4-dimensional patient models for evidence-based biological optimization of radiotherapy

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

der Fakultät für Mathematik und Physik
der Eberhard-Karls-Universität zu Tübingen

vorgelegt von

Matthias Söhn
aus Bremen

2009

Tag der mündlichen Prüfung:	30. April 2009
Dekan:	Prof. Dr. Wolfgang Knapp
1. Berichterstatter:	Prof. Dr. Dr. Fritz Schick
2. Berichterstatter:	Prof. Dr. Heinz Clement

Zusammenfassung

Die vorliegende Arbeit umfasst die drei Grundpfeiler evidenzbasierter „4-dimensionaler“ Strahlentherapie: biologische Dosis-Wirkungs-Modelle, probabilistische Deformationsmodelle der Patientengeometrie sowie Methoden zur Evaluation und Optimierung der vom sich bewegenden Gewebe akkumulierten Dosis in statistischen Sinne. Im Zuge dessen werden klassisch verwendete statische Patientenmodelle durch neuartige dynamische Patientenmodelle ersetzt.
Die Anwendung derartig präziser und umfassender Strahlentherapie-Konzepte führt zu komplexen, fundamental andersartigen Dosisverteilungen. Dies ist nur dann medizinisch unbedenklich, wenn entsprechende biologische Modelle für die Dosisoptimierung im Patienten vorhanden sind. Modelle, die die Wahrscheinlichkeit von Nebenwirkungen in gesundem Gewebe beschreiben, wurden auf klinische Daten grosser, an Prostata- und Rektalkrebs behandelter Patientengruppen angewendet. Dadurch konnten Modellparameter für chronische Rektal- und akute Dünndarm-Nebenwirkungen bestimmt werden. Mögliche systematische Fehler solcher Modelle, die in Form spezifischer Dosis-Volumen-Korrelationen auf die Behandlungsmethode zurückzuführen sind, werden diskutiert und quantitativ mittels einer Hauptkomponentenanalyse (PCA, von *engl.* principal component analysis) von Dosis-Volumen-Histogramm-(DVH)-Daten untersucht.
Es wurden zwei Ansätze zur „4-dimensionalen“ geometrischen Modellierung für zwei in der Strahlentherapie auftretende Unsicherheiten entwickelt. Für zufällige geometrische Unsicherheiten erweist sich die Anwendung einer PCA auf Stichproben von Organgeometrien als effiziente Methode, statistische Ersatzmodelle abzuleiten, die die komplexen biomechanischen Vorgänge patientenindividueller Organbewegung in Form von Deformations-Eigenmoden zusammenfassen. Diese Methodik wurde auf das für Beckenbestrahlungen relevante Problem interner Organbewegungen von Prostata, Rektum und Blase angewendet. Zur Modellierung quasi-periodischer Atembewegungen wurde ein schneller und genauer Algorithmus zur elastischen Bildregistrierung entwickelt. Als solcher dient er als zentrales Element beim Erstellen probabilistischer Thoraxmodelle, die auf 4D-atemkorrelierten CT-Bildinformationen basieren. Der Algorithmus führt starre Registrierungen für eine grosse Zahl kleiner Bild-Teilbereiche durch. Dabei kommt ein neues, physikalisch motiviertes Regularisierungsschema auf Basis einer automatischen Klassifizierung der lokalen Registrierungsgüte zum Einsatz.
Schliesslich wird eine klinisch praktikable Implementierung von 4D-intensitätsmodulierter Radiotherapie (IMRT) präsentiert, welche die biologische Optimierung des Erwartungswertes der Dosis im bewegten Gewebe realisiert. Dies wird durch explizite Optimierung in mehreren Geometrieinstanzen erreicht. Der Algorithmus benutzt einen anspruchsvollen semi-analytischen Nadelstrahlalgorithmus in Verbindung mit einem Monte-Carlo Dosisberechnungsverfahren, um die physikalische Dosisdeposition in den beweglichen Volumenelementen der Gewebematrix dynamischer Patientengeometrien präzise zu bestimmen. Anhand eines Planungsbeispieles für die Bestrahlung bewegter Lungentumore unter freier Atmung wird gezeigt, daß 4D-IMRT ähnlich gute Ergebnisse wie atemgetriggerte Bestrahlung erzielt. Es kann deshalb als effizientere Alternative zu atemgetriggerter Be-

strahlung ohne verlängerte Behandlungszeiten angesehen werden. Zusätzlich wird eine klinisch relevante Anwendung für 4D-Evaluation in Gestalt einer Studie beschrieben, in der verschiedene intensitätsmodulierte Planungsmethoden für Prostatabestrahlung mit Photonen und Protonen hinsichtlich ihrer Robustheit gegenüber Organbewegung verglichen werden.
Zusammenfassend zeigen die in dieser Arbeit vorgestellten Entwicklungen, daß evidenzbasierte 4D-Strahlentherapie eine umfassende und leistungsfähige Verallgemeinerung gegenwärtiger, auf statischen Patientengeometrien beruhenden Verfahren ist. Dies ist ein wichtiger Schritt hin zu einem neuen Paradigma, welches dem dynamischen Charakter der Bestrahlung von Patienten Rechnung trägt.

Summary

The present work comprises the three cornerstones of evidence-based "4-dimensional" radiotherapy: biological dose-effect models, probabilistic patient geometry deformation models and methods for evaluation and optimization of accumulated dose-to-moving-tissue in a statistical sense. For this purpose, commonly used static patient models are substituted by novel dynamic patient models.

The application of such precise and comprehensive radiotherapy concepts results in complex, fundamentally different dose distributions. This is only safe if adequate quantitative biological models for dose optimization in the patient exist. Normal tissue complication probability (NTCP) models were applied to clinical data of large patient populations treated for prostate and rectal cancer, and parameters are derived for chronic rectal and acute small bowel toxicity. Potential biases of such models introduced by the treatment technique in terms of specific patterns of dose-volume correlations are discussed and quantitatively investigated with a principal component analysis (PCA) of dose-volume histogram (DVH) data.

Two 4-dimensional geometric modelling approaches for two major types of uncertainties encountered in radiotherapy were developed. For random geometric uncertainties, a PCA of organ geometry samples proves an efficient method to create statistical surrogate models that summarize the complex biomechanical processes involved in patient-individual organ motion in terms of deformation eigenmodes. The method was applied to the problem of internal motion of prostate, rectum and bladder as relevant for pelvic radiotherapy. For modelling of quasi-periodic respiratory motion, a fast and accurate deformable registration algorithm was developed as central tool for the creation of probabilistic thorax models based on 4D respiration correlated CT image information. The algorithm performs rigid registrations for a large number of small subregions and introduces a new, physically motivated regularization scheme on the basis of an automatic assessment of local registration quality.

Ultimately, a clinically feasible implementation of 4D-intensity modulated radiotherapy (IMRT) is presented that realizes biological optimization of the expected dose-to-moving-tissue by explicit optimization in multiple instance geometries. The algorithm makes use of an advanced semi-analytical pencil-beam algorithm in combination with Monte Carlo dose computation to accurately determine the physically deposited dose in the moving tissue elements of dynamic patient geometries. Based on a planning example of free-breathing treatment of moving lung tumors, 4D-IMRT is shown to perform similar to gated treatment and thus constitutes an more efficient alternative to gating without prolonged treatment times. Finally, a clinically relevant application of 4D-evaluation is described in terms of a study comparing the robustness against organ motion of different intensity modulated prostate planning approaches for photon and proton irradiation.

Summarizing, the developments presented in this work render evidence-based 4D-radiotherapy as comprehensive and powerful generalization of current static patient geometry approaches. This is an important step to a new paradigm that accounts for the dynamic nature of patient treatment.

Contents

List of abbreviations

AIC	Akaike Information Criterion
AP	Anterior-Posterior
ART	Adaptive Radiotherapy
BEV	Beam-Eye-View
CBCT	Cone-Beam CT
CC	Cranio-Caudal *or* Correlation Coefficient
CRT	Conformal Radiotherapy
CT	Computed Tomography
CTV	Clinical Target Volume
DVH	Dose-Volume Histogram
EUD	Equivalent Uniform Dose
FE(M)	Finite Element (Method)
GTV	Gross Target Volume
H&N	Head-And-Neck
IGRT	Image-Guided Radiotherapy
IMAT	Intensity Modulated Arc Therapy
IMPT	Intensity Modulated Proton Therapy
IMRT	Intensity Modulated Radiotherapy
IS	Inferior-Superior (used synonymous to CC)
ITV	Internal Target Volume
LL	LogLikelihood
MC	Monte Carlo
MI	Mutual Information
MLC	Multi-Leaf Collimator
MRI	Magnetic Resonance Imaging
NMI	Normalized Mutual Information
NTCP	Normal Tissue Complication Probability
OAR	Organ at Risk
(fs)PB	(Finite Size) Pencil Beam
PC	Principal Component *or* Prostate Cancer
PCA	Principal Component Analysis
pdf	Probability Density Function
PI	Portal Image
PRV	Planning Organ-At-Risk Volume
PTV	Planning Target Volume
RCCT	Respiratory-Correlated CT
RL	Right-Left
RT	Radiotherapy
SI	Superior-Inferior
SSD	Source-Surface Distance
TEV	Tissue-Eye-View
TCP	Tumor Control Probability
4D-CT	Four-Dimensional CT (used synonymous to RCCT)

Chapter 1

Introduction

The history of therapeutic use of radiation dates back over hundred years. Nowadays radiotherapy is one of the three main modalities in the fight against cancer and contributes to more than 50% of all cancer treatments – as single modality or in combination with surgery and chemotherapy. Thus technical as well as methodical advances in this field clearly have the potential for an overall improvement of treatment outcome in terms of survival rates and quality of life.

The fundamental challenge of radiotherapy is treatment of the tumor with sufficient dose while minimizing harm to adjacent normal tissues due to unavoidable co-irradiation. For a long time, significant improvement of the tradeoff between these goals was hampered mainly by technical limitations preventing tight shaping of the radiation dose to the tumor region. In the last two decades developments in treatment and imaging technology along with the dramatic increase of computer power enabling virtual treatment simulation and -optimization pushed the frontier of possibilities in favor of more precise irradiation and thus higher therapeutic doses. One of the key concepts is the idea of irradiation with *modulated* instead of homogeneous photon radiation intensity (intensity modulated radiotherapy, IMRT [15, 14, 16]), which is a powerful technique for tailoring the radiation dose to a predefined region with previously unprecedented precision. In the near future the clinical introduction of other types of radiation with physically favorable dose deposition characteristics such as protons (intensity modulated proton therapy, IMPT [31]) or heavy ions promises further improvements in this direction.

The biological dose-response of tumors and normal tissues is relatively steep, which is why uncertainties in dose translate into relatively large loss of tumor control or increase of complication risk. Thus it is of vital importance for successful and save treatment to minimize uncertainties in the applied dose. One of the factors here are *dosimetric uncertainties* that are are introduced by the challenges encountered for dose calculation of complex, irregular shaped radiation fields in presence of large density inhomogenieties in the patient body. In the last decade, this has been successfully addressed by the introduction of advanced dose calculation methods into clinical practice, ultimately with Monte-Carlo algorithms as most precise dose calculation methods for dosimetric treatment simulation [20].

A major problem in radiotherapy are *geometric uncertainties* of various nature. Inevitably, the exact position and/or shape of the tumor and adjacent normal tissues (organs-at-risk, OARs) at time of treatment differs from the geometric configuration in the pre-treatment image information used to generate a patient's treatment plan. Typical sources of uncertainties are differences in patient positioning relative to the radiation equipment during the treatment course, as well as internal organ motion due to breathing or digestion activities. The magnitude of all relevant geometric uncertainties has to be regarded already in the process of treatment planning to prevent serious deterioration of tumor dose. In the common approach [1, 2], the tumor volume as seen in a single pre-treatment CT scan (clinical target volume, CTV) is enlarged by a *margin*, and the resulting larger *planning target volume* (PTV) is irradiated instead. In this way, coverage of the CTV is sought to be guaranteed even in the presence of geometric uncertainties. A large margin, however, goes at the expense of irradiating additional normal tissue which in turn limits the possible therapeutic dose to the PTV. As knowledge about individual uncertainties is incomplete prior to treatment, typically generic population-based margins are applied based on unspecific information about uncertainties in a population of patients. Thus, within the PTV-concept margin reduction by individualization or reduction of uncertainties is an important means to facilitate tumor dose-escalation. In recent years, this has driven major developments in imaging technology, enabling tumor localization prior to or during treatment sessions (image-guided radiotherapy, IGRT), optionally combined with methods to adaptively individualize margins based on image information that becomes available during the treatment course (adaptive radiotherapy, ART [72, 73]).

Remaining random uncertainties and tumor motion during the irradiation itself still necessiate non-vanishing margins within the PTV-concept. Especially in the presence of large geometric variations, planning based on a static patient geometry and PTV region is an insufficient model of reality, posing major conceptual constraints to attempts at further dose-escalation. An important example in this context is radiotherapy of lung tumors, which still fails frequently. It is the premise of this PhD work that *new geometric modelling and optimization concepts* that capture patient geometry as a dynamic entity are indispensable to unlock the full potential of radiotherapy. The basis of this is the development of methods for creating *"4-dimensional" patient models* which explicitly incorporate geometric uncertainties of the 3-dimensional patient geometry in terms of individual deformation characteristics. Ultimately, this enables new treatment planning concepts based on the *accumulated dose-to-moving-tissue* for tumor and OARs, which are the biologically relevant quantities as opposed to the static PTV- and OAR-doses employed in the conventional PTV-concept as mere surrogates.

The radiotherapy concepts described above introduce new types of largely inhomogeneous dose distributions characterized by steep dose gradients between tumor and adjacent OARs and a significant redistribution of dose within the surrounding normal tissues. This deviation from longstanding clinical practice carries the danger of undesirable side effects if dose-response of complications is not well understood, highlighting the importance of reliable models of normal tissue dose-response relationships. In this sense, *save dose-escalation* using new radiotherapy techniques has to be accomplished with the concept of

evidence-based biological optimization, which aims to translate biological knowledge and past clinical experience into quantitative biological models for dose optimization.

With these premises in mind, this PhD work adresses the following three key issues:

- **Evidence-based biological modelling (chapter 2):** Extracting quantitative information about normal tissue dose-response from clinical data is demanding and requires large sets of patient data with long follow-up time and a sufficient variability of dose patterns in the clinical dataset. Great care has to be taken with dose-response models for the prospective use with new radiotherapy techniques, as such models can only be based retrospectively on past clinical data. Differences in clinical practice in terms of dose prescription and treatment technique lead to differences in the characteristics of dose distributions and their correlated variability, which can induce a model bias. In **section 2.1** this effect is investigated in detail based on the example of rectal side-effects in the treatment of prostate cancer. In **section 2.2** clinical data of 319 prostate and 152 rectal cancer patients are used to adapt and compare a number of biological models for chronic rectal bleeding and acute diarrhea, respectively, as important dose limiting side effects in pelvic radiotherapy.

- **4-dimensional patient models (chapter 3):** The ultimate method for modelling tissue deformations is the use of biomechanical models explicitly describing physical properties of the different tissue materials [70]. This, however, is challenging not only due to incomplete knowledge of tissue parameters and the vast amount of variables involved to describe deformation fields, but also because of the complex and often unknown boundary conditions driving deformations in real patients. Radiotherapy is mainly faced with either *random* or *quasi-periodical* deformable organ motion, and surrogate models for both phenomena have been developed in this work as efficient alternatives to full biomechanical modelling. **Section 3.1** introduces an approach to model random deformable uncertainties of 3-dimensional organ geometries. The method employs a method from multidimensional statistics – principal component analysis (PCA) – to determine patient individual 'deformation modes' which serve as a statistical model of correlated organ deformations between treatment sessions, and was applied to patient data of prostate, bladder and rectum. **Section 3.2** presents a new approach of deformable image registration for direct image-based modelling of quasi-periodical respiratory motion and deformations based on time-resolved '4-dimensional' image datasets of the thorax (respiratory correlated computed tomography, RCCT). This allows construction of patient-individual probabilistic models of the thorax region. Apart from this, the algorithm can also be applied to multimodality deformable registration problems of other organ sites encounterd in radiotherapy.

- **4-dimensional radiotherapy planning (chapter 4):** The development of 4-dimensional patient models facilitates new radiotherapy concepts that implement

"4-dimensional" optimization and evaluation based upon the accumulated dose-to-moving-tissue instead of static dose surrogates as in the PTV-concept. Such a 4D-IMRT optimization method and its application to lung planning is presented in **section 4.1**. The algorithm performs concurrent optimization in multiple RCCT-geometries, thereby explicitly incorporating respiratory motion into planning. Dedicated dose calculation methods are employed for realistic modelling of dose-buildup and -scattering in the thorax region. The final plan dose is calculated using a 4D-Monte Carlo (MC) algorithm as currently most precise dose calculation method in the presence of large tissue inhomogeneities. Based on a patient example, it could be shown that 4D-IMRT optimization outperforms common PTV-based planning approaches in terms of therapeutic tumor dose and efficiency of application. In **section 4.2** an example for 4D-evaluation is given in terms of a study comparing the robustness of IMRT and IMPT against organ motion. The methodology allows retrospective comparison of different radiotherapeutic approaches based on the accumulated dose-to-moving-tissue for simulated treatment courses.

Summarizing, the methods developed in this PhD work appreciate that patients cannot be regarded as static in the process of dose optimization. Application of new, more precise radiotherapy concepts leads to fundamentally different dose distributions, which is only safe if quantitative biological models for dose optimization exist. Accordingly, this thesis provides the following three cornerstones of *4-dimensional biological optimization*: biological models, probabilistic deformation models and methods for evaluation and optimization of accumulated dose in a statistical sense.

Chapter 2

Evidence-based biological modelling

Radiotherapy techniques like IMRT, IMPT and "4D radiotherapy" concepts constitute powerful tools to control the physical dose distribution with a large number of degrees of freedom, and undoubtedly bear potential of improved cure rates through higher therapeutic doses. Normal tissue doses will nevertheless deviate from established practice. *Safe dose-escalation* obviously requires a profound understanding of the dose-response of normal tissues to the largely inhomogeneous dose distributions produced by such new techniques. In this sense, *evidence-based biological modelling* intends to translate clinical experience about dose-response into quantitative biological models for dose optimization.

Clinical data from the past provide a valuable source of information about normal-tissue dose-response. However, retrospective biological modelling on such datasets is fraught with practical challenges and problems. Firstly and most obviously, datasets with a large number of patients and toxicity events, as well as sufficiently long follow-up times are necessary to draw statistically reliable conclusions. For many organs only few of such datasets have been published, and the comparability of different studies is additionally hampered by differences in the schemes used for grading toxicities and the treatment techniques. In fact, the latter deserves special attention when trying to extract quantitative information about dose-response for prospective clinical use with new treatment techniques, as every dose prescrition scheme and treatment technique induces specific patterns of dose-volume correlations in patient populations which can introduce a model bias. This has been investigated in detail in this PhD work on the example of rectal dose-response in prostate radiotherapy as presented in section 2.1.
Section 2.2 summarizes the results of restrospective clinical studies for chronic rectal and acute small bowel complications as examples for biological modelling in terms of normal tissue complication probability models (NTCP). Such empiric or semi-empiric models parameterize the vast information about inhomogeneous dose distributions and corresponding outcome data from large patient populations into few-parametric models that assign a single probability value to an individual treatment plan. This enables evidence-based ranking of alternative plans in the planning process according to their predicted complication risk. The relevance of such biological models for dose optimization is discussed in section 2.3.

2.1 Correlated dose variability and retrospective biological modelling

A first step in complication modelling is the reduction of 3D dose distributions to dose-volume histograms (DVHs). Based on this, correlations of parameters derived from the DVHs and reported toxicities are investigated. An inherent problem is the influence of the treatment technique on the results of modelling. For a given patient population, the treatment technique induces correlations between DVH bins of different dose levels owing to the interaction of the given beam directions and shapes with the variability of the patient geometries. As this can deteriorate the prospective use of such models for other patient populations treated using different techniques, such correlations should be considered when interpreting and comparing the results of different studies.
The method at hand to analyze the correlated variability of DVH shapes in a given patient population is *principal component analysis (PCA)*. In this PhD work, PCA was applied to rectal wall DVHs of a population of prostate cancer patients that were treated with a four-field box 3D-CRT technique (Söhn *et al.* 2007 [59], see Appendix A).

2.1.1 Principal component analysis (PCA) of DVH data

Due to the very nature of DVHs of a patient population treated with the same treatment technique the volume values $V_1, \ldots, V_p$ of the p DVH-dosebins are highly correlated, which implies that the underlying dimensionality of this multivariate statistical problem is actually much smaller than p. By calculating the eigenvectors of the covariance matrix, PCA allows to find the statistically independent *'eigenmodes'* of variability, which are ordered by their statistical importance in terms of the corresponding *eigenvalues*. An eigenmode is a p-dimensional vector defining differential volume change for each dosebin, thereby representing correlated DVH-shape change. Mathematically, the p eigenmodes define a new coordinate system, in which each DVH can be (exactly) represented by p new coordinates $c_1, \ldots, c_p$ ("principal components", PCs). In other words, each DVH can be represented as a linear combination of all p modes weighted by the corresponding values of $c_1, \ldots, c_p$. If the overall variability of the dataset is clearly dominated by few PCs, this offers the possibility to approximately represent the most important features of individual DVH morphology by the first few parameters $\{c_1, \ldots, c_M\}$, with $M \ll p$. Further details can be found in Appendix A. The use of PCA in a different context (geometric modelling) is described in sect. 3.1.

Fig. 2.1a shows the rectal wall DVHs of 319 patients treated at the William Beaumont Hospital (Royal Oak, MI) with a *four-field box 3D-CRT technique*. The first three eigenvectors resulting from a PCA of the DVHs are shown as fig. 2.1b together with the corresponding eigenvalues. For this population, the first eigenmode represents correlated variability of the DVH dataset around the mean DVH in the range ∼5-75Gy, and thus the first PC essentially describes the variation of the fraction of rectal wall volume which is inside the irradiated volume. This is why the first PC correlates with dosimetric mea-

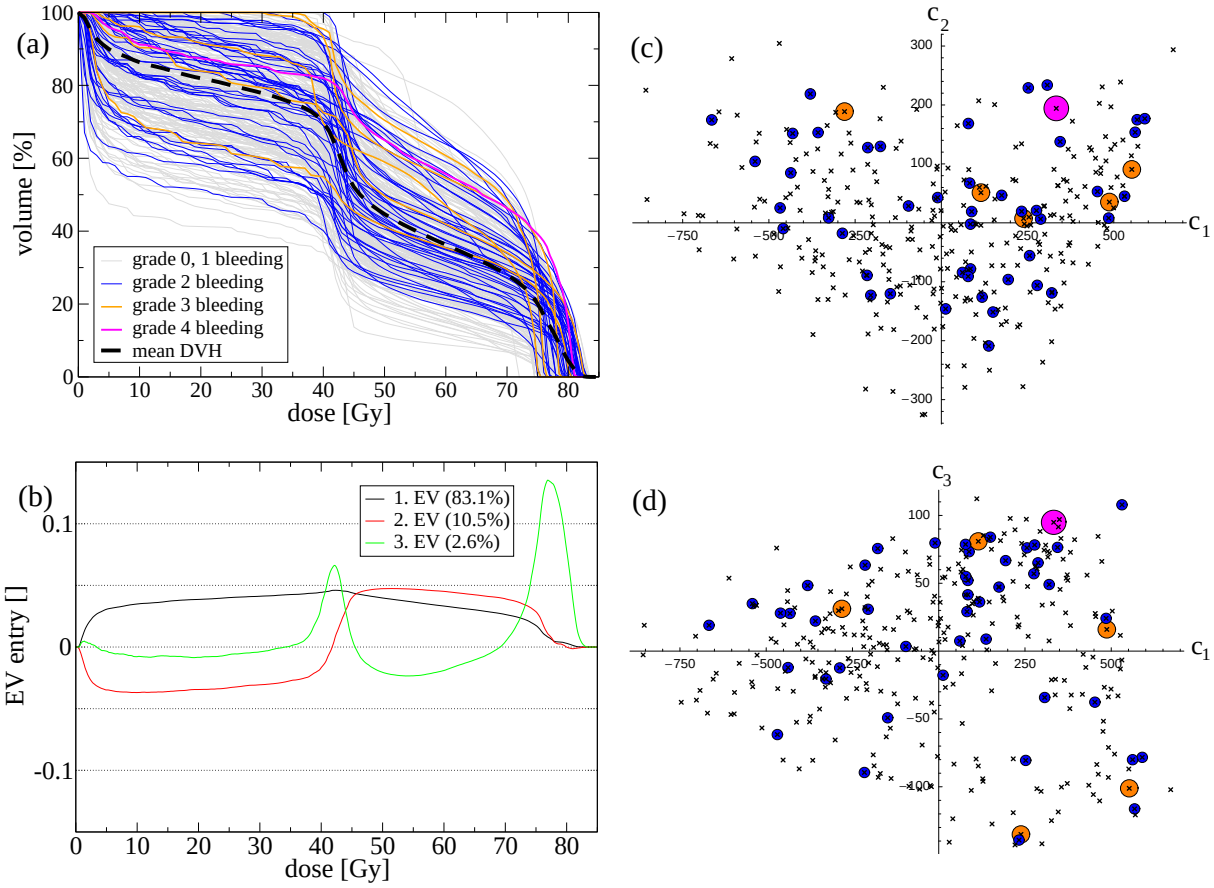

Figure 2.1: (a) Rectal wall dose-volume histograms (DVHs) of 319 patients treated with a four-field box 3D-CRT technique, plotted together with the population-mean DVH. (b) First three eigenvectors resulting from a principal component analysis (PCA) of the DVH dataset. (c)/(d) Scatterplots of the first principal components (PCs), c_2 vs. c_1 and c_3 vs. c_1, for the 319 DVHs. The color coding in (a), (c) and (d) indicates the grade of chronic rectal bleeding developed by the patients. The scatterplots in (c) and (d) suggest increased toxicity of grade ≥ 2 if combinations of the PCs are simultaneously large. (adapted from [59])

sures like the mean dose or V_{60} as shown in appendix A. The second eigenvector describes anticorrelated variability of the low and high dose region (∼5-40/50-75Gy). Thus, the second PC is related to the fraction of rectal wall volume which is inside the 4-field overlap region (∼presciption dose and less) and therefore not in the 2-field overlap region (∼half the presciption dose and less). The third mode describes correlated variability of volumes receiving intermediate doses (∼40-45Gy) and doses above ∼70Gy. It essentially results from variations of the prescribed 4-field box dose itself, which obviously also affects the dose in the 2-field overlap. This variability of dose levels in the population is a consequence of the particular dose prescription scheme applied to this specific patient population. As can be expected from the pattern of the third eigenmode in fig. 2.1b, the third principal component shows correlation to the maximal dose (appendix A).

The spectrum of eigenvalues is clearly dominated by few values: about 93.6% (96.2%) of the DVH shape variability can already be described by the first two (three) PCs. Scatterplots showing the distribution of the first PCs (c_2 vs. c_1 and c_3 vs. c_1) for the 319 DVHs are plotted as fig. 2.1c and 2.1d. The PCs itself – irrespective of toxicity – do not show significant clustering, which corresponds to a fairly regular distribution of DVH shapes for the given eigenvectors in visual agreement to fig. 2.1a. This supports the applicability of PCA, which to a certain extent is based on the assumption that the underlying distribution of input data is approximately Gaussian.

2.1.2 DVH-PCA and toxicity

Investigation of the correlations of principal components and toxicity gives insight into the specific influence of the treatment technique on toxicity modelling.

Points corresponding to DVHs associated with chronic rectal bleeding of grade ≥ 2 were marked colorcoded in the scatterplots in fig. 2.1c and 2.1d. Severe toxicity of grade ≥ 3 appears to be associated with positive values of the first two PCs, corresponding to increased volumes especially in the high dose region in the 4-field overlap as discussed above. Concerning toxicity of grade ≥ 2, the scatterplots suggest increased toxicity rates if combinations of PCs are simultaneously large. This was analyzed by multivariate logistic regression for the first two and three PCs. As shown in appendix A, both models revealed significant correlations of combinations of PCs and toxicity. The corresponding linear combinations of eigenvectors give rise to 'compound modes', which express the DVH pattern variability (relative to the population-mean DVH) that causes largest changes of toxicity according to the respective logistic regression model.

Thus, in principle the DVH-PCA approach could be applied to clinical treatment planning by penalizing DVH patterns that deviate from the population-mean DVH according to the compound modes. However, such a phenomenological 'DVH-PCA based NTCP model' has to be used with caution. as it is limited to 4-field techniques only and does not allow extrapolation to techniques such as IMRT with many more degrees of freedom and thus larger variability of dose distributions.
Limited extrapolability, however, is a drawback shared to some extend by all NTCP models. In this context it is illustrative to distinguish DVH-shape variability in a patient

population imposed by the treatment technique from the 'biologically relevant' variability in dose-levels, i.e. the variability that leads to significantly different risks of toxicity. If (one or more) dominating PCs show correlation with toxicity (as for the patient population used in the study above), the treatment technique obviously induces 'biologically relevant' variability. Conversely, if none of the dominating PCs correlates to toxicity, basing NTCP models on such a population might lead to misinterpretations. This risk is particulary pertinent for directly dose-volume based NTCP models (eg. cutoff-dose models, see next section) as will be discussed in sect. 2.3

2.2 Biological models for optimization

A *normal tissue complication probability (NTCP) model* assigns a complication probability for an organ at risk to a generally inhomogeneous dose-distribution. This enables evidence based ranking of alternative plans in the planning process according to their predicted complication risk. The functional form of such a model can be based on a mechanistic description of biological processes, or might be designed to result in a phenomenological 'fit of the data'. The models regarded in this work are of the following general form:

- First, a summary measure μ is calculated from the dose distribution. The quantity μ serves as a ranking function by imposing an order among individual plans according to their complication risk.
- Then, a function, NTCP(μ), which assigns complication probabilities to the values of the summary measure is defined. Such a function is required to continuously map μ to the interval $[0, 1]$, while preserving the ranking imposed by the numerical values of the summary measure. This is afforded by the class of sigmoid-type functions.

Two important types of NTCP-models can be distinguished: *Dose-volume based models* use a single DVH-parameter (eg. the volume V_D irradiated to a certain dose-level D) as summary measure. This gives rise to dose-volume constraints in classical, i.e. non-biological treatment planning, which are widely used in current clinical practice.
In contrast, *EUD-like models* define an *equivalent uniform dose*, $\mathrm{EUD} = f^{-1}\big(\sum_i \nu_i f(D_i)\big)$, as surrogate parameter calculated using all bins (ν_i, D_i) of a DVH, where the form of the (monotonic) function f depends on the model. A number of models define EUD as Lebesque a-norm of the dose, i.e. as the following power-law relationship [43]:

$$\mathrm{EUD}_{PL} = \left(\sum_i \nu_i D_i^a\right)^{1/a} \tag{2.1}$$

Here $a \in [1 \ldots \infty]$ is a parameter associated with the strength of the *volume effect* for the organ under consideration, which reflects quantitatively how organs react to inhomogeneous dose distributions in terms of high doses to small volumes. Two types of normal tissue behavior are distinguished in the context of biological modelling as extremes: Organs

with a large volume effect show relatively large tolerance to partial overdosage (*parallel organs*, eg. lung). The other extreme are so-called *serial organs* which are sensitive to partial overdosage, i.e. high doses are only tolerated if confined to small subvolumes. In this case the dose-volume parameter a expresses the specific increase in overall dose-tolerance with reduction of the partial volume receiving high doses. Such organs are formally modelled as series ('chain') of independent functional subunits, where the (local) destruction of single subunits leads to functional obliteration of the whole organ. Typical examples for serial organs are spinal cord and other neural structures which are characterized by large a-values (eg. spinal cord: $a \approx 20$), rendering the maximum dose as dosimetric parameter most relevant for toxicity for $a \to \infty$.

In this work, retrospective clinical studies for two predominantly serial organs – rectum and small bowel – were conducted, aiming to determine biological parameters of these organs as relevant for pelvic radiotherapy.

2.2.1 Late rectal complications in prostate radiotherapy

In this study (Söhn *et al.* 2007 [60], see Appendix B), one dose-volume based and five EUD-like NTCP models for chronic rectal bleeding of grade ≥ 2 were applied to a population of 319 prostate cancer patients treated with a 3D-CRT technique to doses between 70.2 and 79.2Gy (same population as used in sect. 2.1). In the following, the results of the study are briefly summarized based on two of the considered models as representative examples.

Cutoff-dose model In this phenomenological dose-volume based model the summary measure μ is given by the proportion V_{D_c} of the OAR receiving doses equal to or above a (cutoff) dose level D_c. For given D_c, logistic regression was used to test for correlation of V_{D_c} and toxicity. Thus, in terms of the general NTCP model scheme presented above, the NTCP function is given by the two-parametric (sigmoid-type) logistic function

$$\mathrm{NTCP}_{logistic}(\mu) = \frac{1}{1 + \exp\left(-\beta_0 - \beta_1 \mu\right)} \tag{2.2}$$

Poisson-EUD model This model utilizes mechanistic concepts to describe predominantly serial tissue dose-response and derives the following NTCP-function based on Poissonian statistics [8]:

$$\mathrm{NTCP}_{poisson}(\mathrm{EUD}_{PL}) = 1 - \exp\left[-\ln 2 \cdot \left(\frac{\mathrm{EUD}_{PL}}{D_{50}}\right)^{a}\right], \tag{2.3}$$

with D_{50} as dose causing 50% complication probability and the volume-effect parameter a as steepness parameter of this sigmoid-type function. The power-law EUD, eq. (2.1), serves as summary measure in this model.

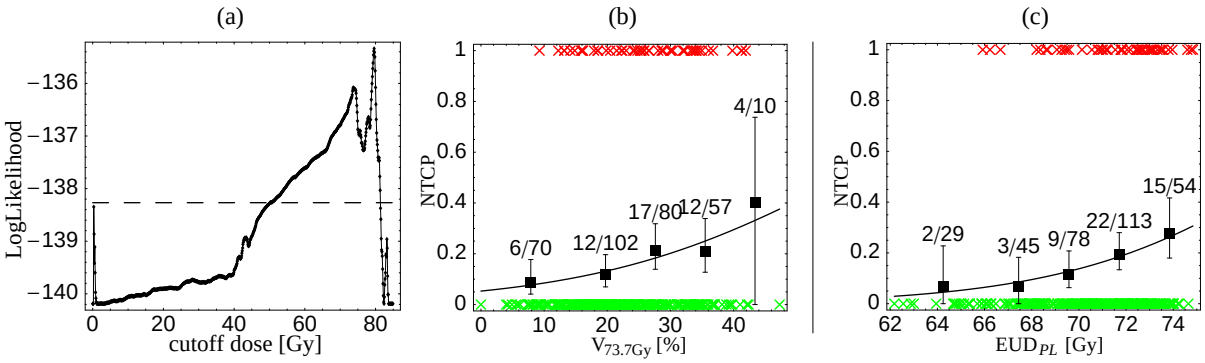

Figure 2.2: (a)/(b) Cutoff-dose logistic regression model and (c) Poisson-EUD model for chronic rectal bleeding of Grade ≥ 2: (a) Values of the LogLikelihood (LL) in dependence of the cutoff dose D_c; models with LL above the dashed horizontal line show a significant correlation ($\alpha < 0.05$) between V_{D_c} and toxicity. (b)/(c) Predicted probability of bleeding according to (b) the cutoff-dose logistic regression model with $D_c = 73.7$Gy, plotted as a function of the relative volume receiving ≥ 73.7Gy, and (c) the Poisson-EUD model, plotted as a function of the equivalent uniform dose EUD_{PL}, eq. (2.1). The '×'-symbols represent toxicity (1/0 for patients with/without toxicity, respectively). (from [60])

The models were fitted to the patient data using maximum likelihood estimation [23]. This method allows assessment of the statistical significance of a model fit by the likelihood-function L, which is usually specified as LogLikelihood in terms of its natural logarithm.
Fig. 2.2 shows the fit results for the two models under consideration. Concerning the cutoff-dose model, fig. 2.2a shows the LogLikelihood values of different logistic regression model fits when varying the cutoff dose D_c. For all dose levels in the range $D_c \sim$ 50-80Gy a significant correlation ($\alpha < 0.05$) of the relative volume irradiated with doses $\geq D_c$ and chronic rectal bleeding of grade ≥ 2 was found. As will be discussed in sect. 2.3, this characteristics is closely related to the specific treatment technique applied for the patient population under consideration. As example for a model fit, fig. 2.2b depicts the NTCP curve for the local maximum at $D_c = 73.7$Gy. The NTCP curve resulting from the Poisson-EUD model is shown in fig. 2.2c. As obvious by comparison with the observed complication rates this EUD-based model describes the data very well, and a clear correlation of EUD and toxicity could be observed. The optimal fit was found for a volume-effect parameter of $a = 13.5 \pm 3.8$ and a reference dose of $D_{50} = 78.5 \pm 0.6$Gy.

2.2.2 Acute Small Bowel reactions in pelvic radiotherapy

The dose-volume relationship of acute small bowel reactions (diarrhea of grade ≥ 3) was studied on a population of 152 rectal cancer patients treated predominantly with a three-field 3D-CRT technique (Appendix C). In this study, a cutoff-dose logistic regression model (see sect. 2.2.1) was used to test for correlations of V_{D_c} (absolute small bowel vol-

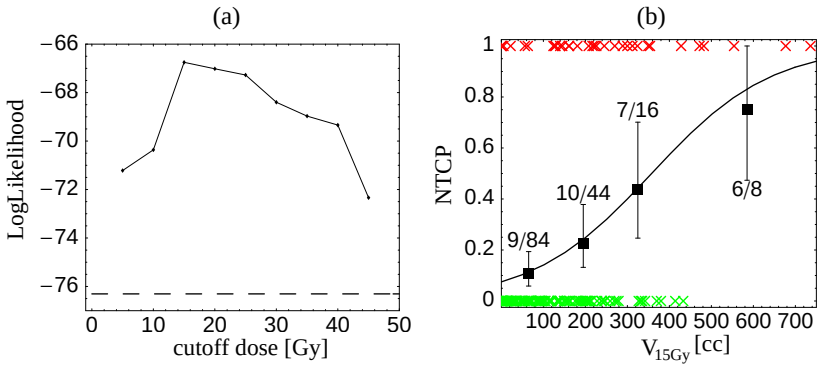

Figure 2.3: Cutoff-dose logistic regression model of acute diarrhea of grade ≥ 3: (a) Values of the LogLikelihood in dependence of the cutoff dose D_c; all models reach a significance level of $\alpha < 0.05$ as marked by the dashed horizontal line as threshold. (b) Predicted probability of diarrhea according to the model for $D_c = 15$Gy, plotted as a function of small bowel volume receiving at least 15Gy.

ume receiving doses above a cutoff dose D_c) and toxicity.
Fig. 2.3 depicts the corresponding fit results. According to this model, significant correlations of V_{D_c} and toxicity exist for all dose-levels D_c in the range 5-45Gy for the population under consideration. The most significant model was found for $D_c = 15$Gy (fig. 2.3a). The corresponding NTCP curve is plotted in fig. 2.3b, showing a steep increase of toxicity for larger small bowel volumes receiving doses of 15Gy and more. Similar findings applied when only the pre- and postoperatively irradiated patients were considered as subgroups (Appendix C). Thus for patients treated with similar 3D-CRT techniques it is advisable to minimize the volume V_{15Gy}. For patients treated with different treatment techniques, however, other dosimetric measures might be more relevant as discussed in the following.

2.3 Discussion and Conclusions

The studies about chronic rectal and acute small bowel complications presented in this work provide valuable quantitative information about dose-response relationships for evidence-based dose optimization of future patient treatment in pelvic radiotherapy.

Concerning rectal complications, the results clearly confirm a volume effect for chronic rectal bleeding of grade 2 or worse which can be described very well with EUD-like models. The volume effect parameter of the power-law EUD, eq. (2.1), was determined to be in the order of $a \approx 12$-13, which was recently confirmed by two other studies based on large, independent patient populations (Fellini *et al.* 2008 [18]: $a \sim 11.8$ (1119 patients); Tucker *et al.* 2008 [63]: $a \sim 12.5$ (1023 patients)). Thus, a volume-effect parameter of $a \approx 12$-13 can be regarded as well established for prospective EUD-based treatment planning.
Apart from this, correlations of single dose-levels with toxicity were investigated in this work in terms of a cutoff-dose logistic regression model. This directly dose-volume based

model generally fitted the data worse, but still found significant correlation of rectal wall relative volume above single cutoff-dose levels D_c in the range ∼50-80 Gy and toxicity for the patient population under investigation. This can be clearly traced back to the 4-field box technique used, which induces correlations of all DVH dosebins in this dose range as shown by the first two eigenvectors resulting from a PCA of the DVH data (sect. 2.1.1). In fact, the finding that V_D correlates to bleeding for all doses $D \sim$ 50-80Gy is in contradiction to results of some other studies that reported significant correlations for $D \sim$ 40-50Gy [24], 40.9Gy [64] or 60Gy [74]. In these studies a six-field conformal technique was used for all or part of the treatment/population, which induces different correlations of the DVH dosebins among the patients, and is the likely reason for these differences. Generally speaking, if a treatment technique generates correlations in the dose distributions of a population, the biological significance of certain dosimetric variables may be attached to correlated, yet biologically insignificant variables. These latter variables "lend" their biological significance from the truly significant parameters to which they are correlated. This risk is particulary pertinent if dose-volume constraints (i.e. cutoff-dose models) are assessed for their clinical significance. PCA of DVH data is a powerful method to reveal such correlations (sect. 2.1).

Concerning acute small bowel complications, the planned value of V_{15Gy}, i.e. the small bowel volume receiving doses above 15Gy, was found most relevant for developing diarrhea of grade 3 or worse in a population of rectal cancer patients predominantly treated with a three-field 3D-CRT technique (pelvic treatment to 45Gy). Thus, treatment planning with similar three-field techniques may use this dosimetric parameter as constraint to prevent excessive small bowel reactions to irradiation. Of note, for such techniques the fraction of small bowel volume that is covered by only one of the three treatment beams is closely related to the value of V_{15Gy}. Following the discussion above, this suggests that the observed correlations of V_{15Gy} and toxicity may be influenced by the three-field technique used for the study's patient population. In this context, two studies by other authors (Roeske *et al.* 2008 [49] and Fiorino *et al.* 2008 [19]) that used different treatment techniques (IMRT and four-field box/IMRT), found the high dose levels as most relevant for developing acute diarrhea. As these studies are also different in other aspects (toxicity endpoints; volume definition; treatment concepts), a final conclusion about the truly 'biologically relevant' dosimetric parameters for this kind of toxicity necessitates further studies, such as applying EUD-based models to small bowel data.

