compute mean and variance via BMC yields fig. 3.3.

To this point, it was presented how Gaussian process regression can be used to augment DVHs. The idea is to understand DVHs as functions of dose

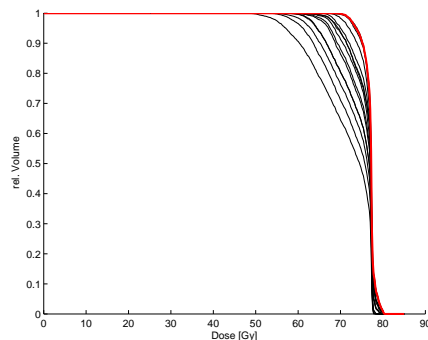


Figure 3.4: Prostate DVHs for 15 CTs taken during treatment with the DVH of the planning geometry in red.

and uncertainty parameters and that as such, they can be learned and substituted by GPs. It was shown that the substitution has two major advantages. Firstly, while classical DVHs are CT bound, GPs can estimate a DVH for previously unseen patient geometries, i.e. uncertainty realizations. Hence, the uncertainty space may be traversed freely resulting in a family of DVHs. Secondly, using the trained GP, quantile information may be obtained by integrating out the uncertainty parameters. The required computations can be carried out analytically via *Bayesian Monte Carlo*, i.e. in place of numerical integration methods, if the input uncertainty distribution are Gaussian.

While statistical information may already be extracted from fig. 3.3, it still somewhat lacks a basic property that the plot fig. 3.1(b) exhibits, specifically asymmetry. The normality assumption entails that the mean coincides with the mode and that uncertainty intervals are equal on both sides of the mean. In the next section, BMC will be extended to quantitatively predict the level of asymmetry.

3.3 Accounting for Asymmetry

Perturbations of the DVH typically do not lead to a symmetrical distribution around the mean as implied by the normality assumption. Consider, for example, the fact that a DVH that is at unity for a given dose can only decrease under perturbations, i.e. not all of the organ volume is contained entirely in the high dose region anymore. The DVH plotted red in fig. 3.4 is the DVH of the planning geometry. Notice that between 55 Gy and 65 Gy, the DVH moves toward lower volumes only. Consequently, the ensuing asymmetry should be quantitatively taken into account. Thus, the grade of

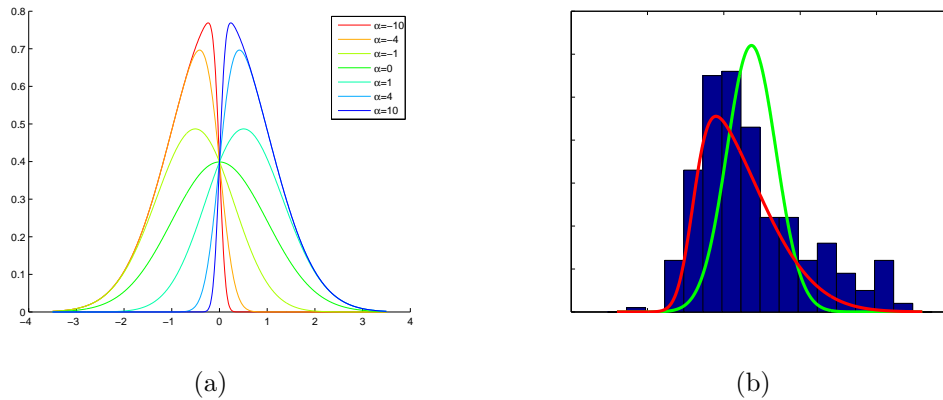


Figure 3.5: (a) Probability density functions of the skew normal distribution for difference choices of the shape parameter α . (b) The histogram of the DVHs in fig. 3.1(b) at 42Gy. Green is the normal approximation based on mean and variance. To take the asymmetry into account the skew normal distribution is used (red).

the error bounds is greatly improved by including the skewness,

$$\text{skew}_{\mathbf{x}}[h(d, X)] = \frac{\mathbf{E}_{\mathbf{x}}[(h(d, X) - \mathbf{E}_{\mathbf{x}}[h(d, X)])^3]}{\sqrt{\text{var}_{\mathbf{x}}^3[h(d, X)]}} \quad (3.4)$$

in addition to mean, (3.2), and variance, (3.3). Typically, the computation of error bounds is based on the Gaussian distribution. However, the Gaussian distribution does not allow for the inclusion of skewness. Hence, a distribution is needed that is based on the first three moments to compute error bars that reflect the asymmetry in the underlying data. In the upcoming paragraphs, a distribution is presented that can be parameterized in terms of mean, variance *and* skewness.

3.3.1 Skew Normal Distribution

Given mean and variance, error bars are typically based on the quantiles of the normal distribution. The inclusion of the skewness, however, demands a distribution that includes mean, variance as well as skewness directly or indirectly as parameters. The extension of the normal distribution to non-zero skewness has been presented by the authors of [Azz85]. With the density function of the normal distribution, $\phi(x)$, and its cumulative distribution function, $\Phi(x)$, the skew normal distribution density can be written as

$$f(x) = 2\phi(x)\Phi(\alpha x) \quad (3.5)$$

where α is the shape parameter. Fig. 3.5(a) shows the density, (3.5), for several α . For $\alpha = 0$ the normal distribution is recovered, for $\alpha > 0$ the distribution is right skewed and for $\alpha < 0$ the it is left skewed. Via the transformation $x \mapsto \frac{x-\xi}{\omega}$ location ξ and scale ω parameters are added. The transformed density now reads

$$f(x) = \frac{2}{\omega} \cdot \phi\left(\frac{x-\xi}{\omega}\right) \cdot \Phi\left(\alpha \frac{x-\xi}{\omega}\right). \quad (3.6)$$

The values of ξ , ω and α are estimated via the method of moments from mean μ , variance σ^2 and skewness ν . Given the sample skewness ν , the shape parameter α is computed according to

$$|\theta| = \sqrt{\frac{\pi}{2} \frac{|\nu|^{\frac{2}{3}}}{|\nu|^{\frac{2}{3}} + \left(\frac{4-\pi}{2}\right)^{\frac{2}{3}}}}. \quad (3.7a)$$

where $\theta \equiv \frac{\alpha}{1+\alpha}$. With θ and the sample variance σ^2 , the scale parameter ω is

$$\omega = \sqrt{\frac{\sigma^2}{1 - \frac{2\theta^2}{\pi}}}. \quad (3.7b)$$

and similarly the location ξ

$$\xi = \mu - \omega\theta\sqrt{\frac{2}{\pi}}. \quad (3.7c)$$

These equations are obtained by inverting the relations for mean, variance and skewness of the skew normal distribution given in appendix B.2. Fig. 3.5(b) depicts the histogram of the DVHs in fig. 3.1(b) at 42Gy. The Gaussian (green) was fitted to the histogram based on mean and variance, whereas the skewed Gaussian (red) was fitted based mean, variance and skewness of the data.

3.3.2 Extension of Bayesian Monte Carlo to compute the Skewness in Addition to Mean and Variance

In order to compute the skewness of the DVH distribution in each dose point, BMC needs to be extended. Previously, details on the computation of the mean $E_{f|S} [E_{\mathbf{x}}[f(X)]]$ and variance $E_{f|S} [\text{var}_{\mathbf{x}}[f(X)]]$ were given in (2.24) and (2.25), respectively. Similarly, a solution for the skewness will now be derived.

However, instead of computing $\mathbb{E}_{f|S}[\text{skew}_{\mathbf{x}}[f(X)]]$, the latter is *approximated* with a solution for

$$\begin{aligned} \text{skew}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]] &\equiv \mathbb{E}_{\mathbf{x}} \left[\left(\frac{\mathbb{E}_{f|S}[f] - \mathbb{E}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]]}{\sqrt{\text{var}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]]}} \right)^3 \right] \\ &= \frac{\mathbb{E}_{\mathbf{x}}[\mathbb{E}_{f|S}^3[f]] - 3\mathbb{E}_{\mathbf{x}}[\mathbb{E}_{f|S}[f] \text{var}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]]] - \mathbb{E}_{\mathbf{x}}^3[\mathbb{E}_{f|S}[f]]}{\sqrt{\text{var}_{\mathbf{x}}^3[\mathbb{E}_{f|S}[f]]}}, \end{aligned} \quad (3.8)$$

where $\mathbb{E}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]]$ stems from (2.24) and $\text{var}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]]$ from (2.26a). This corresponds to neglecting the GP predictive uncertainty, $\text{var}_{f|S}[f]$ (2.17c). To see this, bear in mind that $\text{skew}_{\mathbf{x}}[\mathbb{E}_{f|S}[f]]$ computes the skewness solely based on the mean of the GP, $\mathbb{E}_{f|S}[f]$ (2.17b). Hence, for GPs with a low predictive uncertainty, the approximation is reasonable.

The only new term in (3.8) that needs to be computed is $\mathbb{E}_{\mathbf{x}}[\mathbb{E}_{f|S}^3[f]]$. With (2.18) this can be rewritten as $\mathbb{E}_{\mathbf{x}}[(k(\bar{\mathbf{x}}, X)^{\top} \cdot \boldsymbol{\beta})^3]$. In the computation of the average, only the kernel depends on \mathbf{x} ,

$$\begin{aligned} \mathbb{E}_{\mathbf{x}}[(k(\bar{\mathbf{x}}, X)^{\top} \cdot \boldsymbol{\beta})^3] &= \int d\mathbf{x} p(\mathbf{x}) (k(\bar{\mathbf{x}}, \mathbf{x})^{\top} \cdot \boldsymbol{\beta})^3 \\ &= \sum_i \sum_j \sum_k \beta_i \beta_j \beta_k \int d\mathbf{x} p(\mathbf{x}) k(\bar{\mathbf{x}}, \mathbf{x}_i) k(\bar{\mathbf{x}}, \mathbf{x}_j) k(\bar{\mathbf{x}}, \mathbf{x}_k) \\ &\equiv \sum_i \sum_j \sum_k \beta_i \beta_j \beta_k \Gamma_{ijk} \end{aligned} \quad (3.9)$$

rendering Γ the only quantity of interest. Under the condition of Gaussian input uncertainty $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$ as well as the usage of the squared exponential kernel (2.14), there exists a closed form solution

$$\begin{aligned} \Gamma_{ijk} &= \omega_0^6 |3\Omega^{-1}\Sigma_{\mathbf{x}} + \mathbb{1}|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^{\top} (2\Omega)^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right) \\ &\quad \exp\left(-\frac{1}{2}(\mathbf{x}_{ij} - \mathbf{x}_k)^{\top} \left(\frac{3}{2}\Omega\right)^{-1} (\mathbf{x}_{ij} - \mathbf{x}_k)\right) \\ &\quad \exp\left(-\frac{1}{2}(\mathbf{x}_{ijk} - \mu_{\mathbf{x}})^{\top} \left(\frac{1}{3}\Omega + \Sigma_{\mathbf{x}}\right)^{-1} (\mathbf{x}_{ijk} - \mu_{\mathbf{x}})\right) \end{aligned} \quad (3.10)$$

where $\mathbf{x}_{ij} \equiv \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j)$, $\mathbf{x}_{ijk} \equiv \frac{1}{3}(\mathbf{x}_i + \mathbf{x}_j + \mathbf{x}_k)$ and \mathbf{x}_l denotes the l -th training point. A detailed derivation of (3.10) is given in appendix A.1.

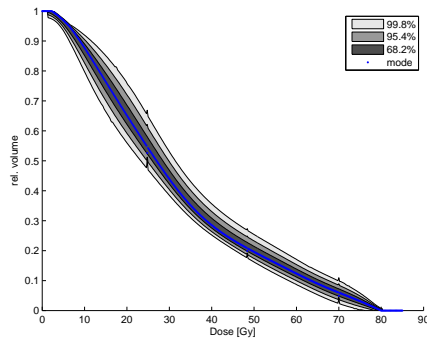


Figure 3.6: DVH with confidence levels. Due to the inclusion of the skewness, the confidence intervals on both sides of the mode are non-symmetric.

The skew normal distribution is used to compute mode and confidence intervals. This was done for the rectum DVHs in depicted in fig. 3.1(a). As opposed to the first analysis which was based only on mean and variance fig. 3.3 the result of the full analysis is shown in fig. 3.6. The asymmetry that is apparent in fig. 3.1(b) is now reproduced, greatly increasing the accuracy of the analysis.

Accounting for a linear offset in the BMC skewness calculation

It was argued in the previous chapter, that the overall fit can be improved if a linear offset is subtracted before the actual GP fit [O’H87]. In essence, this means that the GP is used to solely represent the non-linear parts of the behavior. Hence, with $m(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$, the GP $\mu(\mathbf{x})$, and (3.8), the goal is to compute the skewness of the sum of both,

$$\text{skew}_{\mathbf{x}}[m(X) + \mu(X)] = \frac{\mathbb{E}_{\mathbf{x}} [(m(X) + \mu(X))^3] - \mathbb{E}_{\mathbf{x}}^3 [m(X) + \mu(X)]}{\sqrt{\text{var}_{\mathbf{x}}^3 [\mu(X) + m(X)]}} - \frac{3 \mathbb{E}_{\mathbf{x}} [m(X) + \mu(X)] \text{var}_{\mathbf{x}} [\mu(X) + m(X)]}{\sqrt{\text{var}_{\mathbf{x}}^3 [\mu(X) + m(X)]}}. \quad (3.11)$$

To recover expression for which results were given earlier, (3.11) is decomposed into expectations and variance of $m(X)$ and $\mu(X)$ individually, where

possible. Some algebra yields

$$\begin{aligned} \text{skew}_{\mathbf{x}}[m(X) + \mu(X)] &= \frac{\mathbb{E}_{\mathbf{x}}[\mu^3(X)] + \mathbb{E}_{\mathbf{x}}[m^3(X)]}{\sqrt{\text{var}_{\mathbf{x}}^3[\mu(X) + m(X)]}} \\ &+ \frac{3\mathbb{E}_{\mathbf{x}}[\mu^2(X)m(X)] + 3\mathbb{E}_{\mathbf{x}}[\mu(X)m^2(X)]}{\sqrt{\text{var}_{\mathbf{x}}^3[\mu(X) + m(X)]}} \\ &- \frac{(\mathbb{E}_{\mathbf{x}}[m(X)] + \mathbb{E}_{\mathbf{x}}[\mu(X)])^3}{\sqrt{\text{var}_{\mathbf{x}}^3[\mu(X) + m(X)]}} \\ &- \frac{3(\mathbb{E}_{\mathbf{x}}[m(X)] + \mathbb{E}_{\mathbf{x}}[\mu(X)]) \text{var}_{\mathbf{x}}[\mu(X) + m(X)]}{\sqrt{\text{var}_{\mathbf{x}}^3[\mu(X) + m(X)]}}. \end{aligned} \quad (3.12)$$

The variance of the sum $\mu(X) + m(X)$ is calculated according to (2.30). Except for $\mathbb{E}_{\mathbf{x}}[\mu^2(X)m(X)]$, $\mathbb{E}_{\mathbf{x}}[\mu(X)m^2(X)]$ and $\mathbb{E}_{\mathbf{x}}[m^3(X)]$, solutions for all terms have been given. Detailed derivations for the three missing expressions are given in appendix A.3.

3.4 Determination of Significant Influences

To this point, it was dealt with the overall impact of uncertainty on the DVHs of a given treatment plan. However, it is also of great interest to screen the setup for the influence of each source of uncertainty. With the aid of such information, the practitioner may evaluate whether additional measures for uncertainty reduction (e.g. IGRT) promise to be worthwhile for a specific patient.

To conduct a sensitivity analysis, a certain input uncertainty distribution, e.g. the setup error, is held fixed while the others, e.g. breathing motion amplitude are varied. A superposition with the DVHs where all inputs were varied, quickly reveals which factors have a great impact, and may require special care in planning or treatment.

Based on the techniques that were introduced, this functionality can be implemented very efficiently. In fact, all computations run in real time because the Gaussian process emulator decouples all calculations from the original patient model. In this context, it is important to note that the input distribution $p(\mathbf{x})$ which is subject to variation during the analysis, enters analytically into the computations (2.24),(2.25),(3.8). This reduces the computational effort even further.

The sensitivity of the rectum DVHs, fig. 3.1(a)/fig. 3.6, was examined with respect to uncertain model parameters. In order to identify the most

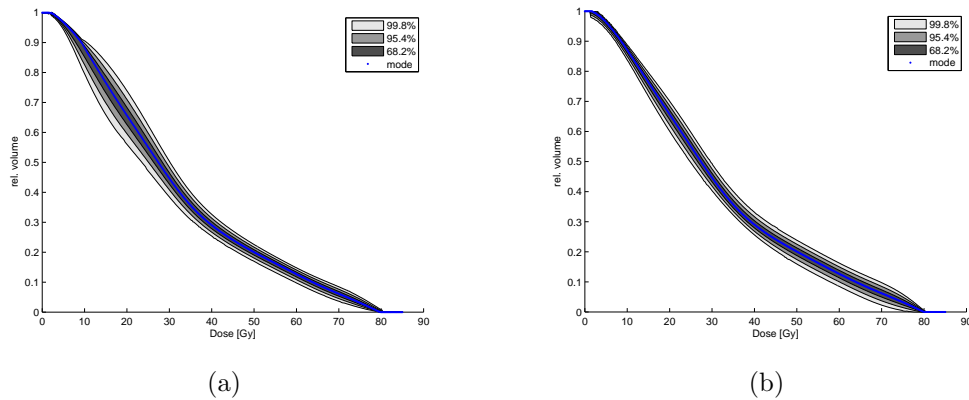


Figure 3.7: Rectum DVH with confidence levels. To screen the influences quantitatively, the (a) first (b) second input was held constant at its nominal value. (a) Notice that fixing the first parameter significantly reduces the spread in the high dose region of the DVH. (b) The second parameter is clearly responsible for the DVH spread in lower doses, pointing towards the fact that the rectum is not moved into the high dose region by this parameter. See also fig. 3.6 for comparison when all inputs are varied.

influential source of errors, they were each held fixed one at a time to their respective nominal value \hat{x} . For brevity, it is assumed that the inputs are independent, i.e. $p(\mathbf{x})$ can be decomposed into $\prod_l p_l(x_l)$. To employ BMC, it is necessary to further restrict the input distribution to Gaussians $p(\mathbf{x}) \equiv \prod \mathcal{N}(x_l | \hat{x}_l, \sigma_l^2)$. The results for the first input being fixed, $p(\mathbf{x}) \equiv p(\mathbf{x} | x_1 = \hat{x}_1)$, at the nominal position are shown in fig. 3.7(a). This parameter is critical because it does influence the high dose irradiation of parts of the rectum. Since these localized high dose exposures play a paramount role in triggering side effects, further control for this patient would be worthwhile. Quite the opposite is the case for the second parameter, fig. 3.7(b). The movement encoded by the latter does not influence high dose exposure. However, it does have an impact on the irradiation of most of the volume. The rectum being a serial organ, it is not expected that this would affect the treatment outcome. For an organ with a different dose response, the biological relevance of both parameters may be reversed.

3.5 Summary and Discussion

The primary goal of this chapter was to incorporate the additional information that becomes available in probabilistic patient models into DVHs in a

meaningful manner. At the core of the introduced techniques lie Gaussian processes (GPs) that are used to substitute DVHs. It was shown that this is useful in three distinct ways. Firstly, the extra information can be propagated into the DVH, enabling the calculation of DVH error bars. These provide an overview of the risks associated with a treatment plan. Secondly, the computed set of DVHs can be correlated with EUDs, which adds information about the biological effect to each DVH. And lastly, the dependencies of the error bars on the different sources of uncertainty may be investigated interactively.

Starting from a small number of patient geometries or other models of geometric setup error, the DVHs are computed for each geometry instance. These DVHs each represent a particular realization of a treatment with its assumed probability. The DVHs, along with the uncertainty parametrization, are used to train a Gaussian process (GP) surrogate model. It replaces the patient model in subsequent calculations. The analytical structure of the GP in conjunction with Gaussian input distributions can be exploited to compute the DVH uncertainty in closed form via Bayesian Monte Carlo (BMC). Because the input probability distributions enter the analysis analytically, they can be varied freely without having to resample geometries and dose calculations from the patient model. This was exploited to implement a sensitivity analysis. To quantitatively account for inherent asymmetry of the distribution of DVHs at a given dose, BMC was extended by the computation of the skewness in addition to mean and variance. The ensuing distribution based on these moments was modelled with the skew-normal-distribution. The parameters of the latter are estimated via the method of moments, keeping the necessary computations at a minimum. Notice that the algorithm is not limited to geometric deviations, but can accommodate all kinds of uncertainties.

With the methods at hand, it was demonstrated that DVH error bars provide more clarity than the DVHs computed on the CTs of the patient alone. Apart from the better overview, quantitative information to facilitate risk assessment for a given treatment plan may also be extracted. DVH error bars prove their usefulness especially in conjunction with sensitivity analysis. By providing uncertainty-source specific DVH error bars, the expert gains valuable information about the likelihood that a certain course of treatment is realized. The introduced method uncovers the otherwise hidden linkage between a given source of uncertainty and its specific effect on the DVH. Labelling DVHs according to the corresponding biological effect further supports the expert in the evaluation of the merits of a treatment plan at hand.

Chapter 4

Spatially Resolved Sensitivity Analysis

In the previous chapters, a new framework for robust dose optimization and evaluation was introduced. The optimization algorithm is to negotiate a compromise between normal tissue sparing and target coverage. If the objectives are mutually contradictory, the algorithm will not be able to achieve all objectives. Quite frequently, the optimum may exhibit underdosage in some target volume or otherwise does not meet the expectations of the expert, hence mandating further refinements. As the compromise of the optimizer may not be easily comprehensible, the need for an analysis of the solution space arises.

The IMRT optimization problem can be addressed using constrained optimization. This means that any solution has to meet a set of constraints while an objective function is minimized. Typically, the setup is chosen such that delivered dose to the tumor is maximized while the dose to the organs at risk (OAR) is limited by the constraints. In case the dose is not satisfying, it is necessary to refine the constraints until the plan is acceptable. In order to do this, the expert needs to know what steps are necessary to tailor the optimum as needed.

Frequently, if constraints are chosen too tightly, the initial plan is not satisfactory due to underdosage in the target. Such cold spots are clinically unacceptable. It is therefore highly desirable to not only determine the active constraints in the target as a whole, as done by sensitivity analysis [ABN02], but also spatially resolved.

This chapter introduces a simple, yet powerful heuristic to identify the set of active constraints for each subvolume of the target not satisfying the desired dose. The proposed method potentially helps to reduce trial-and-error adjustments of the constraints to eliminate the cold spot. The expert

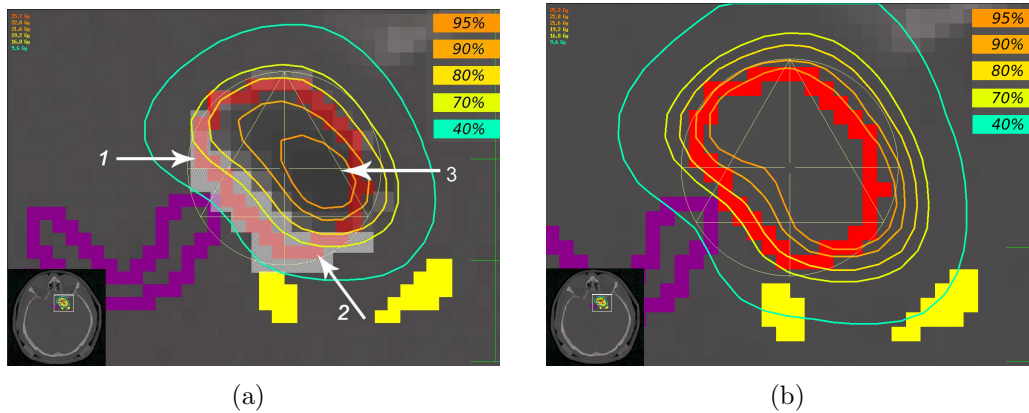


Figure 4.1: (a) Elevated variation grid (grey levels) in zones of conflict. Chiasm (purple) and blood vessels (yellow) are OARs. (b) Optimized plan with PTV (red), chiasm (purple) and blood vessels (yellow). Notice the extended high dose region towards the OARs.

selects an offending region and the algorithm quantifies the impact of the constraints, thus supporting in resolving the conflict at hand.

4.1 The Impact of the Organs at Risk on specific Regions in the Target

In the following, a formula to compute the most influential cost function for any given voxel in the patient model will be derived. For convenience, the IMRT optimization problem (2.28) is restated. The k th constraint is denoted by $g_k(D)$. There is only one objective, $f(D)$. The beamlet weights are denoted by ϕ_j to model the fluence modulation across the beam. The dose distribution is the product of the dose operator T and the beamlet weights ϕ . The element T_{ij} relates the dose deposited in voxel i by beamlet j , thus the dose in voxel i is $D_i = \sum_j T_{ij}\phi_j$. The entire optimization problem can be stated as

$$\text{find } \phi \text{ such that } f(T\phi) \mapsto \min \quad (4.1a)$$

$$\text{while } g_k(T\phi) \leq c_k \quad (4.1b)$$

$$\text{and } \phi_j \geq 0 \quad \forall j, \quad (4.1c)$$

where c_k is the upper limit of constraint k . Typically, IMRT cost functions define a mathematical measure of the dose. The measure of a set of voxels is the sum of its constituent subsets. In that sense, these subsets, labelled

$\hat{f}(D_i)$ and $\hat{g}(D_i)$ respectively, are entirely local quantities, only depending upon the dose in a given voxel i . Notice that objectives as well as constraints can be arbitrary functions of the local dose.

4.1.1 Pointwise Sensitivity Analysis

The algorithm for pointwise sensitivity analysis is based on the gradients, (2.39), used to solve the IMRT optimization problem, (2.28),

$$\frac{\partial f}{\partial \phi_j} = \sum_i \underbrace{\frac{\partial \hat{f}}{\partial D_i}}_1 \underbrace{\frac{\partial D_i}{\partial \phi_j}}_2 \quad (4.2)$$

$$= \sum_i \frac{\partial \hat{f}}{\partial D_i} T_{ij}. \quad (4.3)$$

In the following, the attention is directed towards the first term, which is referred to as first variation of the objective. It states that every voxel i in the volume of interest is assigned a value, since it solely depends upon the local dose at point i by construction. Fig. 4.1(a) depicts the resulting variation grid of the head and neck case example. It consists of 25 cost functions being defined in 18 volumes of interest. The target is located in between the right optical nerve and the chiasm. It comprises $\sim 4.2 \text{ cm}^3$ and was prescribed 18 Gy per fraction. The irradiation geometry consists of 6 non-coplanar arcs with 9 beams each. The setup of the optimization problem included constraints on the maximum equivalent dose (EUD) as well as the maximum dose of the organs at risk. A constraint for the maximum dose for the EUD objective was also imposed. Fig. 4.1(a) indicates that in zones of contradictory objectives, the variation of f is quantitatively elevated. This is intuitively clear as the objective function, in the present case EUD, is highly non-linear.

A non-linearly constrained optimization problem is commonly solved by transformation into an unconstrained subproblem. This is usually done via the method of Lagrange multipliers ($\boldsymbol{\lambda}$). The objective function of the unconstrained problem is the so-called Lagrange function or Lagrangian (L). It consists of the weighted sum of all constraints and objectives. The Lagrange formalism will play a central role in the derivation of the algorithm.

$$L(\boldsymbol{\phi}, \boldsymbol{\lambda}) = f(T\boldsymbol{\phi}) + \sum_k \lambda_k g_k(T\boldsymbol{\phi}) \quad \text{subject to } \boldsymbol{\phi} \geq 0 \quad (4.4)$$

A necessary condition for the Lagrangian, (4.4), at a minimum is that its

first variation must vanish, thus

$$\frac{\partial L}{\partial \phi_j} \stackrel{!}{=} 0 \quad (4.5)$$

$$= \sum_l \frac{\partial \hat{f}(D_l)}{\partial D_l} T_{lj} + \sum_k \lambda_k \sum_m \frac{\partial \hat{g}_k(D_m)}{\partial D_m} T_{mj}. \quad (4.6)$$

Eq. (4.6) can be interpreted as a dose weighted mean of the first variations along the path of the beamlet. It accumulates negative $(\partial \hat{f} / \partial D_i)$ as well as positive $(\partial \hat{g}_k / \partial D_i)$ contributions on its way through the patient. If both contributions compensate, the sum vanishes. In fact, it is the primary goal of the constrained optimizer to alter the multipliers until this equilibrium is achieved. Therefore, in the optimum, any beamlet that sees massive negative contributions by some target structure will also pass through their positive counterparts in the OARs at some point. In consequence, an elevated variation in a given target subvolume entails a counterpart in another region, e.g. organ.

To shed some light on the composition of the dose limiting subvolume, several steps are necessary. The central idea is to exploit the fact that the negative and positive contributions to the variation of the Lagrangian must be equal in the optimum.

Multiplying (4.6) with the dose operator T_{ij} and applying the chain rule reversely yields:

$$\sum_j T_{ij} \frac{\partial f(T\phi)}{\partial \phi_j} = - \sum_k \lambda_k \sum_j T_{ij} \cdot \frac{\partial g_k(T\phi)}{\partial \phi_j}. \quad (4.7)$$

Recall that the dose deposition operator T establishes the link between beamlet weights, scalar quantities and voxels. By applying the same line of thought to gradients instead of beamlet weights, the left side of (4.7) can be understood as the *potential* change of the dose that takes place in a given voxel i , if a *step in the direction of steepest descent of the objective* was taken. If D_i was a cold spot, this is the quantity of interest as an improvement of F means an increase of the dose there. Hence, it is worthwhile to analyze the single components of the sum on the right hand side. Clearly, the impact of a constraint on a voxel in the target is reflected by the magnitude of its contribution to the sum. The term

$$\eta_{ik} = \lambda_k \sum_j T_{ij} \cdot \frac{\partial g_k(T\phi)}{\partial \phi_j} \quad (4.8)$$

Table 4.1: Results of the analysis for three voxels. The queried voxels are marked in fig. 4.1(a).

Constraint	Voxel 1	Voxel 2	Voxel 3
Dose (% prescription)	59	62	101
Chiasm	1	0.23	0.02
Blood vessel	0.85	1	0.04
Max. Dose	0.0	0.0	1

denotes the impact on a given voxel i , induced by a certain constraint c_k . Thus, the expert gains insight into what actions to take in order to address problems with the dose distribution in the target.

Notice that computing (4.8) is possible almost instantaneously in practice. Because the constraint derivative $\frac{\partial g_k}{\partial \phi_j}$ does not contain the voxel index i , it can be precomputed. The dose operator T as well as the Lagrange multipliers λ are also readily available.

The method was applied to the test case mentioned above. Assume that the initial optimization run yielded the dose distribution depicted in fig. 4.1(a). The dosage in the lower left part of the target volume is not satisfactory. Consequently, a voxel was picked in that region to investigate which constraints are dose limiting. The results for three voxels are shown in table 4.1. Because η has no physical interpretation, all values are normalized to the maximum impact. As expected, voxel 1 and 2 receive the highest influence from the abutting OARs. The elevated value of the blood vessel in the distant voxel suggests that its overall target influence is higher compared to the chiasm. After slightly loosening the constraints for chiasm and blood vessel, the dose distribution looks as in fig. 4.1(b).

4.1.2 Perturbation Analysis

In the previous section, a novel method to perform spatially resolved sensitivity analyses was proposed. It allows the expert to take action directly, avoiding cumbersome trial and error. In order to further accelerate the planning process, comprehensive predictions of the outcome of potential changes are useful. In the present context, this translates into predicting the dose change if a constraint was altered. However, because constraints are interdependent, a step in the direction of the steepest decent, ∇g , is not a good approximation of the expected outcome.