Chapter 3

4-dimensional geometric modelling

Radiotherapy is confronted with geometric uncertainties of various nature such as differences in the geometric configuration (position, shape) of the tumor and adjacent organs at time of treatment as compared to planning. This can lead to a degradation of tumor dose and increased toxicity if it is not appropriately compensated for in the process of planning and treatment.

Geometric uncertainties arise from the impossibility of reproducible patient positioning relative to the radiation equipment prior to a treatment session (*setup errors*) as well as *internal organ motion* relative to the patient's bony structure. In this context, *intrafractional uncertainties* refer to geometric changes happening in the timescale of a treatment session, such as internal organ motion caused by breathing. *Interfractional uncertainties* are differences of patient setup and -geometry between the repeated treatment sessions of a multifractionated treatment course. Differences of the patient geometry can have *rigid* and *non-rigid* components. Mathematically, rigid motion can be described relatively simple as combination of a translation and rotation, while modelling of non-rigid geometric changes, i.e. *deformations*, is demanding. In current clinical practice, the influence of deformations is usually neglected, which is a valid approximation for uncertainties like interfractional setup errors in treatment of pelvic or intracranial tumors, and to some extent also for internal motion of certain tumors. Especially internal organ motion, however, shows significant deformable components. Important examples are deformations within the pelvic region caused by inter- and intrafractional changes in bladder and rectum filling, as well as respiratory deformations within the thorax. As a consequence, the moving tissue elements accumulate dose at different positions of the treatment beam during irradiation. Thus, new radiotherapy concepts that aim to incorporate the accumulated tumor and normal tissue dose into optimization and evaluation (chapter 4) inherently have to account for deformations. In this sense "4-dimensional modelling" refers to methods that allow for efficient modelling of inter- and intrafractional deformations of the patient geometry.

Two such methods have been developed in this PhD work to address deformable uncertainties of both *random* and *quasi-periodical* character. Section 3.1 introduces a method to extract the fundamental 'deformation modes' from random organ deformation samples.

Based on the example of interfractional prostate, bladder and rectum deformations, the method is shown to facilitate efficient low-order parametric modelling of such deformable uncertainties. Section 3.2 presents the use of a novel deformable registration method for modelling quasi-periodical respiratory deformations based on RCCT ("4D-CT") datasets.

3.1 Modelling of random deformable uncertainties

An effective method of modelling deformations and correlated organ motion in terms of so-called *eigenmodes* which represent the essential characteristics of individual geometric variations was developed (Söhn *et al.* 2005 [58], see Appendix D). The method is based on a *Principal Component Analysis (PCA)* of random organ geometry samples. The PCA-approach allows for reducing the large dimensionality of deformation data from multiple CT-studies to a few-parametric statistical model of organ deformation.
After a brief introduction into the basic ideas of the method in section 3.1.1, exemplary applications of the method for new planning and evaluation approaches in radiotherapy of prostate cancer are presented in section 3.1.2. Its impact and use in the context of other new radiotherapeutic applications is discussed in section 3.1.3.

3.1.1 Principal component analysis (PCA) of random organ geometry samples

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 is given by the $3M$-dimensional vector

$$\boldsymbol{p}(i) = (\boldsymbol{x_1}(i), \ldots, \boldsymbol{x_M}(i)) \in \Re^{3M}, i = 1 \ldots N \tag{3.1}$$

of positions $\boldsymbol{x_j}(i) \in \Re^3$ of $j = 1 \ldots M$ points distributed over the organ surface or the entire organ. Here 'organ geometry' may refer to a single anatomical organ or a composite geometry of multiple organs, where in the latter case eq. (3.1) is to be understood as collection of points distributed over multiple anatomical organs. For the following it is important that for each point index j the series $\boldsymbol{x_j}(1), \ldots, \boldsymbol{x_j}(N)$ refers to the same anatomical position, i.e. the set of geometry vectors $\{\boldsymbol{p}(i)\}_{i=1\ldots N}$ is a collection of *corresponding points*.

For many phenomena in radiotherapy involving deformations information about the exact anatomical and physical processes that drive the deformations is not available. Thus it is desirable to develop a *surrogate model* of the deformation process. For this, the set of geometry vectors $\{\boldsymbol{p}(i)\}_{i=1\ldots N}$ is assumed to represent random samples from a stochastic process of organ motion/deformation. Obviously, due to anatomical reasons the displacements of the M organ points are highly correlated, which implies that the underlying dimensionality of this multivariate problem is actually much smaller than $3M$. This motivates a surrogate model that follows the idea to substitute biomechanical properties of moving/deforming organs by observed correlations in the motion of tissue elements.

The method at hand to extract information about correlated organ deformations from this $3M$-dimensional problem with N samples is an approach from multivariate statistics, the *Principal Component Analysis (PCA)* [38, 27]. This involves calculation of the *mean geometry* $\bar{\boldsymbol{p}} \in \Re^{3M}$ and the *covariance matrix* $\mathcal{C} \in \Re^{3M \times 3M}$:

$$\bar{\boldsymbol{p}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{p}(i) \tag{3.2}$$

$$\mathcal{C} = \frac{1}{N-1} \sum_{i=1}^{N} \big(\boldsymbol{p}(i) - \bar{\boldsymbol{p}}\big) \cdot \big(\boldsymbol{p}(i) - \bar{\boldsymbol{p}}\big)^T \tag{3.3}$$

Here $() \cdot ()^T$ denotes the outer product of the two $3M$-dimensional vectors. Being interested in correlations in the movement of the M organ points, the covariance matrix is diagonalized. The resulting *eigenvectors* $\boldsymbol{q_l} \in \Re^{3M}$ of the covariance matrix are the mutually uncorrelated modes of deformation, so-called *eigenmodes*. Each eigenvector defines a 3D-vector field of correlated displacements for the M organ points. Each *eigenvalue* quantifies the variance of the N geometry samples in the ($3M$-dimensional) direction of the corresponding eigenmode. Thus the eigenmodes with largest eigenvalues represent the 'principal' deformation modes, that span the space of largest variability.

New geometry samples can be simulated within the low dimensional 'deformation space' by deforming the mean shape with a weighted sum of eigenmodes:

$$\boldsymbol{p}_{simul} = \bar{\boldsymbol{p}} + \sum_{l=1}^{L} c_l \boldsymbol{q_l} \tag{3.4}$$

where the sum runs over the first L dominating eigenmodes, and, according to the theory of PCA, the coefficients $c_l \in \Re$ obey Gaussian distributions with the corresponding eigenvalues as variances. In this way, PCA provides a low dimensional, statistical model of individual organ deformation.

Quantitative evaluation

The potential of the PCA-method to model geometric variability of clinical patient data was investigated based on the example of collective motion/deformation of prostate, bladder and rectum as important model system in radiotherapy of pelvic tumors (see Appendix D).

For this study, data of 4 example patients with $N = 15 - 18$ manually contoured organ geometries each were used. The input data to the PCA-model according to eq. (3.1) were generated using a finite element method. Based on these data, the spectrum of eigenvalues was found to be dominated by only few values, which indicates that deformable motion of prostate/bladder/rectum as composite geometry is governed by only few independent processes that describe patient individual characteristics of internal organ motion and deformation. Specifically, about 80% of the overall geometric variability could be represented by the first 4 eigenmodes according to the PCA model.

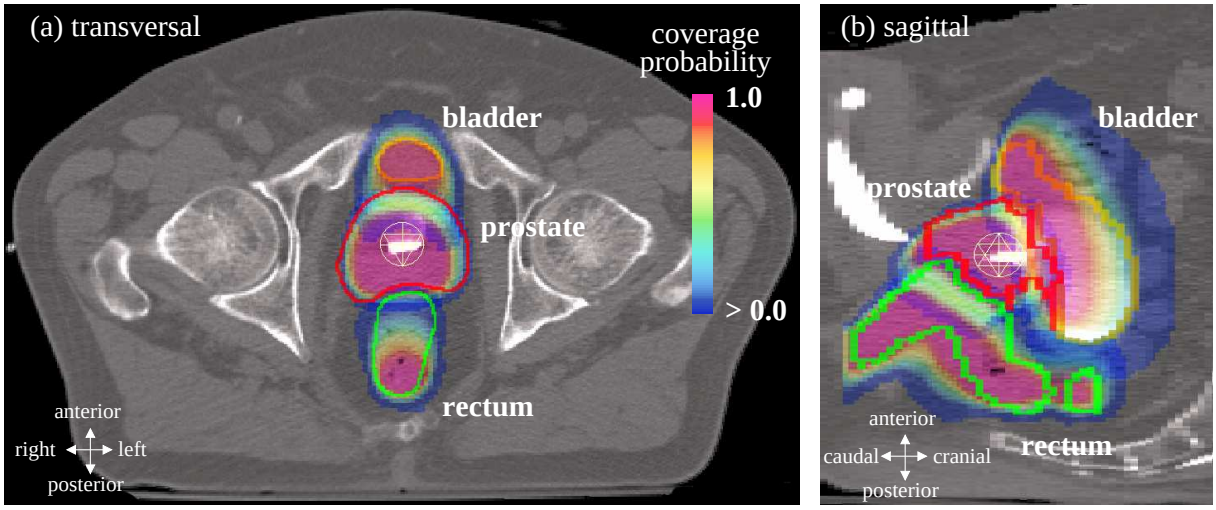

Figure 3.1: (a) transversal and (b) sagittal view of the pelvic region in the planning CT scan for the example patient. The color-overlay shows the coverage probability distributions of internal organ motion for prostate, bladder and rectum as calculated with a PCA-model of the three organs. The contours depict the positions and shapes of the organs as seen in the planning CT.

The capability of the PCA model to represent measured organ geometries according to eq. (3.4) was evaluated by calculating residual errors of organ surface points for a varying number L of dominating eigenmodes. It could be shown that the residual errors decrease fast with increasing number of eigenmodes used in the PCA model. Using $L = 4$ eigenmodes, the organ surface residuals of all four patients were found to be well below 2mm for all three organs.

3.1.2 Applications of the PCA-method in radiotherapy

PCA is an efficient way to statistically model patient individual organ deformation characteristics. The key to this is given by eq. (3.4). This equation describes how synthetic organ geometries not measured previously can be generated, which follow the multidimensional statistics of the deformation samples used as input to the model. Two important applications of this are described in the following: The prediction of coverage probability distributions and the dosimetric evaluation of treatment courses.

Prediction of coverage probability distributions for use in probability-based treatment planning

Mathematically, the *coverage probability distribution* associated with a moving/deforming organ structure is a scalar field of values in the range $[0 \ldots 1]$, which for each point within the patient gives the frequency of being covered by any part of the organ. Coverage probabilities have been used in the field of radiotherapy as advanced method for margin

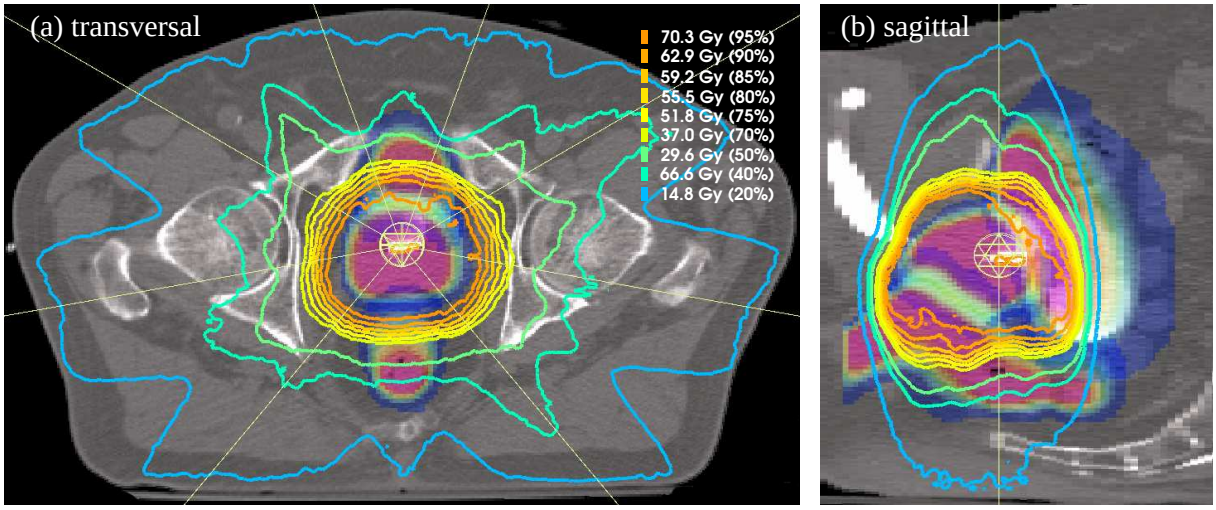

Figure 3.2: IMRT treatment planning incorporating coverage probability as generated with the PCA-model into optimization: (a) transversal and (b) sagittal view of the dose distribution of an example plan for the patient shown in fig. 3.1. The directions of the incident treatment beams are depicted by yellow lines.

definition [61] as well as in the context of probabilistic treatment planning [9, 10]. The latter approach incorporates information about organ specific coverage probabilities into the optimization process by associating an importance factor to each tumor and OAR voxel according to its coverage probability. For the example of prostate radiotherapy, this approach provides a favorable ratio of tumor vs. OAR doses as well as improved plan robustness in a population of patients when compared to conventional margin-based planning [10].

At time of planning the actual patient individual coverage probability is unknown. Thus a surrogate distribution predictive for the treatment course has to be estimated. For this, Baum *et al.* 2006 [10] used five pre-treatment CT scans, and calculated the coverage probability of internal organ motion by directly using the five nominal geometry samples as discrete model of organ motion. In this context, the PCA-method provides an advanced statistical model of organ motion/deformation that can be used for coverage probability generation as demonstrated by Söhn *et al.* 2004 [56] and Price and Moore 2007 [48].

As example for use in prostate radiotherapy, fig. 3.1 shows the coverage probabilities of internal organ motion for prostate, bladder and rectum as generated with the PCA-method. The example patient showed significant geometric variability, which is reflected in the spatially wide-spread coverage probability distributions of all three organs and their large mutual overlap. These distributions were used to create a coverage probability-based prostate treatment plan (fig. 3.2). For this plan a poisson cell-kill EUD of 74Gy in 37 fractions was prescribed to the prostate, see eq. (4.7), with serial EUD constraints to rectum and bladder as dose-limiting organs. In the overlap regions of prostate and

rectum/bladder the trade-off between the conflicting goals of dosimetric tumor coverage and OAR sparing could be solved by the optimizer based on the predicted coverage probabilities of the organs, which results in satisfying tumor coverage even in the presence of considerable internal organ motion in the course of treatment (see also fig. 3.3 in the next subsection). This is an important advantage over conventional margin-based planning, where the considerable overlaps of PTV and rectum/bladder lead to ill-defined trade-off problems during plan optimization, typically preventing sufficient dose coverage of the PTV or excessive dose to the OARs.

Dosimetric treatment course evaluation

Geometric uncertainties as introduced by internal organ motion and other sources of error translate into dosimetric uncertainties at time of treatment planning. The PCA-method provides a powerful tool for 'virtual treatment simulations' with full information about motion of individual organ points by sampling organ geometries $\boldsymbol{p}_{simul} \in \Re^{3M}$ from the 'deformation space' spanned by the patient individual eigenmodes according to eq. (3.4). This facilitates predictive dosimetric evaluation of treatment plans by simulating a multitude of treatment courses and their accumulated organ doses.
Fig. 3.3 shows an example, where a composite PCA-model initialized with $N = 5$ input geometries of internal motion of prostate, bladder and rectum is used to calculate distributions of cumulative EUDs of the three organs for a given dose distribution. For this, 1000 treatment courses with 37 fractions each were simulated by sampling altogether 37.000 composite organ geometries of prostate, bladder and rectum from a PCA-model according eq. (3.4) with $L = 3$ dominating modes. For each of the simulated treatment courses the accumulated organ doses were calculated using the dose distribution of the coverage probability plan described above (fig. 3.2) as example. This allows quantification of the EUD distributions of the three organs (fig. 3.3c) as well as investigation of EUD-correlations in terms of joint distributions (fig. 3.3d). For this example, the coverage probability plan proves robust with respect to the dose to prostate, which shows only small variability around the prescribed EUD of 74Gy. Absolute values of bladder and rectal wall EUDs are acceptable, however, especially for rectal-wall the EUD variability is relatively large due to the large geometric variability of rectum (see also fig. 3.3a).

3.1.3 Discussion and Outlook

The work presented in this PhD thesis is the first application of PCA to model patient-individual organ deformation characteristics. Based on the example of deformable motion of prostate, bladder and rectum, it could be shown that patient-individual deformation characteristics as taken from a multiple CT study can typically be modelled by $\lesssim 4$ dominating eigenmodes (Söhn *et al.* 2005 [58]). This facilitates the creation of low-order parametric statistical organ motion/deformation models, rendering the PCA-method as powerful tool for 'virtual treatment simulations' with widespread potential for use in the context of probabilistic treatment plan optimization and evaluation as well as adaptive

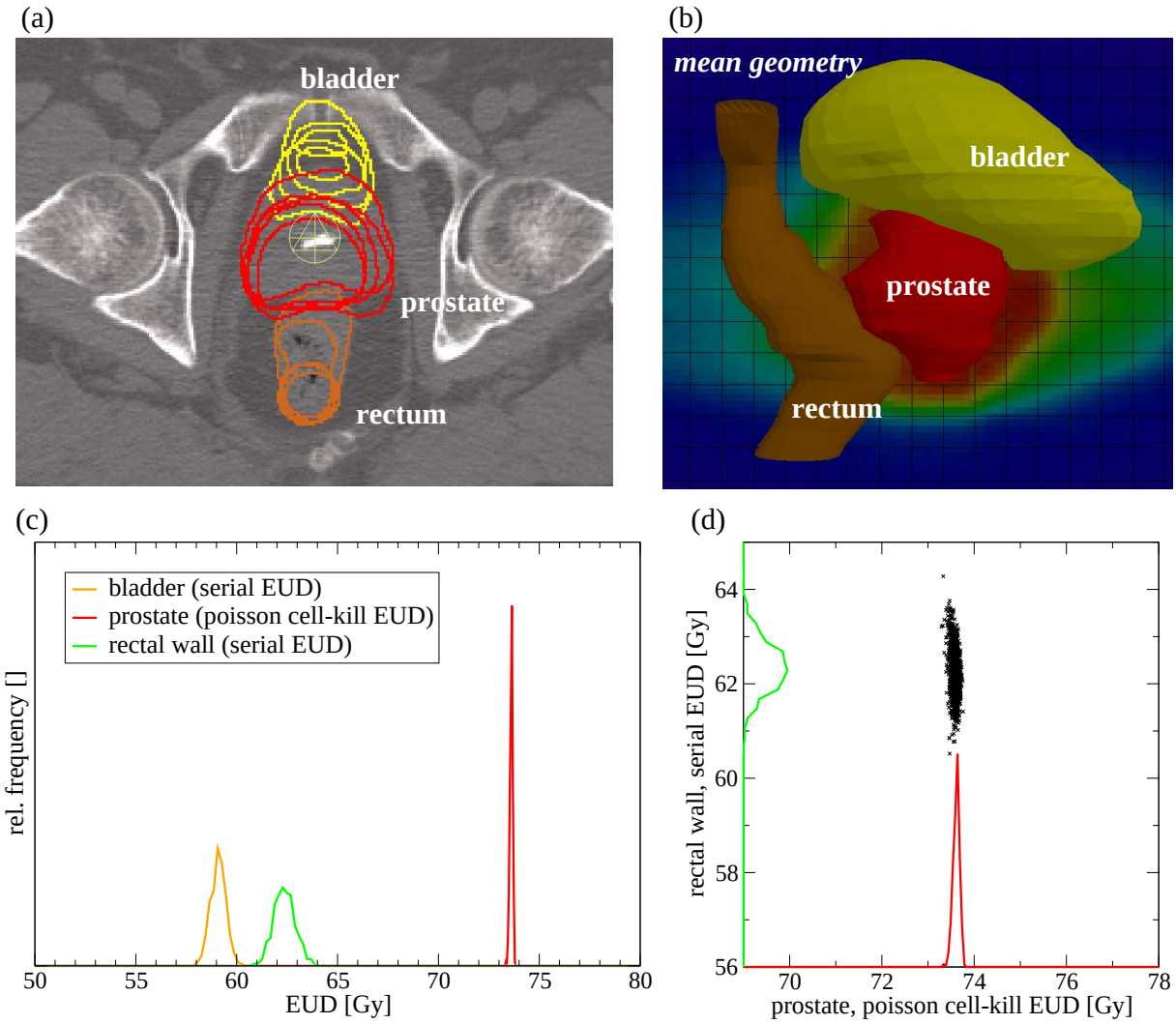

Figure 3.3: Use of the PCA-model for dosimetric treatment course simulation. (a) transversal CT slice of the example patient in figs. 3.1&3.2 with contours for bladder, prostate and rectum as seen in the planning CT and the first four daily pretreatment CTs. These organ geometries were used as input for the PCA-model. (b) The mean geometry of the PCA-model as 3D-view in front of the dose distribution of the coverage probability plan in fig. 3.2. (c) EUD distributions resulting from repeat PCA-based treatment course simulations (1000 simulated treatment courses of 37 fractions each). Information about the internal EUD-correlation structure is readily available in this approach, as shown in the scatterplot (d) for the example of prostate–rectal wall treatment course EUDs. [parameters of EUD calculations: prostate: poisson cell-kill EUD with $\alpha = 0.4$, see eq. (4.7); bladder/rectal wall: serial (power-law) EUD, eq. (2.1), with $a = 8.0/12.0$]

radiotherapy.
Two exemplary applications of this have been presented in section 3.1.2. In the first example, a probabilistic prostate treatment plan was created which utilized a PCA-model of prostate/bladder/rectum for predictive generation of patient-individual coverage probability distributions of the three organs. As second important application of PCA, a dosimetric evaluation of the coverage probability plan as example was performed by repeated calculation of the accumulated dose-to-moving-tissue for a large number of simulated treatment courses. The resulting EUD distributions are a direct measure of plan robustness against organ motion. Apart from this, the method allows spatially resolved visualization of dosimetric uncertainties (not shown) based on dose accumulated within moving tissue elements, as the motion of individual tissue elements is explictly modelled by PCA. Such local dose uncertainty information provides valuable guidance in treatment plan evaluation and robust optimization schemes [36]. The PCA-model in this example used $N = 5$ organ geometry samples as input. This reflects a typical situation encountered in adaptive radiotherapy concepts where a new, adapted treatment plan is to be created based on image information aquired in the first week of treatment [71, 39]. Thus, in robust adaptive radiotherapy schemes, PCA-based dosimetric treatment course evaluation has potential as a central tool to quantify robustness of plan adaptations.

Apart from its application to model random deformable uncertainties as shown for the example of prostate, bladder and rectum, PCA has been proposed for other applications in radiotherapy. Zhang *et al.* 2007 [75] used PCA to model respiratory motion of lung based on displacement fields determined with deformable registration of RCCT-datasets as input. Two dominating modes were found adequate to model lung motion. As respiratory motion is of non-random and non-gaussian nature (section 3.2), the PCA-method had to be adapted to account for the non-gaussian character of this problem. Birkner *et al.* 2007 [13] used PCA to analyze and model patient setup uncertainties in head and neck radiotherapy as seen in portal image (PI) datasets. Such uncertainties were quantified as 2D-shifts of a number of typically 3-5 rectangular PI-subregions, which served as input for PCA. In this way, patient-individual translational, rotational, 'nodding' or 'tilting' modes could be identified. This can serve as quantitative basis for an appropriate choice of margins in different anatomical regions and for different setup methods and -equipment. Papiez and Langer 2006 [45] applied PCA to model target uncertainties in terms of rigid 3D-shifts, thereby generalizing commonly used margin-'recipes' (Stroom *et al.* 1999 [61], van Herk *et al.* 20000 [66]) to the case of correlations between shift-components. Another application of PCA in radiotherapy is model-based organ segmentation [47, 46], where population-based PCA-models of anatomical shape variability [32] are used to guide automatic segmentation algorithms.

A practical issue relevant for future implementation of PCA-methods into clinical workflow is the problem of efficient generation of input data in terms of sets of corresponding points according to eq. (3.1). This can be done with image-based deformable registration methods as in [75], or based on manual organ contours as defined by physicians. In the latter case, finite element methods (FEM) as in [58] can be applied to register the data samples. However, for clinical use numerically less expensive alternative methods

are desirable. This has been addressed by Söhn *et al.* 2004 [56] and Price and Moore 2007 [48], who proposed simplified schemes for generation of registered sets of corresponding points from manually segmentated data. Possible inaccuracies resulting from approximations made in such schemes typically manifest themselves as statistically uncorrelated errors ('noise') which appear only in eigenmodes with small eigenvalues. The problem of input data generation has recently been adressed in detail by Merck *et al.* 2008 [42].

Summarizing, PCA is an efficient method to quantify the essential characteristics of correlated organ motion/deformation in terms of few dominating eigenmodes. These span the patient-individual deformation space of the treatment geometry, which provides an efficient statistical surrogate model of organ mobility with full information about motion of individual tissue elements. Thus PCA has potential as fundamental '4D-tool' for new probabilistic treatment evaluation and optimization schemes based on accumulated dose-to-moving-tissue.

3.2 Modelling of quasi-periodical deformable uncertainties

Respiratory motion is the main process inducing intrafractional quasi-periodical geometrical uncertainties in the field of radiotherapy. Here, novel time-resolved 3D-imaging techniques like respiratory-correlated computed tomography (RCCT) provide "4D" image information with previously unprecedented quality, which enables direct modelling of quasi-periodical respiratory motion based on deformable image registration. For this purpose, a fast model-independent deformable image registration algorithm was developed in this work (Söhn *et al.* 2008 [57], see Appendix E).
After section 3.2.1 gives a brief introduction into the basic ideas of the algorithm, section 3.2.2 shows how this can be used to construct patient-individual probabilistic thorax models for 4D-radiotherapy. The significance of the presented deformable registration algorithm in this context, as well as its potential for other applications is discussed in section 3.2.3.

3.2.1 Model-independent deformable image registration by Local Matching of Anatomical Features

Given a *reference (source) image* and a *target image*, deformable image registration algorithms optimize a set of image transformation parameters based on a similarity measure to align the images. A multitude of algorithms has been proposed [22, 35, 40, 52], which differ in the type of transformation and the similarity measure. Common to the majority of algorithms is that similarity is quantified by *global* similarity measures. As opposed to this, the basic idea of the the algorithm proposed here is to break down the problem of global deformbable registration to (fast) *local rigid registrations* of small subvolumes – denoted as *"featurelets"* –, which are regularly distributed over the image region to

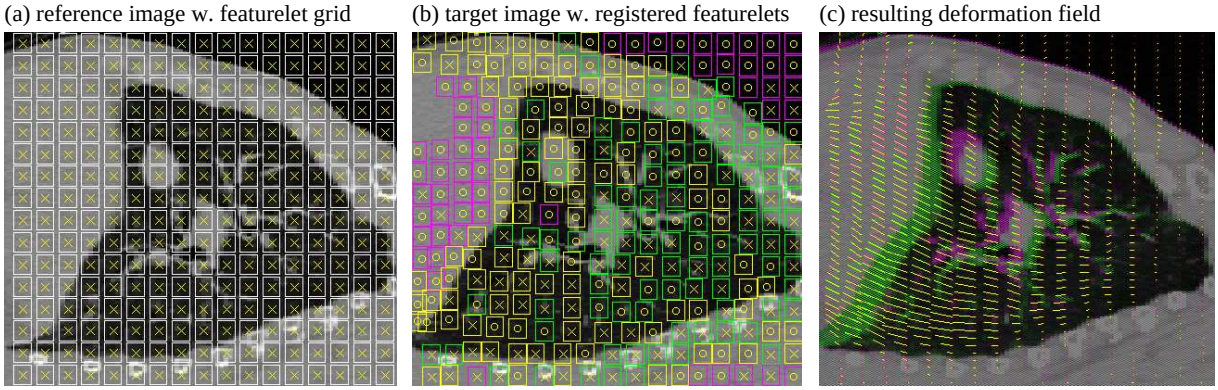

Figure 3.4: Featurelet-based deformable registration for a lung example patient (sagittal view). (a) Grid of reference featurelets in the exhale CT. (b) Registered featurelets in the inhale CT after local registration and relaxation. The colors refer to the results of the local registration quality assessment (see fig. 3.5). (c) Red-green overlay of inhale and exhale CT and the displacement field as calculated by the deformable registration algorithm.

be registered. This concept is motivated by deformable registration of organs like lung which comprise a large number of internal *anatomical features*. Assuming that these *locally* show only a small amount of deformation and thus can be treated as *locally* rigid, featurelet-based deformable registration allows fast, model-independent deformable image registration of both mono- and multimodal registration problems. In the literature, this class of algorithms is also known as template- or block-matching [50, 28, 37].

In short, featurelet-based deformable registration as developed in this work can be described as a five-step algorithm:

Step 1: Initialization of the featurelet grid The image region to be be registered is covered by a regular 3D grid of equally sized rectangular 3D regions ("featurelets") in the reference image (see fig. 3.4a). The featurelet size is a parameter of the algorithm to be chosen according to the typical size of distinct anatomical features. For lung registration, a value of ∼15mm was found to be appropriate as edge length of featurelet regions. This typically results in $\sim 10^3$ featurelets needed to cover the thorax region.

Step 2: Local rigid registration It is assumed that reference and target image are rigidly preregistered such that center of mass shift and bulk rotation vanish. The basic idea of the algorithm is to individually register each reference featurelet to the target image in terms of a local translation, i.e. to determine a 3D-shift vector $\boldsymbol{\delta X}_i$ for each featurelet i such that the similarity between the reference featurelet and the shifted target featurelet is maximized within a predefined search region.
Several similarity measure functions have been proposed for quantification of image similarity [22]. The algorithm shows great flexibility with respect to the similariy

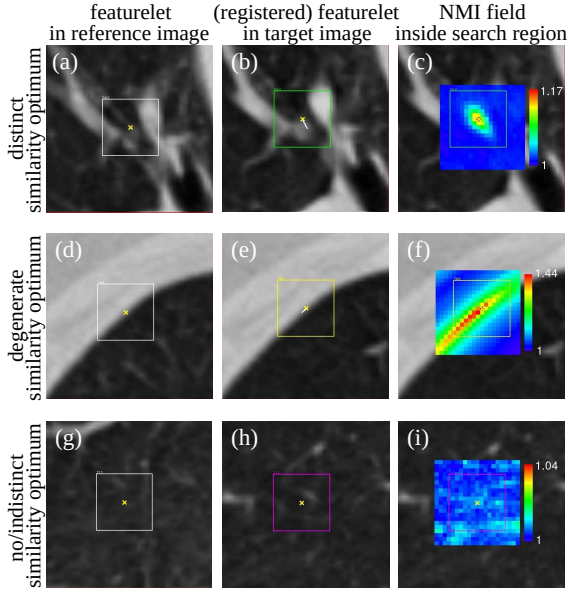

Figure 3.5: Examples for the three typical situations encountered in local rigid registration of featurelet regions: distinct (a-c), degenerate (d-f) or no/indistinct (g-i) similarity optimum. First column: featurelet position in the reference image; second column: (registered) featurelet in the target image; third column: NMI as similarity measure for all possible shifts of the featurelet inside of search region. (from [57])

measure, which may be chosen according to the nature of the registration problem. Generally, the normalized mutual information (NMI) is a robust similarity measure for inter- and intra-modality registration problems. Alternatively, the correlation coefficient (CC) may be used as numerically efficient similarity measure for intra-modality registration.

Step 3: Assessment of local registration quality The local registration quality depends on the amount of image information available inside of a featurelet region. Essentially three situations can be distinguished. In the first, ideal case, the local image information is sufficient to allow unambiguous featurelet registration. Fig. 3.5a-c depicts a typical example for this situation, where the optimal position of the target featurelet is characterized by a unique, distinct optimum of the similarity measure. In the second case, local image information does not allow unique featurelet registration, which is expressed as a degenerate optimum of the similarity measure function (fig. 3.5d-f). In the third case, the featurelet covers a region with low or no image contrast, thus the similarity measure function shows low similarity values for all possible shifts and no clear optimum (fig. 3.5g-i). An automatic categorization scheme to distinguish the three cases has been developed. The scheme is based on the absolute values of the similarity measure within the search region as well as the eigenvalues of the Hessian (curvature) matrix at the similarity measure

optimum (details in appendix E).

Step 4: Relaxation Due to inevitable local registration ambiguities or misregistrations (second and third case described above), a featurelet-based registration ansatz solely based on local image information produces unphysical deformation fields which locally violate injectivity and continuity requirements. This has to be corrected for by imposing additional assumptions which ensure physically meaningful deformation fields. This is realized in the algorithm by moving mis- and ambiguously registered featurelets (second and third case described above) to positions minimizing the overall deformation energy of the featurelet grid. For featurelets with degenerate similarity optimum (second case) the additional constraint to stay within the degenerate optimum is imposed.
Numerically, minimization of the deformation energy was implemented as a *relaxation algorithm*: Assigning a virtual mass m_i to each featurelet, the dynamics of the i-th target featurelet as represented by its center position $\boldsymbol{X}_i$ is subject to the following physically motivated equation of motion:

$$m_i \ddot{\boldsymbol{X}}_i = -D \cdot \dot{\boldsymbol{X}}_i + \boldsymbol{F}^{[i]}_{\text{internal}} + \boldsymbol{F}^{[i]}_{\text{external}} \tag{3.5}$$

Here D is a damping factor, and the dynamics is driven by two types of forces, which are typically referred to as 'internal' and 'external' force [40]. The internal force gives rise to the deformation energy of the featurelet grid. In the present implementation, $\boldsymbol{F}^{[i]}_{\text{internal}}$ for a featurelet i was modelled as resulting force imposed by virtual springs connected to the centers of all direct 3D-neighbors. For featurelets with degenerate similarity optimum the locally available image information enters the relaxation process through the external 'image force', which is represented by the gradient of the similarity measure here. More details about the numerical implementation of the relaxation step are given in appendix E.

Step 5: Calculation of the global deformation field The shift vectors $\boldsymbol{\delta X}_i$ of the target featurelets after local registration and relaxation (fig. 3.4b) represent the displacement field sampled on the regular grid of featurelet centers in the reference image. From this, the deformation field at any position (fig. 3.4c) is obtained by B-spline interpolation [51].

3.2.2 Patient-individual probabilistic thorax models

The respiratory-correlated CT (RCCT) technique [21, 67] allows to reconstruct thorax geometries of different respiratory phases, which provides a straightforward way to represent the dynamics of patient-individual breathing motion in terms of multiple phase CTs. An example for a RCCT dataset of a lung cancer patient is shown in fig. 3.6.

Applications in 4D-radiotherapy that aim to explicitly incorporate quasi-periodical geometrical uncertainties into dosimetric evaluation and optimization (sect. 4.1) require

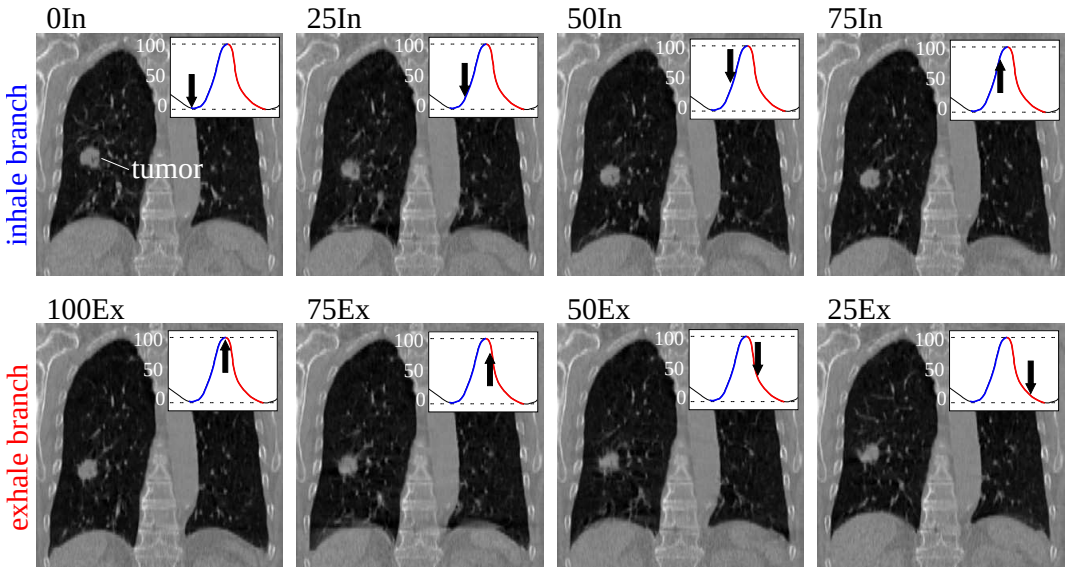

Figure 3.6: Coronal view of a RCCT-dataset, which was reconstructed into eight respiratory phases equally distributed over the full breathing cycle. The RCCT technique allows separate reconstruction of phases in the inhale ('In') and exhale ('Ex') branch of the breathing curve for user-defined respiratory levels in the range $0 \ldots 100$ (0=full expiration, 100=full inspiration). Breathing excursion of the tumor for this example patient is 2.9cm.

patient-individual thorax models with information about motion of individual tissue elements. The central tool for this is deformable registration relating image-voxels of different RCCT-geometries in terms of displacement fields.
The potential of featurelet-based deformable registration (sect. 3.2.1) for creation of accurate thorax models has been evaluated in Söhn *et al.* 2008 [57]. Evaluation based on a virtual thorax phantom with known displacement field showed that ∼95% of all lung voxels had registration residuals $\leq$ 3mm with an average of 1.1 ± 1.2mm. This was supported by data of four patients, where the inhale-exhale registration residuals were found to be 1.6 ± 1.0mm based on anatomical landmarks distributed over the lung.

For calculation of the dose-to-moving-tissue as key quantity of 4D-radiotherapy information about the relative time spend in each breathing phase is required. As can be seen in fig. 3.7a, breathing curves may show significant irregularities over the time-scale of a treatment fraction. This very nature of breathing as *quasi*-periodical process renders deterministic modelling of respiratory motion infeasible. However, a probabilistic description in terms of the probability density function (pdf) of the breathing curve provides information about the relative time spend in each respiratory phase *on average*. Thus, each phase CT of an RCCT-dataset is associated with a weight factor expressing the relative importance of the respective instance geometry to represent a 'snapshot' of the underlying dynamic process (fig. 3.7b). In combination with information about geometric motion of the tissue elements from deformable registration, this constitutes a probabilistic thorax

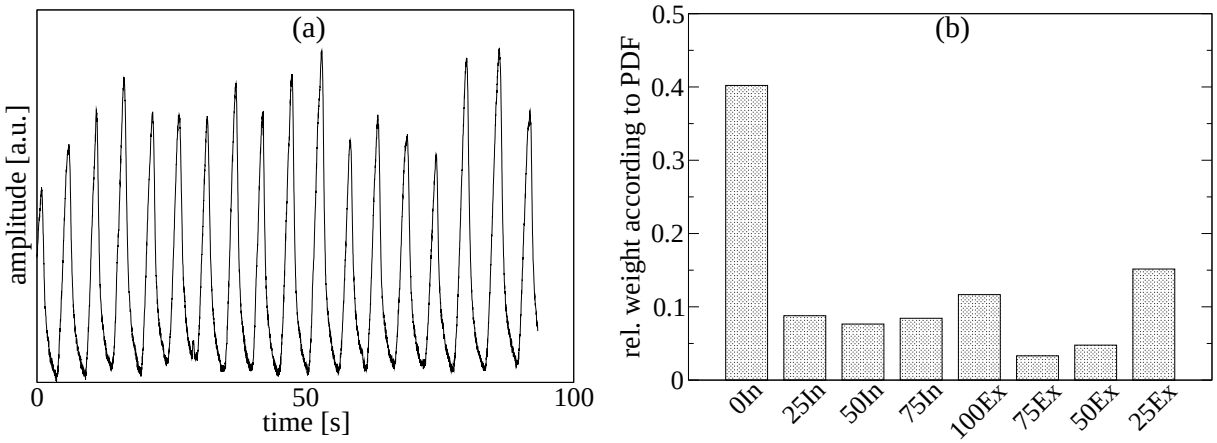

Figure 3.7: (a) Breathing curve of the example patient from fig. 3.6 as recorded during aquisition of the RCCT. (b) Relative time spend in the different breathing phases the RCCT dataset as derived from the pdf of the breathing curve in (a).

model, enabling patient-individual calculation of the *expected* dose-to-moving-tissue (see also sect. 4.1 and appendix F)

3.2.3 Discussion and Outlook

Rigorous biomechanical deformation modelling [70] is difficult as a priori information about tissue properties [17] and tissue boundary conditions [29] is not available or subject to uncertainties. Thus, it is the idea of deformable registration algorithms to complement the limited knowledge of physical properties by image information. The deformable registration algorithm developed in this work follows the concept of dividing the global registration problem into rigid registrations of a multitude of small image subvolumes ("featurelets"). It is the advantage of the presented algorithm that it allows model-independent registration in image regions with sufficient local image information, while it provides physically reasonable interpolation by means of minimization of elastic energy in regions with no or ambiguous image information. This concept proves especially efficient for RCCT-based deformable registration of lung with its numerous internal image features, but has also been successfully applied to intermodality registration of head-and-neck datasets (Söhn *et al.* 2008 [57]).

Deformable registration algorithms are a central tool of "4D radiotherapy". As shown in sect. 3.2.2, deformable thorax models based RCCT datasets employ patient-individual deformation fields to estimate dose-to-moving-tissue. As this involves multiple deformable registrations, algorithmic performance in terms of calculation speed is an important issue with respect to clinical feasibility. In this context it is a strength of the presented featurelet-based algorithm that it allows computational parallelization in a straightforward manner. This facilitates calculation times well below 1 min on recent multiprocessor/-core computer hardware for full registration of two 3D-image datasets, and

thus puts "*online* deformable registration" in reach. This is profitable also for other applications in adaptive radiotherapy such as dose accumulation based on daily treatment images [53, 34], online re-optimization [69] and contour propagation for fast re-contouring of online image datasets or interactive "4D-contouring" on RCCT datasets [68, 33, 76].

Summarizing, the presented algorithm allows accurate, 'almost-online' multi-modality deformable registration. Apart from its use in generation of probabilistic thorax models, it bears promise as versatile basic tool in adaptive and 4D-radiotherapy applications.

Chapter 4

4-dimensional radiotherapy: evaluation and optimization

Conventional radiotherapy planning approaches employ a static patient model based on a single CT scan to define the gross-tumor volume (GTV) and organs at risk (OARs) and to determine the dose distribution. This, however, is only a snapshot of the dynamic processes present in patient treatment. Inevitable geometric uncertainties such as rigid and non-rigid interfractional setup errors or intrafractional organ motion cause discrepancies between this planning CT scan and the treatment geometries. The common method to account for geometric uncertainties is the extension of the clinical target volume (CTV) by a margin, and to ensure coverage of the resulting larger planning target volume (PTV) (ICRU report 50 [1]). With regard to internal organ motion (eg. breathing) the ICRU report 62 [2] additionally introduced the concept of the internal target volume (ITV) as hull of possible tumor positions relative to bony structures. An ITV-to-PTV margin accounts for setup errors. Similarly, the ICRU report 62 [2] proposes the concept of expanding OARs to planning organ at risk volumes (PRVs). It is the fundamental conceptional shortcoming of this conventional approach that treatment plan quality is scored using a static dose distribution evaluated based on static surrogate volumes (PTV, OARs/PTVs) only. This is a rather incomplete model of reality in the presence of large geometric uncertainties, which can both alter the positions of the CTV and OARs relative to the dose and the dose distribution itself.

The term *"4-dimensional radiotherapy"* refers to new treatment planning concepts that explicitly utilize *dynamic* patient models to determine the *accumulated dose-to-moving-tissue* to the moving and deforming CTV/OAR-structures. As opposed to the PTV- and OAR-/PRV-doses considered in static planning, this is the truly relevant quantity for plan optimization and evaluation. Following the notion of 'beam's-eye-view' (BEV) in conventional radiotherapy, the transition from static dose surrogates to the actual accumulated dose-to-moving-tissue is denoted as *'tissue-eye-view' (TEV)*. In TEV, the accumulated dose changes physically – due to changes in the density distribution – as well as physiologically, as tissue elements move away from the planning position. The former can be approximated by explicit dose recalculation in multiple instances of the patient geometry.

The latter requires displacement field information from deformable image registration or statistical deformation models as described in chapter 3. Formally, this is expressed by the following equation for the dose-to-moving-tissue D_{accum} as accumulated from the instance dose distributions D^k of N geometric instances:

$$D_{accum} = \sum_{k=1}^{N} w_k \cdot \tilde{D}^k = \sum_{k=1}^{N} w_k \cdot (\mathcal{W}_k \circ D^k). \tag{4.1}$$

Here the w_k's are instance weight factors to be chosen according to the specific dose accumulation problem, $\tilde{D}^k$ are the instance dose distributions as *warped* to a common reference geometry for the purpose of dose accumulation, and $\mathcal{W}_k$ is the *warping operator* which mediates the coordinate transformation:

$$\tilde{D}(\boldsymbol{x}) = (\mathcal{W} \circ D)(\boldsymbol{x}) = D(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})), \tag{4.2}$$

where $\boldsymbol{d} = \boldsymbol{d}(\boldsymbol{x}) \in \Re^3$ is the displacement field reflecting the local anatomical shifts between the reference and instance geometry.

New methods for optimization and evaluation based on the accumulated dose-to-moving-tissue have been developed in this thesis. Section 4.1 introduces a concept for "4D-IMRT planning" of moving lung tumors. Section 4.2 presents a clinically relevant application of "4D-evaluation" in terms of a study comparing different IMPT/IMRT prostate planning approaches with respect to robustness against organ motion. The challenges and capabilities of such new "4D-radiotherapy" concepts are discussed in section 4.3.

4.1 4D-IMRT planning of lung

A clinically feasible implementation of biological optimization in tissue-eye-view and its application to free-breathing treatment of moving lung tumors is presented in the following (see also Appendix F). The method is based on a patient-individual probabilistic thorax model (sect. 3.2.2) for optimization of the expected dose-to-moving-tissue and incorporates advanced dose calculation methods for realistic modelling of dose deposition in the presence of large tissue density inhomogeneities.
After an introduction into the concept of biological 4D-optimization in section 4.1.1, the potential of the method as compared to a number of conventional static approaches is presented based on an example lung patient with large breathing excursion (section 4.1.2).

4.1.1 Biological optimization in tissue-eye-view

Preliminaries: IMRT dose calculation in a static geometry

In IMRT concepts, the modulation of irradiation fields is expressed as fluence distribution Φ which is numerically represented by a set of fluence weights $\{\phi\}$ of n discrete fluence elements (beamlets or segments). For static dose calculation, the dose D_i to voxel i in the

fixed treatment room coordinate system is given by the weighted sum of contributions from all fluence elements as expressed by the *static dose operator* T:

$$D_i = \sum_{j=1}^{n} T_{ij}\phi_j \tag{4.3}$$

Here T_{ij} is the dose deposited in voxel i by fluence element j, i.e. T_{*j} holds the *static* dose distribution of fluence element j, which is determined based on the density information of a single planning geometry.

Dynamic patient models

According to eq. (4.1), the calculation of the accumulated dose-to-moving-tissue, D_{accum}, is based on a representation of the patient's dynamics by multiple deformable instance geometries. For 4D-radiotherapy of lung, such a *dynamic patient model* has been presented in sect. 3.2.2 in terms of a RCCT-based probabilistic thorax model, which uses a deformable image registration algorithm to determine the warping operators $\mathcal{W}_k$ in eq. (4.1). In this model, the instance weight factors w_k (with $\sum_{k=1}^{N} w_k = 1$) are provided by the probability density function (pdf) of the patient's breathing curve. Thus, D_{accum} in eq. (4.1) becomes an *expected dose-to-moving-tissue*, $\langle D \rangle$.

IMRT dose calculation in tissue-eye-view

Optimization of $\langle D \rangle$ is the aim of 4D-IMRT planning of free-breathing lung treatment. For this, eq. (4.1) can be written in terms of the fluence weights by expressing the instance doses D^k through eq. (4.3):

$$\langle D \rangle_i = \sum_{k=1}^{N} w_k \cdot \sum_{j=1}^{n} (\mathcal{W}_k \circ T^k)_{ij}\phi_j = \sum_{j=1}^{n} \langle T \rangle_{ij}\phi_j \tag{4.4}$$

Here $\langle T \rangle$ is the motion-averaged version of the static instance dose operators T^k with

$$\langle T \rangle_{ij} = \sum_{k=1}^{N} w_k \cdot (\mathcal{W}_k \circ T^k)_{ij}. \tag{4.5}$$

$\langle T \rangle_{*j}$ is the *expected dose-to-moving-tissue* of a single beamlet j (fig. 4.1). For beamlet-based IMRT-optimization of the fluence distribution, $\langle T \rangle_{ij}$ can be precalculated by separate warping of all beamlets according to eq. (4.5), which allows efficient updates of $\langle D \rangle$ according to eq. (4.4) during iterative dose optimization.

Pencil-beam dose calculation algorithm A dedicated pencil-beam algorithm was used for beamlet-based IMRT pre-optimization (Jeleń, Söhn and Alber 2005 [25], see Appendix G; Jeleń and Alber 2007 [26]). The algorithm incorporates lateral and longitudinal inhomogeneity corrections and was commissioned to Monte Carlo dose calculations

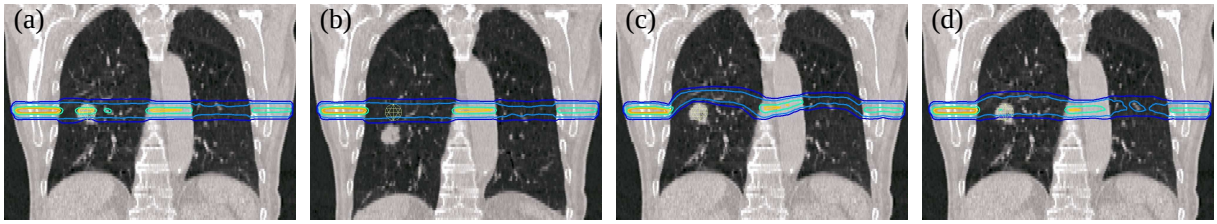

Figure 4.1: Beamlet warping: (a)/(b) beamlet doses calculated in the exhale/inhale CT of an example RCCT dataset, (c) beamlet dose of the inhale CT warped to the exhale reference geometry and (d) expected dose-to-moving-tissue of the beamlet, $\langle T \rangle_{*j}$, accumulated using all 8 CTs from the RCCT dataset according to eq. (4.5). Shown are the isodose lines of 95/70/40/10/5% isocentre dose for this beamlet as calculated with the pencil-beam algorithm.

(see below). As depicted in fig. 4.1a&b, the algorithm models effects like penumbra-broadening from lateral scatter and reduced dose deposition in low-density regions, as well as dose re-buildup in high-density areas.

4D-Monte Carlo (MC) dose calculation A fast Monte Carlo algorithm (XVMC [20, 54]) was used for dose calculation of the segmented plan in the final step of optimization. MC algorithms explicitly model the physical processes of radiation transport and deposition by simulating the trajectories of a large number of particles and their interaction with matter, and thus can be regarded as most precise dose computation method available in the presence of density inhomogeneities.

4D-dose calculations require separate dose computations in each of the N instance geometries. Thus, calculation time scales with N for most dose calculation methods. For MC algorithms, however, the calculation time is determined by the overall number of simulated particle histories n_{hist}. As the statistical accuracy of the accumulated dose solely depends on this number, 4D-MC calculations can be performed with a reduced number of histories in the single instance geometries.
This is accounted for in the algorithm presented here by reducing the number of simulated histories for each instance geometry, $n_{hist,k}$, according to the respective instance weight, i.e. $n_{hist,k} = w_k \cdot n_{hist}$. This leads to noisy instance dose distributions with reduced statistical accuracy (stat. variance $\sigma_k^2 \sim 1/n_{hist,k}$). However, assuming Gaussian propagation of errors when summing up the warped instance doses according to eq. (4.1), the overall stat. variance σ^2 of the accumulated dose distribution is determined by the overall number of histories as desired: $\sigma^2 \sim \sum_k w_k^2 \cdot \sigma_k^2 = 1/n_{hist}$. Thus, 4D-MC allows efficient calculation of the accumulated dose essentially independently of the number of instances.

Constrained biological optimization based on isoeffects

The isoeffect framework was used to incorporate knowledge about the dose-response characteristics of different tissues into optimization in terms of cost functions [4, 6, 7]. In this

concept, two dose distributions are regarded as equivalent with respect to their radiobiological effect to tissue, if their *isoeffects*

$$G = \sum_i \nu_i g\big(\langle D\rangle_i\big) \tag{4.6}$$

are equal. Here $g(D)$ is the *local effect density* for the organ under consideration, and the sum is defined over all organ voxels (ν_i: relative voxel volume). This gives rise to the definition of the corresponding *isoeffective dose* $D_{iso} = g^{-1}\big(G\big)$, which is motivated by the equivalent uniform dose (EUD) [44] and can be seen as *generalized EUD* (see also sect. 2.2). It is the strength of the isoeffect concept that it allows incorporation of both physical and biologically motivated costfunctions and objectives [6]. Subsequently, the isoeffects used for the lung planning example below are briefly introduced.

Tumor control Based on the standard Poisson TCP model, the local effect density is defined as the surviving fraction of clonogenic cells:

$$g_{poisson}(D) = e^{-\alpha D} \tag{4.7}$$

The corresponding isoeffective dose is the poisson cell-kill EUD [44].

Normal tissue parallel complication mechanism Motivated by NTCP models describing radiobiological effects in tissues with parallel complication mechanisms (eg. lung), the local effect density is defined as [4, 7]

$$g_{parallel}(D) = \left(1 + \left(\frac{d_0}{D}\right)^k\right)^{-1} \tag{4.8}$$

with k, d_0 as parameters of the sigmoidal function modelling local dose-response. The corresponding isoeffect can be interpreted as mean damage, i.e. as fraction of the organ that is functionally obliterated.

Maximum dose To avoid target hotspots or to control maximum OAR doses, a physical overdose constraint

$$g_{max}(D) = (D - D_{max})^2 \cdot \Theta(D - D_{max}) \tag{4.9}$$

is defined ($\Theta(\cdot)$: Heaviside step function). The corresponding isoeffective dose is the root mean square (rms) overdosage inside of the volume of interest.

In order to ensure the existence of a solution for a given set of prescriptions, IMRT-optimization was implemented in terms of a *constrained optimization* problem [6, 5], where tumor cell survival is minimized (i.e. tumor control is maximized) while isoeffects to critical structures are constrained to maximum tolerable values:

$$\arg\min_{\Phi} \; G_{poisson}\Big(\big\{\langle D\rangle_i(\Phi)\big\}\Big) \tag{4.10a}$$

$$\text{such that } G_m\Big(\big\{\langle D\rangle_i(\Phi)\big\}\Big) \le V_m\, g_m\big(\langle D_{iso,m}\rangle\big),\; m = 1\ldots M \tag{4.10b}$$

$$\text{and } \phi_j \ge 0 \quad \forall \phi_j \in \Phi \tag{4.10c}$$

Here eq. (4.10b) represents the M isoeffect constraints on the volumes of interest V_m.

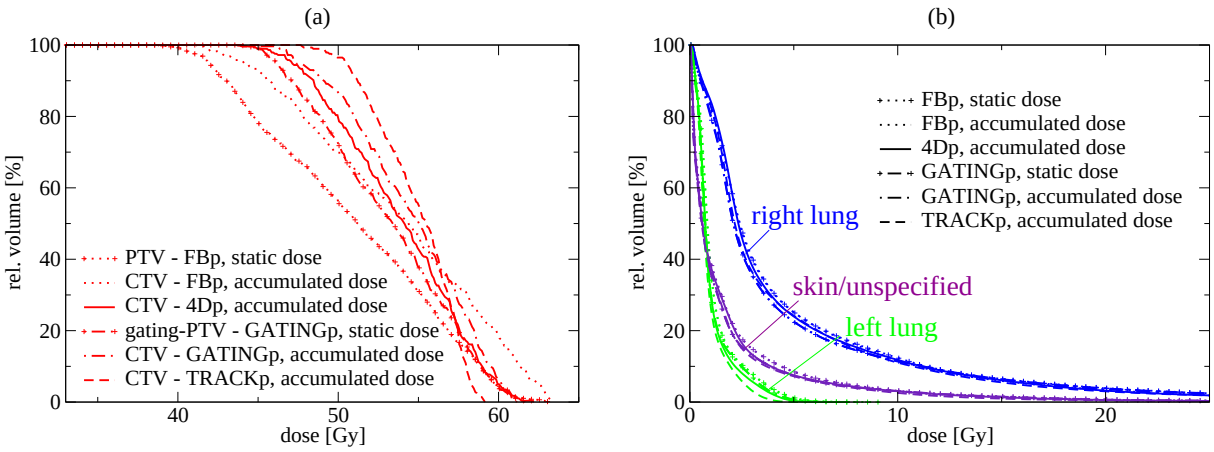

Figure 4.2: DVHs of the different plans: (a) target DVHs; (b) DVHs of the OARs.

4.1.2 Comparison with alternative methods: A planning study

The 4D-IMRT planning algorithm was implemented into the clinical IMRT-planning software HYPERION [3] and its potential for improving free-breathing treatment of moving lung tumors were tested for an example patient with large breathing excursion ($\sim$ 2.9cm) in a planning study. The probabilistic thorax model of this patient was based on eight RCCT phase CTs (shown in fig. 3.6) with the exhale CT as reference geometry for deformable registration and dose accumulation. The instance weight factors were determined from the breathing curve recorded during aquisition of the RCCT (fig. 3.7).

The 4D-plan (abbreviated as "4Dp" in the following) was compared to three other static approaches: a static PTV-plan for free-breathing treatment ("FBp"), where the ITV (hull of all eight CTV positions in the phase CTs) was used as PTV; a static PTV-plan for exhale gating ("GATINGp"), where the PTV encompassed the CTVs of the three phase CTs around the exhale state; and a tracking plan ("TRACKp") as idealized 'gold-standard', for which an optimal static PTV-plan was created separately for each phase CT with the respective CTV as PTV, thereby simulating optimal 'life'-tracking during treatment. These three approaches were optimized based on static dose distributions as calculated on static CT geometries (respective average CTs for FBp and GATINGp; the single phase CTs for TRACKp). To allow for direct dosimetric comparison, the probabilistic thorax model was used to determine the expected dose-to-moving-tissue for the three plans. All plans were subject to the same prescription (see tab. 4.1) and other planning parameters to ensure comparability. More details are given in Appendix F.

The DHVs of the different plans are shown in fig. 4.2, and the corresponding isoeffects and EUDs are summarized in tab. 4.1. All plans show very similar OAR doses as prescribed, however, large differences were found for the target doses. TRACKp provided the best coverage (53.0Gy accumulated CTV-EUD) as could be expected for this idealized planning approach. In contrast, the lowest target dose of all plans was found for FBp with 45.1Gy static EUD to the PTV. Fig. 4.3a depicts the respective static dose distribution

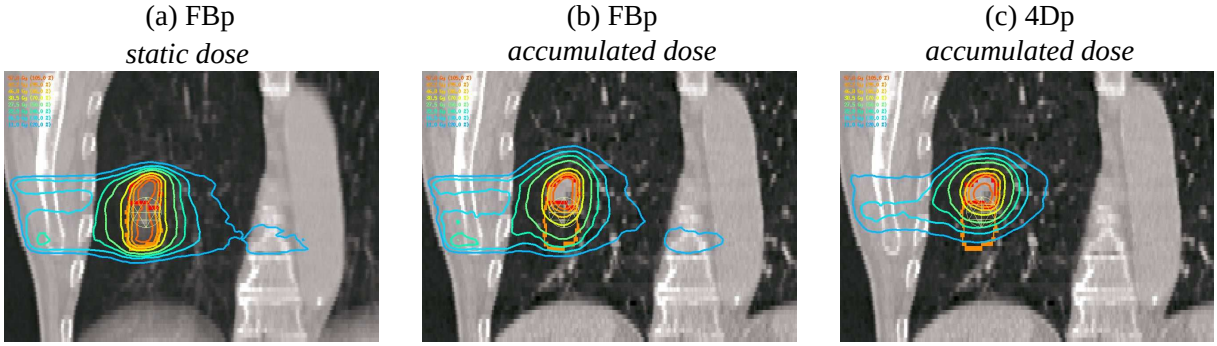

Figure 4.3: Coronal view of the MC-dose distributions of the (a)/(b) static free-breathing PTV-plan and (c) 4D-plan. For the former, (a) shows the static dose distribution as calculated on the average CT, while (b) depicts the accumulated dose $\langle D \rangle$ in tissue-eye-view (exhale CT). Shown are the isodose lines for 57.8/52.2/49.5/44.0/38.5/22.0/16.5/11.0Gy and the contours of the PTV (orange) and the CTV as seen in the exhale CT (red).

	objective	*constraints*				
	target	target	right lung		left lung	skin*
	poisson EUD [Gy]	*rms overdosage >55Gy [Gy]*	*parallel, mean damage [%]*	*rms overdosage >20Gy** [Gy]*	*parallel, mean damage [%]*	*rms overdosage >20Gy [Gy]*
prescription	*55.0*	*2.0*	*5.0*	*0.05*	*1.0*	*0.10*
FBp						
–static	45.1	1.9	5.1	0.05	0.1	0.10
–accum.	47.4	3.2	5.0	0.18	0.1	0.10
4Dp						
–accum.	50.2	2.0	5.0	0.05	0.1	0.10
GATINGp						
–static	49.3	1.9	5.0	0.05	0.1	0.10
–accum.	50.9	2.3	5.0	0.11	0.1	0.10
TRACKp						
–accum.	53.0	1.6	4.7	0.02	0.0	0.04

Table 4.1: Prescribed and resulting isoeffects/EUDs for the different plans. For FBp and GATINGp, the values denoted by 'static' refer to the static dose distributions calculated on the respective average CTs as used for optimization (target: free-breathing-PTV or gating-PTV, respectively). All other values refer to the accumulated dose distribution $\langle D \rangle$ (target: CTV). [*unspecified tissue inside of skin contour; **applied only to voxels with $\geq$2cm distance to the target]

and planning geometry. Due to the large breathing excursion, the PTV encompassed a relatively large region and thus could not be covered well for the given constraints in this approach. The 'real' dose distribution in tissue-eye-view, $\langle D \rangle$, is actually higher by about 2Gy (tab. 4.1), and shows significant target overdosage and dose inhomogeniety. This renders the static PTV-dose as poor predictor of the actual dose-to-moving-tumor. Major differences are also manifested visually (fig. 4.3a,b). As significant dose-builup can happen only inside of the high tissue density of the moving CTV, the high dose region of the FBp is centered around the tumor in TEV. However, considerable spread-out of doses in the range ∼10-30Gy to surrounding lung tissue is the consequence of dose blurring due to breathing motion. As compared to this, the expected dose of 4Dp conforms substantially better to the tumor (fig. 4.3c), as this was explicitly optimized in this approach. Quantitatively, a target coverage of 50.2Gy accumulated CTV-EUD was found for 4Dp (tab. 4.1). Thus, for free-breathing treatment of lung tumors 4D-optimization offers the possibility of improved coverage of the moving target without exceeding overdosage constraints, and provides similar results as for gating (tab. 4.1).

4.2 4D-evaluation of IMPT/IMRT prostate radiotherapy

A study investigating the sensitivity of different static IMPT and IMRT prostate planning approaches with respect to organ movement was performed as part of this PhD project (Appendix H). Organ motion and rectal gas filling with related changes of a patient's density distribution render IMPT particularly prone to a degradation of the actual *accumulated dose-to-moving-tissue* as compared to the static planning dose distribution. The study identified planning strategies that improve plan robustness for both treatment modalities, and thus serves as a clinically relevant example for the application of 4D-evaluation.

IMPT and IMRT treatment plans were evaluated for four patients with an average of 16 CT datasets per patient. The plans were calculated on one CT dataset and recalculated on CTs taken during the treatment course. The resulting dose distributions D^k were then accumulated to a reference geometry (planning CT) according to eq. (4.1) with equal weights to arrive at a measure for the actual dose-to-moving-tissue of the whole treatment course. The displacement fields used for dose accumulation of prostate, bladder and rectal wall were based on a finite element model of these organs [70].
All IMPT/IMRT treatment plans followed the static, PTV-based planning concept currently applied in the clinical IMRT routine in Tübingen university hospital to account for geometric uncertainties. The CTV was defined as hull of the prostate contours from the first three CTs to account for internal organ motion, and a 7mm CTV-to-PTV margin was applied to account for remaining uncertainties. The same hull concept was used for bladder and rectum to arrive at planning-organ-at-risk volumes (PRVs) for these organs.

A number of different planning strategies was investigated. Apart from "standard" IMPT/IMRT plans that used the original planning CT for dose calculation, another set of treatment plans was created with a water density overwrite for rectal gas in the original

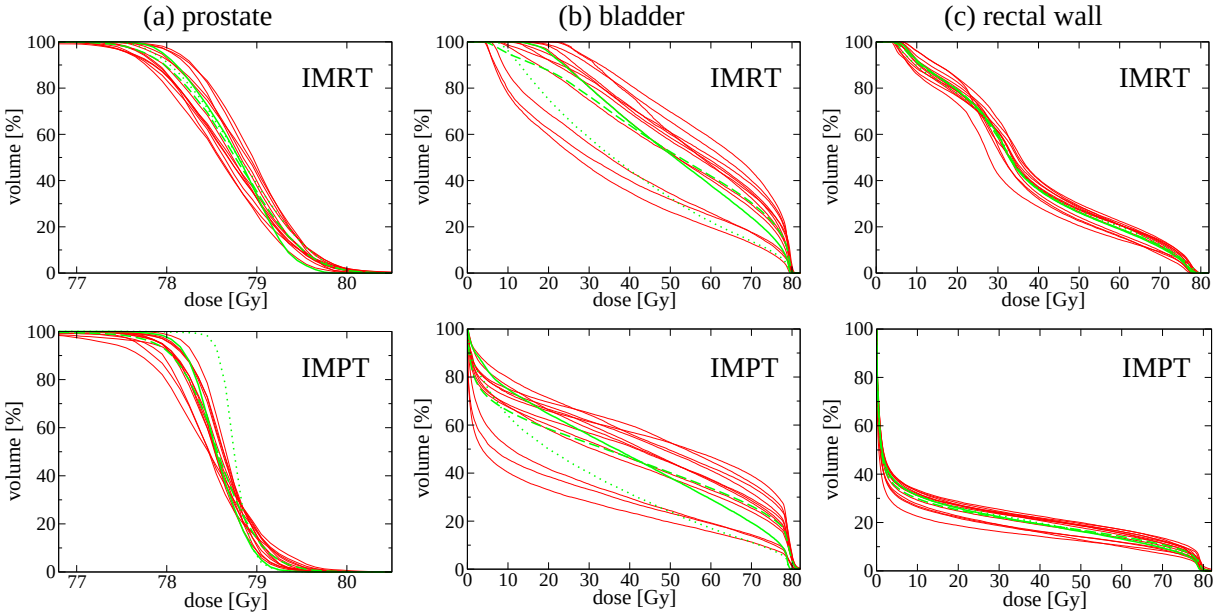

Figure 4.4: DVHs of an IMRT- and IMPT-plan for (a) prostate, (b) bladder and (c) rectal wall for an example patient (patient A of the study in Appendix H). For both plans, the most robust approaches were used, i.e. water density overwrite for rectal gas in the original planning CT and in case of IMPT also the field dependent initial spot weights. In red, the DVHs after recalculation on the single treatment CTs. In dotted green, the accumulated DVH on the first three geometries, representing the expected DVH at time of planning. In solid green, the accumulated DVH on the treatment CTs, representing the actual dose-to-moving-tissue. In dashed green, the mean DVH on the treatment CTs resulting from averaging the single DVHs. See also tab. 4.2 for the corresponding EUD values.

planning CT to estimate the influence of rectal gas filling. Additionally, two different methods of initial spot weight definition were implemented for IMPT (details in Appendix H), as the final 'optimum' spot weight pattern – and its sensitivity to geometric changes – depends on the initial values owing to the high degeneracy of the IMPT optimization problem.

Based on plan recalculation in the treatment CTs and dose accumulation, standard IMPT was found highly sensitive to internal geometric uncertainties. A clinically unacceptable degradation of accumulated GTV doses and large interfraction spread of EUD values was observed especially for patients with rectal gas filling in the planning CT. A similar level of plan robustness against organ movement as for IMRT, however, could be achieved for IMPT if the methods of water density overwrite for rectal gas in the planning CT and field dependent initial spot weights were applied. An improvement of the accumulated dose-to-moving tissue for prostate and rectal wall was observed also for IMRT with the method of water density overwrite for rectal gas. DVHs of an example patient with relatively large organ motion are shown in fig. 4.4 for the two most robust IMPT/IMRT

	patient A			*all patients*		
	prostate EUD [Gy]	bladder EUD [Gy]	rectal wall EUD [Gy]	prostate EUD [Gy]	bladder EUD [Gy]	rectal wall EUD [Gy]
IMRT planned	78.7	61.3	64.4	77.6 ± 1.2	59.9 ± 1.4	63.9 ± 1.9
IMRT realized						
–accumulated	78.7	65.6	63.5	77.4 ± 1.6	62.2 ± 2.5	64.4 ± 0.8
–mean	78.7 ± 0.1	67.1 ± 3.2	63.6 ± 1.2	–	–	–
IMPT planned	78.7	59.8	64.4	77.7 ± 1.1	58.5 ± 1.1	63.3 ± 2.0
IMPT realized						
–accumulated	78.5	63.8	63.7	77.4 ± 1.4	60.8 ± 2.1	63.7 ± 0.8
–mean	78.5 ± 0.1	66.2 ± 3.3	64.6 ± 1.1	–	–	–

Table 4.2: Planned and realized EUDs for the example patient in fig. 4.4 and averages for all four patients considered in the study in Appendix H. Here 'planned' denotes the EUD accumulated on the three organ geometries used for PTV/PRV definition, representing the expected EUD based on a pretreatment estimate of motion. Further, 'realized – accumulated' denotes the accumulated EUD on the treatment CTs, representing the actual dose-to-moving-tissue. Finally, 'realized – mean' was calculated by averaging the EUD values on the single treatment geometries without deformable registration for dose accumulation. [parameters of EUD calculations: prostate: poisson cell-kill EUD with $\alpha = 0.4$, see eq. (4.7); bladder/rectal wall: power-law EUD, eq. (2.1), with $a = 8.0/12.0$]

plans. The EUD-values are given in tab. 4.2. Both plans are similarly robust against interfractional changes with respect to prostate and rectal wall doses, while bladder doses show a relatively large spread (red DVH curves in fig. 4.4; standard deviations given under 'realized – mean' in tab. 4.2). Accordingly, the accumulated dose-to-moving tissue (solid green) for bladder deviates significantly from its planned value (dotted green), reflecting systematic geometric differences of bladder configuration at time of planning as compared to the treatment geometries. For prostate and rectum, differences between realized and planned doses were much smaller, reflecting again the robustness of both plans in this respect. Similar findings apply to the other patients, as the population-averages in tab. 4.2 show.
It should be mentioned that comparable results were found if the dose distributions D^k on the treatment CTs were not explicitly recalculated, but approximately substituted by the planning dose distribution $D^k \sim D^{plan}$ in the process of dose-accumulation according to eq. (4.1) (see Appendix H).

The calculation of the actual accumulated dose-to-moving tissue relies crucially on deformable organ registration. If displacement field information is not available, straightforward surrogates for accumulated doses are mean DVHs and EUDs resulting from averaging the DVH and EUDs of the single treatment geometries. Here the present study showed systematic differences between such mean metrics (dashed green DVH curves in fig. 4.4; values given under 'realized – mean' in tab. 4.2) and the accumulated values. The

accumulated DVHs determined based on deformable organ registration had less extreme doses on both sides for both IMRT and IMPT, as under- and overdosage averages out on the level of single tissue elements. This cannot be modelled correctly by averaging treatment DVHs or EUDs, as spatial information is lost in these dosimetric measures. The differences were largest for bladder and rectal wall, where the accumulated EUDs were lower by up to 2.9Gy and 1.7Gy, respectively, for the four patients under consideration. Thus, the mean DVHs and EUDs can be used only as conservative estimates.

4.3 Discussion and Outlook

The concept of optimization and evaluation in 'tissue-eye-view' (TEV) considers the dose-to-moving-tissue by accumulating doses of multiple geometric instances of the patient geometry to a reference geometry based on patient-individual geometric 4D-models (chapter 3). It is the strength of such "4D-radiotherapy" approaches, that the two major paradoxes of conventional PTV-based concepts are overcome, which arise as a consequence of using a static geometry as patient model: *(1) The necessity to expand physiological structures to PTV and OAR/PRV volumes to account for geometrical uncertainties.* Doses calculated based on these non-physiological surrogate volumes are not well representative for the actual, biological relevant dose-to-moving tissue especially for large geometric uncertainties and density changes as in the lung planning example (sect. 4.1.2). Moreover, resulting PTV-OAR/PRV overlaps lead to ill-defined optimization problems with mutually conflicting objectives, and thus a loss of control over the tradeoff between target coverage and OAR sparing. *(2) The question which density grid to use for dose-calculation.* This is a problem of conventional static planning especially in the presence of large geometrical changes, where the dose distribution itself dynamically changes as for lung (sect. 4.1.2) or in the presence of rectal gas cavities (sect. 4.2). These problems do not exist in 4D-radiotherapy concepts, which realize prescription and dose-scoring on the non-overlapping physiological structures of CTV and OARs, albeit in multiple instances of the patient geometry, which need to be combined by deformation models.

The presented 4D-planning algorithm (sect. 4.1) combines advantages of similar 4D-radiotherapy concepts presented by other authors in the literature. The first implementation of pdf-based TEV-optimization for real patient data was presented by Birkner *et al.* [11] (see also [12]) in terms of an adaptive framework for prostate radiotherapy, where a few CTs taken during the first days of treatment were used to determine and optimize the expectation values of doses to prostate, bladder and rectum. At variance with the method presented in this PhD work, the dose distribution itself was calculated only in the planning geometry, which is a good approximation for the pelvic region as shown by Birkner [12]. This is supported by the results of the prostate IMRT/IMRT study presented in sect. 4.2, which found explicit dose re-calculation in the instance geometries of minor importance for quantification of the accumulated organ dose and plan robustness of prostate radiotherapy. However, for head-and-neck, lung or breast-cancer cases with large density inhomogeneities and tissue-air interfaces this is not a valid approximation.

McShan *et al.* [41], Trofimov *et al.* [62] and Lin *et al.* [30] presented implementations of pdf-based TEV-optimization, where dose was explicitly calculated for each geometry instance. In contrast to these implementations, the 4D-planning method presented in this PhD work incorporates a 4D-Monte Carlo (MC) algorithm for dose calculation of the segmented 4D-plan. This not only allows for realistic modelling of dose deposition especially in the presence of large tissue inhomogeneities, but also has the great advantage over other dose calculation methods that calculation time is essentially independent of the number of geometry instances (sect. 4.1.1). This renders 4D-MC the method of choice for dose calculation in future 4D-radiotherapy applications.

The 4D-planning approach introduced in this work explicitly incorporates *random* errors into optimization, while systematic errors deserve special attention. As for conventional planning, a major challenge are setup errors which should ideally be avoided through frequent image-guidance, optionally combined with adaptive concepts. Apart from such *geometrical* systematic errors, 4D-planning is additionally prone to model biases introduced by misestimation of the patient's motion-pdf and displacement fields as discussed in detail in Appendix F. Thus methods for quantification of the *sensitivity* of individual treatment plans to uncertainties will be an important topic of future research. Generally, it is desirable to explicitly incorporate measures of *robustness* into the planning process [65, 55], which will constitute the foundation for a widespread introduction of 4D-radiotherapy into clinical practice.

Chapter 5

Conclusions

The introduction of computed tomography into clinical practice made tumor definition and radiotherapy planning based on a 3-dimensional model of the patient possible. In the following years, research in the field of radiotherapy planning concentrated on methods to improve 3D-dose calculation and -conformation to predefined target structures. This was brought to perfection by today's state-of-the-art techniques like IMRT which provide for precise dose conformation to targets of almost arbitrary shape, and even open up the possibility of biologically guided dose painting. Yet, treatment planning is conventionally based on a single CT scan which is of limited precision if inter- or intrafractional geometric uncertainties cause deviations from this snapshot of the patient. Thus, major improvements of radiotherapy planning can only be expected by the development and incorporation of dynamic patient models which facilitate calculation of the dose-to-moving-tissue as central quantity of "4-dimensional radiotherapy". A safe transition to such advanced methods which inevitably produce new types of dose distributions invariably needs to be augmented by a profound understanding of biological dose-response.

Based on these premises, the developments presented in this work address and exemplarily realize the three cornerstones of *evidence-based 4-dimensional radiotherapy*: biological modelling, probabilistic deformation models as well as evaluation and optimization of dose-to-moving-tissue (Tissue-Eye-View).

The influence of the treatment technique on the results of biological modelling has been investigated for chronic rectal toxicity in the treatment of prostate cancer. The method of principal component analysis (PCA) quantitatively reveals treatment specific dose-volume correlations present in population data, and it could be shown that these especially influence direct dose-volume based models. This has to be considered when applying biological models derived from past clinical experience to new planning and treatment methods. As shown consistently by this work and recent related studies, rectal dose-response can be modelled well with EUD-based NTCP-models, and quantitative parameters are given.

Further, 4D-modelling approaches for two major types of geometric uncertainties encountered in radiotherapy – random and quasi-periodic uncertainties – were developed as basis of 4D-optimization and -evaluation. For random uncertainties, efficient statistical

surrogate models of patient-individual organ motion and deformation can be based on a PCA of organ geometry samples. As shown for internal motion of prostate, bladder and rectum, PCA allows creation of compact dynamic models with full information about motion of individual tissue elements. For quasi-periodic respiratory uncertainties, probabilistic thorax models can be derived from 4D-image information. With respect to clinical applicability, a fast and accurate deformable registration algorithm has been developed as central tool for such applications.

Ultimately, an implementation of 4D-IMRT planning for free-breathing treatment of lung is described. The algorithm is based on a probabilistic thorax model and realizes direct, biological IMRT-optimization of the expected dose-to-moving-tissue by explicit optimization in multiple instance geometries. The use of advanced pencil beam and 4D-Monte Carlo dose calculation methods ensures accurate estimation of the dose-to-moving-tissue in all stages of the optimization, and allows calculation of fully-segmented 4D-plans in clinically realistic time frames in the order of an hour on current PC hardware owing to parallel computation. It is shown that this planning concept performs similar to gated lung treatment with respect to accumulated tumor and OAR-doses and thereby provides an efficient alternative to gating in day-to-day delivery without prolonged treatment time.

The more such comprehensive 4D-methods enhance treatment specificity, the more the robustness of plans against model uncertainties will become an important issue of future research. In this context, methods for quantification of the *sensitivity* of individual treatment plans to uncertainties will be of great relevance for clinical implementation and acceptance of 4D-radiotherapy and other high precision concepts, and explicit incorporation of robustness measures into optimization will constitute the basis of further improved and safe patient treatment.

The approaches presented in this work envision the future of a 4D-radiotherapy planning paradigm that appreciates the dynamic character of patient treatment and the intricate biological dose-response relationships in the presence of complex inhomogeneous dose distributions. The refinement of probabilistic dynamic patient models as well as research concentrating on the topic of robust planning will pave the way to a widespread introduction of 4D-radiotherapy into clinical practice.

Bibliography

[1] Prescribing, recording and reporting photon beam therapy. ICRU Report 50. International Commission on Radiation Units and Measurements, Bethesda, MD, 1993.

[2] Prescribing, recording and reporting photon beam therapy (supplement to ICRU report 50). ICRU Report 62. International Commission on Radiation Units and Measurements, Bethesda, MD, 1999.

[3] M. Alber, M. Birkner, W. Laub, and F. Nüsslin. Hyperion: an integrated IMRT planning tool. In W. Schlegel and T. Bortfeld, editors, *Proc. XIIIth Int. Conf. on the Use of Computers in Radiation Therapy*, pages 46–48. Heidelberg: Springer, 2000.

[4] M. Alber and F. Nüsslin. An objective function for radiation treatment optimization based on local biological measures. *Phys. Med. Biol.*, 44(2):470–493, 1999.

[5] M. Alber and R. Reemtsen. Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method. *Optimization Methods and Software*, 22(3):391–411, 2007.

[6] M. Alber. *A Concept for the Optimization of Radiotherapy*. PhD thesis, Fakultät für Physik der Eberhard-Karls-Universität zu Tübingen, 2000.

[7] M. Alber. Normal tissue dose-effect models in biological dose optimisation. *Z. Med. Phys.*, 18(2):102–110, 2008.

[8] M. Alber and F. Nüsslin. A representation of an NTCP function for local complication mechanisms. *Phys. Med. Biol.*, 46(2):439–447, 2001.

[9] C. Baum. *Konzepte zur bildgestützten, adaptiven Bestrahlungsplanung und zur Simulation von Behandlungsverläufen bei fluenzmodulierter Strahlentherapie*. PhD thesis, Fakultät für Physik der Eberhard-Karls-Universität zu Tübingen, 2007.

[10] C. 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(1):27–35, 2006.

[11] M. 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(10):2822–2831, 2003.

[12] M. Birkner. *Bildgestützte, adaptive Bestrahlungsplanung intensitätsmodulierter Strahlentherapie*. PhD thesis, Fakultät für Physik der Eberhard-Karls-Universität zu Tübingen, 2002.

[13] M. Birkner, D. Thorwarth, A. Poser, F. Ammazzalorso, and M. Alber. Analysis of the rigid and deformable component of setup inaccuracies on portal images in head and neck radiotherapy. *Phys. Med. Biol.*, 52(18):5721–5733, 2007.

[14] A. Brahme. Optimization of stationary and moving beam radiation therapy techniques. *Radiother. Oncol.*, 12(2):129–140, 1988.

[15] A. Brahme, J.-E. Roos, and I. Lax. Solution of an integral equation encountered in rotation therapy. *Phys. Med. Biol.*, 27(10):1221–1229, 1982.

[16] Y. Censor, M. D. Altschuler, and W. D. Powlis. A computational solution of the inverse problem in radiation-therapy treatment planning. *Appl. Math. Comp.*, 25(1):57–87, 1988.

[17] Y. Chi, J. Liang, and D. Yan. A material sensitivity study on the accuracy of deformable organ registration using linear biomechanical models. *Med. Phys.*, 33(2):421–433, 2006.

[18] G. Fellin, C. Fiorino, T. Rancati, V. Vavassori, S. Barra, E. Cagna, P. Franzone, P. Gabriele, F. Mauro, and R. Valdagni. Late rectal bleeding after conformal radiotherapy for prostate cancer: NTCP modeling. *Int. J. Radiat. Oncol. Biol. Phys.*, 72(1 Suppl):S332–S332, 2008.

[19] C. Fiorino, F. Alongi, L. Perna, S. Broggi, G. M. Cattaneo, C. Cozzarini, N. D. Muzio, F. Fazio, and R. Calandrino. Dose-volume relationship for acute bowel toxicity for patients treated with pelvic nodal irradiation for prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.* in press.

[20] M. Fippel. Fast monte carlo dose calculation for photon beams based on the VMC electron algorithm. *Med. Phys.*, 26(8):1466–1475, 1999.

[21] E. C. Ford, G. S. Mageras, E. Yorke, and C. C. Ling. Respiration-correlated spiral CT: A method of measuring respiratory-induced anatomic motion for radiation treatment planning. *Med. Phys.*, 30(1):88–97, 2003.

[22] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46(3):R1–R45, 2001.

[23] A. Jackson, R. K. T. Haken, J. M. Robertson, M. L. Kessler, G. J. Kutcher, and T. S. Lawrence. Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int. J. Radiat. Oncol. Biol. Phys.*, 31(4):883–891, 1995.

[24] A. Jackson, M. W. Skwarchuk, M. J. Zelefsky, D. M. Cowen, E. S. Venkatraman, S. Levegrün, C. M. Burman, G. J. Kutcher, Z. Fuks, S. A. Liebel, and C. C. Ling. Late rectal bleeding after conformal radiotherapy of prostate cancer (II): volume effects and dose-volume histograms. *Int. J. Radiat. Oncol. Biol. Phys.*, 49(3):685–698, 2001.