To see this, recall that in the optimum, the first variation of the Lagrangian vanishes, (4.6). If a constraint is altered, its corresponding Lagrange multiplier changes. This leads to a residual gradient ∇g . A step

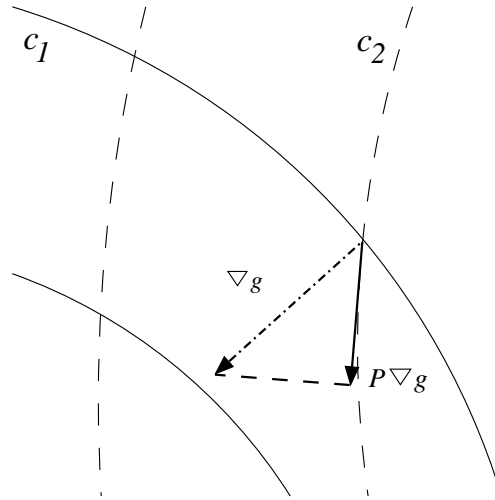


Figure 4.2: Iso constraint lines of constraints c_1 , c_2 (solid, dashed). The system will *not* follow the gradient ∇g (dashed arrow) if c_1 is adjusted, because that would entail a violation of c_2 . The anticipated behavior is approximated by the orthogonal projection $P\nabla g$ (solid arrow).

along the direction of ∇g may violate the remaining constraints. Hence, the constrained optimization algorithm chooses a step that does not entail such violations. The effective direction is the projection $P\nabla g$, only permitting dose changes along the isoconstraint lines of the remaining constraints. This is illustrated in fig. 4.2. Shown are the isoconstraint lines of two constraints. Suppose that the solid constraint c_1 is under investigation. If the gradient (dashed arrow, ∇g) was followed, the other constraint would be violated. To alleviate this problem, the orthogonal projection P of the gradient onto the isoconstraint line (solid arrow, $P\nabla g$) is used. The projected vector, a parallel to the unaltered constraint, is a reasonable approximation for the outcome of the constraint adjustment. Notice that this is a consequence of solving a constrained optimization problem. Notice also that the resulting vector has to be projected onto all constraints, allowing for the treatment of higher dimensional problems.

Fig. 4.3(b) displays the projected dose changes, if a given constraint is relaxed. As anticipated, the dose changes locally where the OAR meets the PTV. This suggests that both constraints have to be loosened in order to achieve a better target coverage. The dose distribution after the adjustments is shown in fig. 4.1(b).

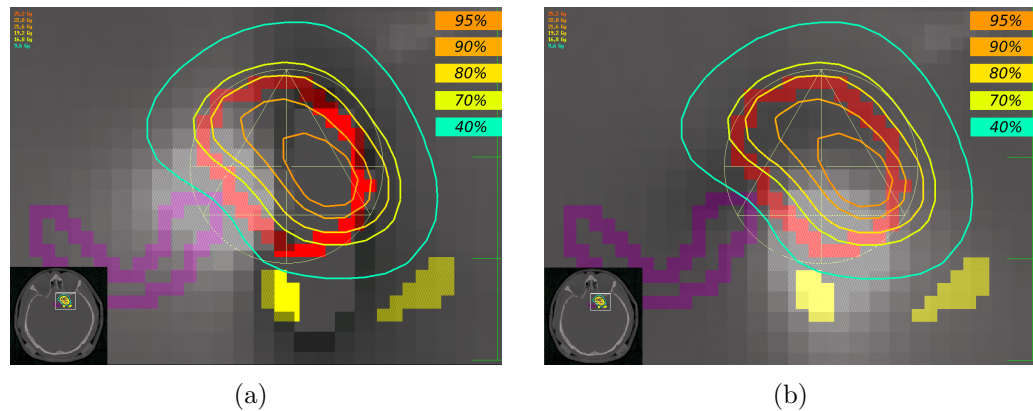


Figure 4.3: (a) chiasm (purple) is loosened (b) blood vessel (yellow) is loosened

4.2 Evaluation and Discussion

In radiotherapy it is crucial to obtain a satisfactory cost-benefit ratio between tumor control and normal tissue sparing. Experience shows that in difficult situations, numerous refinements of the initial plan are necessary. In order to make changes more predictable, [ABN02] put forward a method to assess the impact of certain constraints on the value of the target objective in the optimum, i.e. a sensitivity analysis. This is achieved by analyzing the tangential hyperplane of the Pareto surface in the solution of the constrained optimization problem.

The novelty is the pointwise 3D analysis of impact of each constraint on a cold spot in the target. This is especially useful when several constraints are relevant in the optimization, such that the interplay between the single objectives is not intuitively clear any longer, e.g. fig. 4.4. It was shown that in different areas of the target, different constraints may be limiting the dose. The algorithm enables the expert to quickly screen such scenarios, thus helping to refine the plan in an efficient manner.

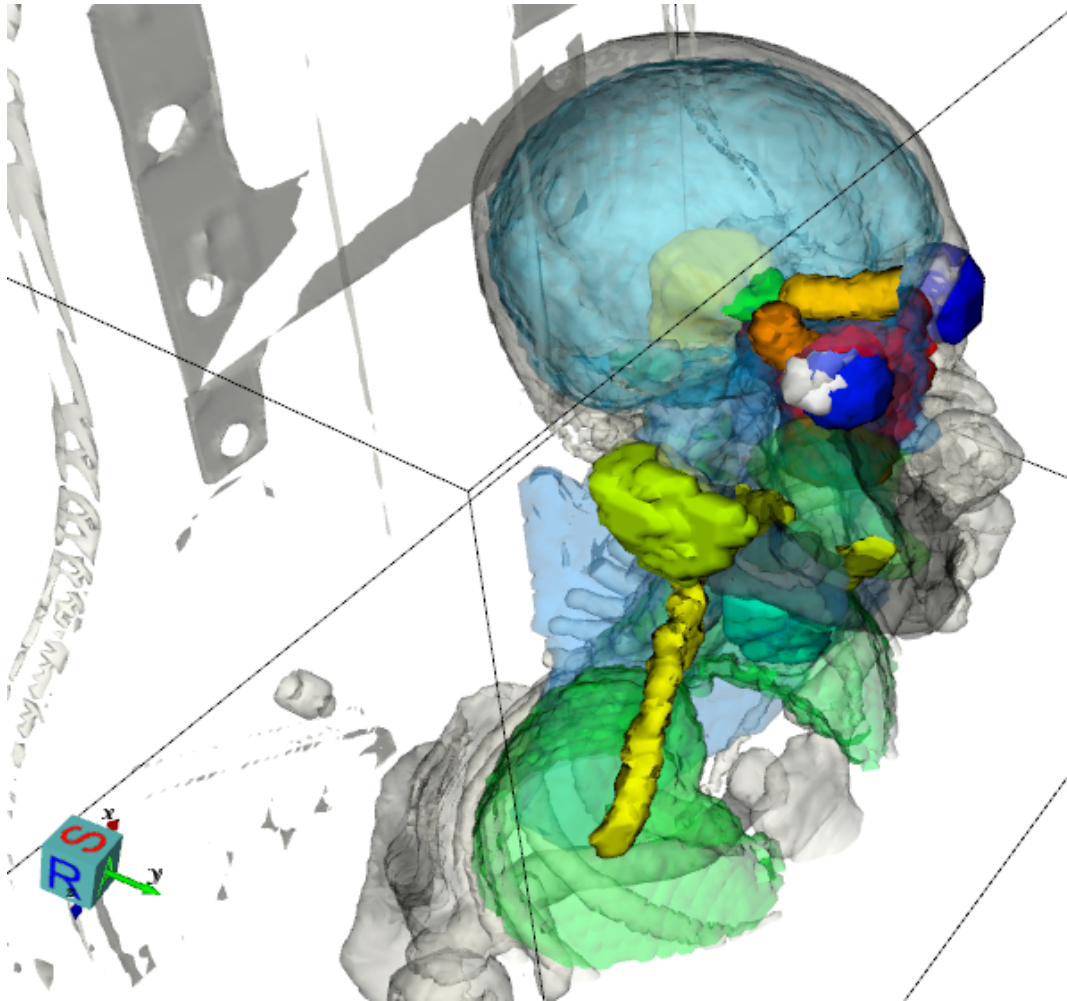


Figure 4.4: 3D reconstruction of many of the volumes of interest in a head and neck case. The interconnections between various costfunctions are hard to overlook intuitively.

Chapter 5

Summary & Conclusions

This thesis addresses one essential problem of radiation therapy: setup error, organ movement along with other inevitable sources of uncertainty can significantly degrade the tumor control probability or increase the risk of complications.

Based on the workflow of the treatment planning process, this thesis is divided into three different parts.

First and foremost, an entirely new approach to account for uncertainties at planning time was presented. The key idea is to quantitatively incorporate the fact that different uncertainty realizations entail different treatment outcomes. In essence, the merit of a treatment plan is no longer quantified by a nominal value but by a treatment outcome distribution. This distribution, reflecting all possible outcomes, is optimized. This means that every scenario in conjunction with its likelihood of being realized is taken into consideration by the optimizer. The first step of putting the idea into practice was to introduce a new formulation of the IMRT optimization problem. This approach is based upon mean and variance of the treatment outcome metric and is mathematically justified by the Chebyshev inequality. Numerically, the front-end for the optimizer is based on a Gaussian approximation of the treatment outcome metric. Normally, the computation of mean and variance would require brute force sampling, but it was shown that the smooth response of the treatment outcome metric to fluctuations in the inputs can be exploited to drastically reduce the number of patient model evaluations. This is made possible by employing Gaussian process regression, a machine learning algorithm. It effectively decouples the patient model, which requires extremely costly computations, from the subsequent analysis. By introducing an emulator between optimizer and patient model, all crucial quantities for the optimization are obtained very efficiently, thus enabling the treatment plan optimization to complete in a clinically feasible timeframe. That

was demonstrated by implementing the algorithm into a clinical treatment planning system. It was evaluated on four prostate cancer patients.

Secondly, at the analysis stage of the treatment plan workflow, an approach to augment classical DVHs to incorporate additional information made available by non-static patient models that are typically used in robust treatment plan optimization was introduced. By employing Gaussian processes, the uncertainty in a DVH under varying conditions can be quantified in real time. The expert is provided with valuable information about the impact of error sources, and the effectiveness of countermeasures against them. To increase the accuracy of the analysis, Bayesian Monte Carlo (BMC) was considerably extended to account for the skewness apart from mean and variance.

Lastly, a tool was introduced that accelerates any prescription refinement steps after a plan has been computed and evaluated. Frequently, the plan resulting from the initial run exhibits spots of insufficient radiation exposure of the target. The introduced 3D sensitivity analysis helps to illuminate the interplay between dose limiting constraints and dose maximizing objectives, thus it helps the expert with the resolution of localized conflicts. While there are already approaches that reveal the interconnection between organs in a treatment plan, the presented approach adds spatial resolution, i.e. instead of treating an organ as a whole, specific subregions can be addressed.

The optimization and evaluation approaches presented in this work are based on an entirely new line of thought. The optimization algorithm (R.-O.B.U.S.T.) is versatile in that it is not bound to any specific source of uncertainty. A proof of principle of the algorithm in conjunction with a deformable biomechanical patient model was given. Combining R.O.B.U.S.T. with other patient models, e.g. rigid shifts to account for setup errors, is very well possible.

The aforementioned proof of principle of R.O.B.U.S.T. was run without dose recalculations. That is, it was implicitly assumed that the patient model deformation entailed only negligible dose deviations. Since the algorithm requires only very few patient model evaluations, dose recalculations are feasible. This is the key to use the algorithm in the realm of heavy particle therapy, as the dose distributions are highly susceptible to geometric variations of the patient.

Despite the fact that the current speed is already within the range of clinical applicability, it can be further improved. Because GP training lies at the core of algorithm, there is still tremendous acceleration potential by using sparse approximations during GP training. Furthermore it may be possible to incorporate the skewness into the optimization. This may be worthwhile because it enables the optimizer to find even better compromises between

normal tissue sparing and tumor control.

To this point, clinical dose optimization is conducted on a static patient model without uncertainty quantification. In contrast, R.O.B.U.S.T. is a feasible method to optimize dose on a probabilistic patient model with quantified uncertainties. Because treatment is inherently a random process, the quality of each goal is reflected by a probability distribution within R.O.B.U.S.T., as opposed to a single number in traditional approaches. Without detailed uncertainty information, the expert has to rely on experience and good faith to ensure that the countermeasures against uncertainties are adequate. With the method presented in this thesis, the effectiveness of countermeasures can be quantified and explicitly incorporated into the optimization for a more robust dose distribution.

Bibliography

- [ABN02] M. Alber, M. Birkner, and F. Nüsslin. Tools for the analysis of dose optimization: II. Sensitivity analysis. *Physics in Medicine and Biology*, 47:N265–N270, 2002.
- [Alb08] Markus Alber. Normal tissue dose-effect models in biological dose optimisation. *Z. Med. Phys.*, 18(2):102–110, 2008.
- [AN99] Markus Alber and Fridtjof Nüsslin. An objective function for radiation treatment optimization based on local biological measures. *Physics in Medicine and Biology*, 44:479–493, 1999.
- [AR07] M. Alber and R. Reemtsen. Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method. *Optimization Methods and Software*, 22:391–411, 2007.
- [Azz85] A. Azzalini. A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, 12:171–178, 1985.
- [BABN04] Christoph Baum, Markus Alber, Mattias Birkner, and Fridtjof Nüsslin. Treatment simulation approaches for the estimation of the distributions of treatment quality parameters generated by geometrical uncertainties. *Physics in Medicine and Biology*, 49:5475–5488, 2004.
- [BABN06] Christoph Baum, Markus Alber, Mattias Birkner, and Fridtjof Nüsslin. Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities. *Radiotherapy and Oncology*, 78:27–35, 2006.
- [BT02] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability*. Athena Scientific, Belmont, MA, 1st edition, 2002.

Bibliography

- [BYA⁺03] Mattias Birkner, D. Yan, M. Alber, J. Liang, and F. Nüsslin. Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation. *Medical Physics*, 30:2822–2831, 2003.
- [CBT06] T. C. Y. Chan, T. Bortfeld, and J. N. Tsitsiklis. A robust approach to IMRT optimization. *Physics in Medicine and Biology*, 51:2567–2583, 2006.
- [CvHMB02] B. C. John Cho, Marcel van Herk, Ben J. Mijneer, and Harry Bartelink. The effect of set-up uncertainties, contour changes, and tissue inhomogeneities on target dose-volume histograms. *Medical Physics*, 29:2305–2318, 2002.
- [CZHS05] Millie Chu, Yuriy Zinchenko, Shane G Henderson, and Michael B Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Physics in Medicine and Biology*, 50:5463–5477, 2005.
- [GRCMS03] A. Girard, C. Rasmussen, J. Candela, and R. Murray-Smith. Gaussian process priors with uncertain inputs – application to multiple-step ahead time series forecasting, 2003.
- [HC08] Francisco Cutanda Henríquez and Silvia Vargas Castrillón. A Novel Method for the Evaluation of Uncertainty in Dose Volume Histogram Computation. *International Journal of Radiation Oncology Biology Physics*, 70:1263–1271, 2008.
- [Jen06] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.
- [KO00] M. C. Kennedy and A. O’Hagan. Predicting the Output from a Complex Computer Code When Fast Approximations Are Available. *Biometrika*, 87:1–13, 2000.
- [LJ01] K. M. Langen and D. T. L. Jones. Organ motion and its management. *International Journal of Radiation Oncology Biology Physics*, 50:265–278, 2001.
- [LLB98] Johan Löf, Bengt K Lind, and Anders Brahme. An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient

- set-up and internal organ motion. *Physics in Medicine and Biology*, 43:1605–1628, 1998.
- [Mac95] David J. C. MacKay. Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6:469–505, 1995.
- [Mac03] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
- [MBC79] M. D. McKay, R. J. Beckman, and W. J. Conover. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21:239–245, 1979.
- [MKVF06] D. L. McShan, M. L. Kessler, K. Vineberg, and B. A. Fraass. Inverse plan optimization accounting for random geometric uncertainties with a multiple instance geometry approximation (MIGA). *Medical Physics*, 33:1510–1521, 2006.
- [MUO06] D Maleike, J Unkelbach, and U Oelfke. Simulation and visualization of dose uncertainties due to interfractional organ motion. *Physics in Medicine and Biology*, 51:2237–2252, 2006.
- [MvHM02] Alan McKenzie, Marcel van Herk, and Ben Mijnheer. Margins for geometric uncertainty around organs at risk in radiotherapy. *Radiotherapy and Oncology*, 62:299–307, 2002.
- [Nea96] Radford M. Neal. *Bayesian Learning for Neural Networks*. Springer-Verlag New York, 1996.
- [NHBT09] T. B. Nguyen, A. C. F. Hoole, N. G. Burnet, and S. J. Thomas. Dose volume population histogram: a new tool for evaluating plans whilst considering geometrical uncertainties. *Physics in Medicine and Biology*, 54:935–947, 2009.
- [Nie97] Andrzej Niemierko. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Medical Physics*, 24:103–110, 1997.
- [O’H87] Anthony O’Hagan. Monte Carlo is Fundamentally Unsound. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 36:247–249, 1987.
-

Bibliography

- [O'H91] Anthony O'Hagan. Bayes-Hermite Quadrature. *Journal of Statistical Planning and Inference*, 29:245–260, 1991.
- [O'H06] Anthony O'Hagan. Bayesian analysis of computer code outputs: a tutorial. Technical report, University of Sheffield, UK, 2006.
- [OO02] Jeremy Oakley and Anthony O'Hagan. Bayesian Inference for the Uncertainty Distribution of Computer Model Outputs. *Biometrika*, 89:769–784, 2002.
- [OO04] Jeremy E. Oakley and Anthony O'Hagan. Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66:751–769, 2004.
- [oRUM93] International Commission on Radiation Units and Measurements. *Prescribing, recording and reporting photon beam therapy*. ICRU, Washington, DC, report 50 edition, 1993.
- [PHR06] T. Pfingsten, D. Herrmann, and C. E. Rasmussen. Model-based Design Analysis and Yield Optimization. *IEEE Transactions on Semiconductor Manufacturing*, 19:475–486, 2006.
- [RDL04] H Edwin Romeijn, James F Dempsey, and Jonathan G Li. A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, 49:1991–2013, 2004.
- [RG01] C. E. Rasmussen and Z. Ghahramani. Occam's razor. In *NIPS 13*, pages 294–300, 2001.
- [RG03] Carl Edward Rasmussen and Zoubin Ghahramani. Bayesian Monte Carlo. In S. Thrun S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 489–496. MIT Press, Cambridge, MA, 2003.
- [RW06] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [SBYA05] M. Söhn, M. Birkner, D. Yan, and M Alber. Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. *Physics in Medicine and Biology*, 50:5893–5908, 2005.

- [SdBHV99] Joep C. Stroom, Hans C. J. de Boer, Henk Huizenga, and Andries G. Visser. Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability. *International Journal of Radiation Oncology Biology Physics*, 43:905–919, 1999.
- [SH06] Joep C. Stroom and Ben J.M. Heijmen. Limitations of the planning organ at risk volume (PRV) concept. *International Journal of Radiation Oncology Biology Physics*, 66:279–286, 2006.
- [SMA08] B. Sobotta, M. Söhn, and M. Alber. Quantifying the Robustness of IMRT Treatment Plans from a Statistical Optimization Model to account for organ deformation. *Radiotherapy and Oncology*, 88 (Supplement 2):S121–S122, 2008.
- [SMA09] B. Sobotta, M. Söhn, and M. Alber. What is the minimum number of geometry samples for a statistical patient model for robust treatment planning? *Radiotherapy and Oncology*, 92 (Supplement 1):S86–S86, 2009.
- [SMA10] B. Sobotta, M. Söhn, and M. Alber. Robust Optimization Based Upon Statistical Theory. *Medical Physics*, 37:4019–4028, 2010.
- [SSSA11] B. Sobotta, M. Söhn, W. Shaw, and M. Alber. On expedient properties of common biological score functions for multimodality, adaptive and 4D dose optimization. *Physics in Medicine and Biology*, accepted, 2011.
- [SS02] Bernhard Schölkopf and Alex Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.
- [SSA09] Matthias Söhn, Benjamin Sobotta, and Markus Alber. Dosimetric Treatment Course Simulation based on a Patient-individual Statistical Model of deformable Organ Motion. *Radiotherapy and Oncology*, 92 (Supplement 1):S65, 2009.
- [SSPA08] B. Sobotta, M. Söhn, M. Pütz, and M. Alber. Tools for the analysis of dose optimization: III. Pointwise sensitivity and perturbation analysis. *Physics in Medicine and Biology*, 53:6337–6343, 2008.

Bibliography

- [SWA09] Matthias Söhn, Martin Weinmann, and Markus Alber. Intensity-Modulated Radiotherapy Optimization in a Quasi-Periodically Deforming Patient Model. *International Journal of Radiation Oncology Biology Physics*, 75:906–914, 2009.
- [SWMW89] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423, 1989.
- [SWN03] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag New York, 2003.
- [TRL⁺05] Alexei Trofimov, Eike Rietzel, Hsiao-Ming Lu, Benjamin Martin, Steve Jiang, George T Y Chen, and Thomas Bortfeld. Temporo-spatial IMRT optimization: concepts, implementation and initial results. *Physics in Medicine and Biology*, 50:2779–2798, 2005.
- [UO04] Jan Unkelbach and Uwe Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Physics in Medicine and Biology*, 49:4005–4029, 2004.
- [UO05] Jan Unkelbach and Uwe Oelfke. Incorporating organ movements in IMRT treatment planning for prostate cancer: Minimizing uncertainties in the inverse planning process. *Medical Physics*, 32:2471–2483, 2005.
- [vHRL02] M. van Herk, P. Remeijer, and J.V. Lebesque. Inclusion of geometric uncertainties in treatment plan evaluation. *International Journal of Radiation Oncology Biology Physics*, 52:1407–1422, 2002.
- [vHRRL00] Marcel van Herk, Peter Remeijer, Coen Rasch, and Joos V. Lebesque. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 47:1121–1135, 2000.
- [vSSLC07] M. von Siebenthal, Gáber Székely, A. Lomax, and Philippe C. Cattin. *Medical Image Computing and Computer Assisted Intervention, MICCAI 2007*, volume 4791 of *Lecture Notes*

- in Computer Science*, chapter Inter-subject Modelling of Liver Deformation During Radiation Therapy, pages 659–666. Springer Berlin / Heidelberg, 2007.
- [WPD01] R. Clint Whaley, Antoine Petitet, and Jack J. Dongarra. Automated Empirical Optimization of Software and the ATLAS Project. *Parallel Computing*, 27(1–2):3–35, 2001.
- [WvdGS⁺07] M.G. Witte, J. van der Geer, C. Schneider, Lebesque J.V., Alber M., and M. van Herk. IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP. *Medical Physics*, 34:3544–3555, 2007.
- [YMS⁺05] Jie Yang, Gig S. Mageras, Spiridon V. Spirou, Andrew Jackson, Ellen Yorke, C. Clifton Ling, and Chen-Shou Chui. A new method of incorporating systematic uncertainties in intensity-modulated radiotherapy optimization. *Medical Physics*, 32:2567–2579, 2005.
- [ÓW06] Arinbjörn Ólafsson and Stephen J Wright. Efficient schemes for robust IMRT treatment planning. *Physics in Medicine and Biology*, 51:5621–5642, 2006.

Acknowledgments

First and foremost, I wish to thank my advisor Dr. Markus Alber. He provided the main guideline for this thesis, while, at the same time, giving me leeway for ideas of my own. His support and invaluable advice shaped this work considerably. Apart from countless contributions to this work, his refreshing views upon reality paired with somewhat unusual metaphors were greatly appreciated.

I would like to express my deepest gratitude to Dr. Matthias Söhn, who accompanied me every step of the way and was always available for fruitful discussions. Many parts of this work thrived on his ideas.

I am deeply indebted to Prof. Dr. Kley, whose support, not only for this thesis, I will never be able to reciprocate. Furthermore, I owe many thanks to Prof. Dr. Dr. Schick for his help and kindness.

To my colleagues:

To Dr. Martin Soukup for many critical discussions and great times at conferences.

To Bettina Frey for patiently explaining countless clinical issues.

To my long term ‘office mate’ Liv Hysing for the great times, inspiring discussions, making my day so often and for simply being in a good mood.

To Dr. Oliver Dohm who enlightened me on the practical side of radiotherapy.

To Philip Oberhammer for managing the teaching of medical technicians and for pointing me toward VTK as well as numerous discussions on software and algorithms.

To Dr. Daniela Thorwarth for her support during the final phase of my work.

To Alexander Horst for *not* letting me build perebak and for providing technical support at my defense.

To David Mönnich for being able to humorously recapture a scene from the Simpsons for any possible occasion.

To my ‘quasi-wife’ Emeline, who enabled me to do this work *and* have a family at the same time by keeping pretty much everything non-work related off my back. All in all, this most certainly required more energy than I will

ever have. She also frequently reminded me that everyone needs to sleep every once in a while and that a break from work every now and then is a good thing. Merci!

I also owe many thanks to free software authors. Without their contribution, this work would not have been possible. This thesis is written in `LATEX` all my files are stored in `Subversion` repositories, and were edited with `emacs`. Amongst *many* others, the `GNU compiler collection`, the `Boost libraries`, `ATLAS` as well as the `FreeBSD` operating system played essential roles for this work. Thanks everyone!

Thanks to my parents and family, who always believed in me.

Appendix A

Calculations

A.1 Solutions for Bayesian Monte Carlo

The goal is to compute

$$\mathbb{E}_{\mathbf{x}}[\mu^a(X)] \quad \text{with } \mu(X) = k(\bar{\mathbf{x}}, X)^\top \cdot \beta \quad (\text{A.1})$$

where $\beta \equiv K^{-1}\mathbf{y} \in \mathbb{R}^{N \times 1}$ and $a = 1, 2, 3$. The results apply for squared exponential kernel, (2.14). The input distribution is Gaussian

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \quad (\text{A.2})$$

Let \mathbf{x}_l denote the l -th training point. A very useful relation for the following derivations is the product of Gaussian densities

$$\mathcal{N}(\mathbf{x}|\mathbf{m}_1, \Sigma_1) \cdot \mathcal{N}(\mathbf{x}|\mathbf{m}_2, \Sigma_2) = c \cdot \mathcal{N}(\mathbf{x}|\mathbf{m}_c, \Sigma_c) \quad (\text{A.3})$$

with

$$\begin{aligned} c &= \mathcal{N}(\mathbf{m}_1|\mathbf{m}_2, (\Sigma_1 + \Sigma_2)) \\ \mathbf{m}_c &= (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mathbf{m}_1 + \Sigma_2^{-1}\mathbf{m}_2) \\ \Sigma_c &= (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}. \end{aligned}$$

For k_0 and k_c in (2.27) we obtain

$$k_0 = \omega_0^2 \quad (\text{A.4})$$

$$k_c = \omega_0^2 |2\Omega^{-1}B + \mathbf{1}|^{\frac{1}{2}}. \quad (\text{A.5})$$

The solutions for z_i , L_{ij} as well as the derivation of Γ_{ijk} , (3.10), are found in the following sections.

A.1.1 $\mathbf{E}_{\mathbf{x}}[\mu(X)]$

$$\mathbf{E}_{\mathbf{x}}[\mu(X)] = \sum_i z_i \beta_i \quad \text{with:} \quad (\text{A.6})$$

$$z_i = \int d\mathbf{x} p(\mathbf{x}) k(\mathbf{x}, \mathbf{x}_i) \quad (\text{A.7})$$

$$= \int d\mathbf{x} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x})^\top \Omega^{-1}(\mathbf{x}_i - \mathbf{x})\right) \cdot \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \quad (\text{A.8})$$

$$= \omega_0^2 (2\pi)^{\frac{D}{2}} |\Omega|^{\frac{1}{2}} \int d\mathbf{x} \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \mathcal{N}(\mathbf{x}|\mathbf{x}_i, \Omega) \quad (\text{A.9})$$

$$= \omega_0^2 (2\pi)^{\frac{D}{2}} |\Omega|^{\frac{1}{2}} (2\pi)^{-\frac{D}{2}} |\Omega + \Sigma_{\mathbf{x}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_{\mathbf{x}} - \mathbf{x}_i)^\top (\Sigma_{\mathbf{x}} + \Omega)^{-1}(\mu_{\mathbf{x}} - \mathbf{x}_i)\right) \quad (\text{A.10})$$

$$= \omega_0^2 |\Omega^{-1} \Sigma_{\mathbf{x}} + \mathbf{1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu_{\mathbf{x}} - \mathbf{x}_i)^\top (\Sigma_{\mathbf{x}} + \Omega)^{-1}(\mu_{\mathbf{x}} - \mathbf{x}_i)\right) \quad (\text{A.11})$$

A.1.2 $\mathbf{E}_{\mathbf{x}}[\mu^2(X)]$

$$\mathbf{E}_{\mathbf{x}}[\mu^2(X)] = \sum_i \sum_j L_{ij} \beta_i \beta_j \quad \text{with:} \quad (\text{A.12})$$

$$L_{ij} = \int d\mathbf{x} p(\mathbf{x}) k(\mathbf{x}, \mathbf{x}_i) k(\mathbf{x}, \mathbf{x}_j) \quad (\text{A.13})$$

$$= \omega_0^4 (2\pi)^{\frac{D}{2}} |\Omega|^{\frac{1}{2}} (2\pi)^{\frac{D}{2}} |\Omega|^{\frac{1}{2}} \int d\mathbf{x} \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \mathcal{N}(\mathbf{x}|\mathbf{x}_i, \Omega) \mathcal{N}(\mathbf{x}|\mathbf{x}_j, \Omega) \quad (\text{A.14})$$

$$= \omega_0^4 (2\pi)^{\frac{D}{2}} |2\Omega^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top (2\Omega)^{-1}(\mathbf{x}_i - \mathbf{x}_j)\right) \int d\mathbf{x} \mathcal{N}\left(\mathbf{x}|\mathbf{x}_{ij}, \frac{1}{2}\Omega\right) \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \quad (\text{A.15})$$

$$= \omega_0^4 |2\Omega^{-1} \Sigma_{\mathbf{x}} + \mathbf{1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}_j - \mathbf{x}_i)^\top (2\Omega)^{-1}(\mathbf{x}_j - \mathbf{x}_i)\right) \exp\left(-\frac{1}{2}(\mathbf{x}_{ij} - \mu_{\mathbf{x}})^\top \left(\frac{1}{2}\Omega + \Sigma_{\mathbf{x}}\right)^{-1}(\mathbf{x}_{ij} - \mu_{\mathbf{x}})\right) \quad (\text{A.16})$$

where $\mathbf{x}_{ij} \equiv \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j)$.

A.1.3 $\mathbf{E}_{\mathbf{x}}[\mu^3(X)]$

The details on the computation of (3.10) is presented in the following.

$$\mathbf{E}_{\mathbf{x}}[\mu^3(X)] = \sum_i \sum_j \sum_k \Gamma_{ijk} \beta_i \beta_j \beta_k \quad \text{with:} \quad (\text{A.17})$$

$$\Gamma_{ijk} = \int d\mathbf{x} k(\mathbf{x}_i, \mathbf{x}) k(\mathbf{x}_j, \mathbf{x}) k(\mathbf{x}_k, \mathbf{x}) p(\mathbf{x}) \quad (\text{A.18})$$

$$= \omega_0^6 (2\pi)^{\frac{D}{2}} |2\Omega^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top (2\Omega)^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right) \\ (2\pi)^{\frac{D}{2}} |\Omega|^{\frac{1}{2}} \int d\mathbf{x} \mathcal{N}\left(\mathbf{x}|\mathbf{x}_{ij}, \frac{1}{2}\Omega\right) \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \mathcal{N}(\mathbf{x}|\mathbf{x}_k, \Omega) \quad (\text{A.19})$$

$$= \omega_0^6 (2\pi)^{\frac{D}{2}} |3\Omega^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top (2\Omega)^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right) \\ \exp\left(-\frac{1}{2}(\mathbf{x}_{ij} - \mathbf{x}_k)^\top \left(\frac{3}{2}\Omega\right)^{-1} (\mathbf{x}_{ij} - \mathbf{x}_k)\right) \\ \int d\mathbf{x} \mathcal{N}\left(\mathbf{x}|\mathbf{x}_{ijk}, \frac{1}{3}\Omega\right) \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \quad (\text{A.20})$$

$$= \omega_0^6 |3\Omega^{-1}\Sigma_{\mathbf{x}} + \mathbf{1}|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top (2\Omega)^{-1} (\mathbf{x}_i - \mathbf{x}_j)\right) \\ \exp\left(-\frac{1}{2}(\mathbf{x}_{ij} - \mathbf{x}_k)^\top \left(\frac{3}{2}\Omega\right)^{-1} (\mathbf{x}_{ij} - \mathbf{x}_k)\right) \\ \exp\left(-\frac{1}{2}(\mathbf{x}_{ijk} - \mu_{\mathbf{x}})^\top \left(\frac{1}{3}\Omega + \Sigma_{\mathbf{x}}\right)^{-1} (\mathbf{x}_{ijk} - \mu_{\mathbf{x}})\right) \quad (\text{A.21})$$

where $\mathbf{x}_{ij} \equiv \frac{1}{2}(\mathbf{x}_i + \mathbf{x}_j)$ and $\mathbf{x}_{ijk} \equiv \frac{1}{3}(\mathbf{x}_i + \mathbf{x}_j + \mathbf{x}_k)$.