[25] U. Jeleń, M. Söhn, and M. Alber. A finite size pencil beam for IMRT dose optimization. *Phys. Med. Biol.*, 50(8):1747–1766, 2005.

[26] U. Jeleń and M. Alber. A finite size pencil beam algorithm for IMRT dose optimization: density corrections. *Phys. Med. Biol.*, 52(3):617–633, 2007.

[27] I. T. Joliffe. *Principal Component Analysis*. Springer, New York, 2002.

[28] Y. H. Lau, M. Braun, and B. F. Hutton. Non-rigid image registration using a median-filtered coarse-to-fine displacement field and a symmetric correlation ratio. *Phys. Med. Biol.*, 46(4):1297–1319, 2001.

[29] J. Liang and D. Yan. Reducing uncertainties in volumetric image based deformable organ registration. *Med. Phys.*, 30(8):2116–2122, 2003.

[30] A. Lin, J. M. Moran, R. B. Marsh, J. M. Balter, B. A. Fraass, D. L. McShan, M. L. Kessler, and L. J. Pierce. Evaluation of multiple breathing states using a multiple instance geometry approximation (MIGA) in inverse-planned optimization for locoregional breast treatment. *Int. J. Radiat. Oncol. Biol. Phys.*, 72(2):610–616, 2008.

[31] A. Lomax. Intensity modulation methods for proton radiotherapy. *Phys. Med. Biol.*, 44(1):185–205, 1999.

[32] C. Lorenz and N. Krahnstöver. Generation of point-based 3D statistical shape models for anatomical objects. *Comput. Vis. Image Underst.*, 77:175–191, 2000.

[33] W. Lu, G. H. Olivera, Q. Chen, M.-L. Chen, and K. J. Ruchala. Automatic re-contouring in 4D radiotherapy. *Phys. Med. Biol.*, 51(5):1077–1099, 2006.

[34] W. Lu, G. H. Olivera, Q. Chen, K. J. Ruchala, J. Haimerl, S. L. Meeks, K. M. Langen, and P. A. Kupelian. Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy. *Phys. Med. Biol.*, 51(17):4357–4374, 2006.

[35] J. B. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–37, 1998.

[36] D. Maleike, J. Unkelbach, and U. Oelfke. Simulation and visualization of dose uncertainties due to interfractional organ motion. *Phys. Med. Biol.*, 51(9):2237–2252, 2006.

[37] U. Malsch, C. Thieke, P. E. Huber, and R. Bendl. An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy. *Phys. Med. Biol.*, 51(19):4789–4806, 2006.

[38] B. F. J. Manly. *Multivariate Statistical Methods: A Primer*. Chapman and Hall/CRC, 3rd edition, 2004.

[39] A. A. Martinez, D. Yan, D. Lockman, D. Brabbins, K. Kota, M. Sharpe, D. A. Jaffray, F. Vicini, and J. Wong. Improvement in dose escalation using the process of adaptive radiotherapy combined with three-dimensional conformal or intensity-modulated beams for prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 50(5):1226–1234, 2001.

[40] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[41] 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). *Med. Phys.*, 32(5):1510–1521, 2005.

[42] D. Merck, G. Tracton, R. Saboo, J. Levy, E. Chaney, S. Pizer, and S. Joshi. Training models of anatomic shape variability. *Med. Phys.*, 35(8):3584–3596, 2008.

[43] A. Niemierko. A generalized concept of equivalent uniform dose [abstract]. *Med. Phys.*, 26(6):1100, 1999.

[44] A. Niemierko. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Med. Phys.*, 24(1):103–110, 1997.

[45] L. Papiez and M. Langer. On probabilistically defined margins in radiation therapy. *Phys. Med. Biol.*, 51(16):3921–3939, 2006.

[46] D. Pasquier, T. Lacornerie, M. Vermandel, J. Rousseau, E. Lartigau, and N. Betrouni. Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 68(2):592–600, 2007.

[47] V. Pekar, T. R. McNutt, and M. R. Kaus. Automated model-based organ delineation for radiotherapy planning in prostatic region. *Int. J. Radiat. Oncol. Biol. Phys.*, 60(3):973–980, 2004.

[48] G. J. Price and C. J. Moore. A method to calculate coverage probabilities from uncertainties in radiotherapy via a statistical shape model. *Phys. Med. Biol.*, 52(7):1947–1965, 2007.

[49] J. C. Roeske, D. Bonta, L. K. Mell, A. E. Lujan, and A. J. Mundt. A dosimetric analysis of acute gastrointestinal toxicity in women receiving intensity-modulated whole-pelvic radiation therapy. *Radiother. Oncol.*, 69(2):201–207, 2003.

[50] P. Rösch, T. Netsch, M. Quist, G. P. Penney, D. L. G. Hill, and J. Weese. Robust 3D deformation field estimation by template propagation. In S. L. Delp, A. M. D. Gioia, and B. Jaramaz, editors, *MICCAI 2000*, volume 1935 of *Lecture Notes in Computer Science*, pages 521–530, 2000.

[51] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Trans. Med. Imaging*, 18(8):712–721, 1999.

[52] D. Sarrut. Deformable registration for image-guided radiation therapy. *Z. Med. Phys.*, 16(4):285–297, 2006.

[53] B. Schaly, J. A. Kempe, G. S. Baumann, J. J. Battista, and J. V. Dyk. Tracking the dose distribution in radiation therapy by accounting for variable anatomy. *Phys. Med. Biol.*, 49(5):791–805, 2004.

[54] M. Sikora, O. Dohm, and M. Alber. A virtual photon source model of an elekta linear accelerator with integrated mini MLC for monte carlo based IMRT dose calculation. *Phys. Med. Biol.*, 52(15):4449–4463, 2007.

[55] B. 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(2 Suppl):S121–S122, 2008.

[56] M. Söhn, M. Birkner, M. Alber, and F. Nüsslin. Modelling patient geometric variation based on dominant eigenmodes of organ deformation. *Radiother. Oncol.*, 73(1 Suppl):S49–S50, 2004.

[57] M. Söhn, M. Birkner, J. W. Y. Chi, D. Yan, B. Berger, and M. Alber. Model-independent, multi-modality deformable image registration by local matching of anatomical features and minimization of elastic energy. *Med. Phys.*, 35(3):866–878, 2008.

[58] M. Söhn, M. Birkner, D. Yan, and M. Alber. Modeling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and evaluation. *Phys. Med. Biol.*, 50(24):5893–5908, 2005.

[59] M. Söhn, M. Alber, and D. Yan. Principal component analysis-based pattern analysis of dose-volume histograms and influence on rectal toxicity. *Int. J. Radiat. Oncol. Biol. Phys.*, 69(1):230–239, 2007.

[60] M. Söhn, D. Yan, J. Liang, E. Meldolesi, C. Vargas, and M. Alber. Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models. *Int. J. Radiat. Oncol. Biol. Phys.*, 67(4):1066–1073, 2007.

[61] J. 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(4):905–919, 1999.

[62] A. 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(12):2779–2798, 2005.

[63] S. L. Tucker, L. Dong, W. R. Bosch, J. Michalski, K. Winter, A. K. Lee, M. R. Cheung, D. A. Kuban, J. D. Cox, and R. Mohan. Fit of a generalized lyman normal-tissue complication probability (NTCP) model to grade ≥ 2 late rectal toxicity data from patients treated on protocol RTOG 94-06. *Int. J. Radiat. Oncol. Biol. Phys.*, 69(3):S8–S9, 2008.

[64] S. L. Tucker, R. Cheung, L. Dong, H. H. Liu, H. D. Thames, E. H. Huang, D. Kuban, and R. Mohan. Dose-volume response analysis of late rectal bleeding after radiotherapy for prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 59(2):353–365, 2004.

[65] J. Unkelbach and U. Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Phys. Med. Biol.*, 49(17):4005–4029, 2004.

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

[67] S. S. Vedam, P. J. Keall, V. R. Kini, H. Mostafavi, H. P. Shukla, and R. Mohan. Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Phys. Med. Biol.*, 48(1):45–62, 2003.

[68] H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung. Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. *Phys. Med. Biol.*, 50(12):2887–2905, 2005.

[69] C. Wu, R. Jeraj, G. H. Olivera, and T. R. Mackie. Re-optimization in adaptive radiotherapy. *Med. Phys.*, 47(17):3181–3195, 2002.

[70] D. Yan, D. A. Jaffray, and J. W. Wong. A model to accumulate fractionated dose in a deforming organ. *Int. J. Radiat. Oncol. Biol. Phys.*, 44(3):665–675, 1999.

[71] D. Yan, D. Lockman, D. Brabbins, L. Tyburski, and A. Martinez. An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 48(1):289–302, 2000.

[72] D. Yan, F. Vicini, J. Wong, and A. Martinez. Adaptive radiotherapy. *Phys. Med. Biol.*, 42(1):123–132, 1997.

[73] D. Yan, J. Wong, F. Vicini, J. Michalski, C. Pan, A. Frazier, E. Horwitz, and A. Martinez. Adaptive modifications of treatment planning to minimize the deleterious effects of treatment setup errors. *Int. J. Radiat. Oncol. Biol. Phys.*, 38(1):197–206, 1997.

[74] A. Zapatero, F. García-Vicente, I. Modolell, P. Alcántara, A. Floriano, A. Cruz-Conde, J. J. Torres, and A. Pérez-Torrubia. Impact of mean rectal dose on late rectal bleeding after conformal radiotherapy for prostate cancer: Dose-volume effect. *Int. J. Radiat. Oncol. Biol. Phys.*, 59(5):1343–1351, 2004.

[75] Q. Zhang, A. Pevsner, A. Hertanto, Y.-C. Hu, K. E. Rosenzweig, C. C. Ling, and G. S. Mageras. A patient-specific respiratory model of anatomical motion for radiation treatment planning. *Med. Phys.*, 34(12):4772–4781, 2007.

[76] T. Zhang, Y. Chi, E. Meldolesi, and D. Yan. Automatic delineation of on-line head-and-neck computed tomography images: Toward on-line adaptive radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 68(2):522–530, 2007.

Acknowledgements

The genesis of this work was accompanied by the support of many friends and colleagues, whom I'd like to express my gratitude here.

First and foremost, I am deeply indebted to my supervisor Dr. Markus Alber. With his great scientific intuition, analytical talents, inspiring visions and – of course – his matchlessly sharp and passionate 'assessment of reality' (in clinics, and in general) he has always been very helpful by pushing my projects into the right direction, while providing the trust and scientific freedom necessary to accomplish real progress. This work would certainly look different without him and the scientific standards he sets in his research group in Tübingen.

I would also like to thank all other former and present members of the Biomedical Physics group in Tübingen, many of whom became more than 'just colleagues'. Thanks for the good times: nice working atmosphere, uncounted coffee-discussions, Neckarmüller-evenings, and much more. I'd like to especially mention and thank Marcin Sikora, with whom I shared so many experiences over the years, that the time in Tübingen will probably stay in my memory almost synonymous with him; similarly Niklas Rehfeld, who I know even longer already from studies in Konstanz and who initially brought my attention to the field of medical physics and Tübingen; Benjamin Sobotta, who to work with and to hang out with is just – *krass!* – fun and made my days so often; Liv Bolstad Hysing as well as Jan Muzik, who made my days probably all the other times, Liv thanks to her bright nature, unique laughter and humor and Jan thanks to his calm and sincere character; Gustav Meedt, who 'made my evenings' in the first months in Tübingen with innumerable deep discussions about the first and the last things, preferably combined with a glass of good whisky; Mattias Birkner, whose ideas, programming skills and help were invaluable for getting started with my "4D"-projects; Christoph Baum, for being the 'living PubMed' of our group; Ula Jeleń, for the good times; Bettina Frey, for doing such a great job in our clinical routine and thereby reducing my clinical workload in the recent month to the minimum that allowed to finally finish up this thesis; and – last, but not least – my long-time colleagues Daniela Thorwarth, Martin Soukup and Oliver Dohm, who in the best sense were and are the 'core' and constant elements of the Biomedical Physics group and shaped its character since I joined it.

Special thanks appertain to Dr. Di Yan and his colleagues from Beaumont Hospital in Detroit, which became my 'second academic home' during my altogether about nine months in his department. I am grateful for what I learned during that time about

medical physics and radiooncology in particular – and the US (science) culture in general. Without the invaluable clinical datasets from – and cooperation with – Beaumont, this PhD-work would have been less '4D' and much less 'evidence-based'. I'd also like to thank Jian Liang, Yuwei Chi and Jian Wang for their help and fruitful cooperations, and the good american and chinese times we had together not only during work. The private background (well, foreground?!) of that time was centered around Sally Kaplan – who rented a room to me in her just-legendary artist-and-more-house – and many friends and interesting people that I got to know thanks to her. Thanks for the great times!

Thanks to my other friends who have nothing to do with medical physics and this thesis – for exactly this, and for their friendship over all the years, which kept me connected to the rest of this much bigger world.

With warmhearted gratefulness I'd like to thank my family. If to anybody, this thesis is dedicated to them. To my parents – without their continuous, life-long promotion, support and just belief-in-me I might have never gone the long way that finally lead to this thesis: *Danke für Alles, was Ihr für mich getan habt!* And to my wife Mary for all her dedication and support especially, but not only, in the last intense months of finishing this thesis – **김매리에게 감사의 마음을 전합니다. 감사합니다** ! – and for her love, which kept and keeps me going and which is what is *actually* important in the end...

Publications

Publications in peer-reviewed journals

- M. Söhn, M. Weinmann and M. Alber: IMRT optimization in a quasi-periodically deforming patient model, accepted for publication in *Int. J. Radiat. Oncol. Biol. Phys.*
- M. Söhn, M. Birkner, Y. Chi, J. Wang, D. Yan, B. Berger and M. Alber: Model-Independent, Multi-Modality Deformable Image Registration by Local Matching of Anatomical Features and Minimization of Elastic Energy, *Med. Phys.* 2008, 35(3), p. 866–878
- M. Söhn, M. Alber and D. Yan: Principal Component Analysis-Based Pattern Analysis of Dose-Volume Histograms and Influence on Rectal Toxicity, *Int. J. Radiat. Oncol. Biol. Phys.* 2007, 69(1), p. 230–239
- M. Söhn, D. Yan, J. Liang, E. Meldolesi, C. Vargas and M. Alber: Incidence of Late Rectal Bleeding in High-Dose Conformal Radiotherapy of Prostate Cancer using Equivalent Uniform Dose-Based and Dose-Volume-Based Normal Tissue Complication Probability Models, *Int. J. Radiat. Oncol. Biol. Phys.* 2007, 67(4), p. 1066–1073
- M. Söhn, M. Birkner, D. Yan and M. Alber: Modeling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and evaluation, *Phys. Med. Biol.* 2005, 50(24), p. 5893–5908

As second author:

- J. Robertson, M. Söhn, D. Yan: Predicting Grade 3 Acute Diarrhea during Radiation Therapy for Rectal Cancer Using a Cutoff-Dose Logistic Regression Normal Tissue Complication Probability Model, accepted for publication in *Int. J. Radiat. Oncol. Biol. Phys.*
- M. Soukup, M. Söhn, D. Yan, J. Liang and M. Alber: A study of the robustness of IMPT and IMRT for prostate against organ movement, accepted for publication in *Int. J. Radiat. Oncol. Biol. Phys.*
- B. Sobotta, M. Söhn, M. Pütz and M. Alber: Tools for the analysis of dose optimization: III Pointwise sensitivity and perturbation analysis, *Phys. Med. Biol.* 2008, 53(22), p. 6337–6343
- U. Jeleń, M. Söhn and M. Alber: A finite size pencil beam for IMRT dose optimization, *Phys. Med. Biol.* 2005, 50(8), p. 1747–1766

Conference contributions

- M. Söhn, D. Yan and M. Alber: 4D stereotactic lung IMRT planning using Monte-Carlo dose calculations on multiple RCCT-based deformable geometries, *Radiother. Oncol.* 2008, 88(2Suppl), p. S96 *[presentation at ESTRO 27, Göteborg]*

- M. Söhn, V. Scheer and M. Alber: Fast Intra- and Intermodal Deformable Registration based on Local Subvolume Matching, *Radiother. Oncol.* 2007, 84(1Suppl), p. S109–S110 *[presentation at the 9th Biennial ESTRO meeting on Physics and Radiation Technology for Clinical Radiotherapy, Barcelona]*

- M. Söhn, M. Alber, E. Meldolesi, C. Vargas and D. Yan: Investigation of EUD- and Dose-Volume Based NTCP Models for Rectal Bleeding, Augmented by Principal Component Analysis of DVH Patterns, *Int. J. Radiat. Oncol. Biol. Phys.* 2006, 66(3Suppl), p. S343–S344 *[poster at the 48th ASTRO annual meeting, Philadelphia]*

- M. Söhn, M. Alber, E. Meldolesi, C. Vargas and D. Yan: Quantitative Modeling of the Incidence of Late Rectal Bleeding in High-Dose Conformal Radiotherapy of Prostate Cancer using EUD- and Dose-Volume Based NTCP Models, *Radiother. Oncol.* 2006, 81(1Suppl), p. S203–S204 *[presentation at ESTRO 25, Leipzig]*

- M. Söhn, D. Yan, J. Liang, L. Kestin, C. Vargas and M. Alber: Influence of Dose Volume Histogram (DVH) Pattern on Rectal Toxicity, *Int. J. Radiat. Oncol. Biol. Phys.* 2005, 63(1Suppl), p. S58–S59 *[presentation at the 47th ASTRO annual meeting, Denver]*

- M. Söhn, M. Birkner, D. Yan and M. Alber: Automatic deformable registration of the lung for 4D respiratory correlated CT (RCCT) datasets, *Radiother. Oncol.* 2005, 76(2Suppl), p. S59 *[presentation at the 8th Biennial ESTRO meeting on Physics and Radiation Technology for Clinical Radiotherapy, Lissabon]*

- M. Söhn, M. Birkner, D. Yan and M. Alber: A Method for Modeling Individual Patient Geometric Variation: Implementation and Evaluation, *Med. Phys.* 2005, 32(6), p. 2143 *[presentation at the 47th AAPM annual meeting, Seattle]*

- M. Söhn, M. Birkner, M. Alber and F. Nüsslin: Modelling patient geometric variation based on dominant Eigenmodes of organ deformation, *Radiother. Oncol.* 2004, 73(1Suppl), p. S49–S50 *[poster at ESTRO 23, Amsterdam]*

As second author:

- M. Alber, G. Nguyen, A. Horst, B. Frey, M. Söhn, U. Ganswindt and C. Belka: The benefit of biological optimization of IMRT for normal tissues: seven years of clinical experience, *Radiother. Oncol.* 2008, 88(2Suppl), p. S97

- U. Ganswindt, G. Nguyen, A. Horst, B. Frey, M. Söhn, M. Alber, M. Bamberg and C. Belka: EUD based IMRT for prostate cancer in 174 Patienten – Toxicities and correlating dose-volume histograms, *Radiother. Oncol.* 2008. 88(2Suppl), p. S324

- B. 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.* 2008, 88(2Suppl), p. S121–S122

- M. Soukup, M. Söhn, J. Liang, D. Yan and M. Alber: Sensitivity to organ movement of prostate IMPT and IMRT, *Radiother. Oncol.* 2008, 88(2Suppl), p. S94

- U. Ganswindt, G. Nguyen, A. Horst, B. Frey, M. Söhn, M. Alber, M. Bamberg and C. Belka: EUD basierte IMRT beim Prostatakarzinom an 174 Patienten – Toxizitäten und korrelierende Dosis-Volumen-Histogramme, *Strahlenther. Onkol.* 2008, 184(1Suppl), p. 20

- J. M. Robertson, M. Söhn and D. Yan: Predicting Grade 3 Acute Diarrhea During Radiation Therapy for Rectal Cancer Using a Cutoff-Dose Logistic Regression NTCP Model, *Int. J. Radiat. Oncol. Biol. Phys.* 2007 69(3Suppl), p. S268–S269

- J. Liang, M. Söhn and D. Yan: Confidence in Offline Image Guided Radiotherapy With Single Adaptive Inverse Planning Modification for Prostate Cancer Management, *Int. J. Radiat. Oncol. Biol. Phys.* 2006, 66(3Suppl), p. S626

- A. Martinez, C. Vargas, E. Meldolesi, M. Söhn, P. Y. Chen, D. Brabbins, L. Kestin, J. Liang and D. Yan: Adaptive Image Guided Radiotherapy (A-IGRT) Allows Delivery of Maximally Tolerable Tumor Dose for Each Prostate Cancer Patient, *Int. J. Radiat. Oncol. Biol. Phys.* 2006, 66(3Suppl), p. S321

- J. M. Robertson, M. Söhn and D. Yan: Principal Component Analysis of Small Bowel Dose Volume Histograms and Acute Diarrhea in the Treatment of Rectal Cancer, *Int. J. Radiat. Oncol. Biol. Phys.* 2006, 66(3Suppl), p. S301–S302

- U. Jeleń, M. Söhn and M. Alber: A finite size Pencil Beam algorithm for IMRT dose optimization based on Monte Carlo calculations, *Radiother. Oncol.* 2005, 76(2Suppl), p. S95

Appendix A

Principal Component Analysis-Based Pattern Analysis of Dose-Volume Histograms and Influence on Rectal Toxicity

published in
International Journal of Radiation Oncology, Biology and Physics 2007; **69**(1): 230-239

Int. J. Radiation Oncology Biol. Phys., Vol. 69, No. 1, pp. 230–239, 2007

0360-3016/07/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2007.04.066

PHYSICS CONTRIBUTION

PRINCIPAL COMPONENT ANALYSIS-BASED PATTERN ANALYSIS OF DOSE–VOLUME HISTOGRAMS AND INFLUENCE ON RECTAL TOXICITY

MATTHIAS SÖHN, DIPL.PHYS.,* MARKUS ALBER, PH.D.,* AND DI YAN, D.SC.†

*Section of Biomedical Physics, University Hospital for Radiation Oncology, Tübingen, Germany; and †Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, MI

Purpose: The variability of dose–volume histogram (DVH) shapes in a patient population can be quantified using principal component analysis (PCA). We applied this to rectal DVHs of prostate cancer patients and investigated the correlation of the PCA parameters with late bleeding.

Methods and Materials: PCA was applied to the rectal wall DVHs of 262 patients, who had been treated with a four-field box, conformal adaptive radiotherapy technique. The correlated changes in the DVH pattern were revealed as "eigenmodes," which were ordered by their importance to represent data set variability. Each DVH is uniquely characterized by its principal components (PCs). The correlation of the first three PCs and chronic rectal bleeding of Grade 2 or greater was investigated with uni- and multivariate logistic regression analyses.

Results: Rectal wall DVHs in four-field conformal RT can primarily be represented by the first two or three PCs, which describe ~94% or 96% of the DVH shape variability, respectively. The first eigenmode models the total irradiated rectal volume; thus, PC1 correlates to the mean dose. Mode 2 describes the interpatient differences of the relative rectal volume in the two- or four-field overlap region. Mode 3 reveals correlations of volumes with intermediate doses (~40–45 Gy) and volumes with doses >70 Gy; thus, PC3 is associated with the maximal dose. According to univariate logistic regression analysis, only PC2 correlated significantly with toxicity. However, multivariate logistic regression analysis with the first two or three PCs revealed an increased probability of bleeding for DVHs with more than one large PC.

Conclusions: PCA can reveal the correlation structure of DVHs for a patient population as imposed by the treatment technique and provide information about its relationship to toxicity. It proves useful for augmenting normal tissue complication probability modeling approaches.

Prostate cancer, Rectal toxicity, Dose–volume histograms, Principal component analysis, Normal tissue complication probability.

INTRODUCTION

Advances in modern radiotherapy (RT), such as three-dimensional conformal RT (3D-CRT) and intensity-modulated RT, have enabled accurate tailoring of dose distributions to target volumes and better sparing of adjacent normal structures, thereby facilitating increased target doses. Safe dose escalation, however, requires reliable information about normal tissue complications and its dependence on dose and volume.

For prostate RT, the essential dose-limiting organs are the bladder and rectum, with chronic rectal bleeding as one of the most relevant side effects. A first step to correlating complications with the applied doses is a reduction of the 3D dose distributions to dose–volume histograms (DVHs) at the expense of losing information about location. Several studies have found significant correlations between the parameters derived from rectal DVHs and the incidence of bleeding (1–8). However, the studies differ in the parameters used as summary measures of the DVH. In one approach, correlations of single DVH features such as the maximal dose or the value of a single DVH point (*i.e.*, the volume V_{D_c} receiving doses greater than a cutoff dose level D_c or the dose D_{V_c} to a cutoff volume V_c) and toxicity are investigated (1, 2, 5–8). In another approach, a comprehensive surrogate value such as an effective volume (9, 10), the mean dose, or another generalized equivalent uniform dose (EUD) (11, 12) is calculated and tested for its correlation with toxicity (3, 6–8). Some of these studies aimed to quantify complication risk

Reprint requests to: Matthias Söhn, Dipl.Phys., Section of Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany. Tel: (++49) 7071-2986061; Fax: (++49) 7071-295920; E-mail: Matthias.Soehn@med.uni-tuebingen.de

Supported in part by Deutsche Krebshilfe e.V. Grant 106280 and National Institutes of Health Grant RO1 CA091020.

Presented in part at the 47th Annual Meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO), Denver, CO, October 16–20, 2005.

Conflict of interest: none.

Received Sept 20, 2006, and in revised form Feb 14, 2007. Accepted for publication April 14, 2007.

in terms of normal tissue complication probability (NTCP) models and determine the corresponding model parameters (3, 4, 6, 8).

An inherent problem of outcome modeling, especially of direct dose–volume-based approaches (*i.e.*, models that consider only single DVH features), is the influence of the treatment technique on the results of modeling. For a given patient population, the treatment technique used induces correlations between DVH bins of different dose levels owing to the interaction of the given beam directions and shapes with the variability of the patient geometries. Because this can deteriorate the prospective use of such models for other patient populations treated using different techniques, these correlations should be considered when interpreting and comparing the results of different studies. Thus, a method that explicitly reveals the correlation structure of a DVH data set is desirable.

Recently, Dawson *et al.* (13) and Bauer *et al.* (14, 15) proposed the use of principal component analysis (PCA) to analyze the partial volume effects of normal tissues to RT and applied it to DVHs of the liver and parotid gland (13) and rectal wall (15), respectively. With this multivariate approach, the correlated variability of DVH shapes in a given patient population can be quantitatively described in terms of ''eigenmodes'', which provide information about the correlation structure inherent in the DVH data set. Moreover, PCA allows characterization of individual DVHs using a few parameters, the ''principal components'' (PCs) (formal definition given below in the subsection ''PCA of DVH data''). Regarding the PCs as a summary measure of individual DVH morphology, correlations with toxicity can be assessed using logistic regression models in a purely phenomenologic manner. However, the value of these models is restricted to the treatment technique for which they were derived.

Bauer *et al.* (14, 15) analyzed two data sets comprising 52 and 119 rectal wall DVHs of patients treated with a six-field 3D-CRT technique with a prescription dose of 70.2 Gy and 75.6 Gy, respectively, and found correlations of some of the dominating PCs with rectal bleeding of Grade 2 or greater.

In the present study, PCA was applied to rectal wall DVHs of 262 prostate cancer patients treated with a four-field box, 3D-CRT adaptive RT (ART) technique, thereby revealing correlations of different dose levels for the population as imposed by this treatment technique. Correlation of the first three PCs and chronic rectal bleeding of Grade 2 or greater was then investigated with logistic regression analysis. The same patient population was used in a parallel study (4), in which six EUD- and dose–volume-based NTCP models were applied to the data. This allowed a comparison of the different approaches with respect to their power to describe the data and make predictions.

METHODS AND MATERIALS

Patient data

The patient data set used comprises 262 prostate cancer patients treated between 1999 and 2002 at the William Beaumont Hospital. This patient data set, with a minimal follow-up time of 1 year (see below), represents a subgroup of a patient population used in a parallel study (5). For additional details, we refer to that study and limit the present description to information relevant for the following investigation. The clinical characteristics of the patient population have been previously described (7, 16).

The patients were part of a phase II dose-escalation study and underwent 3D-CRT with image-guided off-line correction under an ART protocol (17, 18). In brief, a four-field box technique was used for the initial treatment plan of the first week and the adapted plan. The initial planning target volume (PTV) was defined based on the clinical target volume plus a 1-cm uniform margin. For the adapted plan, information from daily portal and computed tomography (CT) imaging was used to form a patient-specific confidence-limited PTV (cl-PTV), thereby considering random and systematic errors as estimated from the first week. Beam apertures for the initial/adapted plan were defined according to the PTV/cl-PTV in the beam's eye view, with a PTV-to-field edge margin of 7 mm everywhere, but 11 mm at the superior and inferior edges of the cl-PTV. The final dose to the cl-PTV was limited by the dose–volume constraints of rectal wall and bladder. For the rectal wall, these were $D_{30\%} = 75.6$ Gy for the minimal dose received by 30% of the target volume and 82 Gy for the minimal dose received by $D_{5\%} = 82$ Gy of the target volume. For each patient, the dose level (minimal cl-PTV dose) was chosen individually so as to meet the rectum and bladder constraints and was one of the following doses as defined by the study protocol: 70.2, 72, 73.8, 75.6, 77.4, and 79.2Gy.

Cumulative DVHs. For DVH calculation, a composite planning dose was used. It included the initial treatment plan for the first week before correcting the systematic error and then the adaptive plan for the rest of the treatment after beam aperture correction according to the cl-PTV. The dose distributions of both plans were calculated using Pinnacle 6.2b (ADAC Laboratories, Milpitas, CA) according to the CT geometry (density information) of the planning CT scan. The overall dose was defined as the sum of the initial and adapted (physical) dose distributions. An in-house–developed software program was used to calculate the DVHs of the rectal wall, which was defined according to the solid rectum contours of the planning CT with 3–4-mm wall thickness. The dose bin size used for calculation of the DVHs was 0.1 Gy, with the volume defined as relative (percentage) volume.

Follow-up information. The toxicity variable regarded in this study was chronic rectal bleeding, for which the grading was determined using the Common Terminology Criteria for Adverse Events (CTCAE), version 3.0 (19). Although the full patient population, as described in our previous study (4), comprised 319 patients, in the present study we considered only those patients with a follow-up time of $\geq$1 year. This resulted in a median clinical follow-up for the remaining 262 patients of 3.2 years (range, 1.0–6.4 years), with an interquartile range of 2.3–4.2 years (25th to 75th percentile).

Principal component analysis of DVH data

Given n observations of p variables, a method from multivariate statistics, PCA, can be used to analyze and describe the correlated variability of the p variables in a data set. First applications of PCA in the field of RT have been proposed only recently and encompass automatic model-based organ segmentation (20) and modeling of organ deformations (21). Regarding the volume values of the dose bins as variables, PCA can also be applied to a set of DVHs to efficiently describe the variance of organ DVH shapes present in a patient population as first proposed by Bauer *et al.* (14, 15) and Dawson *et al.* (13). For our study, we used $p = 850$

variables (dose range, 0–84.9 Gy, with a bin size of 0.1 Gy) for discretization of $n = 262$ rectal wall DVHs.

Because of the very nature of DVHs of a patient population treated with the same treatment technique, the volume values V_1, …., V_p of the p dose bins are highly correlated, implying that the underlying dimensionality of this multivariate statistical problem is actually much smaller than p. By calculating the eigenvectors of the covariance matrix, PCA allows one to find the statistically independent ''eigenmodes'' of variability. Ordered by the statistical importance as given by the corresponding eigenvalues, each of these modes models a different ''dimension'' of DVH shape variability present in the data set. Uncorrelated and erratic solitary features in a DVH give rise to very small or zero eigenvalues. An eigenmode is a p-dimensional vector defining differential volume change for each dose bin, thereby representing correlated DVH shape change. Mathematically, the p eigenmodes define a new coordinate system, in which each DVH can be (exactly) represented by p new coordinates c_1, …., c_p (PCs). In other words, each DVH can be represented as a linear combination of all p modes weighted by the corresponding values of c_1, …., c_p. By definition, the first PC shows the largest statistical variance over the data set of n DVHs, c_2 the second largest, and so forth. Thus, if the overall variability of the data set is clearly dominated by few PCs, this offers the possibility to approximately represent the most important features of individual DVH morphology by the first few parameters (c_1, …., c_M), with $M \ll p$. Additional details can be found in the Appendix.

DVH-PCA and toxicity

The PCA was performed on the DVH data set, irrespective of rectal complications associated with the DVHs. Plotting the resulting first eigenmodes over the dose revealed the major types of correlated DVH shape variability in the data set and the relevant dose ranges for this. The eigenmodes were then used to calculate the PCs of each DVH. The value of the ith PC of a DVH is essentially calculated by summing up the differences of the relative volume values of the DVH and the mean DVH, for each dose bin weighted by the corresponding entry of the ith eigenmode (see Appendix, Eq. A5). Thus, depending on the actual shape of an eigenmode, each PC value has a characteristic dosimetric meaning. Possible correlations of the PCs to widely used dosimetric variables such as the rectal wall relative volume receiving a dose $\geq$D (V_{D}), mean and maximal dose were assessed by linear regression analysis. Of note, there is no reason that such a correlation exists for each of the PCs.

We investigated the correlations of the PC values and chronic rectal bleeding of Grade 2 or greater. Therefore, scatterplots of the dominating PCs, in which each point corresponds to a DVH, were generated, and the points were marked according to the associated complication. Ideally, DVHs with a complication could be distinguished from others by their PC values (*i.e.*, by ''clustering'' of points associated with the complication in such scatterplots and/or by correlation of the PC values and the complication rate).

Logistic regression analysis

The correlations of PCs and toxicity can be analyzed quantitatively by logistic regression analysis. For this, the probability of chronic rectal bleeding, $\mathrm{P_{tox}}$, is defined as a (logistic) function of one or more PCs, which represent certain features of individual DVH shapes (uni- or multivariate logistic regression, see Eqs. 1 and 2, below). If correlations are found by fitting of a logistic regression model to data of a patient population, such a model can be regarded as a phenomenologic NTCP model with $\mathrm{P_{tox}}$ as NTCP function and the values of one or more PCs as a summary measure of the DVH shape and, thereby, the applied rectal dose. Compared to the widely used cutoff-dose and cutoff-volume NTCP models, which consider the position of single DVH points as a summary measure (4, 6), features of the whole DVH are considered in such PCA-based NTCP models. The advantage of regarding only single DVH points is the straightforward interpretability; however, possibly relevant dose–volume information from other parts of a DVH and about its overall pattern are neglected by such models. In contrast, PCA-based models explicitly incorporate this information.

For the univariate case, the logistic regression model is written as follows:

$$P_{tox}(c) = \frac{1}{1 + exp(-\beta_0 - \beta_1 c)} \qquad (1)$$

The variable c denotes a single PC (such as c_1 or c_2), and β_0, β_1 are the parameters to be fitted to the data.

The method of choice for fitting such a model to sparse, dichotomous data (''0'' for bleeding less than Grade 2, ''1'' for bleeding of Grade 2 or greater) is the maximal likelihood estimation (22, 23) (see also, Rancati *et al.* [3]). This method determines the optimal parameters by maximizing the likelihood function L (or the log-likelihood, $\log(L)$ [LL]), which describes the probability of the occurrence of the observed data under the given model. The significance of a model was defined by $p < 0.05$, according to the likelihood ratio test. Uncertainties of the model parameters were assessed using the variance-covariance matrix of the parameters calculated around the maximum of the log-likelihood function as described in Jackson *et al.* (23). The 68% confidence intervals for the parameters can be estimated by the square root of the diagonal elements.

Because more than one eigenmode can signify dose distribution changes that affect the risk of complications, logistic regression analysis with a single PC might not capture the whole picture. The scatterplots (Fig. 2) suggest that a linear combination of the dominant PCs exhibits a stronger correlation with toxicity. This can be analyzed using logistic regression analysis of a higher order:

$$P_{tox}(c_1, \ldots, c_M) = \frac{1}{1 + exp\left(-\beta_0 - \sum_{i=1}^{M} \beta_i c_i\right)} \qquad (2)$$

Depending on the number M of dominant PCs used for logistic regression analysis, the model has $M + 1$ fit parameters.

The goodness of fit estimation followed that off Cox and Snell (22); additional information can be found in Jackson *et al.* (23). Denoting by LL_{obs} the actual (maximized) value of the log-likelihood for the model fitted to the observed patient population, the probability of obtaining a value smaller than LL_{obs} (*i.e.*, a worse fitting model) can be assessed according to the statistical distribution of LL. The latter can be obtained under the assumption of a normal distribution from analytical formulas for the mean $\langle \mathrm{LL} \rangle$ and variance S_{LL} (see Eqs. A4 and A5 in Jackson *et al.* [23]). If this probability is too large, the model ''overfits'' the data; if it is too small, the model does not fit the data well. According to Rancati *et al.* (3), values between 30% and 70% indicate a satisfactory fit.

The presented logistic models differ in the number of parameters. Generally, models with more parameters can be expected to fit the data better, however, at the expense of greater complexity. We used the Akaike information criterion (AIC) (24, 25) (see also Tucker *et al.* [6]) to quantify the tradeoff between a model's quality of fit (associated with the likelihood value) and its complexity

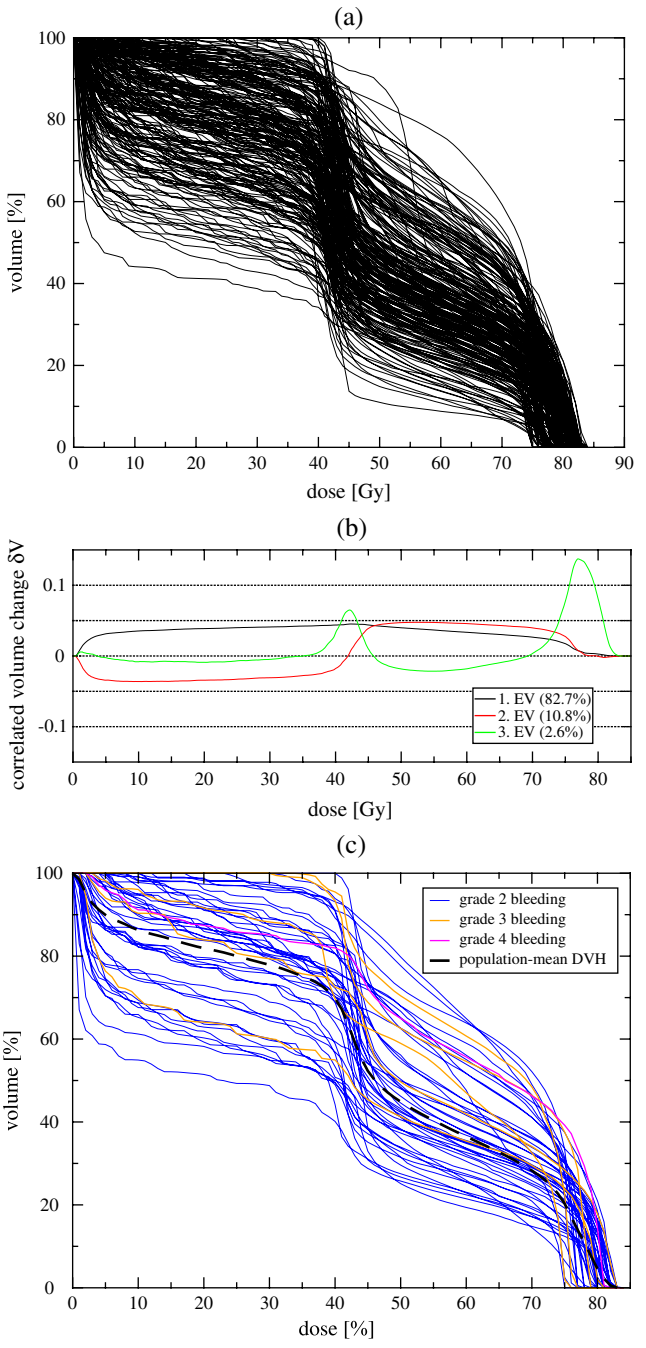

Fig. 1. (a) Rectal wall dose–volume histograms (DVHs) of all 262 patients. (b) First three eigenvectors resulting from principal component analysis of these DVHs. Percentage of data set variability described by the corresponding eigenvectors (see Eq. A6) shown. (c) DVHs of patients with rectal bleeding of Grade 2 or greater (cyan, orange, and magenta indicate Grade 2, 3, and 4, respectively), together with the population-mean DVH (dashed line).

(expressed by the number of model parameters n). It is defined as AIC = -2 LL + 2 n. Models with smaller AIC values are considered to provide a better (in the sense of more efficient) fit to the data than models with larger AIC.

Statistical analysis was performed with Mathematica, version 5.0 (Wolfram Research, Champaign, IL); the PCA itself was implemented in C++.

RESULTS

The observed toxicity rates in the population of 262 patients were as follows: 16.8% (n = 44) had Grade 2 chronic rectal bleeding and 1.9% (n = 5) developed Grade 3 toxicity. One patient (0.4%) had Grade 4 toxicity.

The DVHs of all 262 patients are plotted in Fig. 1a. The first three eigenvectors resulting from PCA of the DVHs are shown as Fig. 1b, together with the corresponding eigenvalues. The first mode represents the correlated variability of the DVH data set around the mean DVH in the range of ∼5–75 Gy, thus DVHs systematically greater than or less than the mean DVH have a positive or negative first PC, respectively (see Eqs. A4 and A5). In brief, the first PC essentially describes the variation of the fraction of the rectal wall volume that is inside the irradiated volume. In contrast, the second eigenvector describes the anticorrelated variability of the low- and high-dose region (∼5–40 vs. 50–75 Gy), that is, the DVHs with increased relative volumes in the high-dose region and simultaneously decreased relative volumes in the low-dose region have a positive second PC (and vice versa). Thus, the second PC is related to the fraction of the rectal wall volume that is inside the four-field overlap region (approximately the prescription dose and less) and therefore not in the two-field overlap region (approximately half of the prescription dose and less). The third mode describes the correlated variability of the volumes receiving intermediate doses (∼40–45 Gy) and doses greater than ∼70 Gy. It essentially results from variations of the prescribed four-field box dose itself (which obviously also affects the dose in the two-field overlap). This variability in dose levels in the population is a consequence of the particular dose prescription scheme of the ART concept.

The spectrum of eigenvalues is clearly dominated by a few values: about 93.5% and 96.1% of the DVH shape variability can already be described by the first two and three PCs, respectively. In contrast, the overall contribution of all greater modes is <4% (data not shown).

Scatterplots showing the distribution of the first PCs (PC2 vs. PC1 and PC3 vs. PC1) for the 262 DVHs are plotted in Fig. 2. The PCs itself (irrespective of toxicity) do not show significant clustering, corresponding to a fairly regular distribution of DVH shapes in the data set and in visual agreement with the data in Fig. 1a. This supports the applicability of PCA, which to a certain extent is based on the assumption that the underlying distribution of input data is approximately gaussian, and large deviations from this property would deteriorate the interpretability of the results of PCA. Generally, DVHs with all PC values around zero have shapes close to the mean DVH (dashed line in Fig. 1c), and a positive PC1 value is associated with an increased percentage of volume in the dose range of ∼5–75 Gy, as discussed previously. Consequently, PC1 correlates highly with the mean dose (Fig. 3a, R = 0.999) and shows a weak correlation with, for example, V_{60}, *i.e.*, the volume receiving a dose of 60 Gy and higher (Fig. 3b, R = 0.89). The PC3 showed a weak correlation to the maximal dose (Fig. 3c, R = 0.91), as could be expected

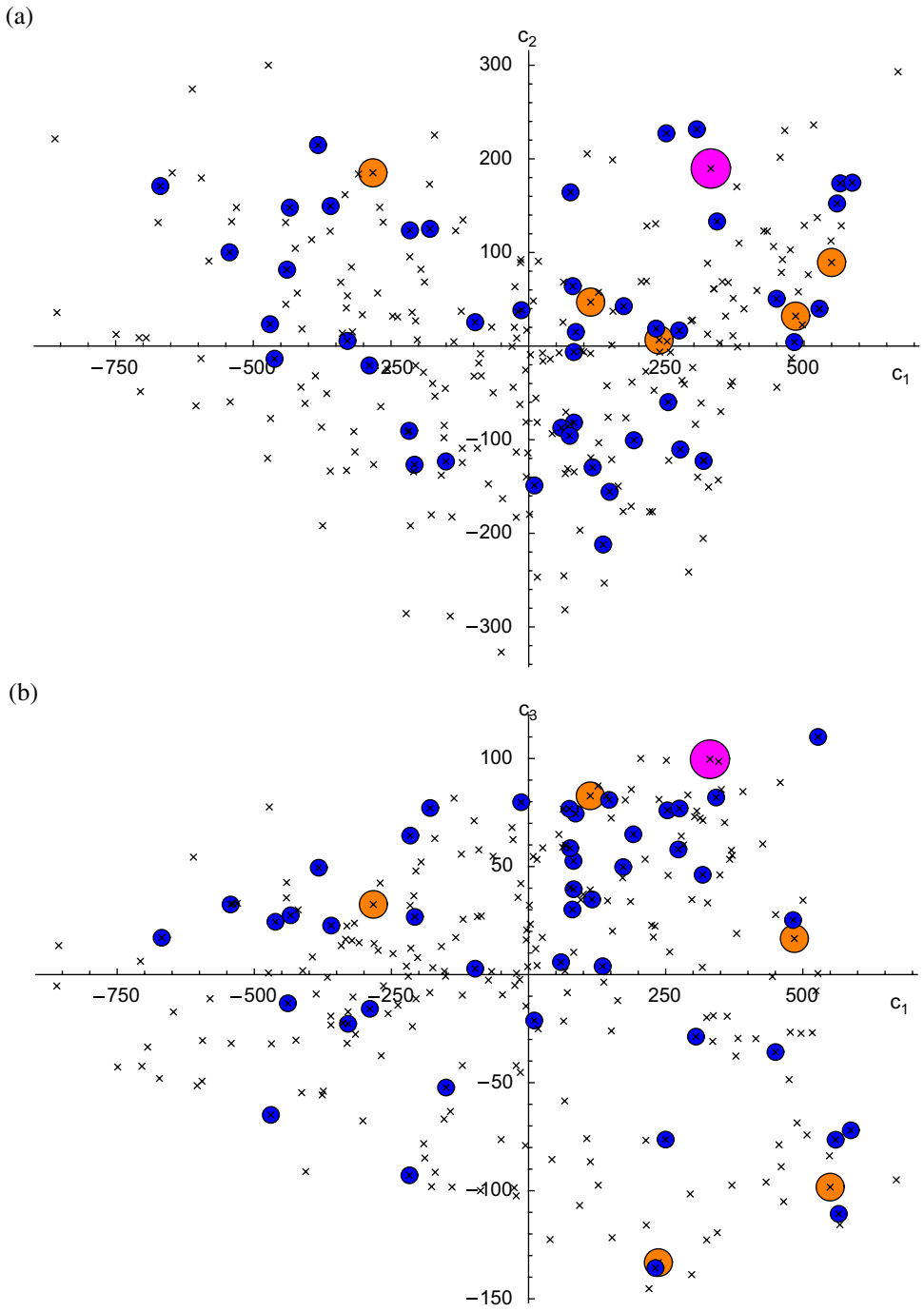

Fig. 2. Scatterplots of the first principal components, (a) c_2 vs. c_1 and (b) c_3 vs. c_1, for 262 rectal wall dose–volume histograms (DVHs). DVHs associated with chronic rectal bleeding of Grade 2 or greater indicated by color coded disks (cyan, orange, and magenta indicate toxicity Grade 2, 3, and 4, respectively).

from the plot of the third eigenvector in Fig. 1b. PC2 did not show a significant correlation with any of the aforementioned dosimetric measures.

Points corresponding to DVHs associated with chronic rectal bleeding of Grade 2 or greater are color coded in Fig. 2. The respective DVHs are shown, together with the population-mean DVH, in Fig. 1c. Severe toxicity of Grade 3 or greater appeared to be associated with positive values of the first two PCs (Fig. 2a), corresponding to increased volumes especially in the four-field overlap region (Fig. 1c). However, the small number of patients with severe toxicity in our population ($n = 6$) did not allow the investigation of this with sufficient statistical quality. While for Grade 2 or greater toxicity no clear clustering was obvious from the scatterplots, the correlation of toxicity with the PCs was analyzed using logistic regression analysis.

Logistic regression analysis. The correlation with toxicity of the first three PCs as single variables was investigated using Eq. 1. According to this analysis, PC2 was the only variable significantly associated with chronic rectal bleeding of

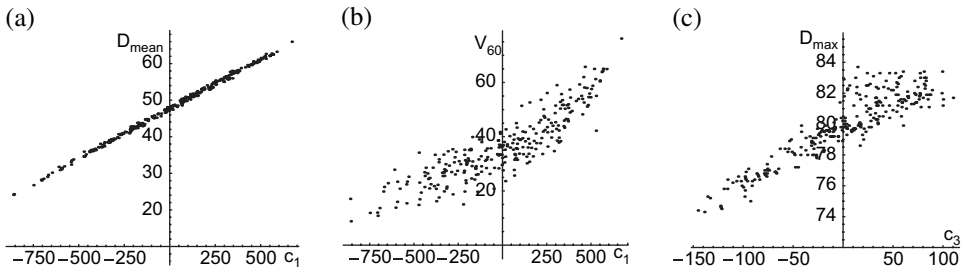

Fig. 3. Correlations between principal components (PCs) and dosimetric measures: (a) first PC and mean dose; (b) first PC and V_{60}; and (c) third PC and maximal dose.

Grade 2 or greater ($p = 0.03$). The parameter estimates and their uncertainties are given in Table 1. The log-likelihood value of this model was $LL_{obs} = -125.5$, and the probability of finding a fit with a smaller log-likelihood was found to be 50.0% (estimated log-likelihood distribution, $\langle LL \rangle \pm S_{LL} = -125.5 \pm 9.1$), indicating that the fit is acceptable.

Although neither PC1 nor PC3 as a single variable showed significant correlation with toxicity, the data presented in Fig. 2 suggest increased toxicity rates if combinations of PCs are simultaneously large. This was analyzed by multivariate logistic regression according to Eq. 2 for the first two or three PCs (Table 1). Both models proved to be significant ($p = 0.047$ and $p = 0.02$, respectively) and provided acceptable fits according to the log-likelihood distribution test.

As example, Fig. 4 shows a scatterplot of the first two PCs, overlayed with the isotoxicity lines of the logistic regression model with PC1 and PC2 as variables. The toxicity gradient has an inclination of $\phi = 78.2°$, *i.e.*, the toxicity rate increases if both c_2 and c_1 increase (see Eq. 2 for positive coefficients β_1, β_2), corresponding to increasing relative volumes in the high-dose region (~45–75 Gy; see Eq. A4 and discussion of eigenmodes above), as could be expected. In Fig. 5, the data from Fig. 4 have been projected to an axis with angle $\phi = 78.2°$, and the resulting toxicity rates were plotted, together with the rates predicted by the multivariate logistic regression model, which allowed additional evaluation of the quality of fit ($\chi^2 = 0.53$; upper limit for 1 degree of freedom, error level $\alpha = 5\%$; $\chi^2 \leq 3.84$).

For the multivariate logistic regression models, the DVH pattern variability (relative to the population-mean DVH), which correlates most significantly to toxicity, can be expressed by ''compound modes.'' These are given by the linear combination of eigenvectors that corresponds to variations of PCs in the direction of the toxicity gradient (*i.e.*, the variations that cause the largest changes in toxicity according to the model). Figure 6 shows the resulting compound modes of the logistic regression models for the first two and three PCs. As discussed above for the model with PC1 and PC2 as variables, the corresponding compound model expresses that increased relative volumes in the dose range of ~45–75 Gy correlate with bleeding. For the model with PC1, PC2, and PC3 as variables, the resulting compound mode was strongly influenced by the third eigenvector because of the comparatively large coefficient β_3 (Table 1), providing a hint of the increased biologic importance especially of the highest doses for the development of bleeding. This was true even though PC3 showed a smaller variability in our patient population than did PC2 (as expressed by the corresponding eigenvalues and visually obvious in Fig. 2); however, because PCA was performed irrespective of toxicity, a smaller variability of a PC in a population does not necessarily imply less importance for describing toxicity. To

Table 1. Estimated parameter values for different logistic regression models

Logistic regression	Parameter estimates (68% CI)	LL	$\langle LL \rangle \pm S_{LL}$	P (%)	AIC
c_1	NS	−126.8	NS	NS	NS
c_2	$\beta_0 = -1.5 \pm 0.1$ $\beta_1 = 2.9 \cdot 10^{-3} \pm 1.0 \cdot 10^{-3}$	−125.5	-125.5 ± 9.1	50.0	254.9
c_3	NS	−126.1	NS	NS	NS
c_1, c_2	$\beta_0 = -1.5 \pm 0.1$ $\beta_1 = 6.0 \cdot 10^{-4} \pm 3.4 \cdot 10^{-4}$ $\beta_2 = 2.9 \cdot 10^{-3} \pm 1.0 \cdot 10^{-3}$	−124.7	-124.7 ± 9.1	50.0	255.3
c_1, c_2, c_3	$\beta_0 = -1.5 \pm 0.1$ $\beta_1 = 6.5 \cdot 10^{-4} \pm 3.4 \cdot 10^{-4}$ $\beta_2 = 3.0 \cdot 10^{-3} \pm 1.0 \cdot 10^{-3}$ $\beta_3 = 5.2 \cdot 10^{-3} \pm 2.0 \cdot 10^{-3}$	−122.9	-122.9 ± 9.1	50.0	253.7

Abbreviations: CI = confidence interval; LL = log-likelihood; $\langle LL \rangle \pm S_{LL}$ = estimated log-likelihood distribution; P = resulting probability P of obtaining smaller LL value than observed; AIC = Akaike information criterion; NS = model not significant.

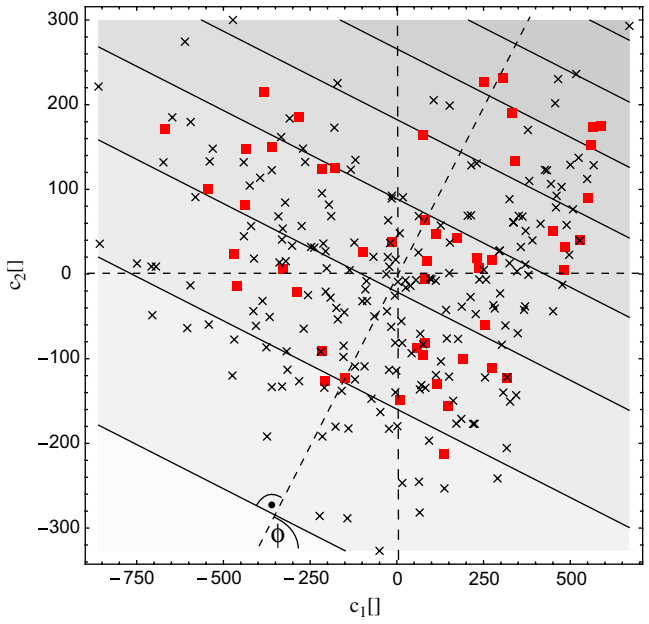

Fig. 4. Scatterplot of first two principal components (PCs), c_2 vs. c_1, for 262 dose–volume histograms (DVHs) (see Fig. 2a); DVHs associated with chronic rectal bleeding of Grade 2 or greater indicated as red squares. Overlayed are isotoxicity lines of a logistic regression model according to Eq. 2 for $M = 2$ variables (for parameter estimates, see Table 1; $\phi = 78.2°$). Isotoxicity lines shown for probability of chronic rectal bleeding (P_{tox}) = 7.5%, 12.5%, …., 42.5%.

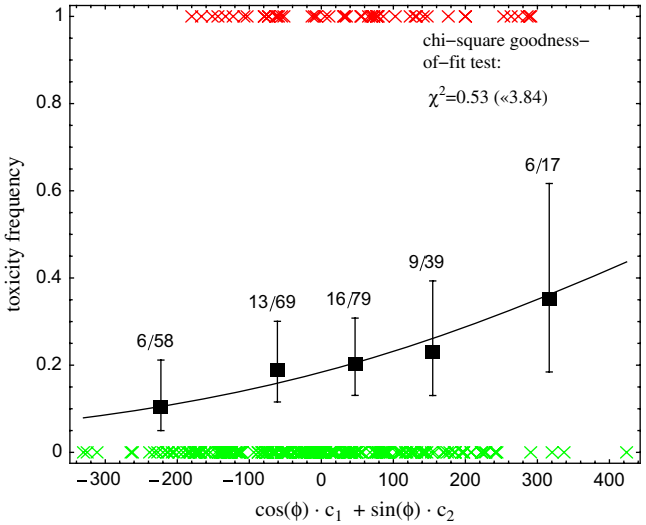

Fig. 5. Data of scatterplot from Fig. 4 projected to an axis with angle $\phi = 78.2°$ (rotation of coordinate system). Symbols represent toxicity (1 and 0 for patients with and without bleeding of Grade 2 or greater, respectively), plotted as function of the abscissa value in the rotated coordinate system. Observed toxicity rates shown in centers of equally sized bins, except for two leftmost and rightmost bins (low incidence region and sample density, respectively), which have been combined. Errors shown are binomial. Solid line represents fit to logistic regression model of Eq. 2.

mention, the local maximum of the compound mode 1+2+3 around ~40–45 Gy can be attributed to the correlations of intermediate doses with the maximal dose as imposed by the four-field box treatment technique, and thus had no biologic interpretation here (in our previous study (4), we found no significant correlation of single dose levels in the range of ~40–45 Gy to bleeding of Grade 2 or greater).

Concerning the log-likelihood (LL), the univariate logistic regression model for c_2 had the lowest LL of the three significant models under consideration (Table 1). Adding c_1 as a variable to the model improved the LL value, without changing the AIC value as a measure of fitting efficiency. However, the three-variate logistic regression model for (c_1, c_2, c_3) leads to both better LL and AIC values, suggesting that the improved fit quality is worth the greater complexity of this model.

DISCUSSION

In this study, we followed the approach of Dawson *et al.* (13) and Bauer *et al.* (14, 15) and used PCA to analyze the rectal wall DVHs and correlation of their PCs with chronic rectal bleeding for a patient population treated with four-field 3D-CRT ART. The eigenvalue spectrum was dominated by few components with an intuitive meaning. This is a consequence of the specific shape of the dose distribution of a four-field box technique, so that the residual uncertainty is generated by the variability in patient geometry.

In a second step, correlations of the PCs and chronic rectal bleeding of Grade 2 or greater were investigated. Logistic regression analysis revealed significant correlation of PC2 as single variable and toxicity. Multivariate logistic regression models with PC1 or PC1 and PC3 as additional variables also proved significant. Thus, patients whose DVH deviates from the mean in the way described by the second eigenmode or the compound modes 1+2 or 1+2+3 have a greater risk of rectal bleeding. The salient feature of all these modes is a greater portion of rectal wall volume in the four-field overlap region. At the very least, this shows that rectal bleeding is subject to a volume effect. Only the most significant compound mode 1+2+3 indicated that the highest dose volume has the greatest impact. It is the benefit of the flexible dose prescription scheme applied in this study that the variability in the maximal dose was large enough to support this finding.

Thus, as the first obvious conclusion for treatment planning using four-field box techniques, the toxicity can be reduced significantly by reducing the rectal volume inside the four-field overlap region, in accordance with the results from other NTCP modeling approaches, as shown in our previous study (4). Generally, image-guided RT or ART schemes prove to be efficient for this purpose. In principle, the DVH-PCA approach could also be applied to clinical treatment planning in a more direct way as offered by Eqs. 1 and 2 as NTCP functions. After calculating the PCs of an individual DVH (see Eq. A5), this phenomenologic NTCP model provides a complication probability for the DVH. This allows for the penalizing of DVH patterns with exceedingly high predicted NTCPs according to this model.

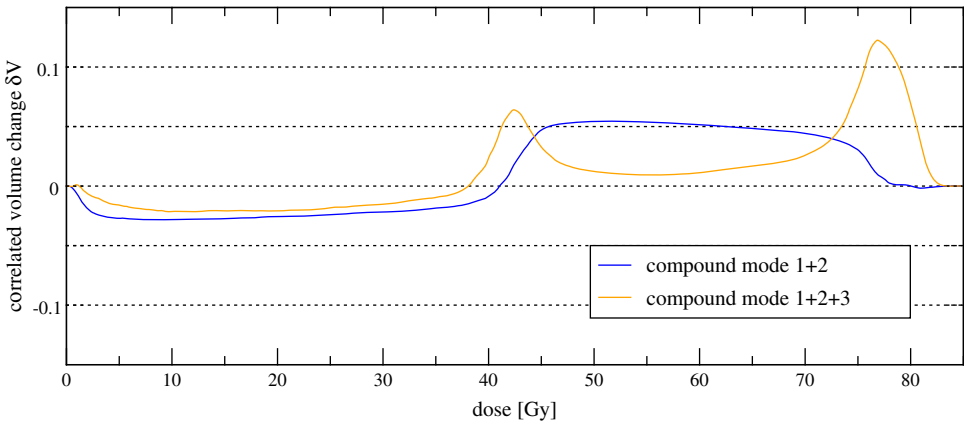

Fig. 6. Normed compound modes of dose–volume histogram (DVH) pattern variability with largest influence on risk of toxicity according to logistic regression models for first two or three principal components.

However, such a ''DVH-PCA–based NTCP model'' should be used with caution. Given the presented eigenvectors and fitting parameters, it is limited to four-field techniques only and does not allow extrapolation to techniques such as intensity-modulated RT with many more degrees of freedom and, consequentially, a larger variability of dose distributions. In a parallel study, different dose–volume- and EUD-based NTCP models were applied to essentially the same patient population (4). These findings revealed that the EUD-based models in particular have similar or better LL and AIC values [the corresponding values for the 262 patients in the present study were as follows: LL = − 123.2 (−123.1) and AIC = 252.3(250.2) for the Logit-EUD and serial reconstruction unit model, respectively (4)] and are thus preferable to describe this data set.

Limited ability to extrapolate, however, is a drawback shared to some extent by all NTCP models. In this context, it is illustrative to distinguish the DVH shape variability in a patient population imposed by the treatment technique from the ''biologically relevant'' variability in dose levels (*i.e.*, the variability that leads to significantly different risks of toxicity). If one or more dominating PCs show a correlation with toxicity (as occurred for our patient population), the treatment technique used obviously induces ''biologically relevant'' variability. Conversely, if none of the dominating PCs correlates with toxicity, basing NTCP models on such a population might lead to misinterpretations.

Moreover, the DVH correlations imposed by the treatment technique affect the results particularly of directly dose–volume-based NTCP models (*e.g.*, cutoff-dose models) and can lead to apparent contradictions in different studies. For example, in our population, significant correlations of V_D with toxicity were found for all dose levels in the range of ∼50–80 Gy, with greatest significance for doses >70 Gy (4). This can be clearly traced back to the four-field box technique used, which induces correlations of all DVH dose bins in the range of ∼50–75 Gy, as shown by the first two PCA eigenvectors. However, the finding that V_D correlates with bleeding for all doses of ∼50–80 Gy is in contrast to the results of some other studies that reported significant correlations for doses of ∼40–50 Gy (2), 40.9 Gy (6), or 60 Gy (8). In all these studies, a six-field conformal technique had been used for all or part of the treatment or population. This treatment technique induces different correlations of the DVH dose bins among the patients, which is the likely reason for these differences. Generally, if a treatment technique generates correlations in the dose distributions of a population, the biologic significance of certain dosimetric variables (*e.g.*, maximal dose, V_D, and so forth, depending on the organ and toxicity under consideration) can be attached to correlated, yet biologically insignificant, variables. These latter variables ''lend'' their biologic significance from the truly significant parameters to which they are correlated. This risk is particularly pertinent if dose–volume constraints (*i.e.*, cutoff dose models) are assessed for their clinical significance.

CONCLUSIONS

Dose-volume histogram-PCA provides valuable information about the correlation structure of DVHs generated by a treatment technique in a patient population and its relationship to toxicity, and thereby proves to be an important tool to augment NTCP modeling approaches.

REFERENCES

1. Boersma LJ, van den Brink M, Bruce AM, *et al*. Estimation of the incidence of late bladder and rectum complications after high-dose (70–78 Gy) conformal radiotherapy for prostate cancer, using dose–volume histograms. *Int J Radiat Oncol Biol Phys* 1998;41:83–92.
2. Jackson A, Skwarchuk MW, Zelefsky MJ, *et al*. Late rectal bleeding after conformal radiotherapy of prostate cancer (II): Volume effects and dose–volume histograms. *Int J Radiat Oncol Biol Phys* 2001;49:685–698.
3. Rancati T, Fiorino C, Gagliardi G, *et al*. Fitting late rectal bleeding data using different NTCP models: Results from an Italian

multi-centric study (AIROPROS0101). *Radiother Oncol* 2004; 73:21–32.
4. Söhn M, Yan D, Liang J, et al. Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* 2007;67:1066–1073.
5. Skwarchuk MW, Jackson A, Zelefsky MJ, *et al*. Late rectal toxicity after conformal radiotherapy of prostate cancer (I): Multivariate analysis and dose-response. *Int J Radiat Oncol Biol Phys* 2000;47:103–113.
6. Tucker SL, Cheung R, Dong L, *et al*. Dose–volume response analysis of late rectal bleeding after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2004;59:353–365.
7. Vargas C, Martinez A, Kestin LL, *et al*. Dose–volume analysis of predictors for chronic rectal toxicity after treatment of prostate cancer with adaptive image-guided radiotherapy. *Int J Radiat Oncol Biol Phys* 2005;62:1297–1308.
8. Zapatero A, García-Vicente F, Modolell I, *et al*. Impact of mean rectal dose on late rectal bleeding after conformal radiotherapy for prostate cancer: Dose–volume effect. *Int J Radiat Oncol Biol Phys* 2004;59:1343–1351.
9. Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method. *Int J Radiat Oncol Biol Phys* 1989;16: 1623–1630.
10. Haken RKT, Martel MK, Kessler ML, *et al*. Use of V_{eff} and iso-NTCP in the implementation of dose escalation protocols. *Int J Radiat Oncol Biol Phys* 1993;27:689–695.
11. Niemierko A. Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Med Phys* 1997;24: 103–110.
12. Niemierko A. A generalized concept of equivalent uniform dose [Abstract]. *Med Phys* 1999;26:1100.
13. Dawson LA, Biersack M, Lockwood G, *et al*. Use of principal component analysis to evaluate the partial organ tolerance of normal tissues to radiation. *Int J Radiat Oncol Biol Phys* 2005;62:829–837.
14. Bauer JD, Jackson A, Skwarchuk M, *et al*. Principal component analysis as applied to volume effects in treatment outcomes for patients with prostate cancer treated with 3D-CRT [Abstract]. *Med Phys* 2004;31:1910.
15. Bauer JD, Jackson A, Skwarchuk M, *et al*. Principal component, Varimax rotation and cost analysis of volume effects in rectal bleeding in patients treated with 3D-CRT for prostate cancer. *Phys Med Biol* 2006;51:5105–5123.
16. Vargas C, Yan D, Kestin LL, *et al*. Phase II dose escalation study of image-guided adaptive radiotherapy for prostate cancer: Use of dose–volume constraints to achieve rectal isotoxicity. *Int J Radiat Oncol Biol Phys* 2005;63:141–149.
17. Martinez AA, Yan D, Lockman D, *et al*. Improvement in dose escalation using the process of adaptive radiotherapy combined with three-dimensional conformal or intensity-modulated beams for prostate cancer. *Int J Radiat Oncol Biol Phys* 2001; 50:1226–1234.
18. Yan D, Lockman D, Brabbins D, *et al*. An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer. *Int J Radiat Oncol Biol Phys* 2000;48:289–302.
19. National Cancer Institute Common Terminology Criteria for Adverse Events v3.0 (CTCAE). Available at *http://ctep.cancer.gov/reporting/ctc_v30.html*. Accessed January 15, 2007.
20. Pekar V, McNutt TR, Kaus MR. Automated model-based organ delineation for radiotherapy planning in prostatic region. *Int J Radiat Oncol Biol Phys* 2004;60:973–980.
21. Söhn M, Birkner M, Yan D, *et al*. Modelling individual geometric variation based on dominant eigenmodes of organ deformation: Implementation and evaluation. *Phys Med Biol* 2005; 50:5893–5908.
22. Cox DR, Snell EJ. Analysis of binary data. 2nd ed. London: Chapman & Hall; 1989.
23. Jackson A, Haken RKT, Robertson JM, *et al*. Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys* 1995;31: 883–891.
24. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest: Academiai Kaido; 1973. p. 267–281.
25. Akaike H. A new look at statistical model identification. *IEEE Trans Autom Control* 1974;19:716–723.
26. Press WH, Flannery BP, Teukolsky SA, *et al*. Numerical recipes in C: The art of scientific computing. 2nd ed. Cambridge: Cambridge University Press; 1992.

APPENDIX

Principal Component Analysis

Given a data set of n DVHs, discretization of the ith DVH results in the p-dimensional vector $\mathbf{X}_i = (V_{i1}, \ldots, V_{ip})$, where V_{ij} is the volume value of the jth dose bin. The values of these p variables are highly correlated, implying that actually much fewer variables are necessary to appropriately describe the data set. PCA is a method to find a statistically motivated transformation of the variables into uncorrelated ''principal components'', which are ordered by their contribution to model the data set variability.

For the purpose of PCA, the mean DVH $\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = (\overline{V}_1, \ldots, \overline{V}_p)$ and the $p \times p$-dimensional covariance matrix

$$\mathcal{C} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})\cdot(\mathbf{X}_i - \overline{\mathbf{X}})^T \qquad \text{(A3)}$$

are calculated, where the column vector $\mathbf{X}_i - \overline{\mathbf{X}}$ describes the difference of the ith DVH from the mean DVH, and $() \cdot ()^T$ denotes the outer product of two vectors. The non-zero off-diagonal elements of this matrix reflect the correlations between the p variables. Diagonalization (in our implementation, the algorithm for diagonalization was taken from Press *et al*. [26]) of $\mathcal{C}$ corresponds to the change to a coordinate system, in which the correlations between the variables vanish. The new coordinate system is defined by the eigenvectors of the covariance matrix (normed, $\|\mathbf{q}_l\| = 1$), which are the statistically independent modes of correlated volume change: $\mathbf{q}_l = (\delta V_{l1}, \ldots, \delta V_{lp})$. Each DVH can be represented by a linear combination of these modes:

$$\mathbf{X}_i = \overline{\mathbf{X}} + \sum_{l=1}^{p} c_l^{[i]}\mathbf{q}_l \qquad \text{(A4)}$$

The coefficients $c_l^{[i]}$ are the new coordinates of the ith DVH, its ''principal components'' (PCs), and are calculated as the following scalar product:

$$c_l^{[i]} = (\mathbf{X}_i - \overline{\mathbf{X}}) \cdot \mathbf{q}_l \tag{A5}$$

The PCs are ordered according to their statistical variance $\sigma^2(c_l)$ over the data set of n DVHs, which is given by the corresponding eigenvalues: $\sigma^2(c_l) = \lambda_l$. The relative eigenvalue $\tilde{\lambda}_l$ then is a measure for the contribution of the lth PC to the overall variability present in the data set:

$$\tilde{\lambda}_l = \lambda_l \Big/ \sum_{j=1}^{p} \lambda_j \tag{A6}$$

The aim of the PCA approach is to describe individual DVHs and their variability by a few dominating PCs, which is approximately possible if a prevailing portion of the overall variability according to Eq. A6 can be described by a few dominating PCs.

Appendix B

Incidence of Late Rectal Bleeding in High-Dose Conformal Radiotherapy of Prostate Cancer using Equivalent Uniform Dose-Based and Dose-Volume-Based Normal Tissue Complication Probability Models

published in
International Journal of Radiation Oncology, Biology and Physics 2007; **67**(4): 1066-1073

Int. J. Radiation Oncology Biol. Phys., Vol. 67, No. 4, pp. 1066–1073, 2007

0360-3016/07/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2006.10.014

CLINICAL INVESTIGATION **Prostate**

INCIDENCE OF LATE RECTAL BLEEDING IN HIGH-DOSE CONFORMAL RADIOTHERAPY OF PROSTATE CANCER USING EQUIVALENT UNIFORM DOSE–BASED AND DOSE–VOLUME–BASED NORMAL TISSUE COMPLICATION PROBABILITY MODELS

MATTHIAS SÖHN, DIPL. PHYS.,* DI YAN, D.SC.,† JIAN LIANG, PH.D.,† ELISA MELDOLESI, M.D.,† CARLOS VARGAS, M.D.,‡ AND MARKUS ALBER, PH.D.*

*Section for Biomedical Physics, University Hospital for Radiation Oncology, Tübingen, Germany; †Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, MI; ‡Radiation Oncology, University of Florida, Gainesville, FL

Purpose: Accurate modeling of rectal complications based on dose–volume histogram (DVH) data are necessary to allow safe dose escalation in radiotherapy of prostate cancer. We applied different equivalent uniform dose (EUD)–based and dose–volume–based normal tissue complication probability (NTCP) models to rectal wall DVHs and follow-up data for 319 prostate cancer patients to identify the dosimetric factors most predictive for Grade ≥ 2 rectal bleeding.

Methods and Materials: Data for 319 patients treated at the William Beaumont Hospital with three-dimensional conformal radiotherapy (3D-CRT) under an adaptive radiotherapy protocol were used for this study. The following models were considered: (*1*) Lyman model and (*2*) logit-formula with DVH reduced to generalized EUD, (*3*) serial reconstruction unit (RU) model, (*4*) Poisson-EUD model, and (*5*) mean dose– and (*6*) cutoff dose–logistic regression model. The parameters and their confidence intervals were determined using maximum likelihood estimation.

Results: Of the patients, 51 (16.0%) showed Grade 2 or higher bleeding. As assessed qualitatively and quantitatively, the Lyman- and Logit-EUD, serial RU, and Poisson-EUD model fitted the data very well. Rectal wall mean dose did not correlate to Grade 2 or higher bleeding. For the cutoff dose model, the volume receiving > 73.7 Gy showed most significant correlation to bleeding. However, this model fitted the data more poorly than the EUD-based models.

Conclusions: Our study clearly confirms a volume effect for late rectal bleeding. This can be described very well by the EUD-like models, of which the serial RU- and Poisson-EUD model can describe the data with only two parameters. Dose–volume–based cutoff-dose models performed worse. © 2007 Elsevier Inc.

Prostate cancer, Rectal toxicity, Normal tissue complication probability, Volume effects, Dose–volume histograms, Equivalent uniform dose.

INTRODUCTION

The essential dose-limiting organs in prostate radiotherapy are the bladder and rectum. One of the most relevant side effects that can significantly compromise a patient's quality of life is chronic rectal bleeding.

Conventional external beam radiotherapy (RT) treatment typically does not allow prostate doses beyond 65 to 70 Gy without an unacceptably high risk of rectal toxicity, although higher tumor doses are favorable for improved tumor control. The possibility of dose escalation beyond 70 Gy to the prostate is based on the volume–effect of rectum, *i.e.,* the observation of increased tolerance to high doses if the high dose region is confined to a small volume. Technically, this becomes feasible because of conformal techniques such as three-dimensional conformal radiotherapy (3D-CRT) or intensity-modulated radiotherapy (IMRT), especially when aided by image-guided adaptive approaches.

Safe dose escalation necessitates accurate quantitative modeling of the volume effect based on the detailed dose–volume information provided by modern treatment planning systems. Numerous studies have established evidence of a significant correlation between parameters derived from rectal dose–volume histograms (DVHs) and toxicity (see

Reprint requests to: Matthias Söhn, Dipl. Phys., Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany; Tel: (+49) 7071-2986061; Fax: (+49) 7071-295920; E-mail: Matthias.Soehn@med.uni-tuebingen.de

Presented in part at the 48th Annual American Society for Therapeutic Radiology and Oncology Meeting (ASTRO), Philadelphia, PA, Nov. 5–9, 2006.

Supported in part by Deutsche Krebshilfe e.V. Grant No. 106280 and National Institutes of Health Grant No. RO1 CA091020.

Conflict of interest: none.

Received Sept 6, 2006, and in revised form Oct 2, 2006. Accepted for publication Oct 9, 2006.

Table 1. Toxicity score for chronic rectal bleeding based on Common Terminology Criteria for Adverse Events (v. 3.0)

Grade	Description
1	Mild hemorrhage/bleeding; intervention (other than iron supplements) not indicated
2	Symptomatic and medical intervention or minor cauterization indicated
3	Transfusion, interventional radiology, endoscopic or operative intervention indicated
4	Life-threatening consequences; perforation/dysfunction requiring urgent intervention

Refs. 1–4 and references therein). However, only a few publications have quantified the risk of rectal complications in terms of normal tissue complication probability (NTCP) models (5–9). Such empiric or semiempiric models parameterize the vast information about inhomogeneous dose distributions and corresponding outcome data from large patient populations into few-parametric models that assign a single probability value to an individual treatment plan. This enables evidence-based ranking of alternative plans in the planning process according to their predicted complication risk. Two important types of NTCP-models can be distinguished. The first are dose–volume–based models, which use a single DVH parameter (*e.g.,* the volume V_D irradiated to a certain dose-level D) for ranking plans according to their complication probability. In contrast, EUD-like models define an equivalent uniform dose, $EUD = f^{-1}(\Sigma_i v_i f(D_i))$, as surrogate parameter calculated using all bins (v_i, D_i) of a DVH, where the form of the (monotonic) function f depends on the model. Besides possible differences in the quality of fit as investigated in this study, it should be mentioned that the choice between dose–volume–based vs. EUD-like models when used for treatment planning, especially IMRT, affects the process of plan optimization (10–12).

In this study, we apply one dose–volume–based and five EUD-like NTCP models for chronic rectal bleeding of Grade ≥ 2 to a population of 319 prostate cancer patients treated with a 3D-CRT adaptive radiotherapy (ART) technique to doses between 70.2 and 79.2 Gy at the William Beaumont Hospital. This is, so far, the largest published, single-institution patient population studied for fitting of rectal NTCP-models. Thus, this study not only provides valuable information for identification of the superior modeling approaches but also statistically well-based estimates of the corresponding model parameters.

METHODS AND MATERIALS

Patient data

Data for 319 prostate cancer patients treated between 1999 and 2002 at the William Beaumont Hospital were used for this study. The characteristics of this patient population have been described in previous studies (3, 13). The patients were part of a Phase II dose-escalation study and underwent 3D-CRT with image-guided off-line correction under an ART protocol.

All patients had one pretreatment planning CT scan, daily portal images to determine and correct for setup errors, four additional CT scans during the first week of the treatment used for individual adaptation of the treatment plan, and weekly CT scans in the following to preclude undetected drifts. The (solid) rectum was contoured on the initial CT scan from the anal verge or ischial tuberosities (whichever was higher) to the sacroiliac joints or rectosigmoid junction (whichever was lower). Rectal wall was defined based on the solid rectum contours with 3- to 4-mm wall thickness.

The ART scheme used has been described elsewhere (14, 15). In short, a four-field box technique with 18 MV photons was used both for the initial treatment plan of the first week and the following adapted plan. In the first week, the patients were treated for a dose of 9 Gy to the target, where the planning target volume (PTV), was generated based on the clinical target volume (CTV), of the initial CT (prostate, or prostate + seminal vesicles) with a population-based margin of 1 cm. For the adapted plan, information from daily portal imaging and the five CT scans available after the first week of treatment were used to estimate setup error and individual prostate mobility, which allowed to define a (generally smaller) patient-specific PTV.

The final dose to the PTV was limited by dose–volume constraints of rectal wall and bladder based on the geometry of the initial planning CT image. For rectal wall these were: (*1*) $D_{30\%} =$ 75.6 Gy, and (*2*) $D_{5\%} = 82$ Gy. The possible dose levels (minimal prostate dose) were chosen under the requirement to meet rectum (and bladder) constraints, and were as follows: 70.2, 72, 73.8, 75.6, 77.4, and 79.2 Gy.

For each patient the dose distributions of the initial and adapted plan were calculated using Pinnacle 6.2b (ADAC Laboratories, Milpitas, CA). An in-house developed software was used to calculate DVHs of the rectal wall. This software used the contours from the initial (planning) CT and calculated the overall dose as sum of initial and adapted (physical) dose distributions. The DVH-dose bin size was 0.1 Gy, with volume defined as relative (percentage) volume irradiated.

The rectal toxicity variable regarded in this analysis is chronic rectal bleeding. The follow-up scheme defined examinations at 3-month intervals during the first 2 years, and every 6 months from the second to the fifth year. As mentioned above, this study is based on the patient population analyzed in Vargas *et al.* (3, 13). However, for the current study, all patient files were re-examined to improve follow-up time. Complications were graded based on the Common Terminology Criteria for Adverse Events (CTCAE) v 3.0 (Table 1). Of the 331 patient datasets, 12 used by Vargas *et al.* could not be used because of technical problems in restoring dose distributions or lost, incomplete, or inconsistent follow-up information. The median clinical follow-up for the remaining 319 patients was 2.8 years (range, 0.1–6.4), with an interquartile range of 1.5 to 4.0 years (25th–75th percentile).

The NTCP models

An NTCP model assigns a complication probability for an organ at risk to a generally inhomogeneous dose distribution. The func-

tional form of such a model can be based on a mechanistic description of biologic processes, or might be designed to result in a phenomenologic fit of the data.

The models considered in this work are of the general forms described below.

First, a summary measure μ, such as the mean dose, an equivalent uniform dose (EUD), or similar, is calculated from the dose distribution. The quantity μ serves as a ranking function by imposing an order among individual plans according to their complication risk.

Next, a function, NTCP(μ), which assigns complication probabilities to the values of the summary measure is defined. Such a function is required (*1*) for continuously mapping μ to the interval [0, 1], while (*2*) preserving the ranking imposed by the numerical values of the summary measure. This leads to the class of sigmoid-type (S-shaped) functions, wherein the following probit, logit, Poissonian, and logistic formulae are used.

We applied six different models to our data, where the endpoint was chosen to be chronic rectal bleeding of Grade ≥ 2. These models differ in the summary measure used and/or the functional form of the NTCP function as described in the following.

Lyman-EUD model

The most widely used phenomenological approach is the family of Lyman models (16–20), which uses the probit function

$$\mathrm{NTCP}_{probit}(\mu) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s\cdot(\mu-\mu_{50})} \exp\left(-x^2/2\right)dx \qquad (1)$$

to map the summary measure μ to the interval [0, 1] of complication probabilities. This integral essentially yields the error function, where the parameter s is the slope of the sigmoid response curve at the steepest point $\mu = \mu_{50}$, for which the NTCP function predicts 50% complication probability. Usually the slope parameter s is replaced by its inverse m according to $s = 1/(m \cdot \mu_{50})$.

Different DVH reduction schemes have been used for defining the summary measure μ, such as an effective volume as in the LKB model (18) or an effective dose (21). In the following, the generalized equivalent uniform dose (22) is used, which defines EUD as Lebesque a-norm of the dose, *i.e.,* in terms of the following power–law relationship:

$$\mu := \mathrm{EUD}_{PL} = \left(\sum_i v_i D_i^a\right)^{1/a} \qquad (2)$$

The sum is calculated over all bins (v_i, D_i) of the differential DVH, and a is a parameter associated with the strength of the volume effect for the organ under consideration (range, $a \in [1 \ldots \infty]$): For $a \to \infty$ the EUD is the maximum dose (*i.e.,* no volume effect), whereas for $a = 1$ Eq. (2) gives the mean dose (large volume effect).

Summarizing, the Lyman-EUD model as used in our study is described by 3 parameters: a, m, and EUD_{50} (usually termed D_{50}).

Logit-EUD model

This model also uses the generalized EUD Eq. (2) as summary measure μ, although it differs from the Lyman-EUD model in the choice of the NTCP function. Here the logit function

$$\mathrm{NTCP}_{logit}(\mu) = \frac{1}{1 + \left(\mu_{50}/\mu\right)^k} \qquad (3)$$

is chosen as sigmoid shape function (6, 10). Its two parameters μ_{50} (*i.e.,* D_{50}) and k are determined by the EUD, which causes a complication rate of 50%, and the slope of the NTCP curve here. Thus, together with the parameter a of the EUD, this model has three parameters.

Serial reconstruction unit model

In contrast to the two previous phenomenologic NTCP models, the serial reconstruction unit model, which has been proposed recently by Alber and Belka (23), arises from certain general assumptions about the biological processes causing normal tissue complications.