A.2 Details of the Incorporation of Linear Offset into the Algorithm for the Variance Computation

In this section, we present the details of the incorporation of the linear offset into the variance computation. Plugging in the definition of the covariance, $\text{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$, along with $m(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$ yields for (2.29)

and (2.30)

$$\begin{aligned} \mathbf{E}_{\mathbf{x}}[\mu(X) + m(X)] &= \mathbf{E}_{\mathbf{x}}[\mu(X)] + \mathbf{a}^\top \mathbf{E}_{\mathbf{x}}[X] + b \\ &= \mathbf{E}_{\mathbf{x}}[\mu(X)] + \mathbf{a}^\top \mu_{\mathbf{x}} + b \end{aligned} \quad (\text{A.22})$$

$$\begin{aligned} \text{var}_{\mathbf{x}}[\mu(X) + m(X)] &= \text{var}_{\mathbf{x}}[\mu(X)] + \mathbf{a}^\top \mathbf{a} \text{var}_{\mathbf{x}}[X] \\ &\quad + 2(\mathbf{E}_{\mathbf{x}}[m(X) \mu(X)] \\ &\quad - \mathbf{E}_{\mathbf{x}}[m(X)] \mathbf{E}_{\mathbf{x}}[\mu(X)]) \\ &= \text{var}_{\mathbf{x}}[\mu(X)] + \mathbf{a}^\top \mathbf{a} \sigma_{\mathbf{x}}^2 \\ &\quad + 2\mathbf{a}^\top (\mathbf{E}_{\mathbf{x}}[X \mu(X)] \\ &\quad - \mu_{\mathbf{x}} \mathbf{E}_{\mathbf{x}}[\mu(X)]). \end{aligned} \quad (\text{A.23})$$

From (A.23) it becomes apparent that the expectation of $X \cdot \mu(X)$ is required,

$$\mathbf{E}_{\mathbf{x}}[X \mu(X)] = \int d\mathbf{x} p(\mathbf{x}) \mathbf{x} \mu(\mathbf{x}) = Q^\top K^{-1} \mathbf{y} \quad (\text{A.24})$$

where Q is a $N \times D$ matrix, N being the total number of trainings points and D is the dimensionality of the data. Hence, an extra term to (2.27) is added

$$\mathbf{q}_l = \int d\mathbf{x} p(\mathbf{x}) \mathbf{x} k(\mathbf{x}, \mathbf{x}_l), \quad (\text{A.25})$$

where \mathbf{q}_l is the l -th column of Q .

With the squared exponential kernel, (2.14) and Gaussian input distribution (A.2), the columns of the matrix Q can also be computed in closed form

$$\mathbf{q}_l = \omega_0^2 (2\pi)^{\frac{D}{2}} |\Omega|^{\frac{1}{2}} \cdot \int d\mathbf{x} \mathbf{x} \mathcal{N}(\mathbf{x} | \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \cdot \mathcal{N}(\mathbf{x} | \mathbf{x}_l, \Omega) \quad (\text{A.26})$$

$$= \omega_0^2 |\Omega|^{\frac{1}{2}} \cdot |\Omega + \Sigma_{\mathbf{x}}|^{-\frac{1}{2}} \cdot \mathbf{E}_{\mathbf{x}}[X] \quad (\text{A.27})$$

$$\begin{aligned} &\exp\left(-\frac{1}{2} [(\mathbf{x}_l - \mu_{\mathbf{x}})^\top (\Omega + \Sigma_{\mathbf{x}})^{-1} (\mathbf{x}_l - \mu_{\mathbf{x}})]\right) \\ &= z_l \cdot \underbrace{((\Omega^{-1} + \Sigma_{\mathbf{x}}^{-1})^{-1} \Omega^{-1} \mathbf{x}_l + (\Omega^{-1} + \Sigma_{\mathbf{x}}^{-1})^{-1} \Sigma_{\mathbf{x}}^{-1} \mu_{\mathbf{x}})}_{\mathbf{E}_{\mathbf{x}}[X]} \end{aligned} \quad (\text{A.28})$$

$$= z_l \cdot (\Omega^{-1} + \Sigma_{\mathbf{x}}^{-1})^{-1} (\Omega^{-1} \mathbf{x}_l + \Sigma_{\mathbf{x}}^{-1} \mu_{\mathbf{x}}) \quad (\text{A.29})$$

$$= z_l \cdot (\Omega^{-1} \Sigma_{\mathbf{x}} + \mathbf{1})^{-1} (\Sigma_{\mathbf{x}} \Omega^{-1} \mathbf{x}_l + \mu_{\mathbf{x}}) \quad (\text{A.30})$$

A.3 Details of the Incorporation of Linear Offset into the Algorithm for the Skewness Computation

The goal is to compute the skewness of the sum of the linear offset and the GP. The unknown expressions in (3.12) are $\mathbf{E}_{\mathbf{x}} [\mu^2(X)m(X)]$, $\mathbf{E}_{\mathbf{x}} [\mu(X)m^2(X)]$ and $\mathbf{E}_{\mathbf{x}} [m^3(X)]$. $\mathbf{E}_{\mathbf{x}} [\mu^2(X)m(X)]$ and $\mathbf{E}_{\mathbf{x}} [\mu(X)m^2(X)]$ can be derived by extending the computation of L and \mathbf{q} in (A.13) and (A.25), respectively. Again, the results given depend on the use of the squared exponential kernel, (2.14), Gaussian input noise, (A.2) and $m(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$.

A.3.1 $\mathbf{E}_{\mathbf{x}} [\mu^2(X)m(X)]$

$$\begin{aligned} \mathbf{E}_{\mathbf{x}} [\mu^2(X)m(X)] &= \mathbf{E}_{\mathbf{x}} [(\mathbf{a}X + b)\mu^2(X)] \\ &= \mathbf{a}^\top \mathbf{E}_{\mathbf{x}} [X\mu(X)^2] + b\mathbf{E}_{\mathbf{x}} [\mu(X)^2] \end{aligned} \quad (\text{A.31})$$

An expression for $\mathbf{E}_{\mathbf{x}} [\mu(X)^2]$ has been given in (A.12). The calculation of $\mathbf{E}_{\mathbf{x}} [X\mu(X)^2]$ adds the expectation of X to (A.13).

$$\mathbf{E}_{\mathbf{x}} [X\mu(X)^2] = \sum_i \sum_j L'_{ij} \beta_i \beta_j \quad \text{with:} \quad (\text{A.32})$$

$$L'_{ij} = \int d\mathbf{x} \mathbf{x} p(\mathbf{x}) k(\mathbf{x}, \mathbf{x}_i) k(\mathbf{x}, \mathbf{x}_j) \quad (\text{A.33})$$

$$= L_{ij} \cdot (2\Omega^{-1} + \Sigma_{\mathbf{x}}^{-1})^{-1} (2\Omega^{-1} \mathbf{x}_{ij} + \Sigma_{\mathbf{x}}^{-1} \mu_{\mathbf{x}}) \quad (\text{A.34})$$

where we have used (A.3) and (A.13).

A.3.2 $\mathbf{E}_{\mathbf{x}} [\mu(X)m^2(X)]$

$$\begin{aligned} \mathbf{E}_{\mathbf{x}} [\mu(X)m^2(X)] &= \mathbf{a}^\top \mathbf{a} \mathbf{E}_{\mathbf{x}} [X^2\mu(X)] \\ &\quad + 2\mathbf{a}^\top b \mathbf{E}_{\mathbf{x}} [X\mu(X)] + b^2 \mathbf{E}_{\mathbf{x}} [\mu(X)] \end{aligned} \quad (\text{A.35})$$

Except for $\mathbf{E}_{\mathbf{x}} [X^2\mu(X)]$, the constituent terms of $\mathbf{E}_{\mathbf{x}} [\mu(X)m^2(X)]$ are given by (A.6) and (A.24).

$$\mathbf{E}_{\mathbf{x}} [X^2\mu(X)] = \int d\mathbf{x} p(\mathbf{x}) \mathbf{x}^\top \mathbf{x} \mu(\mathbf{x}) = \mathbf{u}^\top K^{-1} \mathbf{y} \quad (\text{A.36})$$

where \mathbf{u} is an N -dimensional vector. Similar to the computation of (A.6) and (A.28), \mathbf{u} can be expressed as

$$u_l = z_l \int d\mathbf{x} \mathbf{x}^\top \mathbf{x} \mathcal{N}(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \mathcal{N}(\mathbf{x}|\mathbf{x}_l, \Omega) \quad (\text{A.37})$$

$$= z_l \left((\Omega^{-1} + \Sigma_{\mathbf{x}}^{-1})^{-1} (\Omega^{-1} \mathbf{x}_l + \Sigma_{\mathbf{x}}^{-1} \mu_{\mathbf{x}}) \right)^2 + z_l \text{tr} \left((\Omega^{-1} + \Sigma_{\mathbf{x}}^{-1})^{-1} \right) \quad (\text{A.38})$$

where we have used that

$$E[X^2] = \mu^2 + \text{tr}(\Sigma) \quad (\text{A.39})$$

if X is a Gaussian random variable with mean μ and covariance matrix Σ .

A.3.3 $\mathbf{E}_{\mathbf{x}} [m^3(X)]$

$$\mathbf{E}_{\mathbf{x}} [m^3(X)] = \mathbf{E}_{\mathbf{x}} [\mathbf{a}^3 X^3 + 3\mathbf{a}^2 X^2 b + 3\mathbf{a} X b^2 + b^3] \quad (\text{A.40})$$

$$= \mathbf{a}^3 \mathbf{E}_{\mathbf{x}} [X^3] + 3\mathbf{a} \mathbf{E}_{\mathbf{x}} [X^2] b + 3\mathbf{a} \mathbf{E}_{\mathbf{x}} [X] b + b^3 \quad (\text{A.41})$$

$$= \mathbf{a}^3 (\mu_{\mathbf{x}}^3 + 3\mu_{\mathbf{x}} \Sigma_{\mathbf{x}}) + 3\mathbf{a} (\mu_{\mathbf{x}}^2 + \Sigma_{\mathbf{x}}) b + 3\mathbf{a} \mu_{\mathbf{x}} b + b^3 \quad (\text{A.42})$$

where we have used (A.39) and

$$E[X^3] = \mu^3 + 3\mu\Sigma \quad (\text{A.43})$$

if X is a Gaussian random variable with mean μ and covariance matrix Σ .

Appendix B

Details

B.1 Relating the Average Biological Effect to the Biological Effect of the Average

With the general form of the EUD, (2.7), we can distinguish the following cases.

Serial complication

Given that the serial complication function, $g(D) = D^a$, $a \geq 1$, renders (2.7) a convex function[AN99], it follows from (2.6), that $\text{EUD}(\text{E}[D]) \leq \text{E}[\text{EUD}(D)]$. Hence, considering the average EUD as outcome, (2.5), always *overestimates* the risk for the patient, i.e. the EUD of the mean dose is certain to be lower.

Poisson cell kill

The basic reasoning is similar to the serial complication mechanism. However, (2.7) in conjunction with $g(D) = \exp(-\alpha D)$ is concave. Thus, the relation in (2.6) reverses to $\text{EUD}(\text{E}[D]) \geq \text{E}[\text{EUD}(D)]$. That is, in case of the target we *underestimate* the cell kill by taking the average EUD, (2.5).

Parallel complication

The parallel complication function, $g(D) = (1 + (D_{\text{tol}}/D)^b)^{-1}$, where D_{tol} is the tissue tolerance dose, is convex for $D \leq D_{\text{tol}}$ and concave for $D \geq D_{\text{tol}}$. In consequence, the above reasoning cannot be directly applied. However, typical prescriptions demand that large parts of the organ are kept intact, hence most of the organ volume is irradiated with $D < D_{\text{tol}}$. For this organ

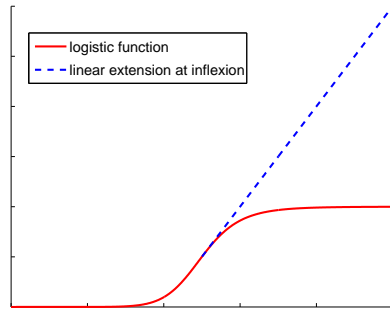


Figure B.1: The parallel complication cost function (logistic function) is linearly extended at the inflexion point to be convex on its entire support. This is useful to relate the mean value of the cost function to the value of the costfunction based on the mean dose.

subvolume, the estimate (2.5) is, again, an upper bound. For the part of the organ that may be sacrificed, the average EUD underestimates the risk. But keeping in mind that (2.5) *overestimates* the risk of the functional reserve being destroyed while *underestimating* the damage to tissue we are willing to sacrifice, this seems acceptable.

If a *safe* upper bound is desired the parallel cost function may be linearly extended at its inflexion point. This ensures convexity for all doses D . This is shown in fig. B.1.

B.2 Mean, Variance and Skewness of the Skew Normal Distribution

Given the probability density function of the skew normal distribution (3.6) with location parameter ξ , scale parameter ω and shape parameter α ,

$$\text{the mean is} \quad \xi + \omega\theta\sqrt{\frac{2}{\pi}} \quad , \quad (\text{B.1a})$$

$$\text{the variance} \quad \omega^2 \left(1 - \frac{2\theta^2}{\pi}\right) \quad \text{and} \quad (\text{B.1b})$$

$$\text{the skewness} \quad \frac{4-\pi}{2} \frac{\left(\theta\sqrt{\frac{2}{\pi}}\right)^3}{\left(1-\frac{2\theta^2}{\pi}\right)^{\frac{3}{2}}} \quad (\text{B.1c})$$

with $\theta \equiv \frac{\alpha}{1+\alpha}$.

Appendix C

Tools for the analysis of dose optimization: III. Pointwise sensitivity and perturbation analysis

published in

Physics in Medicine and Biology 2008; **53**: 6337-6343

Tools for the analysis of dose optimization: III. Pointwise sensitivity and perturbation analysis

B Sobotta, M Söhn, M Pütz and M Alber

Abt. Medizinische Physik, Radiologische Uniklinik, Universität Tübingen,
72076 Tübingen, Germany

E-mail: benjamin.sobotta@med.uni-tuebingen.de

Received 6 June 2008, in final form 1 September 2008

Published 20 October 2008

Online at stacks.iop.org/PMB/53/6337

Abstract

The major challenge in intensity-modulated radiotherapy planning is to find the right balance between tumor control and normal tissue sparing. The most desirable solution is never physically feasible, and a compromise has to be found. One possible way to approach this problem is constrained optimization. In this context, it is worthwhile to quantitatively predict the impact of adjustments of the constraints on the optimum dose distribution. This has been dealt with in regard to cost functions in a previous paper. The aim of the present paper is to introduce spatial resolution to this formalism. Our method reveals the active constraints in a target subvolume that was previously selected by the practitioner for its insufficient dose. This is useful if a multitude of constraints can be the cause of a cold spot. The response of the optimal dose distribution to an adjustment of constraints (perturbation) is predicted. We conclude with a clinical example.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Dose optimization is a vital ingredient in intensity-modulated radiotherapy (IMRT) planning. The purpose of the optimization algorithm is to assist the expert in negotiating a compromise between normal tissue sparing and target coverage. If the objectives are mutually contradictory, the algorithm will not be able to achieve all objectives. In order to resolve these conflicts, methods are needed that enable the expert to quickly screen the problem.

In order to quantify the merit of a given dose distribution, cost functions are employed in the optimization (Webb 2000, Wang *et al* 1995, Bortfeld *et al* 1997, Raphael 1992). Depending on the desired dose–response characteristics, many different types have been devised (Romeijn *et al* 2004, Censor *et al* 2006). Whereas physical cost functions act solely on the delivered dose, their biological counterparts strive to take biological effects into account (Alber and

Belka 2006, Alber 2008). A widely used concept is equivalent uniform dose (EUD) (Niemierko 1997, Alber and Nüsslin 1999, Wu *et al* 2002, Thieke *et al* 2003). Ólafsson *et al* (2005) additionally provide an overview regarding the general development of cost functions. The particular choice of cost functions for certain organs is of importance as it gives rise to different amounts of dose inhomogeneity. Quite frequently, the optimum may exhibit underdosage in some target volume or otherwise does not meet the expectations of the expert, hence mandating further refinements. As the compromise of the optimizer may not be easily comprehensible, the need for an analysis of the solution space arises.

One possible approach is to assign a weight factor to each objective, indicating its priority to the optimizer. Unfortunately, there exists no straightforward method to determine the weights prior to the optimization since this requires an understanding which criteria are competing and non-competing. This leads to a trial and error process that involves multiple optimization runs. Multicriterial optimization is a promising approach to reduce the number of iterations required (Monz *et al* 2008, Hoffmann *et al* 2006). This technique treats each objective separately, giving rise to a vector-valued objective function. One way to judge the grade of a solution vector is to use Pareto optimality. A point on the so-called Pareto surface is characterized by the fact that no criteria can be improved without worsening another. In practice, the Pareto surface is approximated by generating a variety of plans (Craft *et al* 2007). From that surface, the practitioner then selects the best plan for the case at hand (Thieke *et al* 2007).

Alternatively, the IMRT optimization problem can be addressed using constrained optimization (Alber 2000, Alber and Reemtsen 2007). This means that any solution has to meet a set of constraints while an objective function is minimized. This translates into the delivery of a prescribed dose to the tumor while limiting the dose deposited in organs at risk (OAR). Clearly, the target dose is limited only by the imposed constraints. In case the dose is not satisfying, it is necessary to refine the constraints until the plan is overall acceptable. In order to do this, the expert needs to know what steps are necessary to tailor the optimum as needed (Alber *et al* 2002). Varieties of the constrained optimization approach have been introduced by Jee *et al* (2007) and Wilkens *et al* (2007).

For the sake of simplicity, assume that all normal tissue objectives have been given maximum priority so that only the target underdosage reflects the conflicting objectives. More often than not, the initial plan is not satisfactory due to underdosage in the target. Such cold spots are clinically unacceptable. It is therefore highly desirable to not only determine the active constraints in the target as a whole, as done by sensitivity analysis (Alber *et al* 2002), but also in certain subregions. In this paper we present a simple yet powerful heuristic to identify the set of active constraints for each subvolume of the target not satisfying the desired dose. The proposed method potentially helps to reduce human iteration loops, where the constraints are adjusted to eliminate the cold spot. The expert selects an offending region and the algorithm quantifies the impact of the constraints, thus supporting the practitioner in resolving the conflicts at hand.

2. Methods and materials

2.1. Prerequisites

Before elaborating on the algorithm, we briefly review a generic approach to IMRT optimization. The k th constraint, which is a function of the dose distribution D , is denoted by $G_k(D)$. For the sake of simplicity, assume that there is only one objective, $F(D)$. Typically, these cost functions define a mathematical measure of the dose. The measure of a set of voxels

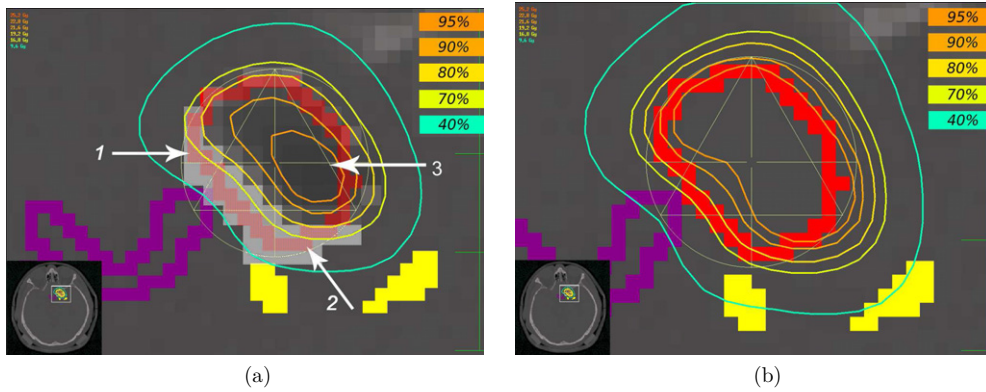


Figure 1. Left panel: optimized plan with underdosage region. Right panel: optimized plan with adjusted constraints. (a) Elevated variation grid (gray levels) in zones of conflict. Chiasm (purple) and blood vessels (yellow) are OARs. (b) Optimized plan with PTV (red), chiasm (purple) and blood vessels (yellow). Note the extended high dose region towards the OARs.

is the sum of its constituent subsets. In that sense, these subsets, labeled $f(D_i)$ and $g(D_i)$ respectively, are entirely local quantities, only depending upon the dose in a given voxel i . Hence, cost functions structurally resemble volume integrals. Note that objectives as well as constraints can be arbitrary functions of the local dose (Alber 2008, Romeijn *et al* 2004).

The incident beams are discretized into so-called beamlets or bixels. Each one of those beamlets is assigned a weight ϕ_j to model the fluence modulation across the beam. The dose distribution is the product of the dose operator T and the beamlet weights ϕ . The element T_{ij} relates the dose deposited in voxel i by beamlet j , thus the dose in voxel i is $D_i = \sum_j T_{ij}\phi_j$. The entire optimization problem can now be stated as

$$\text{find } \phi \text{ such that } F(T\phi) \mapsto \min \quad (1a)$$

$$\text{while } G_k(T\phi) \leq c_k \quad (1b)$$

$$\text{and } \phi_j \geq 0 \forall j, \quad (1c)$$

where c_k is the upper limit of constraint k . It is self-evident that the optimum depends upon \mathbf{c} . Consequently, the constraints need to be adjusted if the result does not meet the requirements on F .

In order to solve the above optimization problem efficiently, the gradients of the objectives are needed. Applying the chainrule yields

$$\frac{\partial F}{\partial \phi_j} = \sum_i \underbrace{\frac{\partial f}{\partial D_i}}_1 \underbrace{\frac{\partial D_i}{\partial \phi_j}}_2 \quad (2)$$

$$= \sum_i \frac{\partial f}{\partial D_i} T_{ij}. \quad (3)$$

In the following, we direct our attention toward the first term, which we refer to as first variation of the objective. It states that every voxel i in the volume of interest is assigned a value, since it solely depends upon the local dose at point i by construction. Figure 1(a) depicts the resulting variation grid of our head and neck case example. It consists of 25 cost

functions being defined on 18 volumes of interest. The target is located in between the right optical nerve and the chiasm. It comprises $\sim 4.2 \text{ cm}^3$ and was prescribed 18 Gy per fraction. The irradiation geometry consists of 6 non-coplanar arcs with 9 beams each. The setup of the optimization problem included constraints on the maximum equivalent dose (EUD) as well as the maximum dose of the organs at risk. A constraint for the maximum dose for the EUD objective was also imposed. Figure 1(a) indicates that in zones of contradictory objectives, the variation of F is quantitatively elevated. This is intuitively clear as the objective function, in our case EUD, is highly nonlinear.

A nonlinearly constrained optimization problem, (1), is commonly solved by transformation into an unconstrained subproblem. This is usually done via the method of Lagrange multipliers (λ). The objective function of the unconstrained problem is the so-called Lagrange function or Lagrangian (L). It consists of the weighted sum of all constraints and objectives. The Lagrange formalism will play a central role in the derivation of our algorithm.

$$L(\phi, \lambda) = F(T\phi) + \sum_k \lambda_k G_k(T\phi) \quad \text{subject to } \phi \geq 0. \quad (4)$$

A necessary condition for the Lagrangian, (4), at a minimum is that its first variation must vanish, thus

$$\frac{\partial L}{\partial \phi_j} \stackrel{!}{=} 0 \quad (5)$$

$$= \sum_l \frac{\partial f(D_l)}{\partial D_l} T_{lj} + \sum_k \lambda_k \sum_m \frac{\partial g_k(D_m)}{\partial D_m} T_{mj}. \quad (6)$$

Equation (6) can be interpreted as a dose weighted mean of the first variations along the path of the beamlet. It accumulates negative ($\partial f/\partial D_i$) as well as positive ($\partial g_k/\partial D_i$) contributions on its way through the patient. If both contributions compensate, the sum vanishes. In fact, it is the primary goal of the constrained optimizer to alter the multipliers until this equilibrium is achieved. Therefore, in the optimum, any beamlet that sees massive negative contributions by some target structure will also pass through their positive counterparts in the OARs at some point. In consequence, an elevated variation in a given target subvolume entails a counterpart in another region, e.g. organ.

2.2. Algorithm

2.2.1. Pointwise sensitivity analysis. To shed some light on the composition of the dose limiting subvolume, several steps are necessary. The central idea is to exploit the fact that the negative and positive contributions to the variation of the Lagrangian must be equal in the optimum.

Multiplying (6) with the dose operator T_{ij} and applying the chain rule reversely yields

$$\sum_j T_{ij} \frac{\partial F(T\phi)}{\partial \phi_j} = - \sum_k \lambda_k \sum_j T_{ij} \cdot \frac{\partial G_k(T\phi)}{\partial \phi_j}. \quad (7)$$

Bear in mind that the dose deposition operator T establishes the link between beamlet weights, scalar quantities and voxels. By applying the same line of thought to gradients instead of beamlet weights, the left side of (7) can be understood as the *potential* change of the dose that takes place in a given voxel i , if a *step in the direction of steepest descent of the objective* was taken. If D_i was a cold spot, this is the quantity we are interested in as an improvement of F

Table 1. Results of our analysis for three voxels. The queried voxels are marked in figure 1(a).

Constraint	Voxel 1	Voxel 2	Voxel 3
Dose (% prescription)	59	62	101
Chiasm	1	0.23	0.02
Blood vessel	0.85	1	0.04
Maximum dose	0.0	0.0	1

means an increase of the dose there. Hence, it is worthwhile to analyze the single components of the sum on the right-hand side. Clearly, the impact of a constraint on a voxel in the target is reflected by the magnitude of its contribution to the sum. By

$$\eta_{ik} = \lambda_k \sum_j T_{ij} \cdot \frac{\partial G_k(T\phi)}{\partial \phi_j} \quad (8)$$

we denote the impact on a given voxel i , induced by a certain constraint c_k . Thus, the expert gains insight into what actions to take in order to address problems with the dose distribution in the target.

Note that computing (8) is possible almost instantaneously in practice. Because the constraint derivative $\frac{\partial G_k}{\partial \phi_j}$ does not contain the voxel index i , it can be precomputed. The dose operator T as well as the Lagrange multipliers λ are also readily available.

We applied the method to the test case mentioned above. Assume that our initial optimization run left us with the dose distribution depicted in figure 1(a). The dosage in the lower left part of our target volume is not satisfactory. Consequently, we pick a voxel in that region to investigate which constraints are dose limiting. The results for three voxels are shown in table 1. Because η has no physical interpretation, all values are normalized to the maximum impact. As expected, voxels 1 and 2 receive the highest influence from the abutting OARs. The elevated value of the blood vessel in the distant voxel suggests that its overall target influence is higher compared to the chiasm. After slightly loosening the constraints for chiasm and blood vessel, the dose distribution looks as in figure 1(b).

2.2.2. Perturbation analysis. In the previous section, we proposed a novel method to perform spatially resolved sensitivity analyses. It allows for the expert to take action directly, avoiding cumbersome trial and error. In order to further accelerate the planning process, comprehensive predictions of the outcome of potential changes are useful. In the present context, this translates into predicting the dose change if a constraint was altered. However, because constraints are interdependent, a step in the direction of the steepest decent, ∇G , is not a good approximation of the expected outcome.

To see this, recall that in the optimum, the first variation of the Lagrangian vanishes, (6). If a constraint is altered, its corresponding Lagrange multiplier changes. This leads to a residual gradient ∇G . A step along the direction of ∇G may violate the remaining constraints. Hence, the constrained optimization algorithm chooses a step that does not entail such violations. The effective direction is the projection $P\nabla G$, only permitting dose changes along the isoconstraint lines of the remaining constraints. This is illustrated in figure 2. The isoconstraint lines of two constraints are shown. Suppose that the solid constraint c_1 is under investigation. If we were to follow the gradient (dashed arrow, ∇G), the other constraint would be violated. To alleviate this problem, we use the orthogonal projection P of the gradient onto the isoconstraint line (solid arrow, $P\nabla G$). The projected vector, a parallel to the unaltered constraint, is a reasonable approximation for the outcome of the constraint adjustment. Note

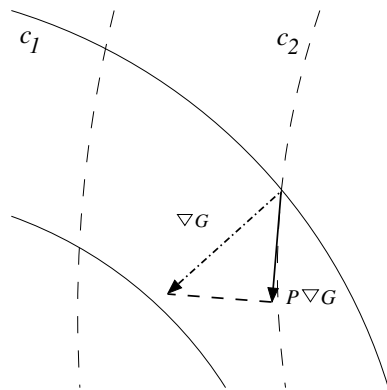


Figure 2. Isoconstraint lines of constraints c_1 , c_2 (solid, dashed). The system will *not* follow the gradient ∇G (dashed arrow) if c_1 is adjusted, because that would entail a violation of c_2 . The anticipated behavior is approximated by the orthogonal projection $P\nabla G$ (solid arrow).

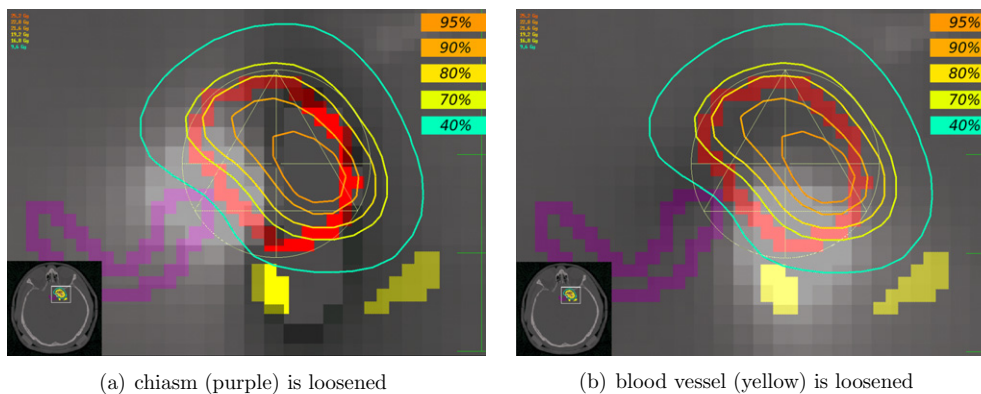


Figure 3. Dose change prediction (by gray level) if (a) the chiasm or (b) the blood vessel constraint is loosened. The projected gradients of the objective function, $P\nabla F$ are shown. Note the local dose increase around the OAR in question.

that this is a consequence of solving a constrained optimization problem. Note also that the resulting vector has to be projected onto all constraints, allowing for the treatment of higher dimensional problems.