The model regards radiation induced complications as the consequence of local failure of dynamic repair processes. As an assumption, the latter is attributed to the finite range of the repair mechanisms, which finds its correlate in the model by the description of finite-sized reconstruction units and their microscopic dose–response. Borrowing analogies from thermodynamics and statistical physics, the authors derive the following expression to describe the macroscopic dose–response in terms of the NTCP for homogeneous irradiation of the partial volume V of an organ with the dose D:

$$\mathrm{NTCP}_{SRU}(V,\ D) = 1 - \exp\left(-V\exp\left(\sigma\left(D - D_0\right)\right)\right) \qquad (4)$$

where σ is an organ specific sensitivity parameter and D_0 is a reference dose.

For inhomogeneous dose distributions an equivalent uniform dose, which would give the same macroscopic dose–response when applied homogeneously to the whole organ ($V = 1$), can be defined as

$$\mathrm{EUD}_{SRU} = \frac{1}{\sigma}\log\left(\sum_i v_i \exp\left(\sigma D_i\right)\right) \qquad (5)$$

Consequently, the NTCP function then reads:

$$\mathrm{NTCP}_{SRU}\left(\mathrm{EUD}_{SRU}\right) = 1 - \exp\left(-\exp\left(\sigma\left(\mathrm{EUD}_{SRU} - D_0\right)\right)\right) \qquad (6)$$

Summarizing, the serial reconstruction unit model has the two parameters σ and D_0 to be fitted. Note that in contrast to the previously described models, which have the volume effect parameter a of the power-law EUD as a third parameter, here the sensitivity parameter σ is inherently coupled to the same value in the EUD and NTCP function.

Poisson-EUD model

Similarly to the serial reconstruction unit model, this model uses mechanistic concepts to describe predominantly serial tissue dose–response. Assuming that complication is a consequence of local

dose–response of noninteracting subunits, the following NTCP-function can be derived based on Poissonian statistics (24):

$$\mathrm{NTCP}_{poisson}(\mathrm{EUD}) = 1 - \exp\left[-\left(\frac{\mathrm{EUD}}{\mathrm{D}_0}\right)^a\right] = 1 - \exp\left[-\ln 2 \cdot \left(\frac{\mathrm{EUD}}{\mathrm{D}_{50}}\right)^a\right] \quad (7)$$

with a reference dose D_0 (or a dose D_{50} causing 50% complication probability) and a volume–effect (steepness) parameter a. The EUD is given by Eq. (2), where, according to this model, the exponent of EUD and the steepness parameter of the NTCP function have the same value. Thus, unlike the Lyman- and Logit-EUD models, the Poisson-EUD model has only two parameters.

Mean dose logistic regression model

An association of the rectal mean dose with chronic rectal bleeding has been reported by some authors (7, 8). To test for such an association based on our data, we used logistic regression as a standard method from statistics for this purpose. In terms of the general NTCP model scheme presented above, here the NTCP function is given by the two-parametric logistic function

$$\mathrm{NTCP}_{logistic}(\mu) = \frac{1}{1 + \exp(-\beta_o - \beta_1 \mu)} \quad (8)$$

with the mean dose $D_{mean} = \sum_i v_i D_i$ as summary measure μ. Note that D_{mean} is a special case of the EUD, Eq. (2), for fixed parameter $a = 1$. Thus, this model has only the two parameters β_0 and β_1.

Cutoff dose logistic regression model

In classical, *i.e.,* nonbiological treatment planning approaches, dose limitation to organs at risk (OAR) is usually implemented in terms of dose–volume constraints. For a given treatment-technique, the most relevant dose level(s) predictive for toxicity can be determined by retrospectively fitting a sigmoid type NTCP function to outcome data of a patient population.

In this phenomenological approach, the summary measure μ is given by the proportion V_{D_c} of the OAR receiving doses equal to or above a (cutoff) dose level D_c. In the present study, V_{D_c} is regarded as relative volume, formally: $V_{D_c} = \sum_{i|(D_i \geq D_c)} v_i$, where v_i is the discretized form of the differential DVH, and only dose bins i with $D_i \geq D_c$ are used in the sum.

For given value $\mathrm{D_c}$ we used logistic regression Eq. (8) to test for correlation of V_{D_c} and chronic rectal bleeding, resulting in two fit parameters β_0 and β_1. This was systematically repeated for all possible D_c up to 85 Gy in increments of 0.1 Gy to assess the significance of such a logistic regression model for different cutoff doses. Thus, altogether the model has the three parameters D_c, β_0, and β_1.

Fitting procedure

The method of choice for fitting such models to sparse, dichotomous response data (0/1, if patient shows bleeding of Grade < 2 or ≥ 2, respectively) is maximum likelihood estimation (Jackson *et al.* (25) and references therein; see also Ref. 6). In this method the optimal model parameters are determined such as to maximize the probability of occurrence of the observed data, which is given by the so-called likelihood function L. Because of its smallness, numerically this is usually implemented as maximization of the natural logarithm $\ln(L)$, the LogLikelihood (LL). In our implementation, the software package Mathematica version 5.0 (Wolfram Research Inc., Champaign, IL) was used. To reduce calculation time, the DVH discretization was changed to 0.5 Gy when fitting the Lyman-, Logit-, Poisson-EUD, and serial reconstruction unit models.

Uncertainties of the model parameters were assessed using the variance–covariance matrix of the parameters calculated around the maximum of LL as described in (25). The 68% confidence intervals for the parameters can be estimated by the square root of the diagonal elements.

The goodness of fit of each individual model has been quantified in two ways. The first method follows Jackson *et al.* (25). If the actual (maximized) value of the LL for the NTCP model fitted to the observed patient population is denoted by LL_{obs}, the probability P of obtaining a value smaller than LL_{obs} (*i.e.,* a worse fitting model) purely by chance can be assessed based on the statistical distribution of LL. This can be obtained from analytical formulae for the mean $\langle \mathrm{LL} \rangle$ and variance S_{LL} (see Eq. A3 and A4 in Jackson *et al.)* under the assumption of a normally distributed LL. If this probability turns out to be too large, the model "overfits" the data; if it is too small, the model does not fit the data well. According to Rancati *et al.* (6) values between 30% and 70% indicate a satisfactory fit.

Additionally, a chi-square goodness-of-fit test was performed for each model: A histogram of the observed patient data was calculated, and the resulting group complication rates $f_i^{(obs)}$ were compared with the corresponding rates $f_i^{(fit)}$ predicted by the respective NTCP model by determining $\chi^2 = \sum_N (f_i^{(obs)} - f_i^{(fit)})^2 / f_i^{(fit)}$, which approximately follows a chi-square distribution with $N-1-n$ degrees of freedom (N = number of histogram bins; n = number of model parameters).

Intermodel comparison: Akaike information criterion

A widely used measure allowing to compare competing models is the AIC (26, 27); see also (7). Generally, for fits of different models to a given dataset a larger likelihood value indicates a better fit to the data. The AIC quantifies the tradeoff between the model's quality of fit (associated with the likelihood value) and its complexity (expressed by the number of model parameters n), and is defined as $\mathrm{AIC} = -2\mathrm{LL} + 2n$. Models with smaller AIC values are considered to provide a better (in the sense of more efficient) fit to the data than models with larger AIC.

RESULTS

Of the 319 patients, 45, 5, and 1 (14.1%, 1.6%, and 0.3%) showed chronic rectal bleeding of Grade 2, 3, and 4, respectively. Thus, altogether 51 patients (16.0%) showed chronic rectal bleeding of Grade ≥ 2, which is the endpoint for the following model fits.

EUD-based models, mean dose model

Parameter estimates and their 68% CIs for the Lyman-, Logit-, Poisson-EUD, and serial reconstrunction unit model are given in Table 2. The plots of the corresponding NTCP-curves are shown in Fig. 1a to 1d. As obvious in these plots by comparison with the observed complication rates, all four models fit the data very well. Quantitatively, this is

Table 2. Chronic rectal bleeding grade ≥ 2: Estimated parameter values for the six normal tissue complication probability models, observed LogLikelihood (LL) values, estimated LL distribution, and resulting probability of obtaining a smaller LL value than observed

Model	Parameter estimates (68% CI)	LL	$\langle LL \rangle \pm S_{LL}$	P (%)
Lyman-EUD model	$a = 11.9 \pm 3.8$ $m = 0.108 \pm 0.027$ $D_{50} = 78.4 \pm 2.1$	−134.5	−134.6 ± 10.5	50.4
Logit-EUD model	$a = 12.1 \pm 3.8$ $k = 15.4 \pm 4.5$ $D_{50} = 78.1 \pm 2.1$	−134.5	−134.5 ± 10.5	50.0
Serial RU model	$\sigma = 0.179 \pm 0.047$ $D_0 = 80.6 \pm 0.9$	−134.5	−135.6 ± 10.6	54.0
Poisson-EUD model	$a = 13.5 \pm 3.8$ $D_{50} = 78.5 \pm 0.6$	−134.5	−135.6 ± 10.6	54.1
Mean dose model	—	−138.9	—	—
Cutoff dose model				
D_c = 73.7 Gy	$\beta_0 = -2.88 \pm 0.34$ $\beta_1 = 0.050 \pm 0.013$	−136.1	−136.1 ± 10.7	50.0
D_c = 79.6 Gy	$\beta_0 = -2.10 \pm 0.16$ $\beta_1 = 0.068 \pm 0.015$	−135.3	−135.3 ± 10.7	50.0

Abbreviations: EUD = equivalent uniform dose; RU = reconstruction unit. Dash indicates that model is not significant.

manifest by the small values of χ^2: For example, the fit of the serial reconstruction unit model (Fig. 1c) resulted in a χ^2 of 0.03, where the upper limit of χ^2 for an acceptable fit is 5.99 according to Chi-square statistics (2 degrees of freedom; $\alpha = 0.05$). As described later here, the goodness of fit was also determined based on the LL, $\langle LL \rangle$ and S_{LL}. These values together with the resulting probability P of obtaining a smaller LL are given in Table 2. Again, for all four models P indicates acceptable fits.

Concerning the mean dose logistic regression model, it turns out that D_{mean} does not significantly correlate to chronic rectal bleeding of Grade ≥ 2 ($p = 0.11$). The worse fit quality of this model in comparison with the four above-mentioned EUD-based models is also expressed by its significantly lower LL value (Table 2). Thus, this model is not considered further in this study.

Cutoff dose model

Figure 2a shows the LL values of different logistic regression model fits when varying the cutoff dose D_c. For all dose levels in the range $D_c \sim 50$ to 80 Gy, a significant correlation ($\alpha < 0.05$) of the relative volume irradiated with doses ≥ D_c and chronic rectal bleeding of Grade ≥ 2 was found for our patient population. The curve has maxima at $D_c = 73.7$ Gy and 79.6 Gy. Figure 2b depicts the model fit for $D_c = 73.7$ Gy: As obvious, the fit quality is slightly worse than for the EUD-models. However, both the value of P in Table 2 and a χ^2 of 1.06 (upper limit according to Chi-square statistics [1 degree of freedom, $\alpha = 0.05$]: 3.84) show that the fit is acceptable. Formally, this is also the case for the model with $D_c = 79.6$ Gy (plot not shown; $\chi^2 = 2.24 < 3.84$). However, as only a part of our patient population receives doses above 79.6 Gy to nonvanishing volumes of rectal wall, the distribution of the summary measure $V_{79.6\ Gy}$ itself is strongly shifted to small values, thereby compromising the statistical strength of the patient dataset in the region of larger $V_{79.6\ Gy}$.

Comparison of the models

Both visually from Fig. 1 and according to the LL values in Table 3, the four EUD-based NTCP models fit the data of chronic rectal bleeding of Grade ≥ 2 equally well. However, as the Akaike information criterion shows, both two-parametric NTCP models (serial reconstruction unit model and Poisson-EUD model) have the lowest AIC-values (Table 3) and thus, can be considered to fit the data most efficiently.

Compared with these models, the AIC-values of the two cutoff-dose models ($D_c = 73.7$ Gy and 79.6 Gy) are considerably larger, which is caused by both the smaller LL values (indicating worse fit quality itself) and the larger number of model parameters. Thus the EUD-based models provide a quantitatively better and more efficient description of our dataset.

DISCUSSION AND CONCLUSIONS

In this study, six dose–volume–based or EUD-like NTCP models were fitted to late rectal bleeding data of a large, consistently treated patient population (319 patients), thereby aiming to identify the most accurate approach for quantifying the risk of chronic rectal bleeding based on planned dose distributions of rectum and allowing statistically robust estimation of the corresponding model parameters.

Our results clearly confirm the volume effect for chronic rectal bleeding of Grade 2 or worse. Quantitatively, this can be described very well with the four EUD-like models,

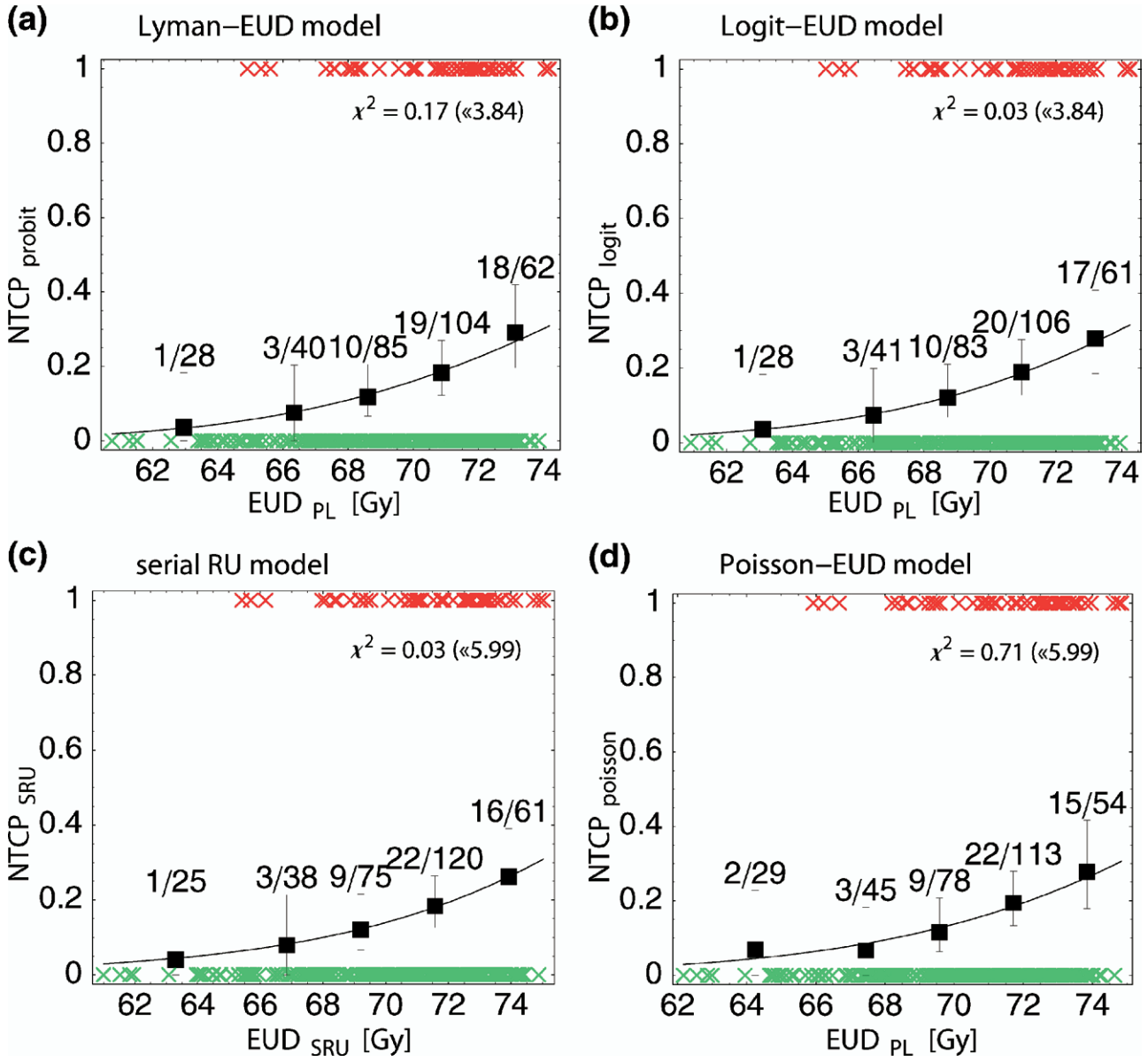

Fig. 1. Predicted probability of chronic rectal bleeding of Grade ≥ 2 according to the different equivalent uniform dose–based normal tissue complication probability (NTCP) models: (a) Lyman–EUD model Eq. (1); (b) Logit-EUD model Eq. (3); (c) serial reconstruction unit model Eq. (6); (d) Poisson-EUD model Eq. (7). NTCP is plotted as function of the corresponding equivalent uniform dose, EUD_{PL} Eq. (2) or EUD_{SRU} Eq. (5). The x-symbols represent reply toxicity (1/0 for patients with/without toxicity, respectively). For each NTCP-model, the observed toxicity rates are shown in centers of equally sized bins (except for two bins in the low incidence region, which were combined). Errors shown are binomial. χ^2 of the fit and the upper threshold according to Chi-square statistics ($\alpha = 5\%$) are given for each model.

where the serial reconstruction unit and the Poisson-EUD model fitted the data most efficiently according to the smallest AIC values, as they need only two parameters to describe the dataset. Thus according to our data, these two models can be considered to provide the most concise approach to quantifying the risk of chronic rectal bleeding of Grade 2 or worse. The cutoff-dose model, a directly dose–volume based model, generally fitted the data worse, but still found significant correlation of rectal wall relative volume above single cutoff-dose levels D_c in the range ~ 50 to 80 Gy (most significant for D_c = 73.7 and 79.6 Gy) for our patient population. Mean dose did not correlate to late rectal bleeding of Grade ≥ 2.

For the three models incorporating the power–law EUD Eq. (2), *i.e.,* the Lyman-, Logit-, and Poisson-EUD model, the parameter describing the volume effect was found to be in the order of $a \approx 12$ (Table 2). Accounting for uncertainties in parameter estimation, this is consistent to Burman *et al.* (5), who found $n \approx 0.12$ (different definition of volume effect parameter: $a = 1/n$, *i.e.,* $a = 8.3$). It is also in agreement with the data published by Skwarchuk *et al.* (28) which yield $a = 10.3$ (fit based on Fig. 3 of the cited paper).

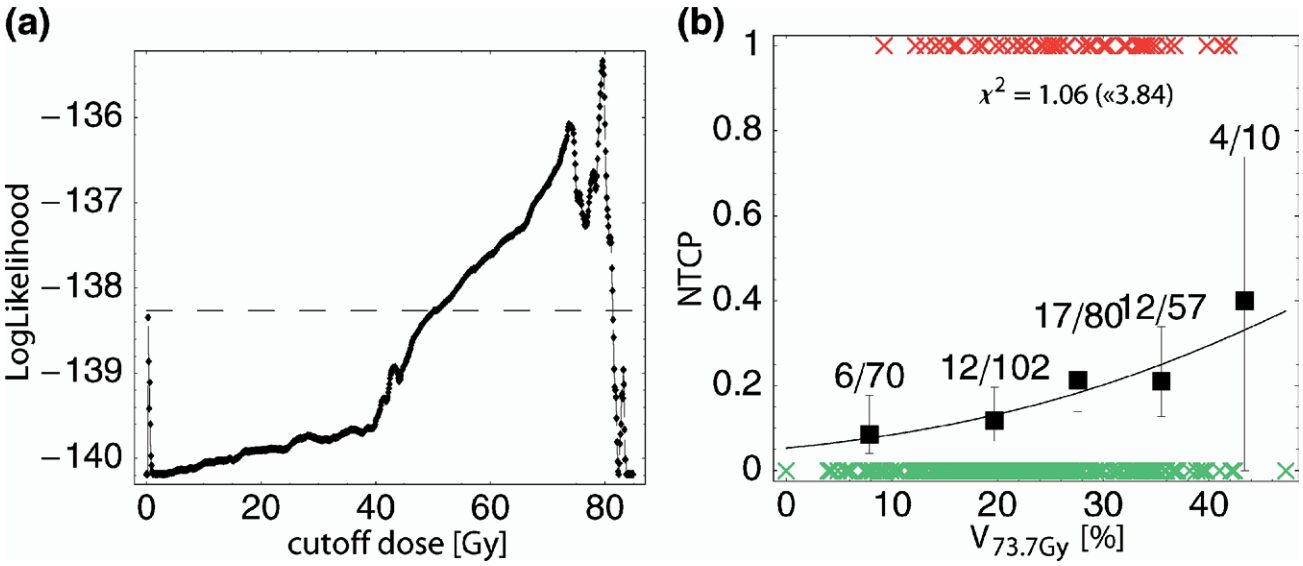

Fig. 2. Cutoff-dose logistic regression model of chronic rectal bleeding Grade ≥ 2: (a) Values of LogLikelihood (LL) in dependence of the cutoff dose D_c; models with LL above the dashed horizontal line reach a significance level of $\alpha < 0.05$. (b) Predicted probability of bleeding according to the model for D_c = 73.7 Gy (local maximum of LL according to (a)), plotted as a function of the relative volume receiving ≥ 73.7 Gy.

However, it is in contradiction to findings of Rancati *et al.* (6) (a = 4.3, 68% CI ~ 3.6—5.6) and Tucker *et al.* (7) (a = 0.3, with large uncertainty a ~ 0–32.3 [95% CI]).

Because of the ART-protocol used, which defined different prescription dose levels depending on individual organ geometry and mobility, the patient population shows a wide range of DVH shapes and thus large variability of dose–volume combinations, which is advantageous for robust fitting of NTCP models. However, possible statistical biases are introduced by the treatment technique and specific characteristics of the patient population itself.

This is indeed the case as becomes most evident for the results of the cutoff-dose model, Fig. 2a: The (local) maxima of the LL at D_c = 73.7 and 79.6 Gy are strongly influenced by the different prescription dose levels defined by the ART protocol (70.2, 72, 73.8, 75.6, 77.4, and 79.2 Gy), whereas the finding that all models with D_c in the range ~ 50 to 80 Gy are formally significant can be traced back to the four-field box treatment technique used, which induces correlations of all DVH dose-bins in this range. Quantitatively, such correlations can be assessed by principal component analysis (29, 30); a detailed investigation of the use of DVH– principal component analysis in the context of NTCP modeling will be presented in a subsequent publication.

Table 3. Comparison of the normal tissue complication probability models using the Akaike information criterion (AIC value)

Model	n	LL	AIC
Lyman-EUD model	3	−134.54	275.1
Logit-EUD model	3	−134.51	275.0
Serial RU model	2	−134.53	273.1
Poisson-EUD model	2	−134.55	273.1
Mean dose model	2	−138.89*	—
Cutoff dose model (73.7 Gy)	3	−136.08	278.2
Cutoff dose model (79.6 Gy)	3	−135.34	276.7

Abbreviations: EUD = equivalent uniform dose; LL = LogLikelihood; n = number of model parameters; RU = reconstruction unit.

* Model is not significant ($p > 0.05$).

In this context, a comparison with results from other publications is elucidating. Both Jackson *et al.* (2) and Tucker *et al.* (7) found significant correlations of intermediate doses ~ 40 to 45 Gy with toxicity in contradiction to our findings. This is likely because of differences in the treatment technique (6-field 3D-CRT vs. 4-field box in combination with 6-field 3D-CRT boost), which induces specific correlations between dose-bins. Thus an important conclusion is that results of DVH-based models like the cutoff-dose model are superimposed by characteristics of the treatment technique and patient population, which compromises inferences about radiobiological effects and extrapolability of results gained from a certain treatment technique to others.

With regard to the mean dose model, our results are in contrast to the studies of Zapatero *et al.* (8), Tucker *et al.* (7), who found a correlation of mean dose and chronic rectal bleeding of Grade ≥ 2. Again, this might be because of differences in the treatment technique or the grading scheme used; in this context, it should be mentioned that Grade ≥ 3 bleeding correlated to mean dose for our population (p = 0.045; data not shown).

With respect to the magnitude of the volume effect for rectum, our results can be compared with values published for other organs (5): Lung ($a \approx 1.1$) and liver ($a \approx 3.1$) are typical organs with a large volume effect, while *e.g.,* spinal cord shows only a small volume effect ($a \approx 20.0$). Thus, for our

data, the EUD-like models suggest a rather small volume effect of rectal bleeding ($a \approx 13.5$ for the Poisson-EUD model, Table 2). As illustration of this, a plan should not have more than 2.4% (1.7%) of rectal volume irradiated with 80 Gy (82 Gy) to have the same EUD (and thus same NTCP) as an otherwise identical plan with 72 Gy to 10% of the volume.

If the presented EUD-based models are used for dose optimization, *e.g.,* for IMRT dose escalation, it is always preferrable to err toward smaller volume effects. In consequence, a EUD-based optimization of prostate RT with the parameters derived here will severely penalize high dose regions in the rectum, perhaps for some patients to an extent that prevents satisfying PTV coverage. Hence, the data suggest that safe dose escalation to the prostate can be achieved only by image guided and adaptive strategies to reduce the extent of the PTV.

REFERENCES

1. Boersma LJ, van den Brink M, Bruce AM, *et al.* Estimation of the incidence of late bladder and rectum complications after high-dose (70–78 Gy) conformal radiotherapy for prostate cancer, using dose–volume histograms. *Int J Radiat Oncol Biol Phys* 1998;41:83–92.
2. Jackson A, Skwarchuk MW, Zelefsky MJ, *et al.* Late rectal bleeding after conformal radiotherapy of prostate cancer (II): Volume effects and dose–volume histograms. *Int J Radiat Oncol Biol Phys* 2001;49:685–698.
3. Vargas C, Martinez A, Kestin LL, *et al.* Dose–volume analysis of predictors for chronic rectal toxicity after treatment of prostate cancer with adaptive image-guided radiotherapy. *Int J Radiat Oncol Biol Phys* 2005;62:1297–1308.
4. Peeters STH, Lebesque JV, Heemsbergen WD, *et al.* Localized volume effects for late rectal and anal toxicity after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2006;64:1151–1161.
5. Burman C, Kutcher GJ, Emami B. Fitting of normal tissue tolerance data to an analytic function. *Int J Radiat Oncol Biol Phys* 1991;21:123–135.
6. Rancati T, Fiorino C, Gagliardi G, *et al.* Fitting late rectal bleeding data using different NTCP models: Results from an Italian multi-centric study (AIROPROS0101). *Radiother Oncol* 2004;73:21–32.
7. Tucker SL, Cheung R, Dong L, *et al.* Dose–volume response analysis of late rectal bleeding after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2004;59:353–365.
8. Zapatero A, García-Vicente F, Modolell I, *et al.* Impact of mean rectal dose on late rectal bleeding after conformal radiotherapy for prostate cancer: Dose–volume effect. *Int J Radiat Oncol Biol Phys* 2004;59:1343–1351.
9. Peeters STH, Hoogeman MS, Heemsbergen WD, *et al.* Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: Normal tissue complication probability modeling. *Int J Radiat Oncol Biol Phys* 2006;66:11–19.
10. Wu Q, Mohan R, Niemierko A, *et al.* Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *Int J Radiat Oncol Biol Phys* 2002;52:224–235.
11. Bär W, Schwarz M, Alber M, *et al.* A comparison of forward and inverse treatment planning for intensity-modulated radiotherapy of head and neck cancer. *Radiother Oncol* 2003;69:251–258.
12. Bos LJ, Schwarz M, Bär W, *et al.* Comparison between manual and automatic segment generation in step-and-shoot IMRT of prostate cancer. *Med Phys* 2004;31:122–130.
13. Vargas C, Yan D, Kestin LL, *et al.* Phase II dose escalation study of image-guided adaptive radiotherapy for prostate cancer: Use of dose–volume constraints to achieve rectal isotoxicity. *Int J Radiat Oncol Biol Phys* 2005;63:141–149.
14. Martinez AA, Yan D, Lockman D, *et al.* Improvement in dose escalation using the process of adaptive radiotherapy combined with three-dimensional conformal or intensity-modulated beams for prostate cancer. *Int J Radiat Oncol Biol Phys* 2001;50:1226–1234.
15. Yan D, Lockman D, Brabbins D, *et al.* An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer. *Int J Radiat Oncol Biol Phys* 2000;48:289–302.
16. Lyman JT. Complication probability as assessed from dose–volume histograms. *Radiat Res* 1985;104:S13–S19.
17. Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method. *Int J Radiat Oncol Biol Phys* 1989;16:1623–1630.
18. Kutcher GJ, Burman C, Brewster L, *et al.* Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. *Int J Radiat Oncol Biol Phys* 1991;21:137–146.
19. Lyman JT, Wolbarst AB. Optimization of radiation therapy, III: A method of assessing complication probabilities from dose–volume histograms. *Int J Radiat Oncol Biol Phys* 1987;13:103–109.
20. Lyman JT, Wolbarst AB. Optimization of radiation therapy, IV: A dose–volume histogram reduction algorithm. *Int J Radiat Oncol Biol Phys* 1989;17:433–436.
21. Mohan R, Mageras GS, Baldwin B, *et al.* Clinically relevant optimization of 3-D conformal treatments. *Med Phys* 1992;19:933–944.
22. Niemierko A. A generalized concept of equivalent uniform dose [Abstract]. *Med Phys* 1999;26:1100.
23. Alber M, Belka C. A normal tissue dose response model of dynamic repair processes. *Phys Med Biol* 2006;51:153–172.
24. Alber M, Nüsslin F. A representation of an NTCP function for local complication mechanisms. *Phys Med Biol* 2001;46:439–447.
25. Jackson A, Haken RKT, Robertson JM, *et al.* Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys* 1995;31:883–891.
26. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest: Academiai Kaido, 1973: 267–281.
27. Akaike H. A new look at statistical model identification. *IEEE Trans Automatic Control* 1974;19:716–723.
28. Skwarchuk MW, Jackson A, Zelefsky MJ, *et al.* Late rectal toxicity after conformal radiotherapy of prostate cancer (I): Multivariate analysis and dose–response. *Int J Radiat Oncol Biol Phys* 2000;47:103–113.
29. Dawson LA, Biersack M, Lockwood G, *et al.* Use of principal component analysis to evaluate the partial organ tolerance of normal tissues to radiation. *Int J Radiat Oncol Biol Phys* 2005;62:829–837.
30. Söhn M, Yan D, Liang J, *et al.* Influence of dose volume histogram (DVH) pattern on rectal toxicity [Abstract]. *Int J Radiat Oncol Biol Phys* 2005;63(Suppl. 1):S58–S59.

Appendix C

Predicting Grade 3 Acute Diarrhea during Radiation Therapy for Rectal Cancer Using a Cutoff-Dose Logistic Regression Normal Tissue Complication Probability Model

accepted for publication in the
International Journal of Radiation Oncology, Biology and Physics

Predicting Grade 3 Acute Diarrhea during Radiation Therapy for Rectal Cancer Using a Cutoff-Dose Logistic Regression Normal Tissue Complication Probability Model*

John M. Robertson, M.D.†
William Beaumont Hospital, Department of Radiation Oncology, 3601 W. Thirteen Mile Rd, Royal Oak, MI 48073

Matthias Söhn, Dipl.Phys.
Department of Radiotherapy, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany

Di Yan, D.Sc.
William Beaumont Hospital, Department of Radiation Oncology, 3601 W. Thirteen Mile Rd, Royal Oak, MI 48073

Purpose: Understanding the dose volume relationship of small bowel irradiation and severe acute diarrhea may help reduce the incidence of this side effect during adjuvant treatment for rectal cancer.

Methods and Materials: Consecutive patients treated curatively for rectal cancer were reviewed and the maximum grade of acute diarrhea was determined. The small bowel was outlined on the treatment planning CT scan and a dose volume histogram was calculated for the initial pelvic treatment (45 Gy). Logistic regression models were fitted for varying cutoff-dose levels from 5 Gy to 45 Gy in 5 Gy increments. The model with the highest LogLikelihood was used to develop a cutoff-dose normal tissue complication probability (NTCP) model.

Results: There were a total of 152 patients (48% preoperative, 47% postoperative, 5% other), predominantly treated prone (95%) with a three-field technique (94%), and a protracted venous infusion of 5-fluorouracil (78%). Acute grade 3 diarrhea occurred in 21%. The largest LogLikelihood was found for the cutoff-dose logistic regression model with 15 Gy as the cutoff-dose although the models for 20 Gy and 25 Gy had similar significance. According to this model, highly significant correlations ($p < 0.001$) between small bowel volumes receiving at least 15 Gy and toxicity exist in the considered patient population. Similar findings applied to both the preoperatively ($p = 0.001$) and postoperatively irradiated groups ($p = 0.001$).

Conclusion: The incidence of grade 3 diarrhea was significantly correlated with the volume of small bowel receiving at least 15 Gy using a cutoff-dose NTCP model.

Keywords: Rectal cancer, three dimensional treatment planning, acute toxicity, diarrhea, normal tissue complication probability

I. INTRODUCTION

Prospective randomized trials of adjuvant conventionally fractionated combined chemotherapy and radiation therapy (RT) for rectal cancer have found the incidence of severe acute diarrhea to be 12% to 39% for preoperative treatment (1-4) and 18% to 35% for postoperative treatment (1, 4-7). Severe acute diarrhea has contributed to the failure to complete therapy, either due to research defined criteria or patient preference (6, 7). Failure to complete therapy is less of a problem for preoperative treatment, reportedly only about 10% or less (1, 3, 4), but occurred in as many as 20% and 50% of patients enrolled in prospective postoperative treatment trials (1, 6).

Diarrhea in rectal cancer patients is multifactorial (8). In the preoperatively treated patients the presence of an unresected rectal cancer can cause altered bowel habits interpreted as diarrhea. In postoperatively treated patients, the surgery itself has been significantly associated with anal incontinence (9), which would be interpreted as grade 3 toxicity according to the Common Terminology Criteria (10). Irradiation of the rectum itself with essentially no RT given to the small bowel has been reported to occasionally cause diarrhea in patients treated for prostate cancer, presumably due to proctitis (11). Diarrhea is also a common side effect of chemotherapy alone, particularly in regimens including bolus 5-fluorouracil (5-FU) and irinotecan, with up to one third of patients experiencing grade 3 or 4 diarrhea (12). Newer agents, such as targeted therapies, have encountered severe acute diarrhea as the dose limiting toxicity in studies of preoperative treatment for rectal cancer (13).

Efforts to minimize the irradiated volume of small bowel have led to multiple well-known maneuvers, such as placing the patient in a prone position, cradles or "belly-boards" for small bowel

exclusion, treatment with a full bladder, and multiple fields for RT (14). Avoidance of the small bowel has also been used to justify studies of Intensity Modulated RT (IMRT) or Intensity Modulated Arc Therapy (IMAT) (15-18). Previous research has shown that the development of grade 3 diarrhea was significantly associated with the volume of small bowel irradiated for all dose levels (19, 20), although the two studies disagreed with regard to the postoperatively treated patients. In this study, a normal tissue complication probability (NTCP) analysis was applied to a larger group of patients for the incidence of grade 3 diarrhea in terms of a cutoff-dose Logistic Regression model in order to determine the small bowel volume above the respective cutoff-dose level most significant for developing and predicting toxicity.

II. METHODS AND MATERIALS

Consecutive patients treated with curative intent using pelvic RT for biopsy proven rectal cancer were reviewed after obtaining Human Investigation Committee permission. All patients were treated by a single radiation oncologist (JMR). Techniques to exclude the small bowel were routinely used if tolerated by the patient. Typically, this was a prone position on a customized vacuum bag cradle with rigid foam placed under the pubis and chest (21, 22). Patients were instructed in bladder distension techniques and told to have a full bladder for simulation, treatment planning CT scans and for treatment.

Between $1\frac{1}{2}$ to 2 hours before the treatment planning CT scan, patients were given diluted barium (300 ml) orally. Treatment planning CT scans were obtained with transverse images at 3 or 5 mm intervals through the pelvis only or the abdomen and pelvis using a helical CT scanner. These images were transferred to a Pinnacle treatment planning station (ADAC Laboratories, Milpitas, CA) and the entire volume of small bowel was outlined on every CT slice. This included both opacified and unopacified loops, as previous experience has shown that outlining only the opacified areas can fail to identify critical portions of small bowel (23).

Treatment typically used a posterior to anterior and opposed lateral field arrangement unless other field arrangements, such as addition of an anterior to posterior field or IMRT, were necessary to fulfill the homogeneity requirement. The minimum intended dose was 45 Gy at 1.8 Gy per fraction at the isocenter with a maximum inhomogeneity of 105%. All treatment fields used the same border definitions, described in detail before (19, 20) and identical to INT 0114 (6) and INT 0144 (24). The small bowel was routinely excluded from the lateral treatment fields after a dose of 45 Gy and treatment was typically given to at least 50.4 Gy.

The maximum diarrhea toxicity grade was assigned using the NCI Common Terminology Criteria v3.0 (Table 1) (10). Acute diarrhea was generally managed with oral antidiarrheal medications and avoidance of dietary fiber until improvement. Grade 3 diarrhea that failed to improve, was rapid in onset, or occurred at a low dose in high risk patients may also have been managed with field modification and/or a treatment break.

The intended small bowel dose for the initial 45 Gy was used in order to develop a predictive model. The absolute volume of the outlined small bowel in cubic centimeters was compared instead of the percentage as the entire small bowel was not imaged for all patients. Additionally, developing a model that required imaging and volume definition of the entire small bowel may be impractical for routine clinic use. The small bowel dose-volume histogram was calculated for the initial pelvic treatment to 45 Gy with the absolute volume of small bowel to be treated to each dose between 5 and 40 Gy reported at 5 Gy intervals, as was done in the previous experience (19, 20).

To find the most relevant dose level(s) predictive for acute grade 3 diarrhea, logistic regression as a standard method from statistics was used, which provides a phenomenological approach to NTCP modeling in terms of the so-called cutoff-dose logistic regression model (25). Here, V_{D_c} is defined as the volume of small bowel irradiated to doses equal to or above a cutoff-dose level D_c. For given value of D_c. the two-parametric logistic function

$$NTCP_{logistic}(V_{D_c}) = \frac{1}{1+\exp(-a-bV_{D_c})} \quad (1)$$

is used to test for correlation of V_{D_c} and toxicity in terms of a logistic regression with a and b as fit parameters. Fitting was performed for all possible values of D_c as available from the small bowel dose-volume histogram, i.e. all values in the range 5 to 45 Gy in increments of 5 Gy.

The method used to fit the models to sparse, dichotomous response data (zero or one, if the patient developed grade less than 3 or grade 3 diarrhea, respectively) was maximum likelihood estimation (25, 26). This method determined the optimal model parameters to maximize the probability of occurrence of the observed data, which is given by the so-called likelihood function L. Because of its smallness, numerically this is usually implemented as maximization of the natural logarithm ln(L), the LogLikelihood. In our implementation, the software package Mathematica version 5.0 (Wolfram Research Inc., Champaign, IL) was used.

Grade	Description
1	An increase of less than four stools per day over baseline or a mild increase in ostomy output compared to baseline.
2	An increase of four to six stools per day over baseline, intravenous fluids indicated for less than 24 hours, or a moderate increase in ostomy output compared to baseline. Not interfering with activities of daily living.
3	An increase of more than six stools per day over baseline, incontinence, intravenous fluids for 24 hours or more, hospitalization, or a severe increase in ostomy output compared to baseline. Symptoms interfere with activities of daily living.
4	Life-threatening consequences (e.g., hemodynamic collapse).

TABLE 1: Acute diarrhea grading as defined by the Common Terminology Criteria for Adverse Events (v3.0) (Adapted from 10)

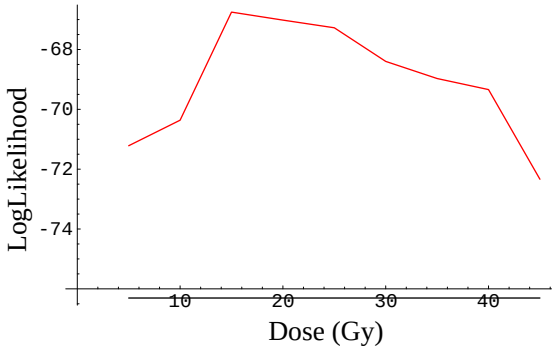

FIG. 1: LogLikelihood for the logistic regression models applied to the small bowel volume of all 152 patients at dose levels from 5 Gy to 45 Gy in 5 Gy increments. The model with the highest significance was found for a cutoff-dose of 15 Gy.

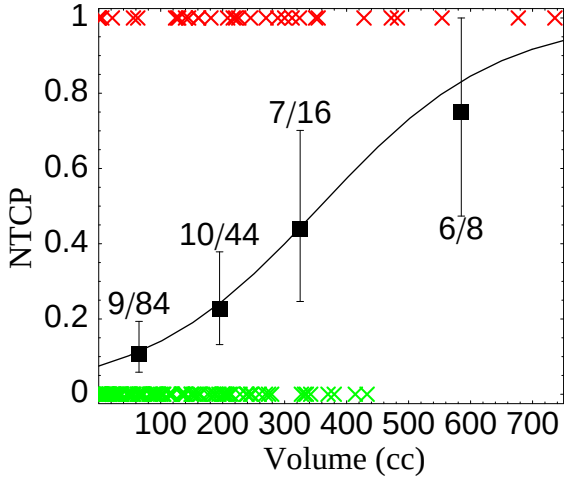

FIG. 2: The normal tissue complication probability (NTCP) of grade 3 diarrhea according to the cutoff-dose logistic regression model compared to the volume of small bowel receiving at least 15 Gy on the treatment planning dose-volume histogram. The red Xs along the top of the graph show the individual small bowel volumes of each patient with grade 3 diarrhea. The green Xs along the bottom show the volumes of patients without grade 3 diarrhea.

III. RESULTS

There were a total of 152 patients (Table 2). Treatment was given in a uniform fashion, with 95% of patients treated prone and 94% treated with a three-field technique. Preoperative treatment was given in 48%, postoperative in 47%, with 5% treated for a local recurrence or as primary treatment. All but two patients received concurrent chemotherapy, which was a continuous infusion of 5-FU in 78%, bolus 5-FU with leucovorin in 13%, capecitabine in 4% and bolus 5-FU in 2%.

Acute diarrhea was grade 0 in 32 patients (21%), grade 1 in 62 patients (41%), grade 2 in 26 patients (17%) and grade 3 in 32 patients (21%). There were no patients with grade 4 or 5 diarrhea.

According to the LogLikelihood values for models with different cutoff-doses at 5 Gy steps between 5 Gy and 45 Gy the cutoff-dose logistic regression model with highest significance was found for 15 Gy, although the models for 20 Gy and 25 Gy were of similar significance (Figure 1).

The cutoff-dose logistic regression model for the volume of small bowel receiving at least 15 Gy was investigated further. The range of small bowel volumes between the minimal and maximal volume was divided into 6 equally-sized bins for histogram construction. The 3 highest volume histogram bins had very few patients and were combined for a total of 4 bins overall (0 to 130 cc, 130 to 260 cc, 260 to 390 cc, and greater than 390 cc). This was used to construct a plot showing the incidence of grade 3 acute diarrhea for each small bowel volume receiving a dose equal to or above the cutoff-dose (Figure 2). When the volume of small bowel receiving 15 Gy or less was 130 cc or less, the NTCP was 11%. The correlation found for the data was highly significant ($p < 0.001$) with an area under

Characteristic	Number patients	
Treatment position	145	Prone
	7	Supine
Treatment technique	143	PA and opposed laterals
	6	AP, PA and laterals
	2	IMRT
	1	Opposed laterals
RT/Surgery Sequence	73	Preoperative treatment
	72	Postoperative treatment
	7	Other (wide local excision or locally recurrent)
Chemotherapy	119	Continuous infusional 5-FU (with leucovorin in 7; with irinotecan in 1)
	20	Bolus 5-FU with leucovorin
	8	Capecitabine
	3	Bolus 5-FU alone
	2	None

IMRT= Intensity Modulated Radiation Therapy
5-FU = 5-fluorouracil

TABLE 2: Treatment Characteristics

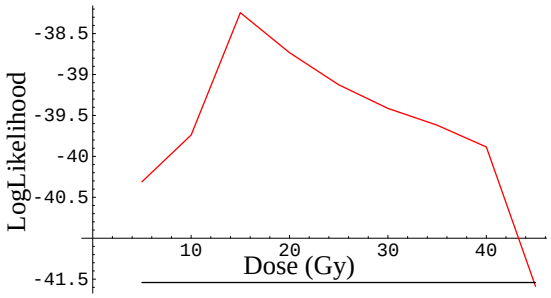

FIG. 3: As Figure 1, but only for postoperatively treated patients.

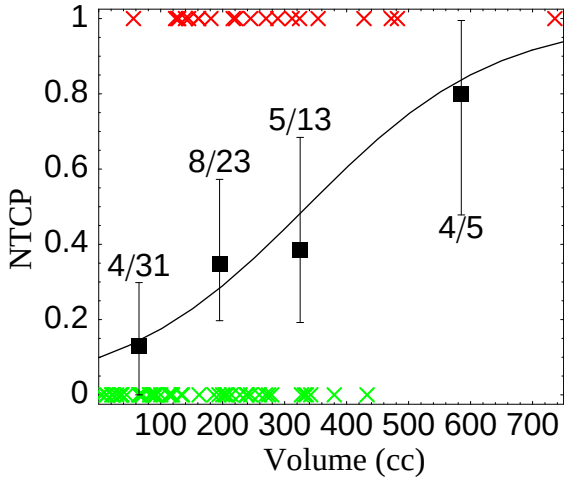

FIG. 4: As Figure 2, but only for postoperatively treated patients.

the receiver operating characteristic curve of 0.71.

The incidence of grade 3 diarrhea was 29% for postoperative treatment (21 of 72 patients) and 14% for preoperative treatment (10 of 73 patients) ($p = 0.03$). The LogLikelihood analysis for postoperative treatment found the highest significance for the 15 Gy dose level (Figure 3). The cutoff-dose logistic regression model at the 15 Gy dose level found a highly significant ($p = 0.001$) relationship (Figure 4). The LogLikelihood analysis for preoperative treatment found the highest significance for the 15 Gy, 20 Gy, and 25 Gy dose levels (Figure 5). The cutoff-dose logistic regression model at the 15 Gy dose level was highly significant ($p = 0.001$). As there were only 10 events in this group, the two largest dose bins (260 to 390 cc, and more than 390 cc) were combined into one bin of more than 260 cc (Figure 6).

IV. DISCUSSION

This was the largest study performed of the relationship of the irradiated small bowel volume and the development of grade 3 diarrhea (19, 20, 27-30). Aside from the total number of patients, this study also had the largest number of grade 3 events, al-

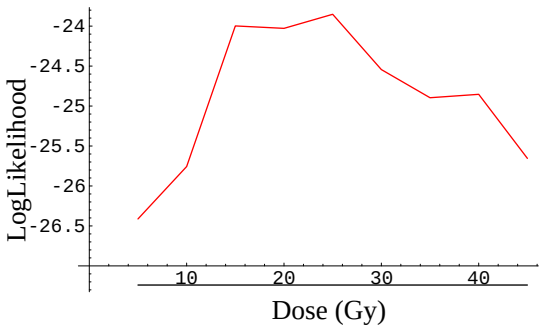

FIG. 5: As Figure 1, but only for preoperatively treated patients.

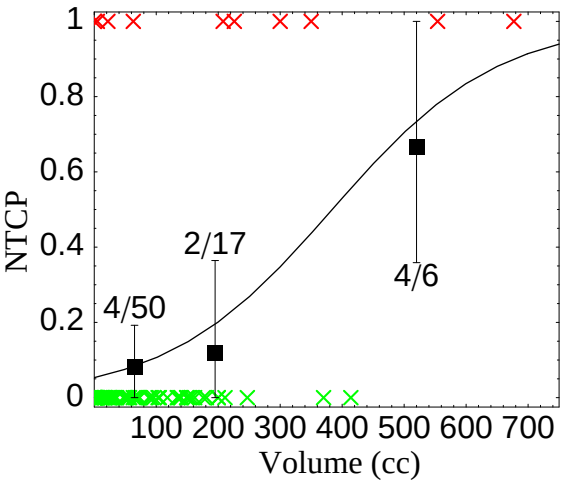

FIG. 6: As Figure 2, but only for preoperatively treated patients. Because there were fewer events than the data shown in Figure 2 and Figure 4, the largest two volume bins were combined into one bin (see text).

lowing the analysis of this endpoint, in contrast to other reports that included grade 2 diarrhea (27-30). Given the importance of grade 3 or higher toxicity in the conduct of clinical trials, with treatment held for grade 3 diarrhea in prospective randomized trials (6), this analysis was probably more appropriate. Although this was the largest number of grade 3 events in publications analyzing the dose-volume relationship of diarrhea, the incidence was in agreement with other trials of preoperative and postoperative therapy (1-7). Additionally, no patients had grade 4 or higher diarrhea, supporting the safety of the management of this patient group.

One particular advantage to this study was that the treatment approach was nearly uniform, with almost all patients treated prone with a three-field technique in a cradle for small bowel exclusion, and concurrent continuous infusional 5-FU. The weekly evaluation, toxicity assignment, and toxicity management were also consistent throughout, as all patients were managed by a single physician. These factors likely increased the uniformity of management and evaluation, reducing these confounding variables as much as possible. As such, the volumetric data determined here can probably be generalized, at least to rectal cancer. Certainly other chemotherapy combinations may alter the incidence and severity of diarrhea, but the volumetric relationship and the necessity to study the lower dose portions of the dose volume histogram would probably still exist.

The relationship between the three-dimensionally determined irradiated small bowel volume and the development of diarrhea has been well established (19, 20, 28, 30). Other than our own research (19, 20) which included patients reported here, two other studies have examined the relationship between three-dimensionally determined small bowel volume and acute diarrhea (28, 30). The first study included 41 patients treated preoperatively for rectal cancer (28) and substantively agreed with the study reported here. All patients were treated in a prone position to 45 Gy with concurrent bolus 5-FU with leucovorin during the first and fifth week. Thirteen patients had grade 2 (9 of 41) or grade 3 (4 of 41) acute diarrhea, using the Common Terminology Criteria v3.0. All small bowel volumes tested, at 5 Gy increments from 5 Gy to more than 42.75 Gy, were significantly associated with the maximum diarrhea scores, although the greatest significance was for the lower dose thresholds of 5 Gy through 30 Gy. The other study in rectal cancer was also supportive of the findings reported here, and included 28 patients treated in a supine position with concurrent 5-FU, leucovorin and oxaliplatin for T4 tumors or locally recurrent disease (30). The volumetric analysis was performed using grade 2 or higher diarrhea that occurred in 18 patients, and reported a relationship at the 15 Gy volume, with an incidence of diarrhea of 52% for patients with more than 150 cc of small bowel compared to only 11% for patients with 150 cc or less.

The utility of the 15 Gy dose level was solidified in the current report. In our previous work, we noted that significant dose volume relationships were found for each dose from 5 Gy to 40 Gy and that the dose volume relationship was nearly flat for the dose levels between 15 Gy and 40 Gy (19, 20). This was likely to be technique related, as the patients were nearly all irradiated using a three-field technique and the dose levels of 15 Gy and above are all linked. The LogLikelihood analysis performed here showed that 15 Gy had the highest significance, although the 20 Gy and 25 Gy dose volume levels were also important. Thus, studies analyzing the relationship of diarrhea and small

bowel volume should include the 15 Gy dose volume in the analysis.

The usefulness of the 15 Gy dose level as an avoidance structure for trials using IMRT or IMAT remains to be determined. A study of helical tomotherapy, without chemotherapy, used a similar dose volume parameter (15 Gy dose volume to 150 cc or less) derived from our previous data and encountered no grade 3 diarrhea, supporting this approach (31). However, because the dose levels are technique dependent, the recommendation to minimize the 15 Gy dose level is preliminary and additional clinical data using IMRT and/or IMAT will be necessary before safe dose and volume recommendations can be finalized.

It is important to recognize that this was not actually the V15, or volume of small bowel receiving at least 15 Gy. In order to derive the actual dose volume parameters, each patient with an event, such as grade 3 diarrhea, would have their individual DVH calculated for the dose and volume at the time point at which they manifested grade 3 diarrhea. This typically occurred at a median dose of 30.6 Gy with a range of 12.6 to 43.2 Gy in our previous analysis (20). The purpose here was not to investigate the mechanism of dose and small bowel toxicity, but to develop a predictive model using pretreatment parameters.

The incidence of grade 3 diarrhea was 29% in postoperatively treated patients which was somewhat higher than the overall 21% incidence reported in a prospective trial of RT with a protracted venous infusion of 5-FU (5). In our previous work, there was some disagreement on the value of small bowel DVH analysis for the postoperative patients. The initial report found a significant relationship of small bowel dose and volume in 40 patients, predominantly comprised of postoperative patients (19). In the follow up report, containing 39 postoperatively treated patients, the small bowel volumes were not statistically significantly different between patients with and without grade 3 diarrhea (20). The current study appears to have settled the issue, with a significant association in 72 patients.

The incidence of grade 3 diarrhea in preoperatively treated patients was 14%, in agreement with other preoperative studies (1-4). The observation of a significant relationship between irradiated small bowel volume and grade 3 diarrhea agreed with our previous work (20). This finding was somewhat unexpected, given the established mobility of the small bowel in preoperatively treated patients (23).

Patients with rectal cancer have a number of potential causes for diarrhea, specifically the tumor itself or rectal surgery, proctitis from RT, and chemotherapy. Studies of diarrhea in gynecologic cancers may be helpful, as there is no rectal tumor or rectal surgery, although RT is typically given to the whole pelvis, usually to a lower dose. Patients with cervix cancer in place should be analogous to the preoperative rectal cancer group as there is often no pelvic surgery, whereas patients with endometrial cancer and resected cervix cancer should be similar to the postoperative low anterior resection rectal cancer group as they are treated postoperatively. Two studies using three dimensionally calculated small bowel volumes for gynecologic cancer patient have been reported (27, 29). The first study analyzed 29 women with cervical cancer, 18 with endometrial cancer and 3 with other gynecologic cancers, approximately one half of whom received concurrent weekly cisplatin (27). All patients were treated in a supine position using IMRT. Grade 2 diarrhea was defined as use of frequent medications and occurred in 14 (28%) of the patients. The small bowel volume that received 25%, 50%, 75% and 100% (corresponding to absolute doses of 11 Gy, 22.5 Gy, 33.75 Gy and 45 Gy, respectively) was found to be nearly associated ($p = 0.06$) for the 25% and 50% volumes but highly significantly associated for the higher dose volumes ($p = 0.009$). Both prior surgery and chemotherapy were not related to the development of diarrhea. On multivariate analysis, the only statistically significant factor was the 100% volume of small bowel. The other study in gynecologic cancers came to different conclusions. This report analyzed the small bowel volume in 80 gynecologic cancer patients and divided them into a group of 49 patients without abdominal surgery and another group of 31 with abdominal surgery (29). Concurrent monthly cisplatin with infusional 5-FU was used in 36 patients in the first group and 14 in the second group, but was not significantly associated with the development of grade 2 or higher diarrhea. Treatment was given to the whole pelvis with the patient in a supine position. Using grade 2 plus 3 diarrhea (defined by the Common Toxicity Criteria v2.0), the low dose volume of small bowel, corresponding to the 16 Gy volume, was most significant for the patients without abdominal surgery, while the full dose volume was most significant for the patients with abdominal surgery. Thus, the gynecologic data was somewhat mixed, with the high dose volume significantly associated with toxicity for all patients in one study but not for postoperative patients in the other study.

An objective measurement of small bowel damage would be of great help in determining these relationships. One candidate is the plasma citrulline level, which is an amino acid that is not incorporated into proteins and is mainly produced by enterocytes of the small bowel (32). The potential role of plasma citrulline levels has been reported

for studies of benign conditions leading to shortened gut or reduced small bowel enterocyte mass as well as myeloablative therapy (33), although it has not been studied specifically for RT.

---

* Presented in part at the 49th Annual Meeting of the American Society for Therapeutic Radiology and Oncology, October 28 - November 1, 2007, Los Angeles, CA.

† Electronic address: jrobertson@beaumont.edu

[1] Sauer R, Becker H, Hohenberger W, *et al.* Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004;351:1731-1740.

[2] Broendengen M, Tveit KM, Berglund A, *et al.* Randomized phase III study comparing preoperative radiotherapy with chemoradiotherapy in nonresectable rectal cancer. *J Clin Oncol* 2008;26:3687-3694.

[3] Bosset J-F, Collette L, Calais G, *et al.* Chemotherapy with preoperative radiotherapy in rectal cancer. *N Engl J Med* 2006;355:1114-1123.

[4] Hyams DM, Mamounas EP, Petrelli N, *et al.* A clinical trial to evaluate the worth of preoperative multimodality therapy in patients with operable carcinoma of the rectum. A progress report of the National Surgical Adjuvant Breast and Bowel Project Protocol R-03. *Dis Colon Rectum* 1997;40:131-139.

[5] Miller RC, Sargent DJ, Martenson JA, *et al.* Acute diarrhea during adjuvant therapy for rectal cancer: A detailed analysis from a randomized intergroup trial. *Int J Radiat Oncol Biol Phys* 2002;54:409-413.

[6] Tepper JE, O'Connell MJ, Petroni GR, *et al.* Adjuvant postoperative fluorouracil-modulated chemotherapy combined with pelvic radiation therapy for rectal cancer: Initial results of Intergroup 0114. *J Clin Oncol* 1997;15:2030-2039.

[7] Wolmark N, Wieand HS, Hyams DM, *et al.* Randomized trial of postoperative adjuvant chemotherapy with or without radiotherapy for carcinoma of the rectum: National Surgical Adjuvant Breast and Bowel Project protocol R-02. *J Natl Cancer Inst* 2000; 92:388-96.

[8] Cherny NI. Evaluation and management of treatment-related diarrhea in patients with advanced cancer: A review. *J Pain Symptom Manage* 2008;36:413-423.

[9] van Duijvendijk P, Slors F, Taat CW, *et al.* A prospective evaluation of anorectal function after total mesorectal excision in patients with a rectal carcinoma. *Surgery* 2003;133:56-65.

[10] The National Cancer Institute Common Terminology Criteria for Adverse Events v3.0 (NCI CTCAE v3.0). Available at: http://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/ctcaev3.pdf. Accessed December 11, 2008.

[11] Nuyttens JJ, Milito S, Rust PF, *et al.* Dose-volume relationship for acute side effects during high dose conformal radiotherapy for prostate cancer. *Radiother Oncol* 2002;64:209-214.

[12] Maroun JA, Anthony LB, Blais N, *et al.* Prevention and management of chemotherapy-induced diarrhea in patients with colorectal cancer: a consensus statement by the Canadian Working Group on Chemotherapy-Induced Diarrhea. *Curr Oncol* 2007;14:13-20.

[13] Czito BG, Willett CG, Bendell JC, *et al.* Increased toxicity with gefitinib, capecitabine, and radiation therapy in pancreatic and rectal cancer: Phase I trial results. *J Clin Oncol* 2006;24:656-662.

[14] Gunderson LL, Russell AH, Llewellyn HJ, *et al.* Treatment planning for colorectal cancer: Radiation and surgical techniques and value of small-bowel films. *Int J Radiat Oncol Biol Phys* 1985; 11:1379-1393.

[15] Nuyttens JJ, Robertson JM, Yan D, *et al.* The influence of small bowel motion on both a conventional three-field and intensity modulated radiation therapy (IMRT) for rectal cancer. *Cancer/Radiotherapie* 2004; 8:297-304.

[16] Guerrero Urbano MT, Henrys AJ, Adams EJ, *et al.* Intensity-modulated radiotherapy in patients with locally advanced rectal cancer reduces volume of bowel treated to high dose levels. *Int J Radiat Oncol Biol Phys* 2006; 65:907-916.

[17] Kim JY, Kim DY, Kim TH, *et al.* Intensity-modulated radiotherapy with a belly board for rectal cancer. *Int J Colorectal Dis* 2007; 22:373-379.

[18] Duthoy W, De Gersem W, Vergote K, *et al.* Clinical implementation of intensity-modulated arc therapy (IMAT) for rectal cancer. *Int J Radiat Oncol Biol Phys* 2004; 60:794-806.

[19] Baglan KL, Frazier RC, Yan D, *et al.* The dose-volume relationship of acute small bowel toxicity from concurrent 5-FU-based chemotherapy and radiation therapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2002;52:176-183.

[20] Robertson JM, Lockman D, Yan D, *et al.* The dose-volume relationship of small bowel irradiation and acute grade 3 diarrhea during chemoradiation for rectal cancer. *Int J Radiat Oncol Biol Phys* 2008; 70:413-418.

[21] Shanahan TG, Mehta MP, Bertelrud KL, *et al.* Minimization of small bowel volume within treatment fields utilizing customized "belly boards". *Int J Radiat Oncol Biol Phys* 1990; 19:469-476.

[22] Huh SJ, Lim DH, Ahn YC, *et al.* Effect of customized small bowel displacement system in pelvic irradiation. *Int J Radiat Oncol Biol Phys* 1998; 40:623-627.

[23] Nuyttens JJ, Robertson JM, Yan D, *et al.* The position and volume of the small bowel during adjuvant radiation therapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2001; 51:1271-1280.

[24] Smalley SR, Benedetti JK, Williamson SK, *et al.* Phase III trial of fluorouracil-based chemotherapy regimens plus radiotherapy in postoperative adjuvant rectal cancer: GI INT 0144. *J Clin Oncol* 2006; 24:3542-3547.

[25] Söhn M, Yan D, Liang J, *et al.* Incidence of late rec-

tal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* 2007;67:1066-1073.

[26] Jackson A, Haken RKT, Robertson JM, *et al.* Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys* 1995;31:883-891.

[27] Roeske JC, Bonta D, Mell LK, *et al.* A dosimetric analysis of acute gastrointestinal toxicity in women receiving intensity-modulated whole-pelvic radiation therapy. *Radiother Oncol* 2003;69:201-207.

[28] Tho LM, Glegg M, Paterson J, *et al.* Acute small bowel toxicity and preoperative chemoradiotherapy for rectal cancer: Investigating dose-volume relationships and role for inverse planning. *Int J Radiat Oncol Biol Phys* 2006;66:505-513.

[29] Huang E-Y, Sung C-C, Ko S-F, *et al.* The different volume effects of small-bowel toxicity during pelvic irradiation between gynecologic patients with and without abdominal surgery: A prospective study with computed tomography-based dosimetry. *Int J Radiat Oncol Biol Phys* 2007;69:732-739.

[30] Gunnlaugsson A, Kjellen E, Nilsson P, *et al.* Dose-volume relationships between enteritis and irradiated bowel volumes during 5-fluorouracil and oxaliplatin based chemoradiotherapy in locally advanced rectal cancer. *Acta Oncol* 2007;46:937-944.

[31] De Ridder M, Tournel K, Van Nieuwenhove Y, *et al.* Phase II study of preoperative helical tomotherapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2008;70:728-734.

[32] Crenn P, Messing B, Cynober L. Citrulline as a biomarker of intestinal failure due to enterocyte mass reduction. *Clin Nutr* 2008;27:328-339.

[33] Lutgens LCHW, Blijlevens NMA, Deutz NEP, *et al.* Monitoring myeloablative therapy-induced small bowel toxicity by serum citrulline concentration. A comparison with sugar permeability tests. *Cancer* 2005;103:191-199.

Appendix D

Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation

published in
Physics in Medicine and Biology 2005; **50**(24): 5893–5908

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **50** (2005) 5893–5908 doi:10.1088/0031-9155/50/24/009

Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation

M Söhn[1], **M Birkner**[1], **D Yan**[2] **and M Alber**[1]

[1] Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany
[2] Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, Michigan 48073, USA

E-mail: matthias.soehn@med.uni-tuebingen.de

Received 14 June 2005, in final form 18 October 2005
Published 7 December 2005
Online at stacks.iop.org/PMB/50/5893

Abstract

We present a method of modelling inter-fractional organ deformation and correlated motion of adjacent organ structures in terms of so-called eigenmodes. The method is based on a principal component analysis (PCA) of organ shapes and allows for reducing the large dimensionality of geometry information from multiple CT studies to a few-parametric statistical model of organ motion and deformation. Eigenmodes are 3D vectorfields of correlated displacements of the organ surface points and can be seen as fundamental 'modes' of the patient's geometric variability. The amount of variability represented by the eigenmodes is quantified in terms of corresponding eigenvalues. Weighted sums of eigenmodes describe organ displacements/deformations and can be used to generate new organ geometries. We applied the method to four patient datasets of prostate/rectum/bladder with $N = 15$–18 CTs to assess interfractional geometric variation. The spectrum of eigenvalues was found to be dominated by only few values, indicating that the geometric variability of prostate/bladder/rectum is governed by only few patient specific eigenmodes. We evaluated the capability of this approach to represent the measured organ samples by calculating the residual errors for the organ surface points, using a varying number of eigenmodes. The distribution of residual errors shows fast convergence with the number of eigenmodes. Using 4 dominating modes, the range of residual errors for the four patients was 1.3–2.0 mm (prostate), 1.4–1.9 mm (rectum) and 1.5–1.9 mm (bladder). Thus, individual geometric variation taken from multiple imaging data can be described accurately by few dominating eigenmodes, thereby providing the most important factors to characterize deformable organ motion, which can assist adaptive radiotherapy planning.

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

1. Introduction

Modern radiotherapy approaches aim to tailor the high dose region to a 3D target volume with high accuracy. However, the quality and toxicity of the actual treatment are compromised by geometrical uncertainties in the treatment process due to motion of the target and its adjacent organs relative to the treatment beams. These uncertainties result from patient set-up errors, i.e. shifts or rotations of the patient's bony structures or external markers relative to the geometry in the planning CT scan, and internal organ motion, i.e. geometric changes relative to the bony structures or external markers. Although the target errors can be reduced significantly with image based set-up methods, the steep dose gradient around the target and inevitable variability of the adjacent organs at risk necessitate quantitative characterization of the remaining uncertainties for planning and evaluation. Rigorous modelling of internal organ motion and deformations is complex and has been addressed only by a few authors.

To simulate internal organ motion including deformations, Mageras *et al* (1996) sampled deformations taken from a database of multiple CT studies and mapped these to the organ contours of the planning CT scan of an individual patient, thereby realizing a statistical population based bootstrap model to study the effects of internal organ motion on treatment parameters. As an obvious shortcoming of this approach only deformations observed in the database of patients can be reproduced. Hoogeman *et al* (2002) simulated rectum motion by deforming the rectum shape of the planning CT scan on a slice-by-slice basis with corresponding probability distributions that were found for a population of reference patients. This allows generation of deformations not present in the group of reference patients on a statistical basis. However, the developed method is limited to studying the influence of rectum shape changes on treatment parameters. Both of the latter two methods model organ shape variability on a population basis, while individual patient characteristics enter the models merely through the organ geometries of the planning CT scan.

Individual organ shape changes as observed in multiple (pre-)treatment CT images were addressed by Yan *et al* (1999), who used a finite element (FE) method incorporating biomechanical properties of the tissues to calculate the displacements of the organ tissue elements for the different CT scans. These displacement fields were used both to reconstruct the dose delivered in the first few fractions and to predict the distribution of the patient specific organ motion. The latter was used to estimate the cumulative dose in the subsequent fractions and adapt the treatment plan accordingly. For the prediction process intermediate geometries were generated based on interpolation between all pairs of measured (pre-) treatment geometries. However, this implicitly implies the assumption of a uniform statistical distribution between the measured geometries. The validity of this approximation has not been investigated by the authors.

So far the most general statistical framework to model deformable organ motion for radiotherapy was proposed by Fontenla *et al* (2001a). Using a set of appropriate parameters to describe observed organ shape changes, the method is based on a non-parametric statistical density estimation of the underlying multivariate probability distribution. The theoretical framework for incorporating both population based and individual deformation as well as set-up information was developed. In an accompanying paper (Fontenla *et al* 2001b) the method was applied to data of prostate cancer patients to model individual organ deformation (without using population based data). Though theoretically appealing, this approach suffers from the problem of slow convergence of the non-parametrically estimated distribution to the 'true' underlying probability distribution if the number of parameters used to describe organ shape changes is too large. Consequently, in the example only 6–12 parameters per organ were used to describe the geometric changes of prostate/bladder/rectum via affine

transformations. The errors introduced by this approximation were not quantified by the authors.

All approaches cited above have the shortcoming that they apply only to single organs of interest, while correlated motion of adjacent organs structures is not taken into account. However, with respect to applications in 4D planning and evaluation this is desirable and thus addressed in this work. We use a method from multivariate statistics, *principal component analysis* (PCA), to quantitatively determine the most important modes of correlated motion and deformation present in a series of (pre-)treatment organ geometries, which were parametrized by sets of corresponding organ surface points. This approach provides a few-parametric statistical model of individual organ deformation. In this paper we give an introduction to the method and test the implementation for data of prostate cancer patients. The accuracy of the PCA organ model to represent the measured organ geometries is evaluated.

2. Theory

2.1. Parametrization of organ geometries

We consider inter-fractional internal organ motion/deformation as given by a series of N (pre-)treatment CT scans of an individual patient. In the following, organ motion/deformation is understood as change of organ geometry as given by its shape. Organ shape is parametrized by the set of positions of M organ surface points: if $\vec{\mathbf{x}}_\mathbf{j}(i)$ denotes the position of the jth point in the ith CT, then the *surface shape vector*

$$\boldsymbol{p}_i = (\vec{\mathbf{x}}_\mathbf{1}(i), \ldots, \vec{\mathbf{x}}_\mathbf{M}(i)) \in \Re^{3M} \tag{1}$$

represents the geometry of the organ in the ith CT ($i = 1, \ldots, N$).

In case simultaneous motion/deformation of more than one organ is to be modelled, the surface shape vectors $\boldsymbol{p}_i^{\text{organ } A} \in \Re^{3M_A}$, $\boldsymbol{p}_i^{\text{organ } B} \in \Re^{3M_B}$, etc, are concatenated to form a vector of dimension $3(M_A + M_B + \cdots)$, which represents the composite geometry of the organs under consideration.

In the following, it is important that each $\vec{\mathbf{x}}_\mathbf{j}$ in (1) corresponds to the same *anatomical* position in each of the N CTs, i.e., the series of *corresponding positions* $\vec{\mathbf{x}}_\mathbf{j}(1), \ldots, \vec{\mathbf{x}}_\mathbf{j}(N)$ provides anatomical point-tracking information. Methods for generation of such datasets based on (pre-)treatment CT scans have been published elsewhere (Yan *et al* 1999) and are not part of the organ model presented in the following. However, we refer to this issue below in section 4.

2.2. Statistical model of organ geometric change based on PCA

It is assumed that the set of surface shape vectors $\{\boldsymbol{p}_i\}_{i=1,\ldots,N}$ can be seen as samples from a random process. Obviously, for anatomical reasons displacements of the M surface points due to organ motion and deformation are highly correlated, which implies that the underlying dimensionality of this multivariate statistical problem is actually much smaller than $3M$. Thus we are interested in correlated displacements of the surface points. For this $3M$-dimensional problem with N samples we use a method from multivariate statistics, principal component analysis (e.g. Manly 2004, Joliffe 2002). This type of approach is known as a *point-distribution model* (PDM) in the literature (Cootes *et al* 1994) and has been applied to problems like population modelling of anatomical shape variability (Lorenz and Krahnstöver 2000) or automatic 3D organ segmentation (Pekar *et al* 2004). In appendix A we present an instructive example of PCA for a two-dimensional dataset.

Decomposition of organ geometry samples into eigenmodes. Applied to the problem of organ geometric change, PCA works as follows. The first two moments of the probability distribution, i.e. the *mean shape vector* $\bar{\boldsymbol{p}} \in \Re^{3M}$ and the *covariance matrix* $\mathcal{C} \in \Re^{3M\times 3M}$, are calculated according to

$$\bar{\boldsymbol{p}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{p}_i \tag{2}$$

$$\mathcal{C} = \frac{1}{N-1}\sum_{i=1}^{N} (\boldsymbol{p}_i - \bar{\boldsymbol{p}}) \cdot (\boldsymbol{p}_i - \bar{\boldsymbol{p}})^T \tag{3}$$

$$= \frac{1}{N-1}\mathcal{P}\mathcal{P}^T. \tag{4}$$

Here the column vector $\boldsymbol{p}_i - \bar{\boldsymbol{p}}$ is called the *centred shape vector*, i.e. the vector describing the displacements of the geometry at time t_i relative to the the mean shape, and $(\,)\cdot(\,)^T$ denotes the outer product of these two $3M$-dimensional vectors. Forming a matrix $\mathcal{P} \in \Re^{3M\times N}$ from the N centered shape vectors, the covariance matrix can be rewritten to the compact form in (4).

In case the probability distribution which governs the assumed random process is a multivariate normal distribution, it is already uniquely characterized by these first two moments. For the problem of internal organ motion/deformation considered here, the exact type of probability distribution is unknown *a priori*. However, considering only the first two moments can still serve as an approximation, where the covariance matrix represents the organ geometric variability.

Correlations of the $3M$ variables are reflected by the existence of nonzero off-diagonal elements of the covariance matrix, implying that the probability distributions of the variables are not independent (see appendix A for a low-dimensional analogue). Diagonalization of the covariance matrix[3] results in *eigenvectors* $\boldsymbol{q}_l \in \Re^{3M}$, which represent statistically independent modes of deformation, the so-called *eigenmodes*. Under the assumption of a multivariate Gaussian distribution these eigenmodes approximately describe the deformation characteristics. Each eigenmode defines a 3D vectorfield of correlated displacements $\delta\vec{\mathbf{x}}_{\mathbf{j}} \in \Re^3$ for the M surface points: $\boldsymbol{q}_l = (\delta\vec{\mathbf{x}}_{\mathbf{1,l}}, \ldots, \delta\vec{\mathbf{x}}_{\mathbf{M,l}})$.

The eigenvectors give a new basis of the $3M$-dimensional parameter space, in which the assumed multivariate normal distribution decomposes into 1D Gaussian distributions along the directions of the eigenvectors. As in the 2D example presented in appendix A, quantitatively each *eigenvalue* λ_l is the statistical variance of the N measured geometry samples projected on the lth eigenvector as new basis vector

$$\sigma_l^2 = \lambda_l. \tag{5}$$

Construction of organ geometries using eigenmodes. The eigenvalues impose an importance ranking on the eigenmodes with respect to the representation of geometric variability. The dominant eigenmodes, i.e., the eigenmodes with largest eigenvalues, are the 'principal' deformation modes, which span the space in which the majority of deformations occur. New geometry samples can be generated by deforming the mean shape by a weighted sum of L dominating eigenmodes

[3] As the number of surface points M is typically in the order of 10^3, direct diagonalization of $\mathcal{C} \in \Re^{3M\times 3M}$ is numerically not feasible; in appendix B a method of efficient numerical implementation is described.

$$\boldsymbol{p} = \bar{\boldsymbol{p}} + \sum_{l=1}^{L} c_l \boldsymbol{q}_l, \qquad \|\boldsymbol{q}_l\| = 1. \tag{6}$$

According to the theory of PCA, the coefficients $c_l \in \Re$ obey Gaussian distributions with the corresponding eigenvalues as variances (cp equation (5)). Thus the dominating eigenmodes can serve as statistical model of individual organ motion/deformation with only a small number of parameters.

3. Quantitative evaluation

The capability of the PCA model to represent the measured organ geometry samples using only few dominating eigenmodes was evaluated. The steps of this procedure are described in the following.

Representation of measured geometries using the PCA-model. For each of the N measured organ geometries $\boldsymbol{p}_i$ (equation (1)), the coefficients c_l in equation (6) are determined so as to get an optimal representation $\boldsymbol{p}_{i,\mathrm{opt}}^{[L]}$ of the respective organ geometry using only L eigenmodes:

$$\boldsymbol{p}_{i,\mathrm{opt}}^{[L]} = \bar{\boldsymbol{p}} + \sum_{l=1}^{L} c_{l,\mathrm{opt}}(i)\boldsymbol{q}_l. \tag{7}$$

The optimal coefficients are supposed to minimize the difference $\left\|\boldsymbol{p}_{i,\mathrm{opt}}^{[L]} - \boldsymbol{p}_i\right\|$, i.e., $\boldsymbol{p}_{i,\mathrm{opt}}^{[L]}$ is given by the projection of the $3M$-dimensional vector $\boldsymbol{p}_i$ on the subspace spanned by the L eigenvectors $\boldsymbol{q}_l$ as basis vectors. The coefficients $c_{l,\mathrm{opt}}(i)$ are therefore calculated as the following scalar product:

$$c_{l,\mathrm{opt}}(i) = (\boldsymbol{p}_i - \bar{\boldsymbol{p}}) \cdot \boldsymbol{q}_l \qquad \text{with} \quad \|\boldsymbol{q}_l\| = 1; \quad l = 1, \ldots, L. \tag{8}$$

Quantization of shape similarity. As a measure for the quality of the representation of the $i = 1, \ldots, N$ measured organ geometries $\boldsymbol{p}_i$ by linear combinations of eigenmodes according to (7), we regarded the shape similarity of $\boldsymbol{p}_i$ and $\boldsymbol{p}_{i,\mathrm{opt}}^{[L]}$. This was quantized as an average local surface distance in the following way.

- For each of the $j = 1, \ldots, M$ surface points of $\boldsymbol{p}_{i,\mathrm{opt}}^{[L]}$, the minimal distance $d_{i,j}^{[L]}$ to the surface of the measured geometry $\boldsymbol{p}_i$ was determined. The search region was centred around the position of the corresponding point $\vec{\mathbf{x}}_{\mathbf{j}}(i)$ (see equation (1)) and limited by the direct neighbours of $\vec{\mathbf{x}}_{\mathbf{j}}(i)$ on the measured surface. $d_{i,j}^{[L]}$ is the *local representation error* or *local residual*.
- For each surface point the results for the N different geometries were averaged, giving the value of the *average local residual* $\bar{d}_j^{[L]}$ for each point:

$$\bar{d}_j^{[L]} = \frac{1}{N} \sum_{i=1}^{N} d_{i,j}^{[L]}, \qquad j = 1, \ldots, M \tag{9}$$

  A histogram of these values gives an overview of the overall quality of representation. Moreover, a colour-coded projection of the values $\bar{d}_j^{[L]}$ onto the organ surface(s) allows localization of the parts of the organ with systematically increased residuals.
- Additionally, the standard deviations $\Delta d_j^{[L]}$ of the local residuals were calculated and visualized correspondingly. A locally high standard deviation together with non-vanishing average local residual reveals the spatial regions of geometric variability that is not described properly by using only L dominating eigenmodes for the PCA model.

- Finally, the average of $\bar{d}_j^{[L]}$ over all M surface points, the *overall* (*mean*) *residual*

$$R^{[L]} = \frac{1}{M} \sum_{j=1}^{M} \bar{d}_j^{[L]}, \tag{10}$$

  as well as the corresponding standard deviation can be calculated as comprehensive measure for the quality of the PCA-model with L eigenmodes.

Convergence characteristics. The overall residual $R^{[L]}$ as defined in (10) was calculated for a varying number L of eigenmodes. This is done separately for each organ under consideration to get organ and patient specific convergence characteristics, which addresses the question of how many dominating eigenmodes are needed to give a sufficient model of individual geometric variability.

Residuals at organ interfaces. With regard to requirements of treatment planning, the quality of modelling correlated geometric variability at organ interfaces, which are usually located in regions with steep dose gradient (e.g. the prostate/rectum- and prostate/bladder-interface), is of special interest. Therefore we define the *organ interface residual* $\tilde{R}^{[L]}$ analog to (10) as mean value of the average local residuals $\bar{d}_j^{[L]}$ of surface points of two adjacent organs in the interface region. For our study, the points on both respective organ surfaces were chosen following the criterion that the distance to the other organ was $\leqslant 4$ mm (based on the mean geometry $\bar{\boldsymbol{p}}$).

4. Materials

We applied the method to datasets of four prostate cancer patients, who were treated at the William Beaumont Hospital in supine position. Each patient received $N = 15$–18 CT scans. The planning CT scan was obtained with 20 cc urethra contrast. The repeat CT scans were aquired on the first four treatment days and subsequently twice a week immediately before or after treatment delivery.

The repeat CT scans were matched on the planning CT scan using bony anatomy, thereby reducing the geometric variability, which is to be described by the PCA model, to internal organ motion/deformation. The contours of prostate, bladder and rectum were manually delineated on a slice-by-slice basis by the same physician for each patient CT-dataset. Contours of the rectum start at the anus and end at the position at which the rectum turns into the sigmoid colon.

The sets of corresponding points, forming the surface shape vectors of the organ geometries as described in section 2.1, were taken from datasets which were generated by the *finite element method* (FEM) from Yan *et al* (1999) applied to the above CT/contour data. This 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 described in Liang and Yan (2003), the (initial) boundary point correspondence was determined by assuming that the points were uniformly distributed along the contours starting from the most posterior borderline of each organ. Optionally, the positions of the boundary points can be optimized by iteratively reducing the energy of the deformed organ mesh between repeated FEM steps, thereby improving the registration accuracy (see Liang and Yan 2003). However, for the data used in this study the latter, numerically intense step has not been performed.

Extraction of the tissue elements on the organ surface typically resulted in several thousand surface points. To ensure a constant surface density of points for all organs under consideration,

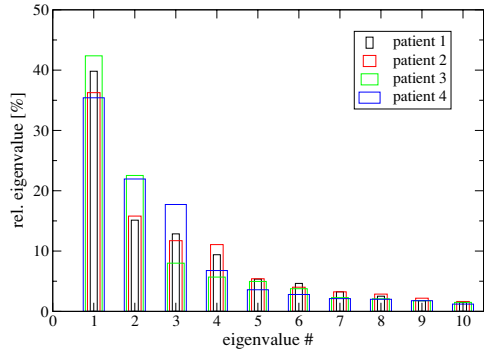

Figure 1. Spectra of eigenvalues for the four patient datasets (relative values: sum of all eigenvalues normalized to 100%).

the number of surface points was reduced accordingly, if necessary. For the four patients considered in this study this resulted in $M = \sim$350–680 surface points for prostate, ∼740–960 for rectum (outer surface) and ∼640–820 for bladder. Finally, these point datasets were triangulated, which allowed 3D visualization of the organ surfaces.

5. Results

Performing a simultaneous PCA for the ∼1500–2000 surface points used to describe the geometry of all three organs takes a few seconds on a 2 GHz PC.

The spectra of eigenvalues are shown in figure 1. For all four patients these are clearly dominated by only a few eigenvalues, indicating that the geometric variability of the measured organ samples is concentrated in few deformation modes only.

Figure 2 gives an exemplary visualization of the first dominating eigenmodes of one of the patients. This patient had large bladder motion/deformation present in the measured geometry samples, which also manifests itself in the deformation fields given by the dominating eigenmodes. In the first two eigenmodes, bladder deformation is associated with correlated movement and deformation of prostate and rectum, whereas in the third eigenmode prostate undergoes only minor geometric changes. Thus for this patient the third eigenmode models bladder variability which is uncorrelated to motion of the prostate, whereas the first two modes can be interpreted as 'bladder-filling driven' correlated motion/deformation of all three organs.

The character of the single eigenmodes differs significantly for different patients (data not shown) and depends on the individual overall geometric variability of a patient and the relative magnitude of variability among the organs. Thus a general association of a specific eigenmode with variability of a specific organ is not possible.

Figures 3 and 4 show detailed data of prostate/rectum/bladder residuals for the PCA-models of two different patients, where the number L of dominating modes used to represent the measured organ geometries was varied. Both the histograms of the average local residuals $\bar{d}_j^{[L]}$ as defined in equation (9) and the values of the overall residuals $R^{[L]}$, equation (10), show enhanced decrease with the first few dominating eigenmodes, where the order of magnitude of the residuals and the convergence characteristics is patient and organ specific.

Of the three organs under consideration, the geometric variability of bladder is the largest, as can be seen in diagrams (f), where the value of $R^{[0]}$ is a measure for the variability present in the dataset itself (for $L = 0$ effectively the differences of the measured samples to the mean

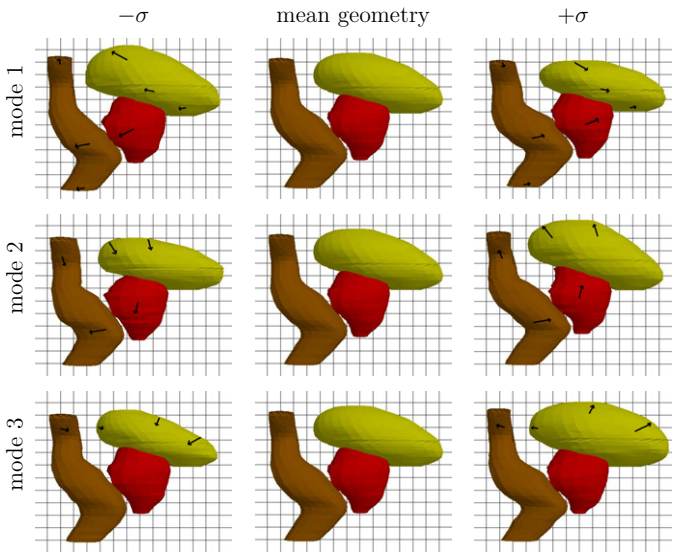

Figure 2. Visualization of the displacement fields given by the first three eigenmodes of patient 4 (first eigenmode in first row etc). Shown are sagittal 3D views of prostate (red), rectum (brown) and bladder (yellow). The centre column gives 3D views of the mean geometry $\bar{\boldsymbol{p}}$, whereas the left/right colums are deformations of the mean geometry by the respective normed eigenvector $\boldsymbol{q}_l$ according to $\boldsymbol{p} = \bar{\boldsymbol{p}} \pm \sigma_l \cdot \boldsymbol{q}_l$ with the variance σ_l^2 given by the eigenvalue λ_l. The arrows illustrate the direction and relative magnitude of the local surface motion/deformation. Input of the prostate/rectum/bladder-PCA model for patient 4 were $N = 18$ geometry samples; number of surface points: $M_{\text{prostate}} = 629$, $M_{\text{rectum}} = 750$ and $M_{\text{bladder}} = 613$. (Animations available as multimedia attachments in online version).

shape are calculated). However, $R^{[L]}_{\text{bladder}}$ reduces significantly to a level of $\lesssim 2$ mm when using only 3–4 eigenmodes to model the variability.