Figure 3 displays the projected dose changes, if a given constraint is relaxed. As anticipated, the dose changes locally where the OAR meets the PTV. This suggests that both constraints have to be loosened in order to achieve a better target coverage. The dose distribution after the adjustments is shown in figure 1(b).

3. Discussion and conclusion

In radiotherapy it is crucial to obtain a satisfactory cost–benefit ratio between tumor control and normal tissue sparing. Experience shows that in difficult situations, numerous refinements of the initial plan are necessary. In order to make changes more predictable, Alber *et al* (2002) put forward a method to assess the impact of certain constraints on the value of the

target objective in the optimum, i.e. a sensitivity analysis. This is achieved by analyzing the tangential hyperplane of the Pareto surface in the solution of the constrained optimization problem.

In the spirit of the previous paper, the present work introduces spatial resolution into this framework. The novelty is the pointwise analysis of impact of each constraint on a cold spot in the target. This is especially useful when several constraints are relevant in the optimization, such that the interplay between the single objectives is not intuitively clear any longer. It was shown that in different areas of the target, different constraints may be limiting the dose. The algorithm enables the expert to quickly screen such scenarios, thus helping to refine the plan in an efficient manner.

References

- Alber M 2000 A concept for the optimization of radiotherapy *PhD Thesis* University of Tübingen
- Alber M 2008 Normal tissue dose-effect models in biological dose optimisation *Z. Med. Phys.* **18** 102–10
- Alber M and Belka C 2006 A normal tissue dose response model of dynamic repair processes *Phys. Med. Biol.* **51** 153–72
- Alber M, Birkner M and Nüsslin F 2002 Tools for the analysis of dose optimization: II. Sensitivity analysis *Phys. Med. Biol.* **47** N265–70
- Alber M and Nüsslin F 1999 An objective function for radiation treatment optimization based on local biological measures *Phys. Med. Biol.* **44** 479–93
- Alber M and Reemtsen R 2007 Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method *Optim. Methods Softw.* **22** 391–411
- Bortfeld T, Stein J and Preiser K 1997 *Proc. 12th Int. Conf. on the Use of Computers in Radiotherapy* pp 1–4
- Censor Y, Bortfeld T, Martin B and Trofimov A 2006 A unified approach for inversion problems in intensity-modulated radiation therapy *Phys. Med. Biol.* **51** 2353–65
- Craft D, Halabi T, Shih H and Bortfeld T 2007 An approach for practical multiobjective IMRT treatment planning *Int. J. Radiat. Oncol. Biol. Phys.* **69** 1600–7
- Hoffmann A, Siem A, den Hertog D, Kaanders J and Huizenga H 2006 Derivative-free generation and interpolation of convex Pareto optimal IMRT plans *Phys. Med. Biol.* **51** 6349–69
- Jee K W, McShan D L and Fraass B A 2007 Lexicographic ordering: intuitive multicriteria optimization for IMRT *Phys. Med. Biol.* **52** 1845–61
- Monz M, Küfer K, Bortfeld T and Thieke C 2008 Pareto navigation—algorithmic foundation of interactive multicriteria IMRT planning *Phys. Med. Biol.* **53** 985–98
- Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10
- Ólafsson A, Jeraj R and Wright S J 2005 Optimization of intensity-modulated radiation therapy with biological objectives *Phys. Med. Biol.* **50** 5357–79
- Raphael C 1992 Mathematical modelling of objectives in radiation therapy treatment planning *Phys. Med. Biol.* **37** 1293–311
- Romeijn H E, Dempsey J F and Li J G 2004 A unifying framework for multi-criteria fluence map optimization models *Phys. Med. Biol.* **49** 1991–2013
- Thieke C, Bortfeld T, Niemierko A and Nill S 2003 From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning *Med. Phys.* **30** 2332–9
- Thieke C, Küfer K, Monz M, Scherrer A, Alonso F, Oelfke U, Huber P, Debus J and Bortfeld T 2007 A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation *Radiother. Oncol.* **85** 292–8
- Wang X H, Mohan R, Jackson A, Leibel S A, Fuks Z and Ling C C 1995 Optimization of intensity-modulated 3D conformal treatment plans based on biological indices *Radiother. Oncol.* **37** 140–52
- Webb S 2000 *Intensity-Modulated Radiation Therapy (Medical Science Series)* (Bristol: Institute of Physics Publishing)
- Wilkens J J, Alaly J R, Zakarian K, Thorstad W L and Deasy J O 2007 IMRT treatment planning based on prioritizing prescription goals *Phys. Med. Biol.* **52** 1675–92
- Wu Q, Mohan R, Niemierko A and Schmidt-Ullrich R 2002 Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose *Int. J. Radiat. Oncol. Biol. Phys.* **52** 224–35

Appendix D

Robust Optimization Based Upon Statistical Theory

published in

Medical Physics 2010; **37**: 4019-4028

Robust optimization based upon statistical theory

B. Sobotta,^{a)} M. Söhn, and M. Alber

Abt. Medizinische Physik, Radiologische Uniklinik, Universität Tübingen, 72076 Tübingen, Germany

(Received 26 October 2009; revised 23 April 2010; accepted for publication 7 June 2010; published 13 July 2010)

Purpose: Organ movement is still the biggest challenge in cancer treatment despite advances in online imaging. Due to the resulting geometric uncertainties, the delivered dose cannot be predicted precisely at treatment planning time. Consequently, all associated dose metrics (e.g., EUD and maxDose) are random variables with a patient-specific probability distribution. The method that the authors propose makes these distributions the basis of the optimization and evaluation process.

Methods: The authors start from a model of motion derived from patient-specific imaging. On a multitude of geometry instances sampled from this model, a dose metric is evaluated. The resulting pdf of this dose metric is termed outcome distribution. The approach optimizes the shape of the outcome distribution based on its mean and variance. This is in contrast to the conventional optimization of a nominal value (e.g., PTV EUD) computed on a single geometry instance. The mean and variance allow for an estimate of the expected treatment outcome along with the residual uncertainty. Besides being applicable to the target, the proposed method also seamlessly includes the organs at risk (OARs).

Results: The likelihood that a given value of a metric is reached in the treatment is predicted quantitatively. This information reveals potential hazards that may occur during the course of the treatment, thus helping the expert to find the right balance between the risk of insufficient normal tissue sparing and the risk of insufficient tumor control. By feeding this information to the optimizer, outcome distributions can be obtained where the probability of exceeding a given OAR maximum and that of falling short of a given target goal can be minimized simultaneously.

Conclusions: The method is applicable to any source of residual motion uncertainty in treatment delivery. Any model that quantifies organ movement and deformation in terms of probability distributions can be used as basis for the algorithm. Thus, it can generate dose distributions that are robust against interfraction and intrafraction motion alike, effectively removing the need for indiscriminate safety margins. © 2010 American Association of Physicists in Medicine.

[DOI: [10.1118/1.3457333](https://doi.org/10.1118/1.3457333)]

Key words: probabilistic planning, uncertainty, outcome distribution, robust treatment

I. INTRODUCTION

In the presence of deviations from the planning patient geometry, there will be a difference between applied and planned dose. Hence, the delivered dose and derived treatment quality metrics become random variables.¹⁻⁴

Evidently, the statistical nature of the treatment also propagates to the treatment quality parameters, e.g., dose volume histogram (DVH) points or equivalent uniform dose (EUD).⁵ Any one of these metrics becomes a random variable, having its own patient-specific probability distribution. This distribution indicates the likelihood that a certain value of the metric will be realized during a treatment session. Because the outcome distribution is sampled only a few times, namely, the number of treatment sessions, it becomes important to consider the width of the distribution in addition to the mean value.^{2,6,7} The wider a distribution becomes, the more samples are necessary to obtain reliable estimates of the mean. In the treatment context, this translates into the need to minimize the distribution width of the quality metrics to arrive at the planned mean of the metric with a high probability (Fig. 1). This is to assure that the actual prescription is met as closely as possible in every course of treatment.

The standard approach to address this problem is to extend the clinical target volume (CTV) to obtain a planning target volume (PTV).⁸ Provided that the margin is properly chosen,^{9,10} the dose to any point of the CTV remains almost constant across all fractions, leading to a narrow distribution of possible treatment outcomes (Fig. 2). However, the PTV concept only addresses the target, not the organs at risk (OARs). The addition of margins for the OARs was suggested by McKenzie *et al.*,¹¹ leading to planning risk volumes (PRVs). This increases the inevitable overlap region of mutually contradictory plan objectives even more.¹²

In order to overcome these problems, numerous approaches have been suggested¹³⁻²³ that explicitly incorporate the magnitude of possible geometric errors into the planning process. Depending on the authors' view on robustness, and thereby the quantities of interest, the proposed methods are diverse.

The method of coverage probabilities can be regarded as the most basic extension of the PTV concept in that it assigns probability values to a classic or individualized PTV/PRV to indicate the likelihood of finding the CTV/OAR there.¹⁰ Baum *et al.*¹³ address the problem of overlapping volumes of

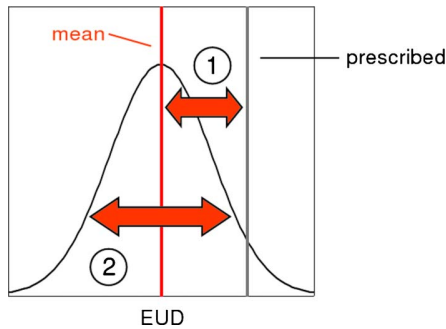


FIG. 1. The important properties of a treatment outcome distribution. (1) The mean should be close to the prescribed dose. (2) The distribution should be as narrow as possible because patients are treated in only a few treatment sessions.

interest by incorporating a coverage probability based on several CTs into the optimization process. Through these occupancy maps, the specific motion of the organs in essence replaces the more unspecific margins. Even though the problem of overlapping volumes is alleviated, it is not solved. The respective probability clouds of the organs may still very well have common regions, since there is no notion of correlated motion. A related concept, namely, the optimization of the expectation of tumor control probability and normal tissue complication probability was investigated.²¹ Note that coverage probability is still based on a static patient model, i.e., the dose is evaluated in dose space, not relative to the moving organs.¹⁴

In contrast, other authors use deformable patient models to track the irradiation of every single volume element (voxel). This means that while dose scoring and evaluation is done in a reference geometry of the patient, the dose is calculated on the deformed model. Chu *et al.*¹⁵ and Ólafsson and Wright¹⁶ propose a method that models organ movement as uncertainties in the dose matrix. In essence, their method addresses the robustness of the dose on a per voxel basis;

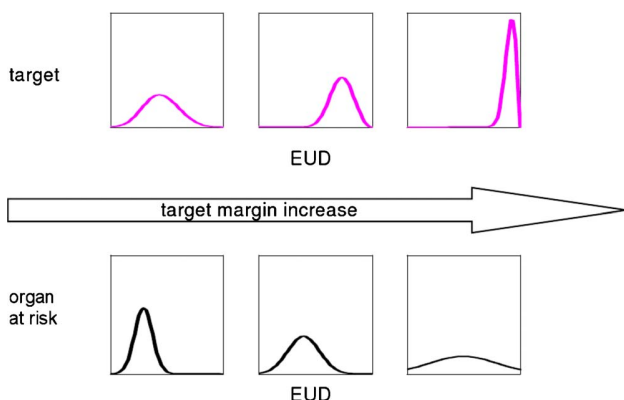


FIG. 2. The treatment outcome distribution (e.g., in terms of EUD) plotted against target margin width. If the target margin becomes larger, the target distribution becomes more and more located (top three panels). The contrary is the case regarding the organ at risk distribution. With increasing target margin, this distribution widens toward higher doses as the gradient moves into organ space.

thus, their method is restricted to physical objective functions. Birkner *et al.*¹⁴ brought forth a method of accounting for uncertainties by substituting the nominal dose in each voxel of a reference geometry with an effective dose. The effective dose is obtained by averaging the doses in the voxel over several geometry instances. This also allows for the use of biological cost functions. A similar approach with the addition of dose recomputation for each geometry instance has been presented.^{18,19,22}

Naturally, all the above methods require precise information about the patient movement prior to treatment. Due to the limits of available information, these deformation models themselves provide for another source of errors. A method to take these errors of the model of motion into account has also been suggested.¹⁷

In consequence, even concepts that go beyond margins to alleviate the impact of geometric variation have to deal with uncertain outcome metrics.

We propose a general probabilistic optimization framework, robust optimization based upon statistical theory (R.O.-B.U.S.T.), for any type of cost function that pays tribute to the statistical nature of the problem by controlling the treatment outcome distribution of target and OARs alike. Any treatment parameter, for which the uncertainty is quantifiable, can be seamlessly incorporated into our framework. By actively controlling the outcome distribution width during the optimization process, the finite amount of treatment fractions is considered. Here lies one of the major differences to Unkelbach and Oelfke.⁶ While their work aims to optimize mean and variance of the dose in each voxel, we strive to control the distribution of the treatment outcome metric as a whole.

When evaluating a treatment plan, the expert is provided with valuable additional information to facilitate the inevitable tradeoff between tumor control and normal tissue sparing, ultimately resulting in a robust custom tailored plan for each individual patient.

II. METHODS AND MATERIALS

II.A. Prerequisites

As the exact patient geometry configuration during treatment is unknown, the delivered dose becomes a statistical quantity. Due to the statistical nature of the delivered dose, all associated quality metrics become random variables, in the following denoted by Y for objectives and Z for constraints. Because the geometric uncertainties are specific for every individual patient, the probability distribution of the quality metrics for any given plan is also patient specific. Hence, it is worthwhile to consider the shape of the outcome metric distribution $p(y)$ instead of just a nominal value y for the outcome.

We postulate two basic properties, that any outcome distribution of a robust plan should possess (Fig. 1). First of all, it is important that the planned value of the metric and the mean of the distribution coincide, i.e., the delivered dose should resemble the planned one. If the treatment had *very* many fractions, this condition would suffice to ensure that

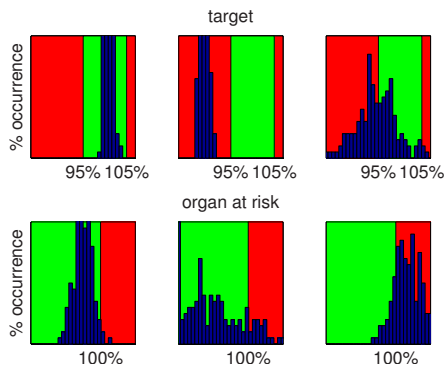


FIG. 3. Exemplary outcome distributions in relation to the aspired treatment outcome region. The top row illustrates the two-sided target cost function (e.g., the area of acceptable treatment outcomes could range from 95% to 105% of the prescribed dose). The distribution in the left panel is entirely located within the acceptable region, indicating that the patient would receive the proper dose during all courses of treatment. While the distribution in the middle panel is sufficiently small, its mean is too low, pointing toward an underdosage of the target. The distribution on the right panel is too wide and its mean too low. In practice, this means that there is a significantly reduced chance that the patient will receive proper treatment. The bottom panels show the one-sided OAR cost function (the area of acceptable outcomes ranges up to 100%). From left to right, the complication risk increases. Note that even though the mean in the middle panel is lower than on the leftmost panel, the overall distribution is less favorable.

the proper dose is delivered. In this case, sessions with a less than ideal delivery would be compensated by many others. However, in a realistic scenario, a failed treatment session cannot be compensated entirely. It is therefore also necessary to take the outcome distribution width into consideration. By demanding that the outcome distribution should be as narrow as possible, it is ensured that any sample taken from it, i.e., any treatment session, is within acceptable limits. This issue becomes more serious for fewer treatment sessions.

Consequently, any robust treatment optimization approach should keep track of the respective outcome distributions for the involved regions of interest. In case of the target, it is desirable to keep the distribution narrow and peaked at the desired dose. The organ at risk distributions, however, should be skewed toward low doses and only a controlled small fraction should be close or above the threshold. This ensures that in the vast majority of all possible treatment courses, the irradiation of the organ at risk is within acceptable limits (Fig. 3).

The R.O.B.U.S.T. algorithm extends our approach to intensity modulated radiotherapy optimization.²⁴ We employ the method of constrained optimization to maximize dose to the target while maintaining reasonable organ at risk dose levels. In this context, the k th constraint, which is a function of the dose distribution D , is denoted by $g_k(D)$. For the sake of simplicity, assume that there is only one objective, $f(D)$. All cost functions can be physical (maxDose or DVH) or biological (EUD or gEUD). The incident beams are discretized into beamlets or bixels. Each one of these beamlets is assigned a weight ϕ_j . The dose distribution is the product of the dose operator T and the beamlet weights ϕ . The ele-

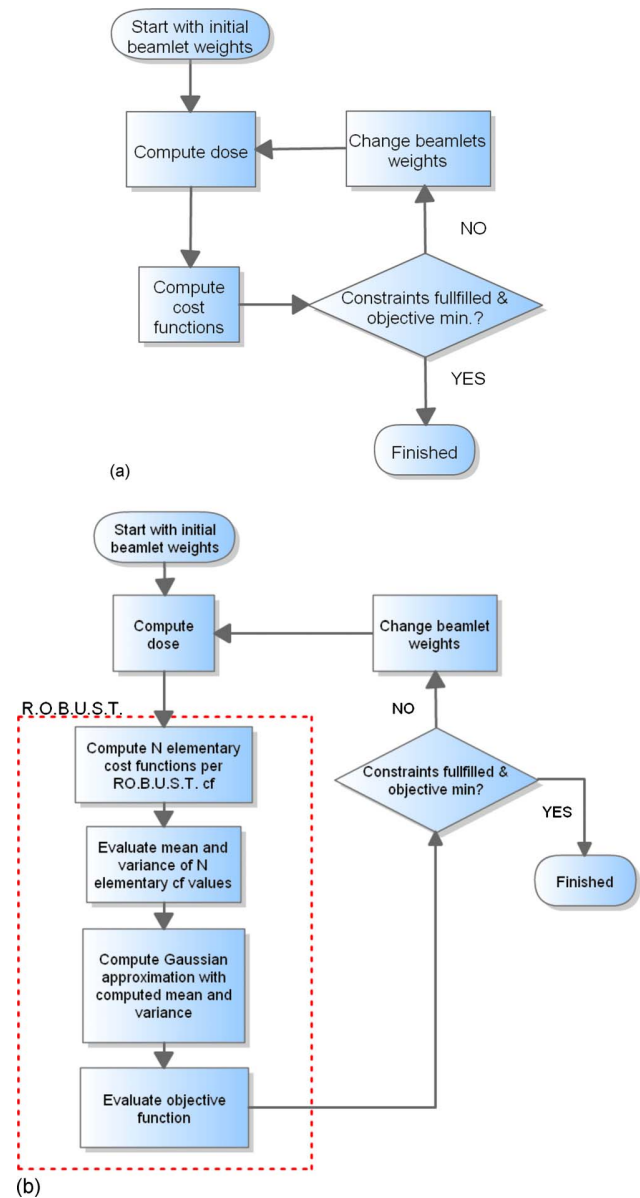


FIG. 4. (a) Constrained optimization in conjunction with classic cost functions. (b) Our probabilistic approach. Notice that the algorithm is entirely concealed from the optimizer, hence no alterations are necessary.

ment T_{ij} relates the dose deposited in voxel i by beamlet j , thus the dose in voxel i is $D_i = \sum_j T_{ij} \phi_j$. The entire optimization problem can now be stated as

$$\text{find } \phi \text{ such that } f(T\phi) \mapsto \min, \quad (1a)$$

$$\text{while } g_k(T\phi) \leq c_k, \quad (1b)$$

$$\text{and } \phi_j \geq 0 \quad \forall j, \quad (1c)$$

where c_k is the upper limit of constraint k . It is evident that the optimum depends on c . An outline of the algorithm is shown in Fig. 4(a).

II.B. Algorithm

The course of treatment is subject to several sources of uncertainty (e.g., setup errors). Suppose each of these sources is quantified by a geometry parameter (e.g., displacement distance), its value indicating a possible realization during treatment. The precise value of any of these geometry parameters is unknown at treatment time, making them statistical quantities. Thus as random variables X , they have associated probability distributions $p(\mathbf{x})$, indicating the probability that a certain value \mathbf{x} is realized. These input distributions need to be estimated prior to planning based either on pretreatment data, data acquired during the course of treatment, or population data.

The initial step is to obtain reliable estimates of the mean and the variance of the outcome distribution $p(y)$ given the input uncertainty $p(\mathbf{x})$. First a number of samples $\mathbf{x}_1, \dots, \mathbf{x}_N$ is drawn from the uncertainty distributions. Then, for all those geometry instances the outcome metric $y_k = f(D, \mathbf{x}_k)$, $0 \leq k \leq N$ is computed. If this is repeated sufficiently often, i.e., for large N , the treatment outcome distribution $p(y)$ can be obtained and alongside an estimate of the probability mass M present in the desired interval \mathbf{I}_{tol} .

$$M[Y] = \int d\mathbf{x} p(\mathbf{x}) \begin{cases} 1 & \Leftrightarrow f(\mathbf{x}) \in \mathbf{I}_{\text{tol}} \\ 0 & \Leftrightarrow f(\mathbf{x}) \notin \mathbf{I}_{\text{tol}} \end{cases} \quad (2)$$

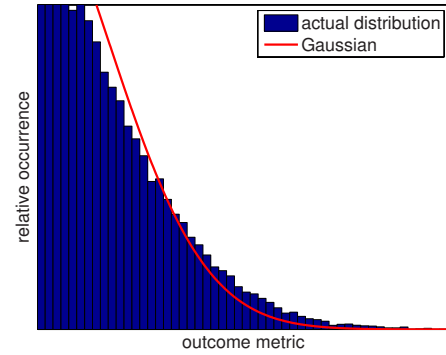
The interval \mathbf{I}_{tol} is the green area in Fig. 3. However, there are two major problems with this approach. First and foremost, for dependable estimates, this method becomes prohibitively expensive, especially with dose recalculations involved. Second, the quantity of interest, Eq. (2), cannot be computed in closed form, an essential requirement for efficient optimization schemes.

Typical prescriptions in radiotherapy demand that the interval contains large parts of the probability mass, i.e., 95% of all outcomes must lie within. That means that the exact shape of the actual distribution loses importance. Instead of trying to obtain the entire outcome distribution, we assume that its tails can be approximated by a Gaussian. The idea is shown in Fig. 5. Given this assumption, it is sufficient to compute mean $E_{\mathbf{x}}[Y]$ and variance $\text{var}_{\mathbf{x}}[Y]$. It is important to note that even if $p(y)$ does not resemble a normal distribution, the above approximation is still reasonable in the context of the proposed prescription scheme [Eq. (2)]. It is sufficient to assume that the tails of $p(y)$ (two-sided for target volumes, one-sided for OARs) can be described by the tails of the Gaussian $\mathcal{N}(y|E_{\mathbf{x}}[Y], \text{var}_{\mathbf{x}}[Y])$. This approximation becomes more powerful for fractionated treatments because of the *central limit theorem*.

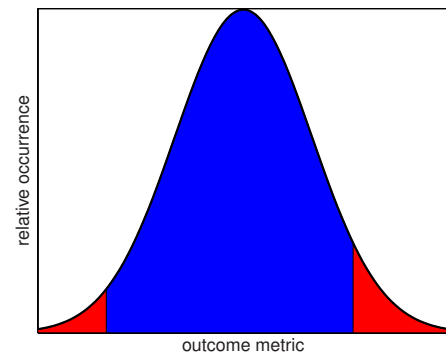
Incorporating the assumption above, we can state that the probability mass enclosed in the interval of interest can be computed analytically

$$M[Y] \approx \int_{\min \mathbf{I}_{\text{tol}}}^{\max \mathbf{I}_{\text{tol}}} dy \mathcal{N}(y|E_{\mathbf{x}}[Y], \text{var}_{\mathbf{x}}[Y]), \quad (3)$$

which is crucial for the formulation of an objective function for the optimizer. The new optimization problem reads



(a)



(b)

FIG. 5. (a) The tails of the actual outcome distribution are approximated with a Gaussian with identical mean and variance to enable us to formulate a cost function in closed form. (b) In case of the target, the area of the tails of the Gaussian, depicting undesirable outcomes, is minimized. In other words, the optimization aims to squeeze the treatment outcome distribution into the set interval.

$$\text{maximize } M[Y], \quad (4a)$$

$$\text{subject to } 1 - M[Z_k] \leq c_k, \quad (4b)$$

$$\text{and } \phi_j \geq 0 \quad \forall j. \quad (4c)$$

It is desirable to quantify the error introduced by the approximation in Eq. (3). The *Chebyshev inequality* provides for an estimate. Loosely speaking, it states that if the variance of a random variable is small, then the probability that it takes a value far from its mean is also small. If Y is a random variable with mean μ and variance σ^2 , then

$$\mathbf{P}(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad k \geq 1 \quad (5)$$

limits the probability of a sample being drawn k standard deviations away from the mean, *independently* of the density $p(y)$. For the problem at hand, k is calculated by determining how many standard deviations fit in the prescription interval \mathbf{I}_{tol} . If most of the probability mass is already concentrated in the desired area, its spread is relatively small compared to the extent of the acceptable region. This means that k is large, and by virtue of Eq. (5), the probability of a sample lying outside \mathbf{I}_{tol} is low. This holds *regardless* of what the real outcome distribution may look like. Since the Gaussian

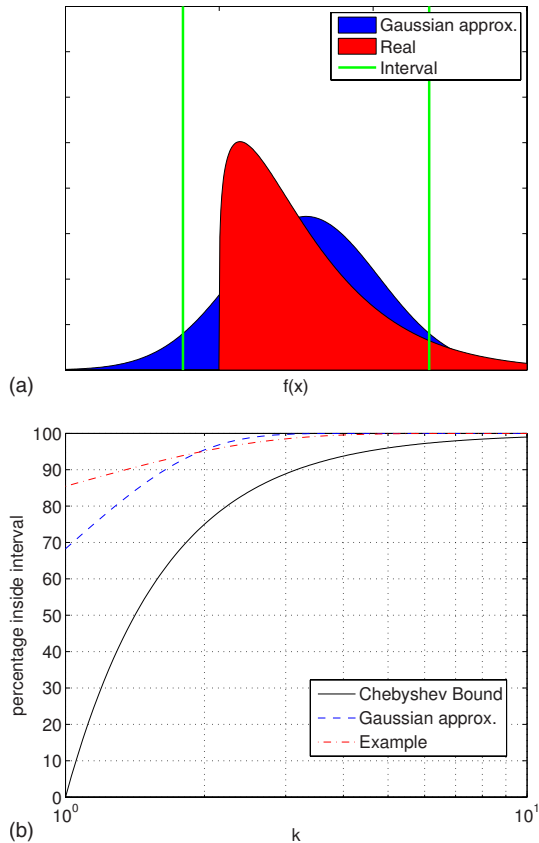


FIG. 6. (a) An arbitrary distribution (foreground) is approximated by a Gaussian (background) with identical mean and variance. (b) The error made by the Gaussian approximation is bounded from below by the Chebyshev inequality. The more probability mass is enclosed in the desired interval (increasing k), the tighter the bound becomes.

approximation tries to squeeze the probability mass into the interval, it points the optimizer into the right direction, independently of the actual shape of $p(y)$. Further refinements regarding the error bounds are possible, but beyond the scope of this paper.

In order to visualize the quality of the Chebyshev inequality, consider Fig. 6(a). Suppose the red area reflects the outcome distribution. The blue Gaussian included in the figure has identical mean and variance. The green bars are the upper and lower limits of the desired interval.

Figure 6(b) shows the behavior of the Chebyshev bound as a function of the interval width, i.e., k in Eq. (5). The red dashed curve is the area enclosed in the interval for our exemplary pdf. For illustrational purposes, we also included the Gaussian approximation, which serves as objective function for the optimizer. As the optimizer squeezes increasing amounts of the Gaussian inside the interval, it also inevitably forces the actual distribution inside. Notice that the actual distribution depends on the intermediate treatment plans of the iterative optimization, and hence constantly changes its shape. The more the dose distribution approaches the optimization goal, the less probability mass stays outside the prescription interval and the better the Gaussian approximation and the tighter the Chebyshev bound become.

The final algorithm is illustrated in Fig. 4(b). Its incorporation into an existing optimizer is straightforward, since the method can be encapsulated. In Fig. 4(b), the entire algorithm is encased in the red box which acts as a single cost function to the outside. In summary, the algorithm works as follows:

- (1) For each of the N geometry instances, the cost function f is computed, $y_k = f(D, \mathbf{x}_k)$, $1 \leq k \leq N$.
- (2) Mean $E_{\mathbf{x}}[Y]$, and variance $\text{var}_{\mathbf{x}}[Y]$ of those N samples are determined.
- (3) Use mean and variance to approximate a Gaussian $\mathcal{N}(y|E_{\mathbf{x}}[Y], \text{var}_{\mathbf{x}}[Y])$ (Fig. 5).
- (4) Compute the objective function, Eq. (3), and return the value to the optimizer.

II.B.1. Accounting for fractionation

We have mentioned earlier that the spread of the outcome distribution becomes increasingly important as we move toward smaller fraction numbers. This is substantiated by the *law of large numbers* in probability theory.²⁵ In this section, we quantify the previous observation and augment our cost function accordingly.

In the present context, the final outcome can be regarded as the sample mean (for details regarding this approximation, please see the Appendix) of all n sessions

$$A_n = \frac{\sum_{i=1}^n Y_i}{n}. \quad (6)$$

The random variable A_n has mean $E_{\mathbf{x}}[A_n] = E_{\mathbf{x}}[Y] \equiv \mu$ and variance $\text{var}_{\mathbf{x}}[A_n] = \text{var}_{\mathbf{x}}[Y]/n \equiv \sigma^2/n$. Again, we employ the Chebyshev inequality, Eq. (5), to obtain an approximation for the probability mass inside the prescription interval. With $k \equiv \varepsilon/\sigma$ and the mean and variance of A_n

$$\mathbf{P}(|A_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad (7)$$

is obtained, where ε is the maximum error we are willing to tolerate. We observe that the probability of a treatment violating the requirements, i.e., the realized outcome is more than ε away from the prescription, is proportional to the variance of a single fraction σ^2 , and goes to zero as the number of treatment fractions n increases. Further refinements can be made by reintroducing our Gaussian approximation. The accuracy of the approximation increases with the number of fractions n , as indicated by the central limit theorem.²⁵ It states that the sum of a large number of independent random variables is approximately normal. The actual convergence rate depends on the shape of Y_i . For example, if Y_i was uniformly distributed, $n \sim 8$ would suffice.²⁵ The ensuing probability mass is computed analogously to Eq. (3),

$$\begin{aligned}
M[A_n] &\approx \int_{\min \mathbf{I}_{\text{tol}}}^{\max \mathbf{I}_{\text{tol}}} dy \mathcal{N}(y | E_{\mathbf{x}}[A_n], \text{var}_{\mathbf{x}}[A_n]) \\
&= \int_{\min \mathbf{I}_{\text{tol}}}^{\max \mathbf{I}_{\text{tol}}} dy \mathcal{N}\left(y | E_{\mathbf{x}}[Y], \frac{\text{var}_{\mathbf{x}}[Y]}{n}\right). \quad (8)
\end{aligned}$$

By explicitly including fractionation effects in the cost function, Eq. (8), the influence of the variance is regulated. That is, we *quantitatively* account for the fact that a given number of fractions offers a certain potential for treatment errors to even out. Note that in the limit of very many fractions, we recover our initial statement that the mean value alone suffices to ensure proper treatment.

III. RESULTS

In this section, we compare R.O.B.U.S.T. to the PTV method and a method that controls the dose variance in each point of a volume, proposed by Unkelbach and Oelfke.⁷ We utilize a simplified 2D model implemented in MATLAB (The MathWorks, Inc., Natick, MA). It is driven by the constrained nonlinear multivariable function optimizer from the MATLAB optimization toolbox. For brevity, we only consider rigid movement in two directions. Recall that our method is by no means restricted to rigid movement; the inclusion of further sources of uncertainty is deemed unnecessary to illustrate the optimization problem and its solution. Our model consists of one tumor as well as two organs at risk. We employ biological cost functions.²⁶ The Poisson cell kill model was chosen for the target. The organs at risk were chosen to be represented by the parallel and serial complication model, respectively. There are four beams impinging on the target, each having 100 beamlets. To avoid artifacts such as a very high modulation of the beam profiles, a penumbra was also modeled [Jelen *et al.*,²⁷ Eq. (3)]. The input uncertainty was chosen to be independent Gaussians, i.e., $p(\mathbf{x}) = \prod_i p_i(x_i) = \prod_i \mathcal{N}(x_i | 0, \sigma_i^2)$. The variances σ_i^2 were all equal, implying isotropic movement. In order to compute mean and variance, 1000 sample geometries were drawn from X prior to the optimization. To save computation time, these were used throughout the optimization process. Under the assumption of a static dose distribution, the cost functions of each geometry were computed and subsequently mean and variance obtained.

The geometry is depicted in Fig. 7(a). In order to make the setup challenging, the tumor is embedded in the organs at risk. The parallel organ is located to the lower left side of the tumor, whereas the serial organ is on the right side.

III.A. Comparison to PTV based optimization

We computed several plans using the described model optimizer. First of all, we created a plan disregarding any uncertainties, i.e., all prescriptions were to the CTV, called the nominal plan. As expected, the possible treatment outcomes in terms of EUD for the tumor are wide spread [Fig. 7(b)]. Notice that the organ at risk distributions are mostly within the acceptable region.

Next we employed the PTV approach to increase the robustness in the target. The margin was set to one σ . As expected, the target outcome distribution is well localized around the prescription [Fig. 7(c)]. However, compared to the nominal plan, the dose to the adjacent organs is clearly elevated. Moreover, the potentially delivered dose in slightly shifted setups lies greatly beyond the nominal threshold. This emphasizes the fact that the PTV concept as a whole ignores the dose to the nearby organs.

The goal for R.O.B.U.S.T. was to fit the target outcome distribution in between 97% and 103% of the prescribed EUD. At the same time, no more than 10% of all possible outcomes were allowed to exceed the organ at risk constraints. The resulting distributions, along with their Gaussian approximations, are depicted in Fig. 7(d). While the target distribution closely resembles its margin counterpart, the EUDs of the organs at risk are significantly lower. Notice also that the target distribution of R.O.B.U.S.T. is left-skewed. This can be explained by taking a closer look at the OAR Gaussian approximations. Both already exceed the EUD limit by 10% [red part of the Gaussians in Fig. 7(d)], thereby preventing further improvements to the objective.

Table I shows a brief summary of the test runs.

III.B. Comparison to Unkelbach robust optimization scheme

In the following paragraphs, we briefly compare R.O.B.U.S.T. to the method of Unkelbach and Oelfke.⁷ They argue that the variance due to a finite amount of treatment fractions needs to be accounted for and they quantitatively incorporate the number of fractions in their cost function. According to Unkelbach and Oelfke,⁷ Eq. (7), the cost function for the target is

$$f_u(D) = \sum_{i \in \text{CTV}} (E_{\mathbf{x}}[D_i] - D_{\text{CTV}})^2 + \frac{1}{n} (E_{\mathbf{x}}[D_i^2] - E_{\mathbf{x}}^2[D_i]) \quad (9)$$

and for the k th organ at risk, they propose

$$g_{u_k}(D) = \sum_{i \in \text{OAR}_k} (E_{\mathbf{x}}[D_i] - D_{\text{OAR}_k})_+^2, \quad (10)$$

where $(\cdot)_+$ is the one-sided overdose, D_i is the dose to voxel i , and D_{CTV} and D_{OAR_k} are the prescriptions for target and organs at risk, respectively. Notice that not unlike Eq. (8), the influence of the variance decreases as the number of fractions increases in Eq. (9).

While Eqs. (9) and (10) are designed to tightly control the dose delivered to each voxel, the optimization of Eq. (3) and (8) implies control over the outcome distribution of an entire volume. In consequence, our approach allows intervolumetric tradeoffs of the mean dose per voxel, a basic requirement for biological cost functions. Their approach deals exclusively with the voxelwise quadratic deviation cost function $(D - D_0)^2$.

For the following comparison, we set $n=30$ fractions. Each point of the histograms in Fig. 8 consequently represents the average over 30 sample geometries. Note that the target distribution, Fig. 8(d), now lies entirely in the desired

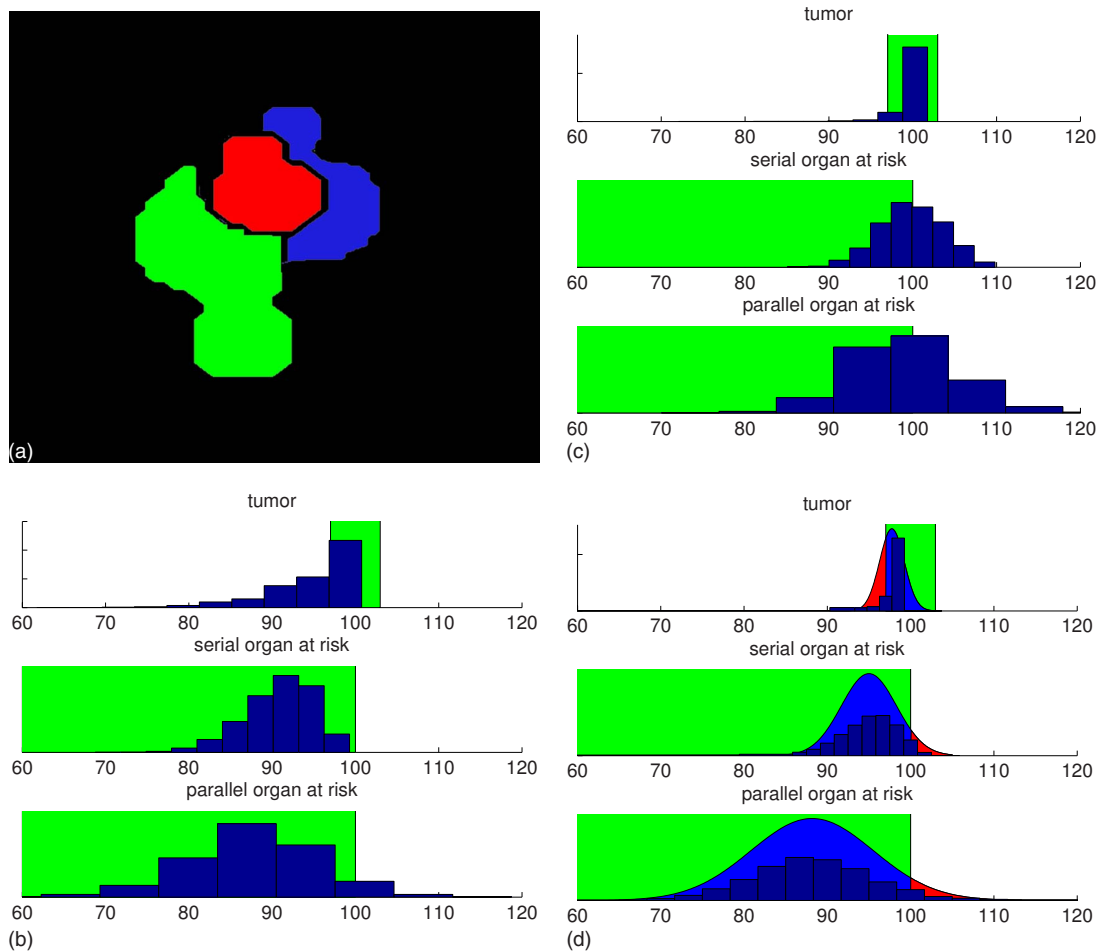


FIG. 7. The model geometry and treatment outcome histograms obtained from several methods. (a) The organ setup. Tumor (center, Poisson model), parallel organ (lower left), and serial organ (right). In the histograms [(b)–(d)], the highlighted area denotes the region of acceptable dosage. The horizontal axes show the percentage of the prescription for the respective organs. Note that these histograms apply for a single fraction. (b) The treatment outcome histogram for the plan without motion compensation. (c) The PTV margin greatly reduces the width of the target distribution, but worsens the organ at risk counterparts. (d) The outcome distributions along with their Gaussian approximation in the probabilistic case (R.O.B.U.S.T.).

interval [as opposed to Fig. 7(d)]. This is because the variance is rescaled by the number of fractions and therefore the OAR constraints allow for a higher mean EUD to the target. Also note that the distributions look Gaussian. This is a direct consequence of the central limit theorem.

For the Unkelbach approach, we initially set D_{CTV} to the target EUD value and D_{OAR_k} were chosen according to their

TABLE I. The first column indicates what fraction of treatment outcomes would be in the 97%–103% prescription interval [Eq. (2)]. The second and third columns show the percentage of treatment outcomes that are over the defined threshold for the serial and parallel organ at risk, respectively.

Goal	Tumor	Serial OAR	Parallel OAR
	All cases 97%–103% presc.	Max. 10% of cases Overconstraint	
No compensation	43%	0%	7%
PTV plan	96%	48%	47%
R.O.B.U.S.T. plan	92%	10%	9%

respective EUD tolerance doses. For better comparability, the optimized dose was slightly rescaled for the serial organ doses to match. The outcome distributions of the Unkelbach method are shown in Fig. 8(c). The target EUD distribution lies outside the desired interval, due to a persistent cold spot that is heavily penalized by the Poisson EUD. It is also significantly wider than its R.O.B.U.S.T. counterpart. Furthermore, the parallel organ is overdosed due to a relatively high mean dose.

We also examined the results of both methods based on the criteria suggested by Unkelbach and Oelfke.⁷ Figure 8(a) shows the DVH of the expected dose in the respective regions of interest. Both methods exhibit underdosage in the target. However, the Unkelbach method enforces stricter tradeoffs between the standard deviation and the mean dose in a voxel, which results in a larger volume being underdosed than by R.O.B.U.S.T., which increased the mean dose slightly to minimize the most extreme cold spots. This is also reflected by the respective EDVH.

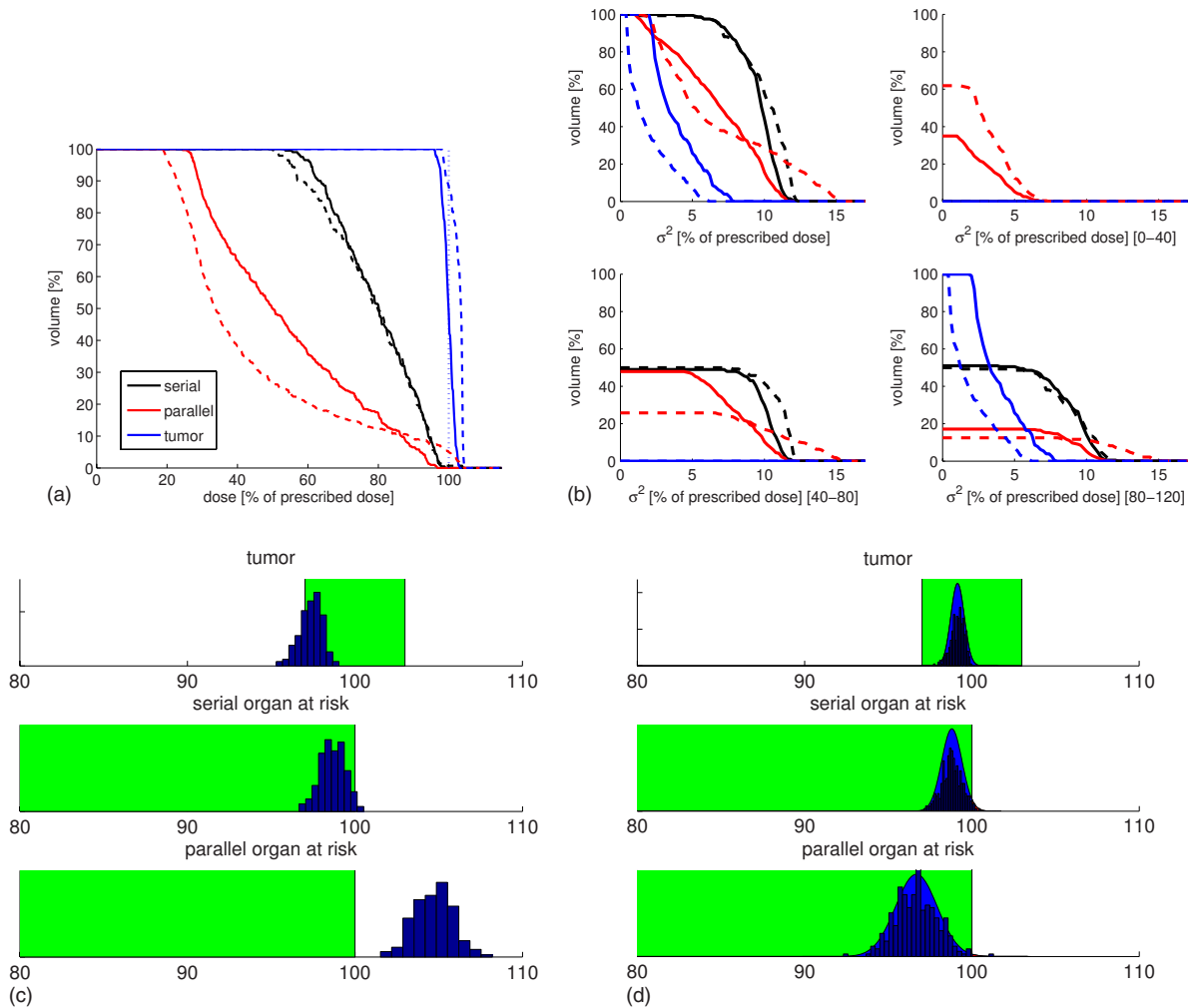


FIG. 8. A brief comparison of the Unkelbach method and R.O.B.U.S.T. The solid lines in (a) and (b) are from the Unkelbach method whereas the dashed lines stem from R.O.B.U.S.T. (a) The DVH of the expected dose. The dotted horizontal line shows the prescription for the target. (b) The standard-deviation-volume-histogram provides insight over the dose variability in each voxel. Note that in SDVHs, better graphs are further left, i.e., less variability in the dose. For more transparency, additionally to the comprehensive SDVH in the top left panel, we included three SDVHs, each for a different dose interval. The EUD treatment histograms for the (c) Unkelbach method and (d) R.O.B.U.S.T. It is important to note that these plots consider 30 fractions. Consequently, each entry in the histograms is the average over 30 randomly drawn sample geometries.

Additionally, Unkelbach and Oelfke⁷ propose to examine the standard-deviation-volume-histogram to visualize the uncertainty of the dose prediction provided by the expectation value. The SDVHs for R.O.B.U.S.T. and the Unkelbach method are shown in the top left plot of Fig. 8(b). R.O.B.U.S.T. succeeds for the target because the standard deviation is consistently lower than its counterpart from the Unkelbach method. The R.O.B.U.S.T. SDVH for the parallel organ, however, shows increased variability for about 30% of the volume. This can also be attributed to the dose gradient in the parallel organ as its cost function penalizes homogeneous irradiation. This can be seen in the top right plot in Fig. 8(b). The maximum standard deviation of the dose to the functional reserve, i.e., the part of the volume in the low dose region (0%–40% of prescribed dose), is similar to its counterpart from the Unkelbach method. The higher dose SDVHs, i.e., the bottom plots in Fig. 8(b), however, show an increased standard deviation which is acceptable because

the irradiated volume may be sacrificed. In essence, R.O.B.U.S.T. shapes the dose distribution such that the functional reserve is kept intact under all circumstances that may occur based on their respective probabilities $p(\mathbf{x})$.

Notice that in the absence of constraints, both methods tend to deliver near identical results (data not shown). However, as soon as constraints come into play, R.O.B.U.S.T. benefits from its larger space of potential solutions, due to the incorporation of nonlocal effects and absence of voxel-wise constraints.

IV. DISCUSSION AND CONCLUSION

We have presented a new approach toward robust treatment plan optimization. Accepting the fact that all outcome metrics are random variables in the presence of geometry uncertainties, it addresses robustness in the sense that under all possible uncertainties, the treatment outcome quality met-

rics should remain within a specified range across all treatment fractions. It focuses on controlling the shape of the outcome distribution instead of a nominal value of the metric. By quantifying the uncertainty of the treatment outcome for each metric, it provides a measure for the robustness of the plan.

If the optimization problem is stated like Eq. (3), precise knowledge of the outcome distribution is not necessary and the Gaussian approximation is possible by virtue of the Chebyshev inequality. This greatly reduces the effort of uncertainty sampling.

Because our method maintains different organ geometry setups, it is able to preserve correlations in geometric shifts. This means that there are no overlapping regions of interest in the patient model. Since it also explicitly takes the shape of the uncertainty distributions of the treatment parameters into account, it can detect a potentially hazardous scenario and assess its likelihood. This way, the plan will neither be overly conservative, nor lenient toward dangerous organ doses.

In this work, we investigated robustness regarding the treatment outcome in terms of EUD. The method relies on sampling the distribution of each source of uncertainty. R.O.B.U.S.T. is very versatile because it can handle all sources of error and is only limited by the available patient uncertainty model.

Accelerations in the evaluation of the R.O.B.U.S.T. cost functions to make the method suitable for clinical problems are possible and are subject of future work.^{28,29}

APPENDIX: DETAILS ON THE EVALUATION OF EUD ACROSS FRACTIONS

We have stated earlier that we regard the final treatment outcome EUD as the average of the EUDs of its constituent sessions. However, the quantity of interest is typically the EUD of the cumulative dose. Taking into account the respective properties of the EUD functions, one can establish a relation between the two quantities. Given a convex function f and a random variable D , *Jensen's inequality*³⁰ establishes a link between the function value at the average of the random variable and the average of the function values of the random variable, namely,

$$f(E[D]) \leq E[f(D)]. \quad (\text{A1})$$

The converse holds true if f is concave. With the general form of the EUD

$$\text{EUD}(D) = g^{-1}\left(\frac{1}{V} \sum_{v \in V} g(D_v)\right), \quad (\text{A2})$$

we can discriminate the following cases.

I. Serial complication

Given that the serial complication function $g(D)=D^a$, $a \geq 1$, renders Eq. (A2) a convex function,²⁶ it follows from Eq. (A1), that $\text{EUD}(E[D]) \leq E[\text{EUD}(D)]$. Hence, considering the average EUD as outcome, Eq. (6), always *overesti-*

mates the risk for the patient, i.e., the EUD of the mean dose is certain to be lower.

II. Poisson cell kill

The basic reasoning is similar to the serial complication mechanism. However, Eq. (A2) in conjunction with $g(D)=\exp(-\alpha D)$ is concave. Thus, the relation in Eq. (A1) reverses to $\text{EUD}(E[D]) \geq E[\text{EUD}(D)]$. That is, in case of the target we *underestimate* the cell kill by taking the average EUD [Eq. (6)].

III. Parallel complication

The parallel complication function $g(D) = (1 + (D_{\text{tol}}/D)^b)^{-1}$ is convex for $D \leq D_{\text{tol}}$ and concave for $D \geq D_{\text{tol}}$. In consequence, the above reasoning cannot be directly applied. However, typical prescriptions demand that large parts of the organ are kept intact, hence most of the organ volume is irradiated with $D < D_{\text{tol}}$. For this organ sub-volume, the estimate Eq. (6) is, again, an upper bound. For the part of the organ that may be sacrificed, the average EUD underestimates the risk. But keeping in mind that Eq. (6) overestimates the risk of the functional reserve being destroyed while *underestimating* the damage to tissue we are willing to sacrifice, this seems acceptable.

In summary, we have shown that using the average EUD as outcome metric for fractionated radiotherapy is viable because it leads to underestimation of tissue damage for the target and overestimation of tissue damage for the organs at risk.

^aElectronic mail: benjamin.sobotta@med.uni-tuebingen.de

¹D. Maleike, J. Unkelbach, and U. Oelfke, "Simulation and visualization of dose uncertainties due to interfractional organ motion," *Phys. Med. Biol.* **51**, 2237–2252 (2006).

²C. Baum, M. Alber, M. Birkner, and F. Nüsslin, "Treatment simulation approaches for the estimation of the distributions of treatment quality parameters generated by geometrical uncertainties," *Phys. Med. Biol.* **49**, 5475–5488 (2004).

³K. M. Langen and D. T. L. Jones, "Organ motion and its management," *Int. J. Radiat. Oncol., Biol., Phys.* **50**, 265–278 (2001).

⁴M. van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque, "The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **47**, 1121–1135 (2000).

⁵A. Niemierko, "Reporting and analyzing dose distributions: A concept of equivalent uniform dose," *Med. Phys.* **24**, 103–110 (1997).

⁶J. Unkelbach and U. Oelfke, "Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions," *Phys. Med. Biol.* **49**, 4005–4029 (2004).

⁷J. Unkelbach and U. Oelfke, "Incorporating organ movements in IMRT treatment planning for prostate cancer: Minimizing uncertainties in the inverse planning process," *Med. Phys.* **32**, 2471–2483 (2005).

⁸International Commission on Radiation Units and Measurements, "Prescribing, recording and reporting photon beam therapy," ICRU Report No. 50 (ICRU, Washington, DC, 1993).

⁹M. van Herk, P. Remeijer, and J. V. Lebesque, "Inclusion of geometric uncertainties in treatment plan evaluation," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 1407–1422 (2002).

¹⁰J. C. Stroom, H. C. J. de Boer, H. Huizenga, and A. G. Visser, "Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability," *Int. J. Radiat. Oncol., Biol., Phys.* **43**, 905–919 (1999).

¹¹A. McKenzie, M. van Herk, and B. Mijnheer, "Margins for geometric uncertainty around organs at risk in radiotherapy," *Radiother. Oncol.* **62**,

- 299–307 (2002).
- ¹²J. C. Stroom and B. J. M. Heijmen, “Limitations of the planning organ at risk volume (PRV) concept,” *Int. J. Radiat. Oncol., Biol., Phys.* **66**, 279–286 (2006).
- ¹³C. Baum, M. Alber, M. Birkner, and F. Nüsslin, “Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities,” *Radiother. Oncol.* **78**, 27–35 (2006).
- ¹⁴M. Birkner, D. Yan, M. Alber, J. Liang, and F. Nüsslin, “Adapting inverse planning to patient and organ geometrical variation: Algorithm and implementation,” *Med. Phys.* **30**, 2822–2831 (2003).
- ¹⁵M. Chu, Y. Zinchenko, S. G., Henderson, and M. B. Sharpe, “Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty,” *Phys. Med. Biol.* **50**, 5463–5477 (2005).
- ¹⁶A. Ólafsson and S. J. Wright, “Efficient schemes for robust IMRT treatment planning,” *Phys. Med. Biol.* **51**, 5621–5642 (2006).
- ¹⁷T. C. Y. Chan, T. Bortfeld, and J. N. Tsitsiklis, “A robust approach to IMRT optimization,” *Phys. Med. Biol.* **51**, 2567–2583 (2006).
- ¹⁸M. Söhn, M. Weinmann, and M. Alber, “Intensity-modulated radiotherapy optimization in a quasi-periodically deforming patient model,” *Int. J. Radiat. Oncol., Biol., Phys.* **75**, 906–914 (2009).
- ¹⁹D. L. McShan, M. L. Kessler, K. Vineberg, and B. A. Fraass, “Inverse plan optimization accounting for random geometric uncertainties with a multiple instance geometry approximation (MIGA),” *Med. Phys.* **33**, 1510–1521 (2006).
- ²⁰J. Yang, G. S. Mageras, S. V. Spirou, A. Jackson, E. Yorke, C. C. Ling, and C.-S. Chui, “A new method of incorporating systematic uncertainties in intensity-modulated radiotherapy optimization,” *Med. Phys.* **32**, 2567–2579 (2005).
- ²¹M. G. Witte, J. van der Geer, C. Schneider, J. V. Lebesque, M. Alber, and M. van Herk, “IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP,” *Med. Phys.* **34**, 3544–3555 (2007).
- ²²A. Trofimov, E. Rietzel, H.-M. Lu, B. Martin, S. Jiang, G. T. Y. Chen, and T. Bortfeld, “Temporo-spatial IMRT optimization: Concepts, implementation and initial results,” *Phys. Med. Biol.* **50**, 2779–2798 (2005).
- ²³J. Löf, B. K. Lind, and A. Brahme, “An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion,” *Phys. Med. Biol.* **43**, 1605–1628 (1998).
- ²⁴M. Alber and R. Reemtsen, “Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method,” *Optim. Methods Software* **22**, 391–411 (2007).
- ²⁵D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, 1st ed. (Athena Scientific, Belmont, MA, 2002).
- ²⁶M. Alber and F. Nüsslin, “An objective function for radiation treatment optimization based on local biological measures,” *Phys. Med. Biol.* **44**, 479–493 (1999).
- ²⁷U. Jeleń, M. Söhn, and M. Alber, “A finite size pencil beam for IMRT dose optimization,” *Phys. Med. Biol.* **50**, 1747–1766 (2005).
- ²⁸B. Sobotta, M. Söhn, and M. Alber, “Quantifying the robustness of IMRT treatment plans from a statistical optimization model to account for organ deformation,” *Radiother. Oncol.* **88**, S121–S122 (2008).
- ²⁹B. Sobotta, M. Söhn, and M. Alber, “What is the minimum number of geometry samples for a statistical patient model for robust treatment planning?,” *Radiother. Oncol.* **92**, S86 (2009).
- ³⁰J. L. W. V. Jensen, “Sur les fonctions convexes et les inégalités entre les valeurs moyennes,” *Acta Math.* **30**, 175–193 (1906).

Appendix E

On expedient properties of common biological score functions for multi-modality, adaptive and 4D dose optimization

published in

Physics in Medicine and Biology 2011; **56**: N123-N129

NOTE

On expedient properties of common biological score functions for multi-modality, adaptive and 4D dose optimization

B Sobotta¹, M Söhn², W Shaw³ and M Alber¹

¹ Section for Biomedical Physics, Radioonkologische Uniklinik, Universität Tübingen, 72076 Tübingen, Germany

² Department of Radiation Oncology, University Hospital Grosshadern, LMU München, 81377 München, Germany

³ Department of Medical Physics, Faculty of Health Sciences, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa

E-mail: benjamin.sobotta@med.uni-tuebingen.de

Received 1 January 2011, in final form 9 March 2011

Published 13 April 2011

Online at stacks.iop.org/PMB/56/N123

Abstract

Frequently, radiotherapy treatments are comprised of several dose distributions computed or optimized in different patient geometries. Therefore, the need arises to compute the comprehensive biological effect or physical figure of merit of the combined dose of a number of distinct geometry instances. For that purpose the dose is typically accumulated in a reference geometry through deformation fields obtained from deformable image registration. However, it is difficult to establish precise voxel-by-voxel relationships between different anatomical images in many cases. In this work, the mathematical properties of commonly used score functions are exploited to derive an upper boundary for the maximum effect for normal tissue and a lower boundary for the minimum effect for the target of accumulated doses on multiple geometry instances.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

The need to compute the combined effect of several dose distributions in distinct geometries arises in a number of applications. For example, some approaches to 4D IMRT optimization optimize dose simultaneously on multiple geometries (Birkner *et al* 2003, McShan *et al* 2006, Söhn *et al* 2009, Trofimov *et al* 2005) by warping the dose in each instance to a reference geometry, thereby realizing optimization in ‘tissue-eye-view’. Also in combined modality treatment, e.g. brachytherapy and external beam therapy, the overall effect has to be computed (Osorio *et al* 2010). Similar problems occur with the dose recomputation in mobile organs,

e.g. prostate on daily cone beam CT (Morin *et al* 2006, Rong *et al* 2010). Since particle therapy is very sensitive to geometry changes, especially in regions with heterogeneous density, doses have to be computed on several geometries and subsequently combined (Soukup *et al* 2009, Knopf *et al* 2010).

All of these methods require a correspondence between the volume elements (voxels) of each geometry instance to add up each dose contribution correctly. In general this requires deformable registration. Given the independently computed doses in each patient geometry, the voxel correspondence is used to warp the individual doses back to some reference geometry. In this reference geometry, the doses are accumulated and evaluated. Unfortunately, there are several issues with this approach. The computation of the deformation fields is fraught with uncertainties and difficult to verify. The computation frequently needs human surveillance and the data set may be incomplete or disturbed by artifacts. Some organs, especially bowel, are difficult to model because of variations in filling and the lack of fiducials. Incomplete knowledge of the extent of movement due to infrequent and sparse imaging may also affect the accuracy of the model. Additionally, special care must be taken to ensure energy conservation during dose warping and accumulation. Dose distributions are frequently analyzed with score functions that are a sum over all volume elements of an organ. In contrast to pointwise scores, a lot of the effort that goes into correct dose accumulation is washed out in the summation. Very often, these score functions are convex, which opens a possibility for very powerful estimations of their true value without going through the process of proper dose accumulation.

This note derives how the mathematical properties of commonly used cost functions, e.g. equivalent uniform dose (EUD) (Niemierko 1997), can be exploited to compute boundaries of the combined effect of the dose distributions in different geometries. These boundaries are computed *without* the need for deformable registration. For the target, lower boundaries will be given, and for the critical structures upper boundaries. Thus, a worst case scenario is obtained that may be used for a quick scoring of a treatment plan at only a fraction of the cost of a full analysis. The method is demonstrated on a prostate case with interfractional movement and a 4D lung case.

2. Methods and materials

Suppose F is a score function and let D_k be the dose distribution of the full treatment applied to the k th geometry instance I_k . For dose accumulation, a reference geometry I_0 is chosen. \tilde{D}_k denotes the warped dose in the reference geometry, i.e. the deformation field from I_k to I_0 applied to dose D_k . Hence, the quantity of main interest, the accumulated dose, is

$$E[\tilde{D}] = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{D}_k \quad \text{with} \quad \tilde{D}_0 \equiv D_0. \quad (1)$$

Note that the accumulated dose is expressed as an average in (1). Average and sum are equivalent because the dose distribution may be rescaled by a constant factor.

Dose warping as well as dose accumulation is a linear operation and therefore the dose integral of D_k over the volume of interest equals that of \tilde{D}_k . This is a direct consequence of energy and volume conservation. By virtue of this, and by way of definition

$$F(D_k) = F(\tilde{D}_k). \quad (2)$$

If no deformation field is available, the single doses D_k cannot be warped back to the reference geometry, and in other words \tilde{D}_k is inaccessible. However, by virtue of *Jensen's inequality* (Jensen 1906) for a convex function F ,

$$F(E[\tilde{D}]) \leq E[F(\tilde{D})] = E[F(D)] \quad (3)$$

can be established. Conversely, for concave functions F , (3) reverses to

$$F(E[\tilde{D}]) \geq E[F(\tilde{D})] = E[F(D)], \quad (4)$$

where $E[F(D)] \equiv 1/N \sum_{k=0}^{N-1} F(D_k)$ is the sum of the score function evaluated for each original geometry instance. The equality relations in (3) and (4) are due to (2).

Convex functions (curved upward) are commonly associated with normal tissues or other dose-limiting objectives. Examples include maximum dose, one-sided dose-limiting quadratic penalty and EUD for serial organs (Alber and Nüsslin 1999), i.e. $f(D) = D^a$, $a \geq 1$. Note that the mean dose is also covered by the inequality because it is synonymous with the serial EUD cost function with $a = 1$. The parallel EUD (Alber and Nüsslin 1999) is an exception because it is neither convex nor concave; however, it can still be handled within the proposed framework, see below. A commonly used cost function for the target volume, the Poisson cell kill EUD (Niemierko 1997, Alber and Nüsslin 1999), is concave.

In practice, this means that in the case of normal tissue the score function value of the accumulated dose ($F(E[\tilde{D}])$) is always smaller than the sum of the score functions of the instance doses ($E[F(D)]$) (3). For the target, the EUD of the accumulated dose is always larger than the sum of the EUDs of the instance doses (4). Hence, using Jensen's inequality, worst case boundaries for the score function of the accumulated dose can be derived without the need for a deformable registration model.

Note that EUD and other cost functions are volume averages and are therefore insensitive to changes of the absolute volume of some organ or tumor. Still, uncertainties in delineation or volume shrinkage can introduce errors into the above inequalities. In order to maintain the validity of the inequality, it is sufficient to formulate a worst case. For normal tissue, this would be equivalent to using the smallest volume of the organ of interest in any geometry instance to normalize EUD in each instance. In case of tumor shrinkage, it is helpful to resort to the mechanistic basis of the model. Tumor EUD is derived from the initial number of clonogenic cells, which can be expressed as the mean cell density times the tumor volume. Under the assumption that no clonogenic cells are lost from the initial population by effects other than those caused by radiation, the primary quantity cell number is preserved despite volume loss. A volume reduction then results in an increase of clonogenic cell density. For the purpose of inequality, the EUDs of different geometry instances may be added although they pertain to differently sized volumes since EUD is independent of cell density. However, if the quantity of interest were the expected number of surviving clonogens instead, the increase in cell density would have to be considered.

The parallel complication function $f(D) = (1 + (D_{\text{tol}}/D)^b)^{-1}$ is convex for doses below the dose D_{tol} and concave for doses above, i.e. it has a point of inflection at $D = D_{\text{tol}}$. In consequence, the above reasoning cannot be directly applied. However, typical prescriptions demand that large parts of the organ are kept intact; hence, most of the organ volume is irradiated with doses below D_{tol} . For this fraction of the volume, (3) applies, i.e. the damage is overestimated. To account for the entire volume of the parallel organ, the cost function can be convexified by linear extrapolation at the point of inflection, see figure 1.

Note that the above relation (3) is *not* valid for dose–volume constraints. Table 1 contains an overview of the properties of commonly used score functions.

3. Results and discussion

The presented method is valuable because it provides a safe upper boundary on normal tissue doses and effects, and a safe lower boundary for tumor EUD. It is applicable even if deformation fields are not available or tumor shrinkage makes warping and accumulation difficult.

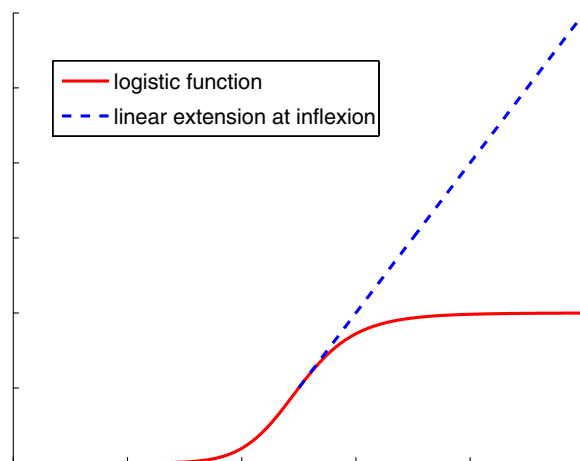


Figure 1. The parallel complication cost function (logistic function) is linearly extended at the inflexion point to be convex on its entire support.

Table 1. Properties of various cost functions.

Cost function	Property
Poisson cell kill EUD	Concave
Parallel EUD	–
Generalized EUD $k \geq 1$ (Serial EUD)	Convex
Generalized EUD $k \leq 1$	Concave
Average dose	Convex and concave
DVH constraints	–
Maximum dose	Convex
Quadratic deviation penalty	Convex

To illustrate one possible application, the method was applied to a prostate cancer case. Target, bladder and rectum (wall) EUDs were evaluated in 18 geometries and their averages computed, table 2. The biomechanical model, which is the basis of the deformation field for the dose accumulation, is based on Yan *et al* (1999). The EUD computations based on the accumulated dose and the estimated boundaries yielded 71.601 Gy (> 70.31 Gy) for the prostate, 59.897 Gy (< 61.28 Gy) for the bladder and 56.675 Gy (< 57.8 Gy) for the rectal wall. As predicted, for the target the average EUD in table 2 is lower than the EUD of the average dose. The contrary is the case for the organs at risk. Resultingly, a useful worst case approximation of the EUDs was obtained without dose accumulation.

To illustrate another application, the method was employed for a 4D lung cancer case. The EUDs for the gross tumor volume (GTV) (Poisson cell kill) and the lung (mean dose) were investigated. The basis were eight 4D CT scan instances taken during a breathing cycle. Each of these instances is associated with an instance weight that was computed from the breathing PDF. It reflects the relative fraction of a breathing cycle that is represented by a given instance. During dose accumulation, the doses of the respective instances were also

Table 2. EUDs for 18 patient geometries and their average.

CT #	Prostate	Bladder	Rectum
	Poisson EUD (Gy) $\alpha = 0.4$	Serial EUD (Gy) $a = 8$	Serial EUD (Gy) $a = 12$
1	71.73	59.24	63.7
2	70.81	58.61	49.56
3	55.28	62.62	42.41
4	71.7	57.21	57.35
5	70.68	62.80	48.75
6	71.7	60.07	60.9
7	71.63	62.49	54.89
8	64.18	63.97	58.16
9	71.76	60.43	59.44
10	71.5	58.14	64.41
11	71.87	63.72	58.39
12	71.79	60.96	61.16
13	71.63	61.04	62.23
14	71.95	63.76	58.08
15	71.88	64.43	59.16
16	71.67	61.48	58.12
17	71.87	61.65	62.31
18	71.88	60.38	61.46
	70.31	61.28	57.80

Table 3. Instance weights and EUDs for GTV and lung from a breathing cycle (inhale (In) exhale (Ex) respiratory levels in the range 0–100) and the weighted sum of the EUD.

Breathing phase	Weight	GTV EUD	Weighted GTV EUD	Lung EUD	weighted lung EUD
		Poisson EUD (Gy) $\alpha = 0.4$	Poisson EUD (Gy) $\alpha = 0.4$	Serial EUD (Gy) $a = 1$	Serial EUD (Gy) $a = 1$
0 In	0.402	48.029	19.312	4.66	1.874
25 In	0.088	48.127	4.229	4.48	0.394
50 In	0.077	49.656	3.800	4.37	0.334
75 In	0.084	49.661	4.190	4.24	0.358
100 Ex	0.117	48.250	5.629	4.29	0.501
75 Ex	0.033	50.718	1.682	4.38	0.145
50 Ex	0.048	49.851	2.379	4.43	0.211
25 Ex	0.152	49.661	7.528	4.51	0.684
			48.749		4.501

weighted. For details, please refer to Söhn *et al* (2009). The EUD of the accumulated dose is 49.835 Gy (> 48.749 Gy) for the GTV and 4.501 Gy (≤ 4.501 Gy) for the lung. Again the prediction in table 3 proves to be a worst case estimate. Note that the EUDs for the lung are equal for both methods by necessity, reflecting energy and volume conservation. This is due

to the fact that the mean dose score function is linear, hence concave and convex at the same time.

Further applications of this relation have been outlined in the introduction, whereby the combination of brachytherapy and teletherapy is of particular usefulness in practice. Here,

$$\text{EUD}(D_{\text{EBT}} + D_{\text{BT}}) \leq \text{EUD}(D_{\text{EBT}}) + \text{EUD}(D_{\text{BT}}) \quad (5)$$

for normal tissues and

$$\text{EUD}(D_{\text{EBT}} + D_{\text{BT}}) \geq \text{EUD}(D_{\text{EBT}}) + \text{EUD}(D_{\text{BT}}) \quad (6)$$

for targets. This work also lays the foundation for robust optimization schemes that require no deformation fields (Sobotta *et al* 2010). The application of Jensen's inequality to the optimization of doses for variable patient geometries is the subject of future work.

4. Conclusions

A method to derive boundaries on the score function of the accumulated dose of multiple patient geometries was presented. By virtue of Jensen's inequality and the mathematical nature of commonly used score functions, in particular EUD and other biological score functions, it is possible to compute a worst case approximation on the combined effect of doses applied to variable geometries. Because the approach circumvents the computation of the accumulated doses by means of deformation fields, it eliminates the need for deformable registration and dose warping.

Acknowledgment

We would like to thank D Yan and J Liang from Beaumont Hospital, Royal Oak, MI, USA, for providing the sample prostate cancer patient and the corresponding deformation fields.

References

- Alber M and Nüsslin F 1999 An objective function for radiation treatment optimization based on local biological measures *Phys. Med. Biol.* **44** 479–93
- Birkner M, Yan D, Alber M, Liang J and Nüsslin F 2003 Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation *Med. Phys.* **30** 2822–31
- Jensen J L W V 1906 Sur les fonctions convexes et les inégalités entre les valeurs moyennes *Acta Math.* **30** 175–93
- Knopf A *et al* 2010 Special report: workshop on 4D-treatment planning in actively scanned particle therapy—Recommendations, technical challenges, and future research directions *Med. Phys.* **37** 4608–14
- McShan D L, Kessler M L, Vineberg K and Fraass B A 2006 Inverse plan optimization accounting for random geometric uncertainties with a multiple instance geometry approximation (MIGA) *Med. Phys.* **33** 1510–21
- Morin O, Gillis A, Chen J, Aubin M, Bucci M K, Roach M and Pouliot J 2006 Megavoltage cone-beam CT: system description and clinical applications *Med. Dosim.* **31** 51–61
- Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10
- Osorio E M V, Hoogeman M S, Teguh D N, Al Mamgani A, Kolkman-Deurloo I K K, Bondar L, Levendag P C and Heijmen B J 2010 Three-dimensional dose addition of external beam radiotherapy and brachytherapy for oropharyngeal patients using nonrigid registration *Int. J. Radiat. Oncol. Biol. Phys.* at press
- Rong Y, Smilowitz J, Tewatia D, Tom W A and Paliwal B 2010 Dose calculation on KV cone beam CT images: an investigation of the Hu-density conversion stability and dose accuracy using the site-specific calibration *Med. Dosim.* **35** 195–207
- Sobotta B, Söhn M and Alber M 2010 Robust optimization based upon statistical theory *Med. Phys.* **37** 4019–28
- Söhn M, Weinmann M and Alber M 2009 Intensity-modulated radiotherapy optimization in a quasi-periodically deforming patient model *Int. J. Radiat. Oncol. Biol. Phys.* **75** 906–14

- Soukup M, Söhn M, Yan D, Liang J and Alber M 2009 Study of robustness of IMPT and IMRT for prostate cancer against organ movement *Int. J. Radiat. Oncol. Biol. Phys.* **75** 941–9
- Trofimov A, Rietzel E, Lu H M, Martin B, Jiang S, Chen G T Y and Bortfeld T 2005 Temporo-spatial IMRT optimization: concepts, implementation and initial results *Phys. Med. Biol.* **50** 2779–98
- Yan D, Jaffray D and Wong J 1999 A model to accumulate fractionated dose in a deforming organ *Int. J. Radiat. Oncol. Biol. Phys.* **44** 665–75

Appendix F

Special report: Workshop on 4D-treatment planning in actively scanned particle therapy - Recommendations, technical challenges, and future research directions

published in

Medical Physics 2010; **37**: 4608-4614

Special report: Workshop on 4D-treatment planning in actively scanned particle therapy—Recommendations, technical challenges, and future research directions

Antje Knopf^{a)}

Center for Proton Therapy, Paul Scherrer Institut, 5232 Villigen, Switzerland

Christoph Bert

GSI Helmholtzzentrum für Schwerionenforschung GmbH, Abt. Biophysik, 64291 Darmstadt, Germany

Emily Heath,^{b)} Simeon Nill, and Kim Kraus

Deutsches Krebsforschungszentrum, 69120 Heidelberg, Germany

Daniel Richter

GSI Helmholtzzentrum für Schwerionenforschung GmbH, Abt. Biophysik, 64291 Darmstadt, Germany

Eugen Hug

Center for Proton Therapy, Paul Scherrer Institut, 5232 Villigen, Switzerland and Proton-Radiotherapy, University of Zuerich, 8006 Zurich, Switzerland

Eros Pedroni, Sairos Safai, and Francesca Albertini

Center for Proton Therapy, Paul Scherrer Institut, 5232 Villigen, Switzerland

Silvan Zenklusen and Dirk Boye

Center for Proton Therapy, Paul Scherrer Institut, 5232 Villigen, Switzerland and ETH, 8092 Zuerich, Switzerland

Matthias Söhn,^{c)} Martin Soukup,^{d)} and Benjamin Sobotta

Section for Biomedical Physics, University Hospital for Radiation Oncology, 72076 Tuebingen, Germany

Antony Lomax

Center for Proton Therapy, Paul Scherrer Institut, 5232 Villigen, Switzerland and ETH, 8092 Zuerich, Switzerland

(Received 19 April 2010; revised 13 July 2010; accepted for publication 13 July 2010; published 10 August 2010)

This article reports on a 4D-treatment planning workshop (4DTPW), held on 7–8 December 2009 at the Paul Scherrer Institut (PSI) in Villigen, Switzerland. The participants were all members of institutions actively involved in particle therapy delivery and research. The purpose of the 4DTPW was to discuss current approaches, challenges, and future research directions in 4D-treatment planning in the context of actively scanned particle radiotherapy. Key aspects were addressed in plenary sessions, in which leaders of the field summarized the state-of-the-art. Each plenary session was followed by an extensive discussion. As a result, this article presents a summary of recommendations for the treatment of mobile targets (intrafractional changes) with actively scanned particles and a list of requirements to elaborate and apply these guidelines clinically. © 2010 American Association of Physicists in Medicine. [DOI: [10.1118/1.3475944](https://doi.org/10.1118/1.3475944)]

Key words: particle therapy, motion, 4D-treatment planning

I. OVERVIEW

The 4D-treatment planning workshop (4DTPW) focused on technical aspects of 4D-treatment planning and delivery in the context of intrafractional motion.¹ Consequences of interfractional motion are, for example, examined in Refs. 2 and 3. The topics which were discussed at the 4DTPW included

- methods to take into account intrafractional, nonrigid anatomical changes and resulting variations in densities through the use of deformable image registration and

dose calculation techniques, together with the inherent uncertainties of these approaches,

- advanced treatment planning approaches including 4D optimization for proton and heavy ion therapy,
- treatment simulations and experimental validation of scanning in the presence of motion to quantify the effects for different delivery strategies, and
- latest developments in beam tracking, gating, and rescanning.

This report is structured in two parts. First, in Sec. II, an overview of the current status of motion mitigation ap-

proaches for scanned particle therapy is given and areas for future investigation are identified. Second, in Sec. III, some suggested research directions (from the point of view of the participants of the workshop) in the field of 4D-treatment planning are summarized. Final comments are given in Sec. IV.

II. CURRENT OPTIONS FOR THE TREATMENT OF TARGET LOCATIONS SITUATED IN TEMPORALLY CHANGING ANATOMICAL SITES

There are many ideas of how to mitigate effects due to temporal changes in anatomy (e.g., breathing). Such changes can be taken into account during planning simply through the use of the internal target volume (ITV) approach⁴ or using the more sophisticated approach of 4D-inverse planning.^{5,6} Alternatively, treatments planned on a static geometry can be delivered in a way that compensates for changes of the anatomy. This approach is followed in beam rescanning, gating, and tracking. For the three latter techniques, the common expectation is that none of the methods represents the ultimate solution. Rather, different approaches have different advantages in different scenarios. Many of the motion management strategies rely on motion information acquired prior to the actual treatment (for example, 4DCT). In this case, one should always be aware that 4DCTs only provide a snapshot of the present motion and the resulting uncertainties at the time of treatment should be taken into consideration. The parameters which should be considered in the choice of the motion management approach include the following:

- The facility specifications such as scanning speed (energy modulation speed, time for lateral position adjustment, and dose rate), the available scanning method (raster-scan, spot-scan, and continuous scan), the deliverable dose resolution, the time latency (in actively controlled beam delivery), and the available imaging tools;
- patient specific parameters such as the target volume and location, surrounding critical structures, motion parameters, clinical status, and patient compliance, and
- the fractionation scheme used.

Below, we summarize the current status of motion management approaches in particle therapy and identify areas for future investigation.

II.A. Margin-based approach

The use of margins alone for motion management is likely not sufficient for scanned particle beams because of interplay effects.⁷ Nonetheless, margins and ITVs could be required also for scanned particle beam treatments since some motion management techniques are of limited precision and thus margins have to cover for the uncertainties. This is especially important for rescanning which does not perform any surveillance of the patient's motion and thus relies on margins. Implementation of the ITV, or any margin approach, using concepts defined for photon beams may well be invalid for protons and heavy ions due to the high sensitivity of the dose distribution to density variations in the entrance channel. Al-

ternatives to a margin definition in geometrical units (e.g., a 5 mm CTV-PTV margin) that is, for example, based on water-equivalent units, approaches incorporating the dependence of the dose on the density variations, are discussed in Refs. 8–11. The development of general margin recipes requires further investigation.

II.B. Rescanning

With rescanning,¹² the ITV/PTV is irradiated multiple times within each field delivery, with the rationale that interplay effects will be smoothed out and thus a homogeneous coverage of the CTV is achieved. However, this will inevitably be at the cost of therapeutic dose in the internal margins that form the PTV. Multiple delivery modes for performing rescanning exist, with the most prominent distinction being whether the scans are first repeated within a 2D energy layer (*slice-by-slice rescanning*) or the whole 3D volume is rescanned (*volumetric rescanning*).⁸ In both modes, different repainting strategies can be employed as, for example, scaled rescanning, for which the beam weight of each scan position is divided by the number of rescans or isolayered rescanning, for which fixed beam weights are applied per rescan.^{13,14} When implementing rescanning, particular awareness has to be given to machine specifications such as the scanning speed, the scanning method, and the dose deliverable resolution which can place limitations on the rescanning approach.

For the determination of the appropriate repainting strategy, the fractionation scheme, number of fields,¹⁵ as well as the number and temporal sequence of the energy layers contributing to a beam position should be considered. For a single-field-uniform-dose plan,¹⁶ all of these parameters have been shown to contribute an intrinsic rescanning effect, which should be considered in the decision of how many rescans and what type of rescanning should be employed. Extensive simulations and experimental work are required to comprehensively evaluate these effects for different patient and plan parameters. Synchronization effects due to the periodicity of target motion and individual rescans, especially for repetitive scaled rescanning, are a potential source of large underdosage in the CTV.^{8,17} Their impact has to be studied more extensively^{8,18} and possible solutions, such as the introduction of random pauses between rescans,¹⁶ have to be investigated.

II.C. Gating

Gating refers to an irradiation technique in which the beam is turned on during specific motion phases only, the so called gating window.^{19,20} Gating thus requires a motion monitoring system that determines the motion phase of the patient during treatment delivery. Reviews on motion monitoring methods are given by Murphy and Evans.^{21,22} Gating has been used for treatment with scattered particle beams for several years.^{19,20,23} For application to scanned beams, the residual motion within the gating window and the scanning process interfere, leading to a residual interplay effect.^{24,25}

To mitigate this interplay effect, a combination of rescanning²⁶ and an increased overlap between neighboring pencil beams have been proposed.²⁴

With respect to treatment planning, gating requires optimization of gating window size and the parameters that mitigate the dosimetric effect of the residual interplay. Robustness analysis to investigate the effect of baseline drifts, limited precision of the motion monitoring system, and variations in patient breathing parameters are also required.

II.D. Beam tracking

Beam tracking for particle therapy refers to the lateral adaptation of pencil beam position, combined with appropriate energy changes to modify the position of the Bragg peak²⁷ in order to compensate for target motion in all directions. Like gating, tracking relies on motion monitoring; ideally, real time 3D imaging should be employed or alternatively pretreatment acquired 4DCT in combination with a motion surrogate. For tracking, it is particularly important to minimize the time lag between the detection of a geometry change and the corresponding adaptation of the beam. Despite an experimental implementation of tracking for carbon beam radiotherapy,^{28,29} beam tracking has not yet been used clinically. Due to the strong dependence of the dose on the radiological depth, the adaptation parameters are currently determined prior to treatment delivery based on 4DCT data, which leads to concerns about variations in patient anatomy and respiratory motion between the 4DCT scan and the time of treatment. By minimizing the time between pretreatment imaging, potential online plan adaptation, and treatment delivery, these uncertainties can be reduced, but there is need to investigate the robustness of beam tracking, particularly over the time scales of fractionated delivery. The main points that should be investigated prior to clinical implementation are compensation of dose changes due to tissue deformation, alternative range adaption methods, robustness against uncertainties in the motion monitoring system, changes in the patient geometry, and respiratory motion variations during treatment delivery. For carbon ion beams, the impact of the relative biological effectiveness (which is also a concern for other motion management strategies) on the compensation vectors has to be investigated. Investigation of the tracking accuracy requirements for proton and heavy ion beams is also needed to provide criteria for future developments in beam control, motion prediction, and target tracking.

Especially for single-field-plans delivered in hypofractionated schema, gating and tracking might be advantageous as in this scenario, the probability to run into serious interplay effects for the rescanning approach is relatively high. In addition, gating and especially beam tracking most likely require smaller internal margins and can thus minimize the irradiation of normal tissues.

II.E. Other approaches

Manipulation of the patient's respiration, for example, by using breath hold techniques, has not yet been investigated for scanned beam delivery. Gantry 2 at PSI is designed to

scan an energy slice ($10 \times 10 \text{ cm}^2$) on a time scale of $\sim 210 \text{ ms}$ and a volume of 1 l within $\sim 6.5 \text{ s}$,³⁰ when no limits on the required dose rate are considered. However, the clinical application of 2 Gy to a volume of 1 l (for a typical beam current of 2.5×10^6 protons/ms, in continuous line-scanning mode and with ten rescans) should lead to a delivery time of approximately 2 min. In this scenario, each rescan could be delivered within a single breath hold. For a detailed discussion on possible scanning modes at the new Gantry 2 at PSI and the consequent delivery time, see Refs. 13 and 14. For hypofractionated radiotherapy, the use of apnea (a short induced suspension of the breathing) or jet-ventilation³¹⁻³³ to immobilize the patient respiration might be clinically justifiable.

In some situations, combinations of the abovementioned motion management methods, such as gated phase-controlled rescanning,²⁶ could be advantageous compared to a single approach. Assuming that the technical requirements for beam tracking, gating, and rescanning can be adequately addressed, the patient specific recipe of Rietzel and Bert⁸ can be followed, which bases the choice of mitigation technique on the motion characteristics. For instance, rescanning is recommended for target motion amplitudes below 5 mm and irregularly moving tumors. Gating, and later beam tracking, on the other hand is recommended for patients with larger motion amplitudes due to the increased CTV dose conformation.

In general, the choice of the management technique for a specific set of patient, machine, and plan parameters has to be investigated much more extensively. Qualitatively, different effects have been reported.^{27,34-37} The severity of motion effects for different treatment scenarios has, however, only been roughly quantified so far. For some of the treatment approaches currently used for stationary target tumors, temporal changes of the anatomy would likely result in only clinically negligible distortions. Furthermore, it is expected that different mitigation techniques have a different sensitivity to variations in motion and deformation.³⁸ Therefore, besides quantitative studies, detailed robustness studies are also needed to ultimately decide on the most suitable 4D-treatment approach, taking into account all the treatment parameters including fractionation scheme and the number of fields delivered per fraction.

III. GENERAL DEMANDS IN THE FIELD OF 4D-TREATMENT PLANNING

III.A. The development of a standardized 4D phantom

Over the past years, many institutes have developed 4D-treatment planning tools that incorporate temporal changes in the anatomy, and which can calculate the dose distributions resulting from different motion management techniques.^{9,39-41} The experimental verification of these 4D dose calculations, however, still remains a problem. In particular, this has been hampered by the lack of a realistic 4D phantom. So far, only a few attempts have been made⁴²⁻⁴⁴ to

construct a phantom satisfying the requirements of the whole 4D community and no commercial product is currently available for such measurements.

For 4D imaging, sophisticated phantoms exist; however, they do not incorporate dosimeters capable of measuring delivered dose.^{45,46} For IMRT, quality assurance tools are available, but are mainly limited to a single ionization chamber^{47–49} and they have not been tested for scanned particle beam delivery, where multiple measurement points are required to detect interplay patterns. For motion monitoring, phantoms have been constructed that enable external surface motion detection, but no internal dose detection has been incorporated.⁴⁸ In addition, motion phantoms have also been used for testing registration software. What is ultimately needed, however, is a commercial product that combines all these features. The currently available in-house solutions are limited in function, often difficult to handle, and are generally not suitable for performing interdepartmental measurements.

Ideally, a 4D phantom for dosimetric verifications should have at least the following specifications. It should

- be capable of being programmed to simulate nonrigid, irregular, and reproducible breathing patterns with varying periods and amplitudes;
- be capable of performing deformations involving realistic density changes;
- consist of tissue equivalent materials suitable for 4DCT and MR imaging;
- be able to simulate realistic surface behavior to extract external surface motion (to perform beam tracking and gating);
- be capable to hold radiochromic or light-shielded radiographic films in multiple positions for 3D dose measurements;
- be equipped with ionization chamber arrays for 4D dose measurements; and
- allow accurate measurements of the deformations (i.e., with implanted markers) at multiple points so that the accuracy of image registration can be evaluated.

The participants of the 4DTPW feel that it is an urgent need to formulate a universal product tender for a sophisticated 4D phantom. If such a tender is supported by many participating institutes, we believe there is a greater chance and motivation for a firm to develop such a phantom.

III.B. Experimental validation of 4D-treatment planning

Given the availability of a realistic 4D-treatment planning phantom, experimental validation is required to benchmark different 4D dose calculation approaches. However, the parameter space in 4D radiotherapy is too large to experimentally validate all possible treatment scenarios. Despite the growing number of particle therapy centers, the available beam time for proton and heavy ion measurements is limited. The participants of 4DTPW therefore propose that standardized experiments, similar to Ref. 25, should be established to efficiently benchmark 4D-treatment planning.

III.C. 4D database

Many institutes have developed their own dedicated 4D-treatment planning tools. Besides an experimental verification, an interinstitutional verification of the different tools would also be desirable. One of the reasons why that has not been attempted until yet is the absence of a commonly usable 4DCT data set(s). Only a few 4D patient CT data sets are available for straightforward public download, for example, the ten lung cases available from the MD Anderson Cancer Center.^{50,51} These images also contain a large number (>100) of manually identified landmarks for the purpose of deformable image registration evaluation. Another thorax data set with vector fields and landmarks is available for download via the Institut National des Sciences Appliquées (INSA) de Lyon.⁵²

For an interinstitutional 4D-treatment planning study, however, we emphasize the need for the establishment of a comprehensive open 4DCT database, containing 4DCT data sets of various target locations, defined volumes, and non-rigid transformation data. Another possibility is the creation of motion libraries to simulate specific motion scenarios following the work by von Siebenthal *et al.*^{53,54} In this approach, motion information acquired over relatively long time periods (1 h) by MRI can be applied to 3DCT data sets to obtain 4DCT data sets. Motion vectors for example cases are already accessible over the web.⁵⁵

A commonly available database would enable interinstitutional comparisons of 4D-treatment plans, excluding the perturbing influence of registration quality and different patient data sets. As a short term solution, it would be useful to define one, “standard” 4D CT data set for such comparisons. This could either be a real patient data set, for example, from the databases mentioned above,^{51,52} or data of an anatomically realistic virtual phantom. For the latter, the 4D NURBS-based cardiac-torso (NCAT) phantom could be used.^{56–59} The NCAT phantom models common patient motions such as the cardiac and respiratory motions in four dimensions and has already been used for 4D Monte Carlo treatment plan studies.⁶⁰ The use of one standardized anatomy would be the first step toward quantitative interinstitutional 4D-treatment plan comparisons.

III.D. Evaluation of different deformable registration methods

Deformable registration is needed for a variety of tasks to establish voxel correspondence, for example, in different respiratory phases. Sequential 3D and 4D imaging highlights the obvious nonrigid nature of motion of the human body, and therefore deformable registration is needed to accommodate the complex physiologic motion and changes in anatomy. Among other applications, it is used for contour propagation, dose accumulation, treatment adaption, dosimetric evaluation, and 4D optimization. One advanced approach enabling fully automatic, multimodality deformable registration was presented during the 4DTPW.⁶¹ Many participants of the 4DTPW have worked on deformable registration related topics.^{9,42,62}

As stated in Ref. 63, several methods of image registration are currently available, both commercially and in research environments. However, algorithmic accuracy has only been tested inconsistently, test cases are usually limited to a narrow range of clinical situations (e.g., imaging modality or disease site), and reported errors vary widely even for methods that are nominally similar. This confounds the ability to objectively benchmark algorithms and hinders the widespread adoption of such algorithms into clinical practice. Although each algorithm has its benefits and limitations, it is likely that the results of different studies could, at least partially, be dependent on the deformable registration algorithm used.

So far, only a few attempts have been made to provide a consistent quantitative metric of accuracy of such registrations which allow for direct comparisons of the various algorithms. One multi-institutional study conducted on a 4D phantom, although showing acceptable results overall, nevertheless had the potential for large errors in regions with large deformations.⁶⁴ A multi-institutional study on patient data highlighted the success achieved in deformable registration for anatomic sites imaged with consistent contrast (i.e., 4DCT of the lung) but also pointed out the work that must be done to ensure that these results are also obtained for multimodality images and anatomy that has less contrast variation (i.e., prostate). Furthermore, it was stated in this work that maximum errors, which were large for some algorithms, and especially their effect on applications (e.g., dose accumulation), should be investigated further.⁶³

It is out of the scope of this workshop to follow these suggestions or to perform more extensive quantifications of different deformable registration algorithms. However, when doing interinstitutional 4D-treatment plan comparisons, one has to be aware that a non-negligible source of differences is due to the different methods used for deformable registration. We therefore propose to establish a quantitative estimate of the influence of different deformable registration algorithms on the dose accumulation in 4D-treatment planning. Thus, in addition to a commonly usable 4DCT database, a standard 4D deformable registration tool could then be established.

III.E. Standards for 4D-treatment plan evaluation

During the workshop, several 4D-treatment planning studies were presented, reflecting the rising number of publications in this field. Besides the problem of different dose calculation codes, patient data sets and deformable registration algorithms used, the quantification of dosimetric effects due to temporal changes in the anatomy on 4D-treatment plans is also an open question. Generally, dose volume histograms and/or mean dose values are reported and D_5 – D_{95} values are calculated to evaluate the dose homogeneity in the target volume. However, combining all effects per plan into one curve or number probably does not reflect the entire clinical picture.^{65,66} Just as important as the existence of dose inhomogeneities is their spatial distribution and how they vary over time, especially in treatment situations affected by

temporal anatomy changes. For example, besides target dose homogeneity, the dose to neighboring organs also has to be evaluated.

Radiation therapy is inherently probabilistic.⁶⁷ One cannot be certain as to whether a tumor will be controlled or whether any given normal tissue will be damaged. These outcomes can be stated only in terms of probabilities, which is particularly the case when tumors are situated in areas of temporally changing anatomy. An overview of how to calculate and present the consequences of temporal changes in the anatomy has been given by Goitein.⁶⁸ To make a clinically relevant plan evaluation in these complex treatment situations, we propose that old concepts such as the calculation of upper-bound and lower-bound dose distributions, also known as worst case or best case dose distributions,⁶⁹ should be revisited and relevant metrics for plan evaluation should be established.

III.F. 4D-treatment plan optimization

Many of the participants of the 4DTPW have also worked on optimization related topics.^{70–74} In 4D-treatment planning, the parameter space is considerably larger than in 3D-treatment planning and the parameter space for the optimization of scanned particle beams is considerably larger than that for IMRT with photons. Treatment plan optimization is therefore expected to play an important role in reducing motion effects at the planning/optimization stage. Generally, 4D-inverse planning refers to methods which explicitly optimize the accumulated dose in the moving tissue (tumor and organs-at-risk) by using multiple CT instances (from a 4DCT) and the respective deformation fields, as well as the probability density function (pdf) of breathing motion.^{5,6} In fact, for photon radiotherapy, it has been shown that 4D-inverse planning results in comparable results to this of beam tracking⁷⁵ and gating.⁷³ For heavy ion/proton radiotherapy, however, only few studies have been performed,⁷⁶ and there are still many issues to be addressed until 4D-inverse planning could be used clinically. A drawback of the current approaches to 4D optimization is that treatment plans are usually optimized for the motion derived from the planning 4DCT, i.e., for one sample probability distribution only, while the breathing pattern and extent typically change on a daily basis. Quantification of the sensitivity of treatment plans to such uncertainties is an important issue especially in scanned particle therapy. Generally, it is desirable to explicitly incorporate measures of robustness into the planning process.^{71,72,77} Methods for robust optimization in the presence of breathing-pdf uncertainties^{74,78} and other uncertainties like setup errors and range uncertainties⁷⁹ have been proposed in this context.

So far, only a few attempts have been made to incorporate the “time” dimension explicitly in the optimization algorithm and to fully exploit it in terms of “optimized tracking,” where the treatment plan is optimized in all phases simultaneously.^{6,80} Furthermore, no consideration of the temporal aspects of the dose delivery has been included in the optimization approaches developed thus far. In a first step,

the magnitude of interplay effects for such first-order robust optimization approaches needs to be evaluated. Indeed, treatment plan optimization in the scope of 4D radiotherapy is an expansive topic itself and we believe deserves a dedicated workshop of its own.

III.G. Image guidance

Apart from the above mentioned topics related to 4D-treatment planning, we would like to emphasize the need for further development of image guidance in particle beam therapy. In-room volumetric imaging (as, for example, cone beam CT) is a fairly mature technology for photon therapy but there is a need to translate this technology to particle radiotherapy to quantify range uncertainties due to density changes or anatomical variations. Compared to photon beams, particle therapy, especially intensity modulated particle therapy, exhibit a higher sensitivity to patient misalignments. Thus, the expected gains from introduction of advanced image guidance in particle therapy is not only limited to intrafractionally moving targets but will most likely also improve the treatment of interfractionally moving as well as stationary targets.

IV. FINAL COMMENTS

This workshop provided a dynamic platform for discussing the current status of 4D-treatment planning in particle beam therapy. As a result, we have tried to identify the most urgent challenges in this field and the likely research opportunities that should be addressed. The topic of 4D-treatment planning is very complex. Clearly, not one group of researchers can solve such challenges on their own, neither in terms of manpower nor in financial regards. For projects like a standardized 4D phantom or a comprehensive 4D database, interinstitutional collaborations are urgently needed. A follow-up of this successful workshop is planned for the end of 2010 at the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany.

^{a)}Electronic mail: antje-christin.knopf@psi.ch

^{b)}Present address: Department of Physics, Ryerson University, Toronto M5B 2K3, Canada.

^{c)}Present address: Radiation Oncology, University Hospital Grosshadern, LMU 80539 Munich, Germany.

^{d)}Present address: CMS GmbH, 79100 Freiburg, Germany.

¹S. Mori, H. M. Lu, J. A. Wolfgang, N. C. Choi, and G. T. Chen, "Effects of interfractional anatomical changes on water-equivalent pathlength in charged-particle radiotherapy of lung cancer," *J. Radiat. Res. (Tokyo)* **50**(6), 513–519 (2009).

²Z. Hui, X. Zhang, G. Starkschall, Y. Li, R. Mohan, R. Komaki, J. D. Cox, and J. Y. Chang, "Effects of interfractional motion and anatomic changes on proton therapy dose distribution in lung cancer," *Int. J. Radiat. Oncol., Biol., Phys.* **72**(5), 1385–1395 (2008).

³M. Engelsman, S. J. Rosenthal, S. L. Michaud, J. A. Adams, R. J. Schneider, S. G. Bradley, J. B. Flanz, and H. M. Kooy, "Intra- and interfractional patient motion for a variety of immobilization devices," *Med. Phys.* **32**(11), 3468–3474 (2005).

⁴ICRU, "Prescribing, recording and reporting photon beam therapy (supplement to ICRU Report 50)," ICRU Report No. 62 (International Commission on Radiation Units and Measurements, Bethesda, MD, 1999).

⁵D. L. Mc Shan, M. L. Kessler, K. Vineberg, and B. A. Fraass, "Inverse plan optimization accounting for random geometric uncertainties with a

multiple instance geometry approximation (MIGA)," *Med. Phys.* **32**(5), 1510–1521 (2005).

⁶A. Trofimov, E. Rietzel, H. M. Lu, B. Martin, S. Jiang, G. T. Y. Chen, and T. Bortfeld, "Temporo-spatial IMRT optimization: Concepts, implementation and initial results," *Phys. Med. Biol.* **50**, 2779–2798 (2005).

⁷P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, C. R. Ramsey, M. B. Van Herk, S. S. Vedam, J. W. Wong, and E. Yorke, "The management of respiratory motion in radiation oncology report of AAPM Task Group 76," *Med. Phys.* **33**(10), 3874–3900 (2006).

⁸E. Rietzel and C. Bert, "Respiratory motion management in particle therapy," *Med. Phys.* **37**(2), 449–460 (2010).

⁹C. Bert and E. Rietzel, "4D treatment planning for scanned ion beams," *Radiat. Oncol.* **2**, 2–24 (2007).

¹⁰M. Engelsman, E. Rietzel, and H. Kooy, "Four-dimensional proton treatment planning for lung tumors," *Int. J. Radiat. Oncol., Biol., Phys.* **64**(5), 1589–1595 (2006).

¹¹M. Van Herk, P. Remeijer, C. Rasch, and J. Lebesque, "The probability of correct target dosage: Dose-population histograms for delivering treatment margins in radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **47**(4), 1121–1135 (2000).

¹²M. Phillips, E. Pedroni, H. Blattmann, T. Boehringer, A. Coray, and S. Scheib, "Effects of respiratory motion on dose uniformity with a charged particle scanning method," *Phys. Med. Biol.* **37**(1), 223–234 (1992).

¹³S. Zenklusen, E. Pedroni, and D. Meer, "Preliminary investigation for developing repainted beam scanning on the PSI Gantry 2," 47th PTCOG Conference, 2008.

¹⁴S. Zenklusen, E. Pedroni, and D. Meer, "A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning for the new Gantry 2 at PSI," *Phys. Med. Biol.* **55**, 1–19 (2010).

¹⁵A. Knopf, E. Hug, and A. Lomax, "Scanned proton radiotherapy for mobile targets—Systematic study on the need and effectiveness of re-scanning in the context of different treatment planning approaches and motion characteristics," *Int. J. Radiat. Oncol., Biol., Phys.* (submitted).

¹⁶T. F. DeLaney and H. Kooy, *Proton and Charged Particle Radiotherapy* (Lippincott Williams and Wilkins, Philadelphia), 2008.

¹⁷J. Seco, D. Robertson, A. Trofimov, and H. Paganetti, "Breathing interplay effects during proton beam scanning: Simulation and statistical analysis," *Phys. Med. Biol.* **54**, N283–N294 (2009).

¹⁸A. Knopf, E. Hug, and A. Lomax, "Re-scanned proton radiotherapy for various plan scenarios—The influence of patient specific parameters like motion amplitude, motion phase and target geometry," in Proceedings of ESTRO 29, 2010.

¹⁹H. D. Kubo and B. C. Hill, "Respiration gated radiotherapy treatment: A technical study," *Phys. Med. Biol.* **41**, 83–91 (1996).

²⁰S. Minohara *et al.*, "Respiratory gated irradiation system for heavy-ion radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **47**, 1097–1103 (2000).

²¹P. Evans, "Anatomical imaging for radiotherapy," *Phys. Med. Biol.* **53**, R151–R191 (2008).

²²M. Murphy, "Tracking moving organs in real time," *Semin. Radiat. Oncol.* **14**(1), 91–100 (2004).

²³H. M. Lu *et al.*, "A respiratory-gated treatment system for proton therapy," *Med. Phys.* **34**, 3273–3278 (2007).

²⁴C. Bert, A. Gemmel, N. Saito, and E. Rietzel, "Gated irradiation with scanned particle beams," *Int. J. Radiat. Oncol., Biol., Phys.* **73**(4), 1270–1275 (2009).

²⁵C. Bert, S. O. Grözinger, and E. Rietzel, "Quantification of interplay effects of scanned particle beams and moving targets," *Phys. Med. Biol.* **53**, 2253–2265 (2008).

²⁶T. Furukawa *et al.*, "Design study of a raster scanning system for moving target irradiation in heavy-ion radiotherapy," *Med. Phys.* **34**, 1085–1097 (2007).

²⁷S. O. Grözinger, C. Bert, T. Haberger, G. Kraft, and E. Rietzel, "Motion compensation with a scanned ion beam: A technical feasibility study," *Radiat. Oncol.* **3**, 3–34 (2008).

²⁸N. Saito, C. Bert, N. Chaudhri, A. Gemmel, D. Schardt, and E. Rietzel, "Speed and accuracy of a beam tracking system for treatment of moving targets with scanned ion beams," *Phys. Med. Biol.* **54**, 4849–4862 (2009).

²⁹C. Bert, N. Saito, A. Schmidt, N. Chaudhri, D. Schardt, and E. Rietzel, "Target motion tracking with a scanned particle beam," *Med. Phys.* **34**(12), 4768–4771 (2007).

³⁰E. Pedroni *et al.*, "The PSI Gantry 2: A second generation proton scanning gantry," *Z. Med. Phys.* **14**(1), 25–34 (2004).

- ³¹R. D. Sanders, "Two ventilating attachments for bronchoscopes," *Del. Med. J.* **39**, 170–192 (1967).
- ³²D. Bohn, "The history of high-frequency ventilation," *Respir. Care Clin. N. Am.* **7**, 535–548 (2001).
- ³³H. Hof, K. Herfarth, and M. Mütter, "Stereotactic single-dose radiotherapy of stage I non-small cell cancer (NSCLC)," *Int. J. Radiat. Oncol., Biol., Phys.* **56**(2), 335–341 (2003).
- ³⁴S. van de Water, R. Kreuger, S. Zenklusen, E. Hug, and A. J. Lomax, "Tumour tracking with scanned proton beams: Assessing the accuracy and practicalities," *Phys. Med. Biol.* **54**(21), 6549–6563 (2009).
- ³⁵S. Mori, T. Yanagi, R. Hara, G. C. Sharp, H. Asakura, M. Kumagai, R. Kishimoto, S. Yamada, H. Kato, S. Kandatsu, and T. Kamada, "Comparison of respiratory-gated and respiratory-ungated planning in scattered carbon ion beam treatment of the pancreas using four-dimensional computed tomography," *Int. J. Radiat. Oncol., Biol., Phys.* **76**(1), 303–312 (2010).
- ³⁶G. D. Hugo, J. Campbell, T. Zhang, and D. Yan, "Cumulative lung dose for several motion management strategies as a function of pretreatment patient parameters," *Int. J. Radiat. Oncol., Biol., Phys.* **74**(2), 593–601 (2009).
- ³⁷J. Seco, G. C. Sharp, Z. Wu, D. Gierga, F. Buettner, and H. Paganetti, "Dosimetric impact of motion in free-breathing and gated lung radiotherapy: A 4D Monte Carlo study of intrafraction and interfraction effects," *Med. Phys.* **35**(1), 356–366 (2008).
- ³⁸M. Soukup, M. Söhn, D. Yan, J. Liang, and M. Alber, "Study of robustness of IMPT and IMRT for prostate cancer against organ movement," *Int. J. Radiat. Oncol., Biol., Phys.* **75**(3), 941–949 (2009).
- ³⁹Y. Y. Vinogradskiy, P. Balter, D. S. Followill, P. E. Alvarez, R. A. White, and G. Starkschall, "Verification of four-dimensional photon dose calculations," *Med. Phys.* **36**(8), 3438–3447 (2009).
- ⁴⁰T. Roland, P. Mavroidis, A. Gutierrez, V. Goytia, and N. Papanikolaou, "A radiobiological analysis of the effect of 3D versus 4D image-based planning in lung cancer radiotherapy," *Phys. Med. Biol.* **54**(18), 5509–5523 (2009).
- ⁴¹G. Starkschall, K. Britton, M. F. McAleer, M. D. Jeter, M. R. Kaus, K. Bzdusek, R. Mohan, and J. D. Cox, "Potential dosimetric benefits of four-dimensional radiation treatment planning," *Int. J. Radiat. Oncol., Biol., Phys.* **73**(5), 1560–1565 (2009).
- ⁴²M. Serban, E. Heath, G. Stroian, D. L. Collins, and J. Seuntjens, "A deformable phantom for 4D radiotherapy verification: Design and image registration evaluation," *Med. Phys.* **35**(3), 1094–1102 (2008).
- ⁴³Y. Y. Vinogradskiy, P. Balter, D. S. Followill, P. E. Alvarez, R. A. White, and G. Starkschall, "Comparing the accuracy of four-dimensional photon dose calculations with three-dimensional calculations using moving and deforming phantoms," *Med. Phys.* **36**(11), 5000–5006 (2009).
- ⁴⁴M. Szegedi, P. Rassiah-Szegedi, J. Hinkle, and B. Salter, *A Proto-Type Design of a Real Tissue Phantom for the Validation of Deformation Algorithm and 4D Dose Calculations*, 51st AAPM Conference, 2009.
- ⁴⁵<http://www.qrm.de/content/products/dynamic.htm>
- ⁴⁶<http://www.cirsinc.com/008A.html>
- ⁴⁷http://www.modusmed.com/quasar_resp_phantom.html
- ⁴⁸<http://www.standardimaging.com/products.php?id=3>
- ⁴⁹http://www.rpdinc.com/html/imrt_dose_verification_phantom.html
- ⁵⁰R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Phys. Med. Biol.* **54**, 1849–1870 (2009).
- ⁵¹<http://www.dir-lab.com>
- ⁵²<http://www.creatis.insa-lyon.fr/rio/pop-i-model>
- ⁵³M. von Siebenthal, G. Székely, U. Gamper, P. Boesiger, A. Lomax, and P. Cattin, "4D MR imaging of respiratory organ motion and its variability," *Phys. Med. Biol.* **52**(6), 1547–1564 (2007).
- ⁵⁴M. Von Siebenthal, G. Székely, A. Lomax, and P. Cattin, *Inter-subject modelling of liver deformation during radiation therapy*, in Proceedings of the 10th International Conference on Medical Image Computing and Computer-Assisted Intervention—Volume Part I, Brisbane, Australia, 2007, edited by N. Ayache, S. Ourselin, and A. Maeder (Lecture Notes in Computer Science, Springer-Verlag, Berlin, Heidelberg, 659–666).
- ⁵⁵<http://www.vision.ee.ethz.ch/4dmri/index.shtml>
- ⁵⁶W. P. Segars, D. S. Lalush, and B. M. W. Tsui, "A realistic spline-based dynamic heart phantom," *IEEE Trans. Nucl. Sci.* **46**(3), 503–506 (1999).
- ⁵⁷W. P. Segars, D. S. Lalush, and B. M. W. Tsui, "Modeling respiratory mechanics in the MCAT and spline-based MCAT phantoms," *IEEE Trans. Nucl. Sci.* **48**(1), 89–97 (2001).
- ⁵⁸W. P. Segars, *Biomedical Engineering* (University of North Carolina Press, Chapel Hill, 2001).
- ⁵⁹http://dmip.rad.jhmi.edu/people/faculty/Paul/Segars_research.htm#simulation
- ⁶⁰R. McGurk, J. Seco, M. Riboldi, J. Wolfgang, P. Segars, and H. Paganetti, "Extension of the NCAT phantom for the investigation of intra-fraction respiratory motion in IMRT using 4D Monte Carlo," *Phys. Med. Biol.* **55**(5), 1475–1490 (2010).
- ⁶¹M. Söhn, M. Birkner, Y. Chi, J. Wang, Y. Di, B. Berger, and M. Alber, "Model-independent, multimodality deformable image registration by local matching of anatomical features and minimization of elastic energy," *Med. Phys.* **35**(3), 866–878 (2008).
- ⁶²Z. Wu, E. Rietzel, V. Boldea, D. Sarrut, and G. C. Sharp, "Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations," *Med. Phys.* **35**(2), 775–781 (2008).
- ⁶³K. K. Brock, "Results of a multi-institution deformable registration accuracy study (MIDRAS)," *Int. J. Radiat. Oncol., Biol., Phys.* **76**(2), 583–596 (2010).
- ⁶⁴K. Rojano et al., "Objective assessment of deformable image registration in radiotherapy—A multi-institution study," *Med. Phys.* **35**(12), 5944–5953 (2008).
- ⁶⁵A. Niemierko and M. Goitein, "Dose-volume distributions: A new approach to dose-volume histograms in three-dimensional treatment planning," *Med. Phys.* **21**(1), 3–11 (1994).
- ⁶⁶M. Goitein, *Radiation Oncology: A Physicist's-Eye View* (Springer, New York, 2007).
- ⁶⁷Journal of the ICRU **7**, 131–134 (2007).
- ⁶⁸M. Goitein, "Organ and tumor motion: An overview," *Semin. Radiat. Oncol.* **14**(1), 2–9 (2004).
- ⁶⁹M. Goitein, "Calculation of the uncertainty in the dose delivered during radiation therapy," *Med. Phys.* **12**, 608–612 (1985).
- ⁷⁰B. Sobotta, M. Söhn, M. Pütz, and M. Alber, "Tools for the analysis of dose optimization: III. Pointwise sensitivity and perturbation analysis," *Phys. Med. Biol.* **53**(22), 6337–6343 (2008).
- ⁷¹B. Sobotta, M. Söhn, and M. Alber, "Quantifying the robustness of IMRT treatment plans from a statistical optimization model to account for organ deformation," *Radiother. Oncol.* **88**, S121–S122 (2008).
- ⁷²B. Sobotta, M. Söhn, and M. Alber, "What is the minimum number of geometry samples for a statistical patient model for robust treatment planning?," *Radiother. Oncol.* **92**, S86 (2009).
- ⁷³M. Söhn, M. Weinmann, and M. Alber, "Intensity-modulated radiotherapy optimization in a quasi-periodically deforming patient model," *Int. J. Radiat. Oncol., Biol., Phys.* **75**(3), 906–914 (2009).
- ⁷⁴E. Heath, J. Unkelbach, and U. Oelfke, "Incorporating uncertainties in respiratory motion into 4D treatment plan optimization," *Med. Phys.* **36**(7), 3059–3071 (2009).
- ⁷⁵P. Zhang, G. D. Hugo, and D. Yan, "Planning study comparison of real-time target tracking and four-dimensional inverse planning for managing patient respiratory motion," *Int. J. Radiat. Oncol., Biol., Phys.* **72**(4), 1221–1227 (2008).
- ⁷⁶M. Söhn, M. Soukup, and M. Alber, "4D lung IMPT planning and evaluation on multiple deformable geometries," 48th PTCOG Conference, 2009.
- ⁷⁷J. Unkelbach and U. Oelfke, "Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions," *Phys. Med. Biol.* **49**, 4005–4029 (2004).
- ⁷⁸T. C. Y. Chan, T. Bortfeld, and J. N. Tsitsiklis, "A robust approach to IMRT optimization," *Phys. Med. Biol.* **51**, 2567–2583 (2006).
- ⁷⁹J. Unkelbach, T. Bortfeld, B. C. Martin, and M. Soukup, "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning," *Med. Phys.* **36**(1), 149–163 (2009).
- ⁸⁰L. Lee, Y. Ma, Y. Ye, and L. Xing, "Conceptual formulation on four-dimensional inverse planning for intensity modulated radiation therapy," *Phys. Med. Biol.* **54**, N255–N266 (2009).

Appendix G

Dosimetric treatment course simulation based on a patient-individual statistical model of deformable organ motion

submitted to

Physics in Medicine and Biology 2011;

Dosimetric treatment course simulation based on a statistical model of deformable organ motion

M Söhn^{1,2}, B Sobotta² and M Alber²

¹Department of Radiation Oncology, University Hospital Grosshadern, LMU München, Marchioninstr. 15, 81377 München, Germany

²Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany

E-mail: `matthias.soehn@med.uni-muenchen.de`

Abstract. We present a method of modelling dosimetric consequences of organ deformation and correlated motion of adjacent organ structures in radiotherapy. Based on a few organ geometry samples and the respective deformation fields as determined by deformable registration, Principal Component Analysis (PCA) is used to create a low-dimensional parametric statistical organ deformation model (Söhn *et al* 2005, PMB 50(24)). PCA determines the most important geometric variability in terms of eigenmodes, which represent 3D-vector fields of correlated organ deformations around the mean geometry. Weighted sums of a few dominating eigenmodes can be used to simulate synthetic geometries, which are statistically meaningful inter- and extrapolations of the input geometries, and predict their probability of occurrence.

We present the use of PCA as versatile treatment simulation tool which allows comprehensive dosimetric assessment of the detrimental effects that deformable geometric uncertainties can have on a planned dose distribution. For this, a set of random synthetic geometries is generated by a PCA model for each simulated treatment course, and the dose of a given treatment plan is accumulated in the moving tissue elements via dose warping. This enables the calculation of average voxel doses, local dose variability, DVH uncertainties, probability distributions of organ equivalent uniform doses (EUDs) and thus of TCP and NTCP, and other dosimetric and biologic endpoints.

The method is applied to the example of deformable motion of prostate/bladder/rectum in prostate IMRT. Applications include dosimetric assessment of the adequacy of margin recipes, adaptation schemes etc., as well as prospective ‘virtual’ evaluation of the possible benefits new radiotherapy schemes.

1. Introduction

Geometric uncertainties caused by patient set-up errors and internal organ motion lead to nontrivial deviations of the actually applied doses to the moving and deforming tumor and organs-at-risk (OARs) as compared to the planned static dose distribution. Methods that allow simulation of such dosimetric uncertainties are vital for an assessment of existing radiotherapy schemes and the potential benefit of new schemes in terms of prospective ‘virtual treatment simulations’. In such simulations planning parameters

like margin sizes, adaptation schemes etc. are systematically varied in order to determine optimal parameters for a given level of geometric uncertainties. Uncertainties caused by patient set-up errors and rigid organ motion can be modelled in terms of shifts and rotations. For this, methods for evaluation of treatment plan quality and dosimetric consequences have been proposed (Killoran *et al* 1996, van Herk *et al* 2002, Baum *et al* 2004, Gordon *et al* 2007). Probability distributions of the translational and rotational parameters can be derived from population data or estimated on a patient-individual level retrospectively. Usually, Gaussian distributions are assumed for the parameters.

Corresponding methods for modelling deformable organ motion are desirable. However, while rigid uncertainties can be described by 6 parameters, organ motion is of a seemingly much higher dimensionality. Therefore, it is not clear a-priori (1) how to parametrize and model deformable organ motion, (2) how to model correlated motion and deformation of adjacent organ structures, (3) how to define probability distributions over and assign probabilities to different deformation states, (4) how to obtain meaningful inter- and extrapolations of observed organ deformation samples, and (5) how to incorporate both patient-individual and population deformation data into a statistical deformation model. In a previous study, it was shown that *Principal Component Analysis* (PCA) can be used to compress the large and unorganized amount of data that results from several observations of organ deformation into a low-dimensional system of base vectors, the eigenmodes (Söhn *et al* 2005). As compared to the widely used approach to model rigid uncertainties by Gaussian distributions in 6-dimensional parameter space, PCA is its natural multidimensional generalization to *non-rigid, deformable organ motion*, which is able to address all of the challenges mentioned above. Organ geometries and -deformations are parametrized in terms of a collection of organ points and -displacement vectors which are used as input to the method. From this, PCA extracts the most important ‘deformation modes’ for the deformation patterns present in a number of input geometries. Each deformation mode is a normalized 3D-vector field, for which PCA provides a quantitative prediction of variability in terms of a Gaussian distribution. It has been shown that a small number of eigenmodes suffices to represent the observed variability of a typical sample set (Söhn *et al* 2005). Weighted superposition of a few dominating deformation modes applied to the mean geometry can be used to simulate new, synthetic geometries, which are statistically meaningful inter- and extrapolations of the input geometries with associated probability of observation. The idea to use PCA for low-dimensional statistical modelling of geometric variability has been applied to the calculation of coverage probabilities (Price and Moore 2007), to respiratory motion as represented by RCCT datasets (Zhang *et al* 2007), and to intra- and interfractional motion of structures in lung radiotherapy (Badawi *et al* 2010). Recently, the method of PCA has been extended to allow population-based deformation modelling (Budiarto *et al* 2011).

In the following, we describe how PCA can be used for dosimetric treatment course simulations including organ deformations. This enables a comprehensive assessment of treatment plan quality in terms of dose-to-moving-tissue for a variety of dosimetric

endpoints including biologic parameters such as EUD. Preliminary results of this research have been presented in Söhn *et al* 2009.

2. Methods

2.1. Statistical model of deformable organ motion based on PCA

2.1.1. Parametrization of organ geometries Given multiple 3-dimensional organ geometries as seen in a series of N CT scans of an individual patient, a straightforward parametrization of a geometry i in terms of an *organ geometry vector* $\mathbf{p}(i)$ is given by

$$\mathbf{p}(i) = (\vec{\mathbf{x}}_1(i), \dots, \vec{\mathbf{x}}_M(i)) \in \mathbb{R}^{3M}, i = 1 \dots N \quad (1)$$

of positions $\vec{\mathbf{x}}_j(i) \in \mathbb{R}^3$ of $j = 1 \dots M$ organ points.

An organ geometry vector \mathbf{p} may refer to a single anatomical organ or a composite geometry of multiple organs, where in the latter case $\mathbf{p}(i)$ in eq. (1) is to be understood as partitioned collection of points distributed over multiple organs. For the following it is important that for each point index j the series $\vec{\mathbf{x}}_j(1), \dots, \vec{\mathbf{x}}_j(N)$ refers to the same *anatomical* position, i.e. the set of geometry vectors $\{\mathbf{p}(i)\}_{i=1 \dots N}$ is a collection of *corresponding points*. Therefore, $\mathbf{p}(i_2) - \mathbf{p}(i_1)$, which is a field of M 3-dimensional vectors, represents the *deformation field* between the organ geometry instances i_1 and i_2 . Any suitable deformable registration method can be used for generating a set of organ geometry vectors $\{\mathbf{p}(i)\}_{i=1 \dots N}$ by applying the respective deformation fields to a reference geometry (say, the organ point positions $\mathbf{p}(1)$ as seen in the planning CT ‘1’). For population-based deformation modelling, *inter-patient* point-correspondence between different organ geometry samples needs to be established. A method to realize this has been described by Budiarto *et al* 2011.

2.1.2. Principal component analysis (PCA) of organ geometric change A comprehensive introduction into the theory of PCA-based statistical modelling of geometric variability can be found in our previous study (Söhn *et al* 2005).

In short, the set of organ geometry vectors $\{\mathbf{p}(i)\}_{i=1 \dots N}$ are regarded as random samples from a stochastic process of organ motion/deformation. Due to anatomical reasons the displacements of the M organ points are highly correlated, which implies that the underlying dimensionality of this multivariate statistical problem is actually much smaller than $3M$. A standard method from multivariate statistics for *dimensionality reduction* is PCA, which models the variability of this dataset as a *multivariate Gaussian distribution*, represented by the first two moments of the probability distribution, i.e. the *mean shape vector* $\bar{\mathbf{p}} \in \mathbb{R}^{3M}$,

$$\bar{\mathbf{p}} = \frac{1}{N} \sum_{i=1}^N \mathbf{p}(i), \quad (2)$$

and the *covariance matrix* $\mathcal{C} \in \mathbb{R}^{3M \times 3M}$,

$$\mathcal{C} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{p}(i) - \bar{\mathbf{p}}) \cdot (\mathbf{p}(i) - \bar{\mathbf{p}})^T, \quad (3)$$

where $() \cdot ()^T$ denotes the outer product of the two $3M$ -dimensional vectors. The covariance matrix \mathcal{C} can be interpreted as ‘generalized standard deviation’ for high-dimensional datasets.

Being interested in correlations in the displacements of the M organ points, the covariance matrix \mathcal{C} is diagonalized: $\tilde{\mathcal{C}} = \text{diag}(\lambda_1, \dots, \lambda_{3M})$. The resulting *eigenvectors* $\mathbf{q}_l \in \mathbb{R}^{3M}$ of the covariance matrix are the mutually independent modes of correlated deformation, also called *eigenmodes*. Each eigenmode represents a 3D-vector field of correlated displacements for the $3M$ organ points. Each *eigenvalue* $\lambda_l \in \mathbb{R}$ quantifies the variance of the N geometry samples $\mathbf{p}(i)$, eq. (1), in the ($3M$ -dimensional) direction of the corresponding eigenmode, and the *relative eigenvalue* $\tilde{\lambda}_l = 100\% \cdot \lambda_l / \sum_{l'=1 \dots 3M} \lambda_{l'}$ is a measure for the fraction of the overall geometric variability modelled by the l -th eigenmode, and thus for its relative importance for representing the geometric variability present in the input samples. Thus, the eigenmodes with largest eigenvalues (the *dominating eigenmodes*) capture the ‘*principal*’ *deformation modes* that span the space of largest dataset variability, i.e. the space in which the majority of deformations occurs.

2.1.3. Construction of organ geometries using eigenmodes New geometry samples can be generated within the low dimensional ‘deformation space’ by deforming the mean shape with a weighted sum of $L \ll 3M$ dominating eigenmodes

$$\mathbf{p}_{simul} = \bar{\mathbf{p}} + \sum_{l=1}^L c_l \cdot \mathbf{q}_l, \quad \|\mathbf{q}_l\| = 1 \quad (4)$$

According to the theory of PCA, the coefficients $c_l \in \mathbb{R}$ are drawn from Gaussian distributions with the corresponding eigenvalues λ_l as variances: $\sigma_l^2 = \lambda_l$. Thus, a simulated organ geometry \mathbf{p}_{simul} is associated with the following realization probability:

$$P(c_1, \dots, c_L) = \prod_{l=1}^L \frac{1}{\sqrt{2\pi\lambda_l}} \exp\left(-\frac{1}{2}c_l^2/\lambda_l\right) = \prod_{l=1}^L P(c_l) \quad (5)$$

2.2. PCA for dosimetric treatment simulations

By repeatedly drawing random sets of coefficients $\{c_1, \dots, c_L\}$ from the Gaussian probability distribution, eq. (5), a multitude of random geometry samples $\{\mathbf{p}_{simul}^{[k]}\}$ (with k as sample index) can be generated according to eq. (4). In this way PCA serves as statistical model of organ motion/deformation with only a small number L of parameters. As this model explicitly provides information about the correlated motion of all organ points/tissue elements, this can be used for dosimetric evaluation of simulated treatment courses and thereby for evaluation of the actual dose coverage of moving

organ structures as described in the following.

Let

$$\mathbf{d}[\mathbf{p}, D] = (d_1, \dots, d_{M_D}) = (D(\vec{\mathbf{x}}'_1), \dots, D(\vec{\mathbf{x}}'_{M_D})) \in \mathbb{R}^{M_D} \quad (6)$$

be the *dose vector* associated with an organ geometry vector \mathbf{p} , i.e. d_j is the dose value of a given dose distribution D at the position $\vec{\mathbf{x}}'_j$ within an organ. In the simplest case these positions coincide with the M organ points represented by the organ geometry vector \mathbf{p} : $\vec{\mathbf{x}}'_j = \vec{\mathbf{x}}_j \quad \forall j = 1 \dots M$. Generally the M_D positions $\vec{\mathbf{x}}'_j$ in eq. (6) can be interpolated from the M organ points, if required by reasons of accuracy, speed etc. of the specific implementation (see results section for an example).

Generating N_{frac} random organ geometries $\{\mathbf{p}_{simul}^{[k]}\}_{k=1 \dots N_{frac}}$ based on the PCA model, the corresponding accumulated dose for a single simulated treatment course with N_{frac} treatment fractions is

$$\mathbf{d}_{accum} = \sum_{k=1}^{N_{frac}} \mathbf{d}[\mathbf{p}_{simul}^{[k]}, D^{[k]}] \quad (7a)$$

$$\approx \frac{1}{N_{frac}} \sum_{k=1}^{N_{frac}} \mathbf{d}[\mathbf{p}_{simul}^{[k]}, D_{planned}]. \quad (7b)$$

Eq. (7a) takes into consideration that the physical dose distribution itself is subject to changes along with the patient's anatomical changes ($D^{[k]}$ denotes the fractional dose distribution of treatment fraction k). To accelerate computations further, this effect can be neglected for photon irradiation of deep seated targets in homogeneous tissue, where the fractional dose distribution $D^{[k]}$ can be represented by the static dose distribution $D_{planned}$ as calculated on the planning CT: $D^{[k]} \approx D_{planned}/N_{frac}$ (Birkner 2002, Birkner *et al* 2003). In this case \mathbf{d}_{accum} can be calculated according to eq. (7b) as will be assumed in the following. Note that eqs. (7a) and (7b) formally represent dose accumulation of fractional doses as *warped* to the moving 'tissue coordinate system'.

In eq. (7b), \mathbf{d}_{accum} is based on a single set $\{\mathbf{p}_{simul}^{[k]}\}_{k=1 \dots N_{frac}}$ of random geometries, which only represents one of many possible realizations of a treatment course with N_{frac} fractions. Thus, \mathbf{d}_{accum} itself is a random variable, the statistics of which is accessible via $S \gg 1$ repeat simulations of single treatment courses according to eq. (7b) by means of a *Monte Carlo (MC) approach*:

$$\text{MC-simulations: } \{\mathbf{d}_{accum}^{[s]}\}_{s=1 \dots S} \quad (8)$$

This quantity is the basis of a variety of informative treatment plan quality measures:

- *voxel-wise dose statistics*: The dose statistics can be evaluated on a per-voxel-/tissue-element-basis from eq. (8) by assessing the distribution $\{d_{accum,j}^{[s]}\}$ of dose values for the M_D dose sampling points. This allows calculation of the *expected value* $\langle d_j \rangle$ and its standard deviation Δd_j for organ point j :

$$\langle d_j \rangle = \frac{1}{S} \sum_{s=1}^S d_{accum,j}^{[s]} \quad (9)$$

$$\Delta d_j = \frac{1}{S-1} \sum_{s=1}^S (d_{accum,j}^{[s]} - \langle d_j \rangle)^2 \quad (10)$$

- *Dose-volume histograms (DVHs)*: The DVH of the expected organ dose distribution as well as DVH-uncertainties induced by organ motion/deformations can be assessed by plotting the DVHs associated with the dose vector $\langle \mathbf{d} \rangle$, eq. (9), resp. the dose vectors of the S simulated treatment courses, eq. (8). Besides the dose values d_j , calculation of a DVH requires information about the *relative organ volume* ν_j associated with each dose sampling point j , which is determined in the mean geometry $\bar{\mathbf{p}}$.
- *EUD statistics: Equivalent uniform doses (EUDs)* and other dosimetric measures derive from the DVH. In this context it is a special strength of the PCA approach that multiple organs can be modelled simultaneously in terms of a composite PCA model as described in sect. 2.1.1. Thus not only single organ EUD-distributions, but also *joint probability distributions* of EUDs for multiple organs are accessible. The following two EUD-definitions are used in this study: The poisson cell-kill EUD derived from the standard Poisson model of TCP (Niemierko 1997),

$$\text{EUD}_{poisson} = -\frac{1}{\alpha} \ln \left(\sum_i \nu_i \exp(-\alpha D_i) \right) \quad (11)$$

with the cell-sensitivity parameter α ; and the generalized EUD as dosimetric measure for serial organs-at-risk (Niemierko 1999),

$$\text{EUD}_{serial} = \left(\sum_i \nu_i D_i^a \right)^{1/a} \quad (12)$$

with the volume-effect parameter $a \geq 1$.

3. Materials

As an example, the method was applied to the dataset of a prostate cancer patient. The patient received a planning CT scan as well as multiple repeat CT scans that were acquired immediately before or after treatment delivery. The repeat CT scans were matched to the planning CT scan using bony anatomy, thereby reducing the geometric variability to be described by the PCA model to internal organ motion/deformation. The contours of prostate, bladder and rectal wall were manually delineated for each CT dataset (fig. 1a).

The sets of corresponding points required as input data for the PCA model were generated by the finite element method (FEM) described in Yan *et al* 1999 as applied to the above contour data (see also Söhn *et al* 2005). This deformable registration method is based on a biomechanical model of tissue and provides the position of each tissue element of the organs in the repeat CT scans, when the positions of corresponding boundary points on the organ surfaces are given as input data. For this method, the organs were decomposed into small tetrahedral subvolumes,

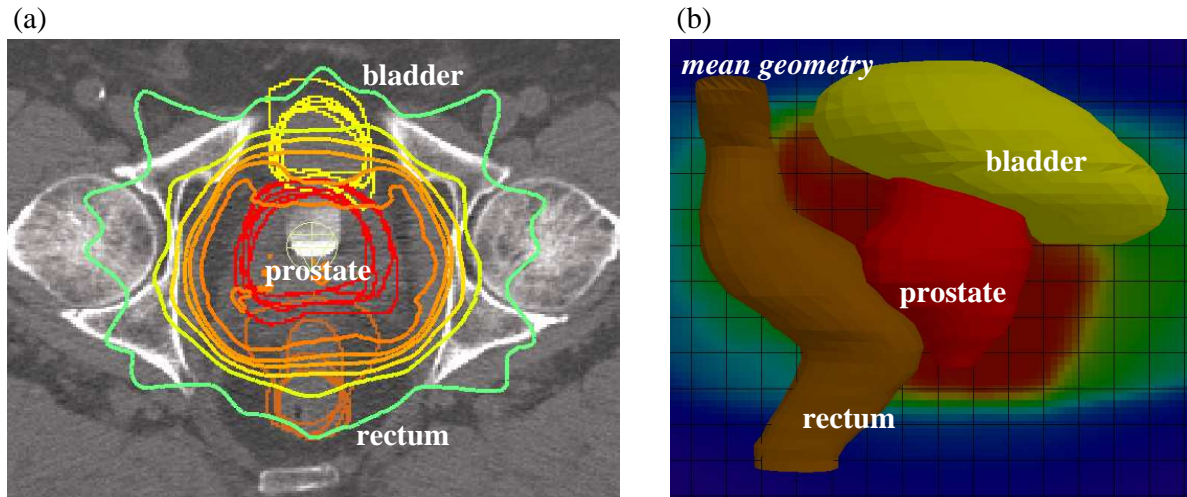


Figure 1. (a) Transversal CT slice of the example patient with contours for bladder, prostate and rectum as seen in the planning CT and the first four daily pretreatment CTs. These organ geometries were used as input for the PCA model. The dose distribution of the example IMRT plan is shown in terms of the 70.0/66.5/63.0/56.0/49.0/35.0 Gy isodose lines. (b) The mean geometry of the PCA model as sagittal 3D-view in front of the dose distribution.

resulting in 30240/19437/38304 tetrahedrons with 6253/3740/8379 nodes representing the prostate/bladder/rectal wall of the example patient. The altogether $M = 18372$ node positions in $N = 5$ CT datasets (planning CT scan plus four repeat CT scans of the first treatment fractions) were used to form composite organ geometry vectors $\{\mathbf{p}(i)\}_{i=1\dots N}$ as input data for a simultaneous PCA of all three organs. Notice, that in this way correlated motion of these three organs is explicitly modelled by the PCA model, which would not be the case when three *separate* PCA models were created for the single organs.

An example IMRT plan was created as basis for dosimetric treatment course simulations (fig. 1). The dose to the PTV (prostate as seen in the planning CT plus 1 cm isotropic margin) was prescribed as 70 Gy poisson cell-kill EUD (eq. (11) with $\alpha = 0.4$) to be applied in 35 fractions under dose constraints to the two relevant organs-at-risk: max. 62 Gy serial EUD (eq. (12) with $a = 12$) to rectal wall, max. 58.5 Gy serial EUD ($a = 8$) to bladder.

4. Results

A simultaneous PCA of the $N = 5$ composite organ geometry vectors representing prostate, bladder and rectal wall of the example patient was performed. This resulted in the following spectrum of relative eigenvalues: 40.7/36.5/13.7/9.1%*. Thus about 90.9% of the overall geometric variability present in the $N = 5$ input samples can be represented by the first three eigenmodes. In the following, we will only consider these

* Note that a PCA model with N input samples has only $N - 1$ eigenmodes with nonzero eigenvalue.

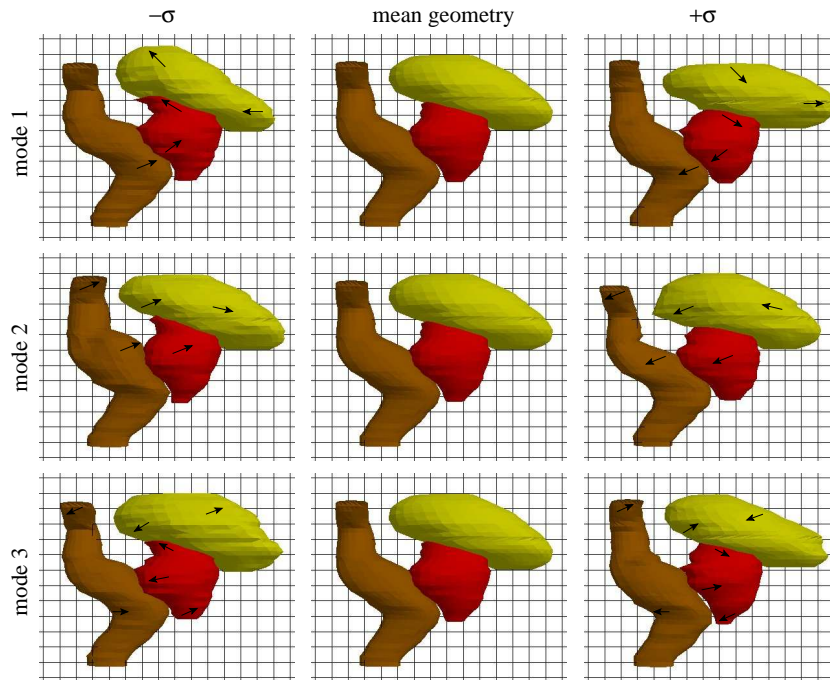


Figure 2. Visualization of the displacement fields given by the first three eigenmodes of the example patient (first eigenmode in first row etc.). Shown are sagittal 3D-views of prostate, rectum and bladder. The center column shows the mean geometry $\bar{\mathbf{p}}$, whereas the left/right columns are deformations of the mean geometry by the respective eigenvector \mathbf{q}_l according to $\mathbf{p} = \bar{\mathbf{p}} \pm \sigma_l \cdot \mathbf{q}_l$ (eq. (4)). The arrows illustrate the local motion/deformation.

three dominating eigenmodes.

Fig. 2 provides a visualization of these eigenmodes. While for this specific example the first and third eigenmode are governed by changes in bladder filling and correlated motion/deformation of prostate and rectum, the second eigenmode is driven by changes in rectum filling. It should be noted in this context that the character of geometric variations represented by *single eigenmodes* is highly patient specific and does not necessarily have clear anatomical interpretation (eg. ‘bladder-filling driven’) for each patient and mode. Only *superpositions* of dominating modes according to eq. (4) define an anatomically meaningful ‘deformation space’.

Dosimetric treatment course simulations for the example IMRT plan were based on the above PCA model with $L = 3$ dominating eigenmodes. For this $S = 1000$ treatment courses of $N_{frac} = 35$ fractions were simulated by repeat sampling of organ geometries based on eqs. (4) and (5). Sampling of the coefficients c_l was restricted to the range $[-2\sigma_l, 2\sigma_l]$. For each sampled geometry, dose was evaluated at the centers of the FEM-tetrahedrons, resulting in $M_D = 30240/19437/38304$ dose points for prostate/bladder/rectal wall. As compared to directly using the $M = 6253/3740/8379$ organ points explicitly modelled by the PCA model in terms of the organ geometry vector \mathbf{p} , this not only provides an increased sampling rate of dose values (and thus higher dosimetric accuracy), but also has the advantage that the corresponding relative

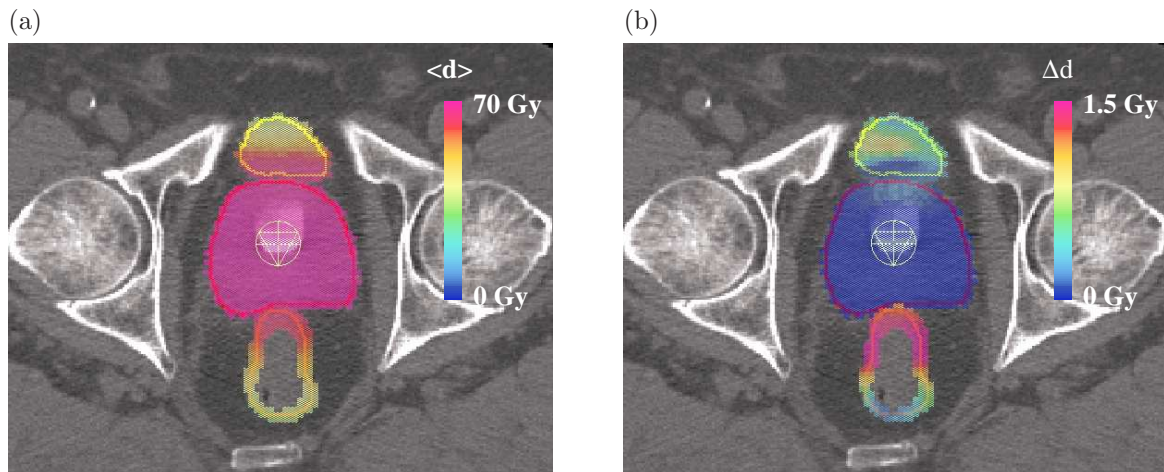


Figure 3. Transversal slice of the planning CT of the patient example in fig. 1, overlaid with the distributions of (a) the voxelwise expected values of dose $\langle d \rangle$ and (b) corresponding standard deviations Δd for prostate (CTV), bladder and rectal wall as resulting from repeat PCA-based dosimetric treatment course simulations ($S = 1000$ simulated treatment courses of $N_{frac} = 35$ fractions). For visualization purposes, the distributions of $\langle d \rangle$ and Δd have been warped to the geometry of the planning CT to match the contours of the respective organs as seen in the planning CT.

organ volumes ν_j needed for DVH- and EUD-calculations (eqs. (11), (12)) are readily available as relative volumes of the tetrahedrons of the FEM-dataset.

Fig. 3 shows the result of these simulations in terms of the voxelwise expected values of dose $\langle d \rangle$, eq. (9), and the corresponding spatial distribution of standard deviations Δd , eq. (10), warped to the geometry of the planning CT. Obviously, the expected values of dose for prostate (CTV) meet the prescribed dose of 70 Gy with slightly increased local dose standard deviations in the region adjacent to the bladder. Thus, at least according to this slice the PTV-margin of 1 cm was sufficient to guarantee correct CTV dosage in presence of organ motion/deformation. As can be expected, the local expected values of dose for regions of the bladder and rectal wall close to the prostate are also in the order of 70 Gy due to the relatively large margin. Interestingly, the local standard deviations of dose do *not* have their maximum directly at the border to prostate but at a distance of ~ 1 cm. This can be understood by the fact that tissue elements in direct neighbourhood to the prostate are almost always residing within the PTV despite organ motion, due to the relatively large PTV margin, while tissue elements that move within the region of the dose gradient have increased standard deviations of local dose.

Fig. 4a shows the DVHs of 300 of the 1000 simulated treatment courses of 35 fractions together with the DVHs of the expected organ dose distributions $\langle d \rangle$ shown in fig. 3a. According to this, the expected dose distribution for the prostate (CTV) proves clinically acceptable target dosage also in presence of organ motion, and the corresponding DVH uncertainties are small (red prostate DVHs of the 300 simulated treatment courses). Similarly, the DVH uncertainties of bladder and rectal wall are relatively small. In contrast, the single-fraction DVH uncertainties (fig. 4b) are

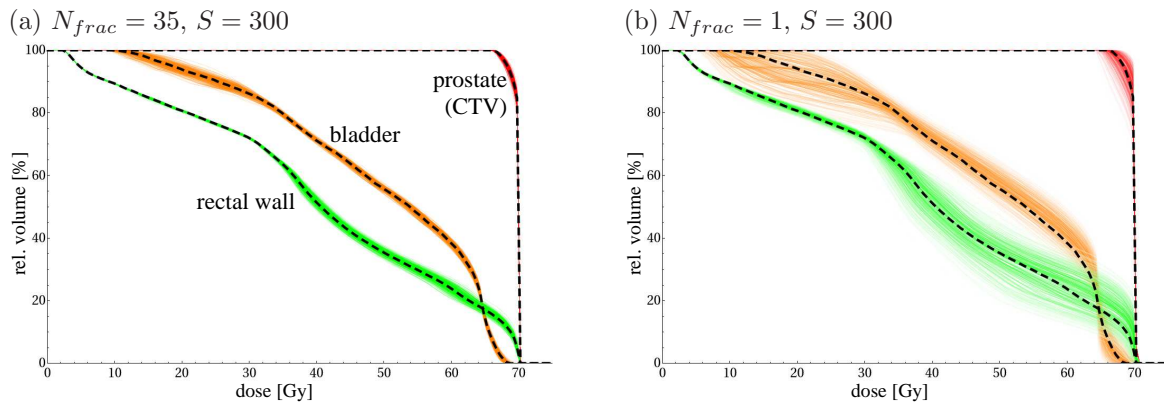


Figure 4. Visualization of DVH uncertainties resulting from repeat PCA-based dosimetric treatment course simulations for the patient example: DVHs of prostate (CTV), bladder and rectal wall as resulting from (a) 300 simulated treatment courses of 35 fractions each and (b) 300 simulated single fractions (rescaled to the dose of a full treatment course for easier visual comparison to (a)). For reasons of visual clarity only 300 DVHs of the 1000 simulated treatment courses are shown. The saturation of the DVH colors in (b) was scaled according to equation (5) to give a visual impression of the relative realization probabilities for the different simulated fractions. The dashed black DVHs are the DVHs of the expected doses $\langle d \rangle$ for the different organs.

considerably larger. A comparison of figs. 4a and b shows that local hot and cold spots in the moving tissue elements cancel out in the accumulated dose to a certain extent, thereby effectively reducing dose uncertainties for $N_{frac} \gg 1$.

Notice, that it is a special strength of the PCA method that the realization probability of simulated geometries and thereby treatment fractions is readily available via eq. (5). This can be used for providing additional visual information about the relative probability of each DVH to be realized as shown in fig. 4b by scaling the DVH-color accordingly. In this way, extreme ‘outliers’ can be easily distinguished from more probable DVHs.

Finally, fig. 5 shows exemplarily how the presented method facilitates evaluation of treatment plan quality in terms of the probability distributions of dosimetric measures like EUDs. According to the single-organ EUD-distributions given in fig. 5a, the prostate (CTV) receives the prescribed EUD of 70 Gy both with high probability and small uncertainty, reconfirming that the PTV-margin was chosen large enough to ensure coverage of the moving CTV. Compared to this, both bladder and rectal wall show increased EUD uncertainties due to organ motion/deformation. Moreover, a systematic difference to the maximal EUD-values used as constraints during planning is present: While rectal wall EUDs stay well below the dose constraint of 62 Gy for 100% of all simulated treatment courses, bladder is overdosed slightly by up to ~ 1 Gy EUD in the majority of the simulated treatment courses.

Simultaneous PCA for more than one organ provides full information about correlated motion/deformation and thus also *correlated organ dosage*. Thus, not only single-organ EUD distributions can be determined, but also joint probability

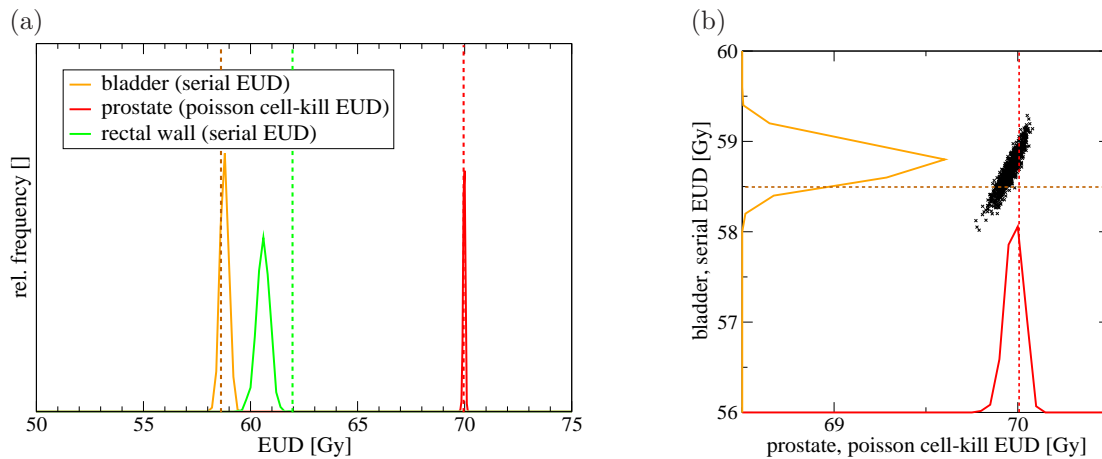


Figure 5. (a) Histograms of the single-organ EUD-distributions for prostate (CTV), bladder and rectal wall as resulting from repeat PCA-based treatment course simulations ($S = 1000$ simulated treatment courses of $N_{frac} = 35$ fractions each). Information about the EUD-correlation structure is readily available in this approach as shown in the scatterplot (b) for the example of prostate vs. rectal wall treatment course EUDs. The dashed lines mark the EUD-values used as prescription/constraints for planning.

distributions of more than one EUD as shown in fig. 5b. This 2D-scatterplot shows the distribution and correlation of EUDs of bladder and prostate for the 1000 simulated treatment courses and reveals the correlation structure of these two dosimetric measures.

5. Discussion and conclusion

Dosimetric effects of geometric uncertainties for a given treatment plan need to be assessed by treatment simulations. To get a realistic picture about the influence of all kinds of geometric uncertainties including the most general case of non-rigid deformable organ motion, the limitations of static patient models and rigid modelling approaches need to be overcome. The PCA-method presented here proves to be a powerful and efficient tool for this. It is shown how PCA facilitates the statistical evaluation of the actual organ doses, i.e. the accumulated doses-to-moving-tissue. Deformable organ motion is modelled in terms of a small number of so-called *eigenmodes*, which are the generalized main ‘directions’ of correlated geometric changes with associated Gaussian probability distributions. This gives rise to the possibility of low-dimensional parametric statistical modelling based on PCA with full information about correlated motion of all organ points/tissue elements of both single and multiple organs.

The current paper presents in detail how such a PCA model can be used for Monte Carlo simulations of treatment courses, enabling a comprehensive assessment of both local and global dosimetric measures. This can be performed in terms of local expected values of (voxel-)dose and their uncertainty, DVHs of the expected dose distributions and corresponding DVH-uncertainties, as well as probability distributions of EUD. Such a

‘virtual treatment simulator’ tool has important applications for evaluation of existing radiotherapy schemes as well as prospective assessment of potential benefits of new radiotherapy schemes. For example, the dosimetric adequacy of margin recipes or plan adaptation schemes can be investigated by performing virtual treatment simulations for a number of patients, which have multiple (conebeam-)CT scans and thus multiple organ deformation samples available as input samples to patient-individual PCA models. Alternatively, corresponding simulations can be based on a population PCA model (Budiarto *et al* 2011).

For prostate (Söhn *et al* 2005) and lung patients (Badawi *et al* 2010) it has been shown that information about patient-individual geometric variability can be typically expressed by 3-5 dominating eigenmodes, which translates into ~ 5 repeat CT scans as minimal imaging requirement. Population-based PCA models typically require more dominating eigenmodes in order to represent a comparable fraction of the overall variability present in the input samples, as shown for deformable motion of prostate/semicle vesicles (Budiarto *et al* 2011). These studies, however, mainly investigated *geometric* modelling accuracy in terms of representation residuals and similar geometric measures. Being interested in the dosimetric consequences of geometric variability, the consequent follow-up question is how this relates to the number of input geometries N and dominating modes of a PCA model with respect to *dosimetric accuracy* and *predictive power*. To a certain extent this will depend on the treatment (planning) method involved (dose gradients etc.) (Gordon and Siebers 2008). A systematic investigation of the presented PCA method for use as a *predictive* model is beyond the scope of this manuscript and will be subject to future research.

Another advantage of the presented method is calculation speed. Thanks to algorithmic optimizations as described in Appendix B of Söhn *et al* 2005, performing the PCA itself (calculation of eigenmodes and -values) is essentially independent of the number of organ points used to form the organ geometry vectors \mathbf{p} and typically takes less than a second on modern PC hardware. For the example case one simulation run of a single treatment course (i.e. dose accumulation for 35 simulated organ geometries of prostate/bladder/rectal wall) took ~ 1 s using in-house developed single-threaded software code. The very nature of Monte Carlo simulations allows parallelization in a straightforward manner, which facilitates calculation times in the order of 1min on recent multiprocessor/-core computer hardware for Monte-Carlo simulations of 1000 treatment courses.

The implementation of the PCA method as presented here uses a *static dose distribution* as basis for dosimetric treatment course simulations (eq. (7b)). This is a valid approximation for eg. the pelvic region with relatively small density inhomogeneities in the vicinity of the moving CTV and OARs. In anatomical regions with large density inhomogenieties (eg. lung, head-and-neck) and especially in particle therapy (protons, heavy ions), however, changes of the physical dose distribution itself along with anatomical motion/deformations invalidate this approximation. For such cases, dosimetric simulations require dose calculation on deformed density distributions

according to eq. (7a). It is a strength of PCA, that it can provide such synthetic density distributions as shown by Zhang *et al* 2007 for the example of 4D-CTs of lung.

To summarize, the presented PCA-based method facilitates the creation of low-dimensional parametric statistical organ motion/deformation models, and thus represents a powerful treatment simulator tool, which allows in-depth assessment of the dosimetric effects of geometric uncertainties in terms of accumulated dose and its uncertainties in the moving and deforming tissue. As such it has important applications in the evaluation of radiotherapy schemes with respect to the adequacy of margin sizes, adaptation schemes etc. in the presence of all types of geometric uncertainties beyond rigid uncertainties.

Acknowledgements

The authors thank D. Yan and J. Liang (Beaumont Hospital, Royal Oak, MI, USA) for kindly providing the example patient dataset incl. deformable registration data as well as fruitful discussions about potential applications of PCA-based geometric modelling.

References

- Badawi A M, Weiss E, Sleeman IV W C, Yan C and Hugo G D 2010 ptimizing principal component models for representing interfraction variation in lung cancer radiotherapy *Phys. Med. Biol.* **37** 5080–91
- Baum C, Alber M, Birkner M and Nüsslin F 2004 Treatment simulation approaches for the estimation of the distribution of treatment quality parameters generated by geometrical uncertainties *Med. Phys.* **49** 5475–88
- Birkner M 2002 Bildgestützte, adaptive Bestrahlungsplanung intensitätsmodulierter Strahlentherapie (PhD thesis University of Tübingen, available under <http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-5429>)
- Birkner M, Yan D, Alber M, Liang J and Nüsslin F 2003 Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation *Med. Phys.* **30** 2822–31
- Budiarto E, Keijzer M, Storchi P R, Hoogeman M S, Bondar L, Mutanga T F, de Boer H C J and Heemink A W 2011 A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases *Phys. Med. Biol.* **56** 1045–61
- Gordon J J, Crimaldi A J, Hagan M, Moore J and Siebers J V 2007 Evaluation of clinical margins via simulation of patient setup errors in prostate IMRT treatment plans *Med. Phys.* **34** 202–14
- Gordon J J and Siebers J V 2008 Evaluation of dosimetric margins in prostate IMRT *Med. Phys.* **35** 569–75
- Killoran J H, Kooy H M, Gladstone D J, Welte F J and Beard C J 1997 A numerical simulation of organ motion and daily setup uncertainties: implications for radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **37** 213–21
- Niemierko A 1997 Reporting and analyzing dose distributions: A concept of equivalent uniform dose *Med. Phys.* **24** 103–10
- Niemierko A 1999 A generalized concept of equivalent uniform dose [Abstract] *Med. Phys.* **26** 1100
- Price G J and Moore C J 2007 A method to calculate coverage probabilities from uncertainties in radiotherapy via a statistical shape model *Phys. Med. Biol.* **52** 1947–65
- Söhn M, Birkner M, Yan D and Alber M 2005 Modeling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and evaluation *Phys. Med. Biol.* **50** 5893–908

- Söhn M, Sobotta B and Alber M 2009 Dosimetric treatment course simulation based on a patient-individual statistical model of deformable organ motion *Radiother. Oncol.* **92**(1 Suppl) S65-6
- van Herk M, Remeijer P and Lebesque J V 2002 Inclusion of geometric uncertainties in treatment plan evaluation *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1407–22
- Yan D, Jaffray D A and Wong J W 1999 A model to accumulate fractionated dose in a deforming organ *Int. J. Radiat. Oncol. Biol. Phys.* **44** 665–75
- Zhang Q, Pevsner A, Hertanto A, Hu Y C, Rosenzweig K E, Ling C C and Mageras G S 2007 A patient-specific respiratory model of anatomical motion for radiation treatment planning *Med. Phys.* **34** 4772–81

Figure captions

Figure 1. (a) Transversal CT slice of the example patient with contours for bladder, prostate and rectum as seen in the planning CT and the first four daily pretreatment CTs. These organ geometries were used as input for the PCA model. The dose distribution of the example IMRT plan is shown in terms of the 70.0/66.5/63.0/56.0/49.0/35.0 Gy isodose lines. (b) The mean geometry of the PCA model as sagittal 3D-view in front of the dose distribution.

Figure 2. Visualization of the displacement fields given by the first three eigenmodes of the example patient (first eigenmode in first row etc.). Shown are sagittal 3D-views of prostate, rectum and bladder. The center column shows the mean geometry $\bar{\mathbf{p}}$, whereas the left/right columns are deformations of the mean geometry by the respective eigenvector \mathbf{q}_l according to $\mathbf{p} = \bar{\mathbf{p}} \pm \sigma_l \cdot \mathbf{q}_l$ (eq. (4)). The arrows illustrate the local motion/deformation.

Figure 3. Transversal slice of the planning CT of the patient example in fig. 1, overlaid with the distributions of (a) the voxelwise expected values of dose $\langle d \rangle$ and (b) corresponding standard deviations Δd for prostate (CTV), bladder and rectal wall as resulting from repeat PCA-based dosimetric treatment course simulations ($S = 1000$ simulated treatment courses of $N_{frac} = 35$ fractions). For visualization purposes, the distributions of $\langle d \rangle$ and Δd have been warped to the geometry of the planning CT to match the contours of the respective organs as seen in the planning CT.

Figure 4. Visualization of DVH uncertainties resulting from repeat PCA-based dosimetric treatment course simulations for the patient example: DVHs of prostate (CTV), bladder and rectal wall as resulting from (a) 300 simulated treatment courses of 35 fractions each and (b) 300 simulated single fractions (rescaled to the dose of a full treatment course for easier visual comparison to (a)). For reasons of visual clarity only 300 DVHs of the 1000 simulated treatment courses are shown. The saturation of the DVH colors in (b) was scaled according to equation (5) to give a visual impression of the relative realization probabilities for the different simulated fractions. The dashed black DVHs are the DVHs of the expected doses $\langle d \rangle$ for the different organs.

Figure 5. (a) Histograms of the single-organ EUD-distributions for prostate (CTV), bladder and rectal wall as resulting from repeat PCA-based treatment course simulations ($S = 1000$ simulated treatment courses of $N_{frac} = 35$ fractions each). Information about the EUD-correlation structure is readily available in this approach as shown in the scatterplot (b) for the example of prostate vs. rectal wall treatment course EUDs. The dashed lines mark the EUD-values used as prescription/constraints for planning.