Modelling rectum variability is more complex, which becomes evident in the comparatively slower convergence characteristics (figures 3 and 4 (c) and (d)). Nevertheless, the overall residual also is in the order of $\lesssim 2$ mm for $L = 4$.

Individual differences in organ variability can be illustrated by means of figures 3 and 4 (a) and (b) for prostate. Patient 1 shows only small prostate motion, thus already the mean shape ($L = 0$) fairly represents most of the sample geometries, and the overall residual does not significantly reduce further for an increasing number of modes L. Contrary to this, patient 4 exhibits large prostate variability. However, this patient shows fast convergence of $R^{[L]}$ for prostate: the variability can essentially be represented already by the first two eigenmodes (cp figure 4(b)). This agrees with the observation from figure 2 that the prostate variability represented by the third eigenmode of patient 4 is of minor magnitude only.

Table 1 summarizes the results of prostate, rectum and bladder for all four patients: using $L = 4$ dominating eigenmodes, all overall residuals are in the range of 1.3–2.0 mm. The residuals $\tilde{R}^{[L]}$ for surface points in the interface regions of prostate/rectum and prostate/bladder, which give a measure for the capability of a multi-organ PCA model to accurately represent correlated geometric variability of these adjacent organs, are shown in table 2 for all four patients: using $L = 4$ dominating eigenmodes, these residuals are in the range of 0.8–1.9 mm (1.3–2.2 mm) for points on the prostate/rectum- (prostate/bladder-) interface, i.e. in the order of the overall residuals for the respective organs.

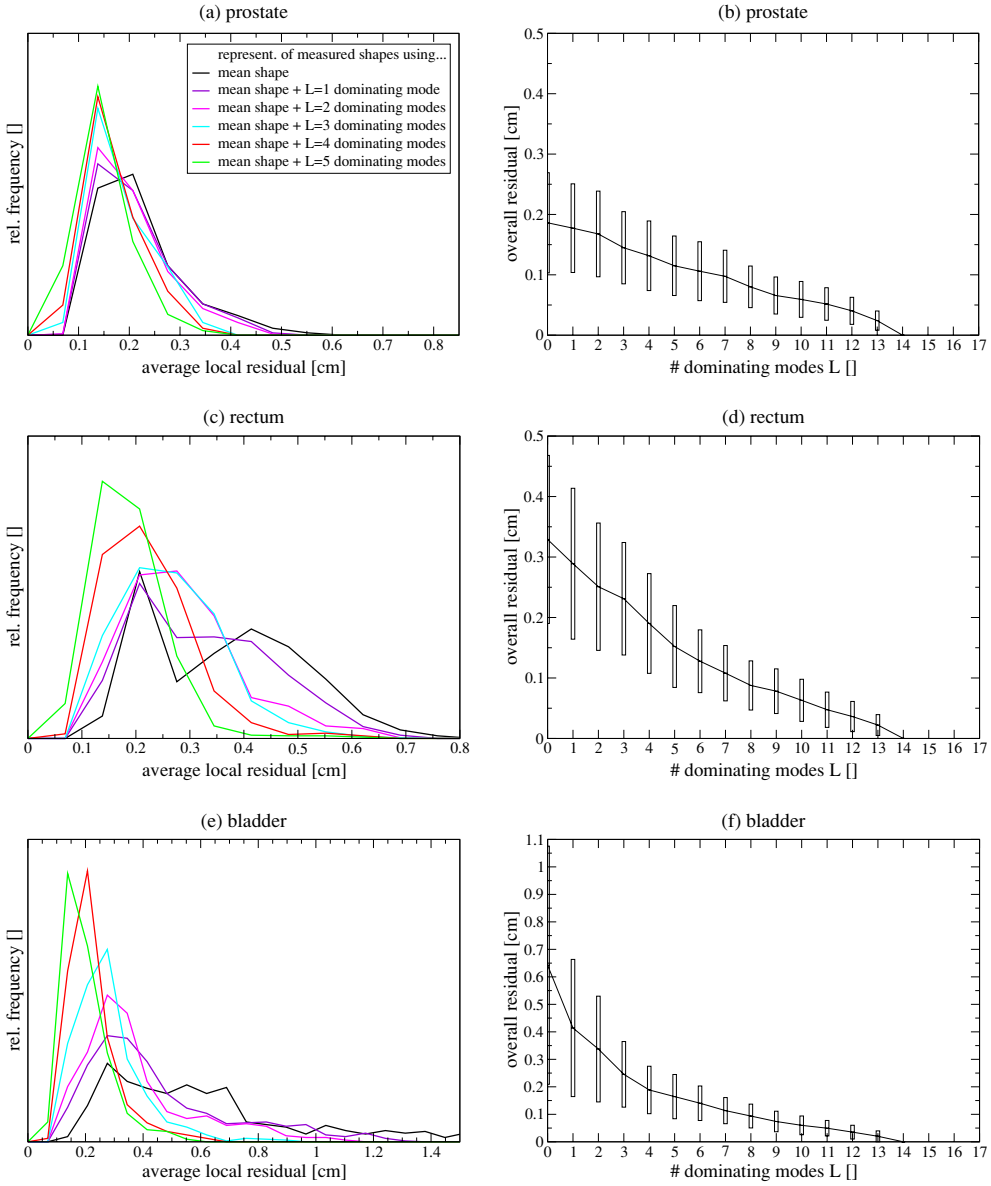

Figure 3. Patient 1: quality of representation of measured geometries using L dominating eigenmodes for the PCA-model. The left column shows histograms of the average local residuals $\bar{d}_j^{[L]}$, equation (9), of prostate ($M = 378$ surface points), rectum ($M = 950$) and bladder ($M = 726$) for $L = 0, \ldots, 5$. Right column: values of the overall residual $R^{[L]}$, equation (10), for varying L; the bars represent the corresponding standard deviations. Input of the prostate/rectum/bladder-PCA model for patient 1 were $N = 15$ geometry samples.

This remaining amount of modelling error is due to the fact that higher eigenmodes were ignored. The residuals mainly reflect the uncorrelated errors/artefacts ('noise') in the dataset, as discussed below. The magnitude of this 'noise' is in the order of $\sim$1 mm as can be estimated by the convergence characteristics of the overall residual $R^{[L]}$ for $L \to N - 1$: it is essentially

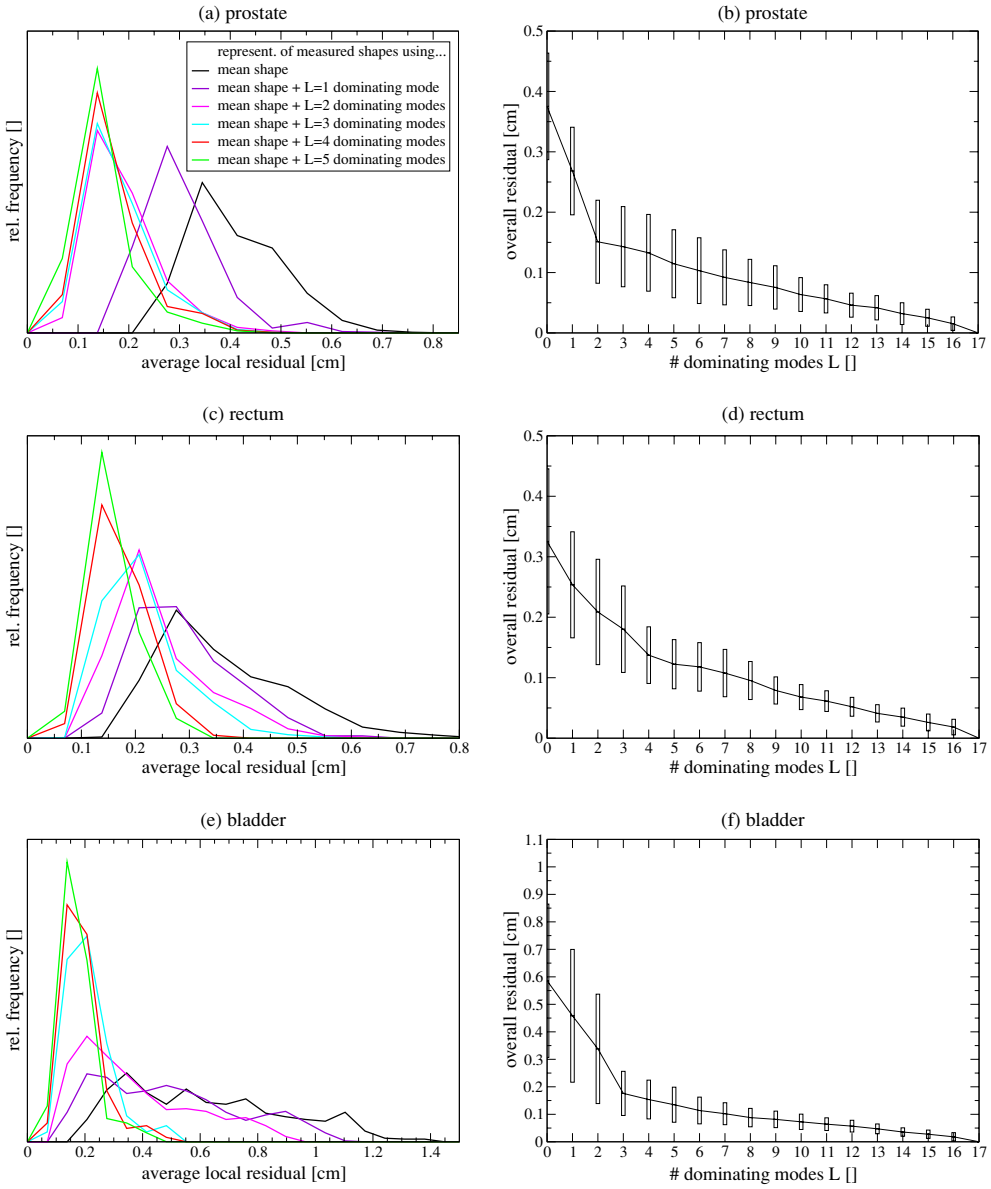

Figure 4. Same as figure 3 with data of patient 4 (number of organ surface points: $M_{\text{prostate}} = 629$, $M_{\text{rectum}} = 750$ and $M_{\text{bladder}} = 613$; $N = 18$ geometry samples used for PCA). Patient 4 has large variability of prostate, which can be described using the first two dominating modes.

a linear decrease for all organs, which is a convergence characteristics to be expected for such 'noise' in the dataset. Another reason is the finite resolution of the organ shape definition (slice distance of adjacent contours as drawn by a physician: 1.5 mm for prostate, 3 mm for rectum/bladder).

Typically, the eigenmodes with small eigenvalue, i.e. modes of higher order ($\gtrsim 5$, patient dependent), show unphysical geometric variability like adverse displacements of adjacent points or organ slices with small magnitude (data not shown). These modes model

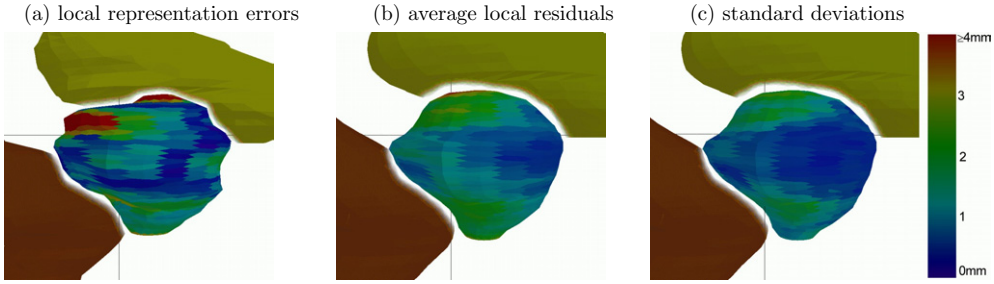

Figure 5. Colour-coded projections of residual data on prostate surface of patient 4 (3D view), $L = 3$ dominating eigenmodes used for PCA-model. (a) Local representation errors $d_{i,j}^{[L]}$ for geometry $i = 14$ (as contoured in CT 14). (b) Average local residuals $\bar{d}_j^{[L]}$, equation (9), projected to the mean shape of prostate. (c) Corresponding standard deviations $\Delta d_j^{[L]}$ of the local representation errors. (Note that to prevent partial covering of prostate in 3D view the adjacent parts of rectum (brown) and bladder (yellow) were clipped.)

Table 1. Overall residuals $R^{[0]}$ and $R^{[4]}$, equation (10), of the organs for all four patients. $R^{[0]}$ is a measure for the geometric variability of the respective organs present in the patient datasets, while $R^{[4]}$ is the overall residual when using 4 dominating eigenmodes for a prostate/rectum/bladder PCA-model. (values in [mm]).

	$R^{[0]}_{\text{prostate}}$	$R^{[0]}_{\text{rectum}}$	$R^{[0]}_{\text{bladder}}$	$R^{[4]}_{\text{prostate}}$	$R^{[4]}_{\text{rectum}}$	$R^{[4]}_{\text{bladder}}$
Patient 1	1.9 ± 0.8	3.3 ± 1.4	6.4 ± 4.3	1.3 ± 0.6	1.9 ± 0.8	1.9 ± 0.9
Patient 2	2.9 ± 0.9	4.2 ± 2.1	4.3 ± 2.8	2.0 ± 0.6	1.8 ± 0.7	1.8 ± 0.7
Patient 3	2.2 ± 0.7	2.6 ± 0.8	5.5 ± 2.6	1.3 ± 0.4	1.5 ± 0.4	1.8 ± 0.6
Patient 4	3.8 ± 0.9	3.3 ± 1.2	5.9 ± 2.8	1.3 ± 0.6	1.4 ± 0.5	1.5 ± 0.7

Table 2. Organ interface residuals $\tilde{R}^{[0]}$ and $\tilde{R}^{[4]}$ for organ surface points on the prostate/rectum resp. prostate/bladder interface (average number of points: 69 (range 52–94) for the prostate/rectum- and 127 (range 101–148) for the prostate/bladder-interface). (values in [mm]).

	$\tilde{R}^{[0]}_{\text{pros-rect}}$	$\tilde{R}^{[0]}_{\text{pros-blad}}$	$\tilde{R}^{[4]}_{\text{pros-rect}}$	$\tilde{R}^{[4]}_{\text{pros-blad}}$
Patient 1	1.8 ± 0.6	2.9 ± 0.8	0.8 ± 0.2	2.0 ± 0.5
Patient 2	3.2 ± 0.5	3.7 ± 1.3	1.9 ± 0.4	2.2 ± 0.8
Patient 3	2.6 ± 0.4	2.9 ± 1.0	1.2 ± 0.3	1.6 ± 0.5
Patient 4	4.9 ± 0.5	3.8 ± 0.7	1.1 ± 0.3	1.3 ± 0.5

(pseudo-)correlations among the surface point displacements which are present in the input dataset due to inevitable uncorrelated errors in establishing point-correspondence for the surface points when initializing the boundary conditions of the FE-model. Another reason are input data artefacts stemming from the process of manual contouring of the organ structures on a slice-by-slice basis, which—for some of the organ shape samples—results in locally highly non-smooth, non-anatomical surface shapes. As modelling of this type of (pseudo) 'geometric variability' is not desirable, these 'noise-modes' of higher order should be ignored for a PCA-model.

As an example, figure 5 visualizes the spatial distribution of residual errors for prostate (patient 4), when $L = 3$ dominating eigenmodes are used to represent measured geometries. Figure 5(a) shows the geometry as contoured in CT 14 with the colour-coded local representation errors ('local residuals'), i.e. the local differences to the optimal representation of this geometry; see equation (7). For this example the surface points with

largest representation error are concentrated in a region where the transversal contour(s) have local contouring errors, as visible in the depicted 3D view. Figures 5(b) and (c) give the spatial distributions of the average local residual and its standard deviation, thereby summarizing the statistics of the local representation errors for all measured geometries ($N = 18$ for patient 4). Again the surface points with increased average local residual are concentrated in certain regions which roughly coincide with the regions of nonzero standard deviation. As discussed in section 3, this allows localization of the regions with local variability which is not completely modelled using only $L = 3$ dominating eigenmodes. The extent of these regions reduces for increasing L (not shown). However, this is not always desirable, as the increased average local residual and corresponding standard deviation may indicate localized contouring uncertainties (e.g. due to badly localizable organ boundaries in the CT image) and might not reflect real anatomical variability.

6. Discussion and conclusion

At first sight, organ motion as represented by displacement vector fields is of a complex nature. However, correlations between the displacements of the tissue elements suggest that the underlying dimensionality of this problem is possibly low. The linear statistical method of principal component analysis can be used to decompose the observed geometric variability into modes of correlated motion with large variability, describing the essential characteristics of individual organ motion/deformation, and with low variability, mainly representing uncorrelated 'noise' in the dataset.

For prostate/rectum/bladder motion the information about *individual* geometric variability as taken from a multiple CT study can typically be expressed by $\lesssim 4$ *patient-specific* dominating eigenmodes. This is in accordance with the finding that geometrical information taken from a few initial CTs ($\sim$5) is already sufficient for a robust individual adaption of the treatment plan (Yan *et al* 2000, Birkner *et al* 2003), which could not be expected, if the underlying statistical process is of large dimensionality. It should be noted here that a PCA-model with N input samples has only $N - 1$ eigenmodes with nonzero eigenvalue; thus at least 5 initial CTs are (mathematically) necessary to give an estimation for the four-dimensional eigenmode space mentioned above. However, a rigorous investigation about the minimal number of initial CTs needed for a robust and accurate prediction of individual geometric variability is beyond the scope of this paper.

PCA could in principle also be used to model single organs only (e.g. motion/deformation of rectum independent of prostate). As would be expected due to the less complex variability to be modelled in this case, plots of the residuals corresponding to figures 3 and 4 show faster convergence (data not shown). Thus single organ PCA-models typically consist of a smaller number of eigenmodes to achieve the same modelling accuracy as a corresponding multi-organ PCA model. Note that this is achieved at the expense of losing information about correlated motion/deformation of the adjacent organs.

The eigenmode approach is a simple way to model dose warping on a statistical basis as needed for 4D-planning and evaluation: for hollow organs, sums of eigenmodes immediately represent the required deformation fields of surface tissue elements; for solid organs, interpolation into the interior from the positions of the organ surface points (Kaus *et al* 2005) or explicity extending the surface shape vector (1) by positions of interior organ points can be used to efficiently calculate approximate deformation fields for multiple geometries, which are defined by the shape space of the dominating eigenmodes.

Possible applications of the eigenmode approach in the field of (adaptive) treatment planning start from equation (6), which describes how new geometry samples from the shape

space given by the dominating eigenmodes can be generated. For example, this might be used for patient-specific prediction of organ dose distributions based on a PCA of geometries of few initial treatment days, or for prediction of coverage probabilities, which have been proposed for indivualized margin definition in the literature (Stroom *et al* 1999). However, a detailed discussion and evaluation of clinical applications is beyond the scope of this paper and has to be subject to future investigations.

In this work the sets of corresponding surface points used as input of the PCA method were taken from finite element data. However, with respect to clinical applications a simpler method starting directly from contour information would be desirable. This has been addressed in a previous study (Söhn *et al* 2004); the resulting dominating eigenvalues and eigenmodes are essentially the same as for FE-input data (not shown). This can be understood by the fact that inaccuracies/differences in generating the simplified point correspondence of the surface points are statistically uncorrelated and thus appear only in eigenmodes with small eigenvalue.

As mentioned in section 5, manual organ delineation on a slice-by-slice basis may lead to unphysical organ shapes used as input samples, which affects the quality of the PCA-model. Data pre-processing methods, such as contour smoothing or algorithms for automatic 3D organ delineation, which use (population-based) PCA-organ shape models (e.g. Pekar *et al* 2004), might help to reduce these types of artefacts, though can be a source of specific errors themselves.

From a statistical viewpoint, these delineation artefacts can be seen as *local shape outlier* or *intra-sample outlier*. A modification of PCA, the so-called *robust PCA* (RPCA), has been developed for handling analogous outlier problems in the field of PCA-based automatic image recognition (De La Torre and Black 2001). Implementation of such an RPCA method could help to overcome the problems associated with contour artefacts on an inherent statistical basis.

Another problem—generally for radiotherapy planning—is inter-sample outliers, i.e., single CTs representing an extreme deformation of the organ geometry compared to the variability shown by all other CTs (e.g. caused by an accidentally non-controlled bladder or rectum filling). Potentially leading to an overestimation of individual variability, the use of such an extreme geometry as input sample for a PCA model may actually not be desired. Also here, RPCA methods developed in the field of computer vision (Xu and Yuille 1995) may help to automatically detect such outliers.

To summarize, PCA of organ geometries is an efficient method to extract the essential information of individual correlated geometric variation from multiple imaging data in a quantitative way, while neglecting uncorrelated 'noise'. Linear combinations of the resulting few dominating eigenmodes span the characteristic deformation space of the treatment geometry of a patient, which can both be used to represent measured geometries with good accuracy as shown and to generate new 'intermediate' geometries on a statistical basis, taking the correlated movement of the adjacent organs into account.

Acknowledgments

The authors wish to acknowledge Dr Jian Liang from Beaumont Hospital, Royal Oak, for his assistence with FEM data handling. This work has been supported in part by Deutsche Krebshilfe e.V. grant no 106 280 (MS, MB) and NIH grant no R01 CA091020.

Appendix A. General idea of principal component analysis (PCA): a 2D example

In this appendix we elucidate the basic features of the PCA-method by means of a low dimensional example.

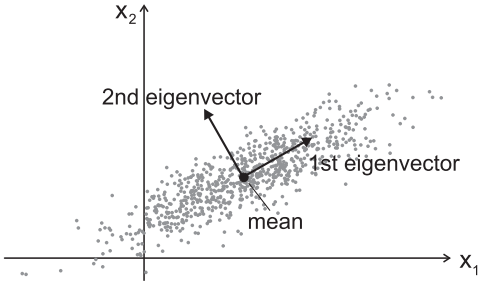

Figure A1. General illustration of the PCA-method applied to a dataset of N repeated measurements of $M = 2$ variables x_1 and x_2: the eigenvectors of the 2×2-dimensional covariance matrix are the principal axes of a multivariate (2D) Gaussian distribution; the corresponding eigenvalues are the variances of correlated change of the variables along these axes.

Suppose $M = 2$ variables x_1 and x_2 obey a multivariate (2D) normal distribution P^{Gauss} with expectation value $\mu = (\mu_1, \mu_2) \in \Re^2$ and correlation matrix $\Sigma \in \Re^{2\times 2}$:

$$P^{\text{Gauss}}(\boldsymbol{x}; \mu, \Sigma) \sim \exp\left[-\tfrac{1}{2}(\boldsymbol{x} - \mu) \cdot \Sigma^{-1} \cdot (\boldsymbol{x} - \mu)^T\right] \tag{A.1}$$

where $\boldsymbol{x} = (x_1, x_2) \in \Re^2$. Note that for the special case of a diagonal correlation matrix, $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2)$, equation (A.1) simplifies to a product of 1D Gaussian distributions: $\sim \exp\left(-(x_1 - \mu_1)^2 / (2\sigma_1^2)\right) \cdot \exp\left(-(x_2 - \mu_2)^2 / (2\sigma_2^2)\right)$.

Figure A1 shows an exemplary scatter plot of a distribution (A.1) with nonzero expectation value and covariances $\Sigma_{1,2} = \Sigma_{2,1} \neq 0$.

If $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)$ are samples from this distribution, the *mean value* $\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1,\ldots,N} \boldsymbol{x}(i)$ and the *empirical covariance matrix*

$$\mathcal{C} = \frac{1}{N-1} \sum_{i=1}^{N} (\boldsymbol{x}(i) - \bar{\boldsymbol{x}}) \cdot (\boldsymbol{x}(i) - \bar{\boldsymbol{x}})^T \in \Re^{2\times 2} \tag{A.2}$$

are (unbiased) estimators for μ and Σ.

Diagonalization of (A.2) results in $\tilde{\mathcal{C}} = \text{diag}(\lambda_1, \lambda_2)$ with the eigenvalues of the covariance matrix as diagonal entries. This corresponds to a rotation of the coordinate system with the eigenvectors as new basis vectors. In the rotated coordinate system the normal distribution can be written as a product of 1D Gaussian distributions, where the variances are given by the eigenvalues of the covariance matrix:

$$\sigma_l^2 = \lambda_l. \tag{A.3}$$

Thus the eigenvectors are the principal axes of the 2D normal distribution as depicted in figure A1, and the eigenvalues are a measure for the variability of the dataset along the directions of the eigenvectors.

The relative magnitude of the eigenvalues provides information about the grade of correlation between the variables which is present in the sample dataset: if $\lambda_1 \gg \lambda_2$ (as in the example of figure A1), the two variables x_1 and x_2 are strongly correlated with large variability along the direction of the first eigenvector, whereas the small variability in direction of the second eigenvector can be interpreted as 'noise'. In this case, the variability of the sample set is already well described by a 1D probability distribution with perfect correlation between the variables. This can be generalized for the case of more than two variables $M > 2$: if the spectrum of eigenvalues is dominated by a few large values, the variability of the dataset is concentrated in the subspace spanned by the corresponding eigenvectors, i.e. the effective dimensionality of the dataset is actually smaller than the number of variables M.

Appendix B. Diagonalization of the covariance matrix: efficient numerical implementation

Applying the PCA method to N organ geometric samples as described in section 2.2 leads to the problem of diagonalizing a $3M \times 3M$ covariance matrix. As the number of points M is typically in the order of 10^3, this is numerically infeasible in terms of computation time.

Generally, the covariance matrix $\mathcal{C} = \frac{1}{N-1}\mathcal{P}\mathcal{P}^T$ (equation (4)) has $3M$ eigenvectors and eigenvalues. However, noting that $\mathcal{C}$ is calculated from only N ($\ll M$) sample vectors, it can be shown that the covariance matrix is of rank $N-1$ (or less) and thus cannot have more than $N-1$ nonzero eigenvalues. In this case the related eigenvalue problem of the much smaller *implicit covariance matrix*

$$\tilde{\mathcal{C}} = \frac{1}{N-1}\mathcal{P}^T\mathcal{P} \in \Re^{N\times N} \tag{B.1}$$

can be solved instead (Murakami and Kumar 1982, Lorenz and Krahnstöver 2000).

- The eigenvalues $\tilde{\lambda}_i$ of $\tilde{\mathcal{C}}$ provide the nonzero eigenvalues of $\mathcal{C}$:

$$\lambda_i = \tilde{\lambda}_i, \qquad i = 1, \ldots, N. \tag{B.2}$$

- The N corresponding eigenvectors $\boldsymbol{q}_i \in \Re^{3M}$ of the correlation matrix $\mathcal{C}$ are calculated from the eigenvectors $\tilde{\boldsymbol{q}}_i \in \Re^N$ of $\tilde{\mathcal{C}}$ according to

$$\boldsymbol{q}_i = \frac{\mathcal{P}\tilde{\boldsymbol{q}}_i}{|\mathcal{P}\tilde{\boldsymbol{q}}_i|}. \tag{B.3}$$

In our implementation the algorithm for diagonalization of the matrix (B.1) was taken from Press *et al* (1992).

References

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

Cootes T F, Hill A, Taylor C J and Haslam J 1994 The use of active shape models for locating structures in medical images *Image Vis. Comput.* **12** 355–66

De la Torre F and Black M J 2001 Robust principal component analysis for computer vision *IEEE ICCV: 2001 Int. Conf. on Computer Vision (Vancouver, Canada)* vol 1 pp 362–9

Fontenla E, Pelizzari C A, Roeske J C and Chen G T Y 2001a Using serial imaging data to model variabilities in organ position and shape during radiotherapy *Phys. Med. Biol.* **46** 2317–36

Fontenla E, Pelizzari C A, Roeske J C and Chen G T Y 2001b Numerical analysis of a model of organ motion using serial imaging measurements from prostate radiotherapy *Phys. Med. Biol.* **46** 2337–58

Hoogeman M S, van Herk M, Yan D, Boersma L J, Koper P C M and Lebesque J V 2002 A model to simulate day-to-day variations in rectum shape *Int. J. Radiat. Oncol. Biol. Phys.* **54** 615–25

Jolliffe I T 2002 *Principal Component Analysis* (New York: Springer)

Kaus M, Pekar V, McNutt T and Bzdusek K 2005 An efficient algorithm for image-based dose deformation and accumulation *Med. Phys.* **32** 1900

Liang J and Yan D 2003 Reducing uncertainties in volumetric image based deformable organ registration *Med. Phys.* **30** 2116–22

Lorenz C and Krahnstöver N 2000 Generation of point-based 3D statistical shape models for anatomical objects *Comput. Vis. Image Underst.* **77** 175–91

Mageras G S, Kutcher G J, Leibel S A, Zelefsky M J, Melian E, Mohan R and Fuks Z 1996 A method of incorporating organ motion uncertainties into three-dimensional conformal treatment plans *Int. J. Radiat. Oncol. Biol. Phys.* **35** 333–42

Manly B F J 2004 *Multivariate Statistical Methods: A Primer* 3rd edn (Chapman and Hall/CRC)

Murakami H and Kumar V 1982 Efficient calculation of primary images from a set of images *IEEE Trans. Pattern Anal. Mach. Intell.* **4** 511–5

Pekar V, McNutt T R and Kaus M R 2004 Automated model-based organ delineation for radiotherapy planning in prostatic region *Int. J. Radiat. Oncol. Biol. Phys.* **60** 973–80

Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1992 *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge: Cambridge University Press)

Söhn M, Birkner M, Alber M and Nüsslin F 2004 Modelling patient geometric variation based on dominant eigenmodes of organ deformation *Radiother. Oncol.* **73** (Suppl. 1) S49–50

Stroom J C, de Boer H C J, Huizenga H and Visser A G 1999 Inclusion of geometrical uncertainties in radiotherapy treatment planning by means of coverage probability *Int. J. Radiat. Oncol. Biol. Phys.* **43** 905–19

Xu L and Yuille A 1995 Robust principal component analysis by self-organizing rules based on statistical physics approach *IEEE Trans. Neural Netw.* **6** 131–43

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

Yan D, Lockman D, Brabbins D, Tyburski L and Martinez A 2000 An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer *Int. J. Radiat. Oncol. Biol. Phys.* **48** 289–302

Appendix E

Model-independent, Multimodality Deformable Image Registration by Local Matching of Anatomical Features and Minimization of Elastic Energy

published in
Medical Physics 2008; **35**(3): 866-878

Model-independent, multimodality deformable image registration by local matching of anatomical features and minimization of elastic energy

Matthias Söhn[a)]
Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

Mattias Birkner
Department of Radiation Therapy, University Hospital for Radiology, Robert-Koch-Strasse 6, 89081 Ulm, Germany

Yuwei Chi, Jian Wang, and Di Yan
Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, Michigan 48073

Bernhard Berger
University Hospital for Radiation Oncology, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

Markus Alber
Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

(Received 16 July 2007; revised 24 October 2007; accepted for publication 21 December 2007; published 12 February 2008)

With respect to the demands of adaptive and 4D-radiotherapy applications, an algorithm is proposed for a fully automatic, multimodality deformable registration that follows the concept of translational relocation of regularly distributed image subvolumes governed by local anatomical features. Thereby, the problem of global deformable registration is broken down to multiple independent local registration steps which allows for straightforward parallelization of the algorithm. In a subsequent step, possible local misregistrations are corrected for by minimization of the elastic energy of the displacement field under consideration of image information. The final displacement field results from interpolation of the subvolume shift vectors. The algorithm can employ as a similarity measure both the correlation coefficient and mutual information. The latter allows the application to intermodality deformable registration problems. The typical calculation time on a modern multiprocessor PC is well below 1 min, which facilitates almost-interactive, "online" usage. CT-to-MRI and CT-to-cone-beam-CT registrations of head-and-neck data sets are presented, as well as inhale-to-exhale registrations of lung CT data sets. For quantitative evaluation of registration accuracy, a virtual thorax phantom was developed; additionally, a landmark-based evaluation on four lung respiratory-correlated CT data sets was performed. This consistently resulted in average registration residuals on the order of the voxel size or less (3D-residuals $\sim 1-2$ mm). Summarizing, the presented algorithm allows an accurate multimodality deformable registration with calculation times well below 1 min, and thus bears promise as a versatile basic tool in adaptive and 4D-radiotherapy applications.
[DOI: 10.1118/1.2836951]

Key words: deformable image registration, multimodality registration, mutual information

I. INTRODUCTION

With the advent of developments in "4D-radiotherapy" applications like adaptive radiotherapy, 4D-planning, etc., and the increasing use of daily image information in the clinical routine, the problem of (deformable) image registration has recently won growing attention in the field of radiotherapy.[1] Special requirements for registration algorithms for use in this field are imposed by the broad spectrum of treatment sites and the need to handle image information from various modalities, as well as limited time-frames for the planning and treatment process. Thus, a deformable registration algorithm should be able to handle both intra- and intermodality registration problems in a model-independent way. In other words, ideally no manual segmentation or other user interaction is required, nor special assumptions or prior knowledge about the organ site to be registered such as atlases are necessary. In view of possible "online" applications, it should allow registration of full data sets within a few minutes with an accuracy appropriate for radiotherapy applications. Ideally, it would be an inherently parallel algorithm to take advantage of recent trends of computer hardware.

Most of the methods proposed for application in radiotherapy stem from the numerous image registration concepts developed before in the medical field[2–4] and other fields.[5]

Generally speaking, the common principle of all algorithms is to determine the optimal values of a certain set of image transformation parameters based on the use of a similarity measure, which allows a rough categorization of the different methods.

Concerning the similarity measure, point-based and voxel-based registration concepts can be distinguished. Point-based methods minimize the distance between segmented homologous features (points, curves, or surfaces), and thus typically require user interaction and/or preprocessing. Finite-element methods are a special type of point-based methods.[6] In contrast, voxel-based methods work based on the image intensity information itself using mathematical similarity measures like the sum of squared distance, correlation coefficient (CC), correlation ratio (CR), or mutual information (MI),[1,2] or iteratively determine deformations based on image gradients like optical flow methods.[7] While CR- and MI-based algorithms can handle both intra- and intermodality registrations, all other voxel-based and point-based methods are restricted to intramodality problems or require preprocessing of the image data sets.

Concerning the type of image transformation used, the algorithms differ in the degrees of freedom (DOF) used for modeling deformations. Deformation fields can be represented in a nonparametric way by B-splines[8] or thin-plate-splines.[9] Here, a similarity measure like MI is optimized by simultaneously adapting the positions of a set of spline control points as variables, thereby facilitating multimodality deformable registration. However, the number of control points (and thus the DOF of the transformation) is practically limited by the computation time for a large number of control points. This has recently been addressed by Pekar *et al.*,[10] who proposed a physics-based parametric approach using a comparatively small number of control points irregularly distributed over the image domain. Another way to speed up calculations was proposed by Shekhar *et al.*,[11] who optimized the global similarity measure by subsequently adapting only small subgroups of B-spline control points, which potentially allows parallelization of the algorithm. In contrast to spline- and other control point-based transformations, algorithms based on free-form deformations (FFD) have "unlimited" DOF by defining each voxel shift as an independent variable, thereby having the full ability to reproduce local deformations. To avoid ill-posedness, FFD algorithms need to be subject to constraints (e.g., smoothness, limited elastic energy) and/or perform regularizations during iterative voxel-based adaptation of the deformation field. Examples for FFD algorithms are the optical flow method,[12] "demons" algorithms,[13] viscous fluid algorithms,[14,15] variational approaches,[4,16] and several others. While FFD algorithms can show good performance for intramodality registration problems,[17–20] the direct application to intermodality problems is generally hampered by performance problems when MI is used instead of simpler and faster similarity measures, or even conceptually impossible as for optical flow.

While the voxel-based algorithms mentioned so far treat deformable registration as a global problem and optimize a global similarity measure for the whole image region to be registered, a few concepts premised on local template/block matching have been proposed as well.[21–23] Lau *et al.*[21] presented a hierarchical approach to individually register small cubic image subregions by local optimization of a similarity measure. The cubic subregions were regularly distributed on a 3D grid and partly overlapped the nearest neighbors. This resulted in a sparsely sampled vector field of shifts for the centers of the subregions, which was regularized by a median filter and interpolated by a Gaussian function to ensure a locally smooth transformation. Malsch *et al.*[23] performed multiple CC-based local registrations of small rectangular subvolumes irregularly distributed over the region to be registered. The global deformation field was determined by a modified thin-plate spline interpolation of the template volumes' shift vectors. As a successful local registration requires sufficient local image information inside a subvolume, a set of anatomical landmarks was automatically identified in a preprocessing step and was used for optimal selection and initial placement of the templates. This algorithm was tested for CT-CT registrations of different organ sites and showed very good accuracy by landmark-based evaluation.

The algorithm presented in this paper shares the concept of fast local subvolume matching with the two aforementioned approaches, and thus allows parallelization of the independent rigid registration problems on multiprocessor/multicore computer systems. By using MI as a local similarity measure, we show that this concept can be extended to intermodality registration problems. In contrast to Malsch *et al.*,[23] a preprocessing step for identification of anatomical features is not necessary for our algorithm, as the subregions are distributed on a regular grid over the image region to be registered. This facilitates simpler interpolation schemes between the shift vectors based on trilinear or B-spline interpolation. In contrast to the regularization scheme used by Lau *et al.*,[21] possible local misregistrations are detected by a novel local match quality assessment scheme and corrected for by minimization of the elastic energy of the displacement field under consideration of image information. The following sections describe the details of the algorithm (Sec. II A) and the validation studies carried out (Sec. II B). Application to different image data sets and quantitative results are presented in Sec. III.

II. MATERIALS AND METHODS

II.A. Registration algorithm

We propose an algorithm for fast deformable registration of organs which comprise a sufficient number of internal anatomical structures. It is assumed that these features are deformed only by a small amount and can be treated as rigid. A typical example here are 4D respiratory-correlated CT (RCCT) data sets of lung, which show numerous internal image features like bronchial and vascular branches/bifurcations, etc.

This facilitates a *featurelet-based* approach of image registration: the problem of global deformable registration for the whole organ is broken down to (fast) individual rigid registrations of small subvolumes ("featurelets"), regularly

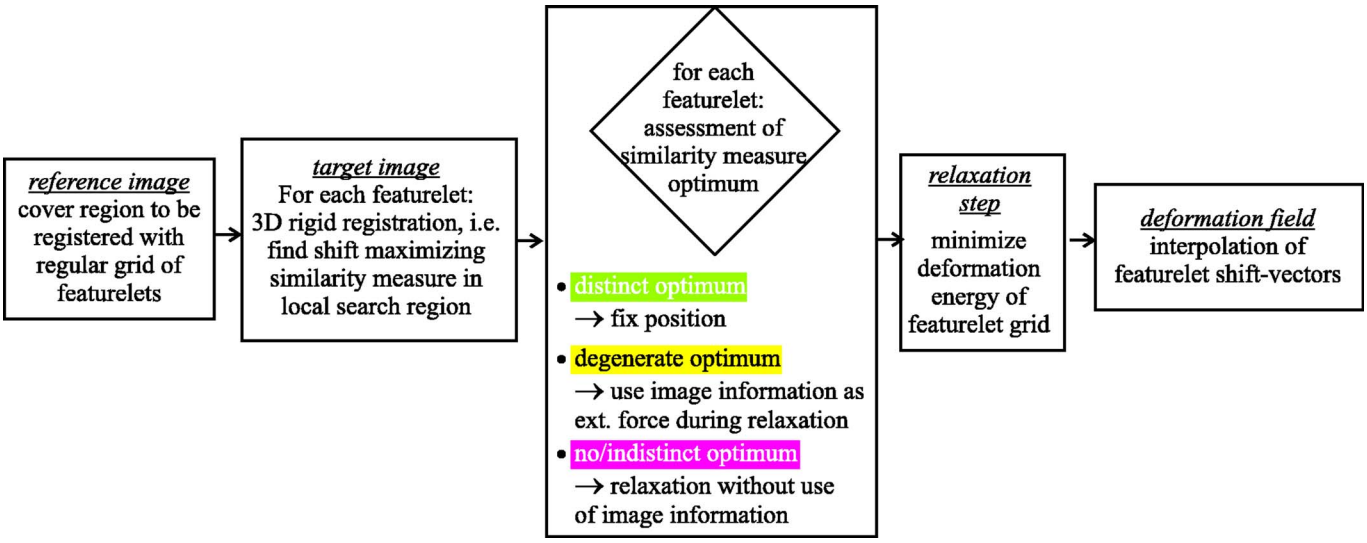

FIG. 1. Flow chart representing the featurelet-based image registration algorithm. The algorithm can be divided into the five steps depicted, which are described in Secs. II A 1, II A 2, II A 3, II A 4, and II A 5.

distributed over the whole region of interest (Secs. II A 1 and II A 2). For a part of the featurelets a local lack of unique anatomical image information or image artifacts leads to ill-defined optimal positions or misregistration of individual featurelets, which have to be identified (Sec. II A 3). In a subsequent step, such ambiguities or misregistrations are corrected for by iterative minimization of the local deformation energy under consideration of local image information if possible (relaxation step, Sec. II A 4). Finally, the global displacement field is calculated from the shifts of the featurelets by interpolation (Sec. II A 5). The algorithm provides model-independent deformable registration, as no specific information about tissue properties is needed and no *a priori* assumptions about the geometry being registered are made.

Figure 1 shows a flowchart of the algorithm; details are presented in the following subsections. The method for evaluation of the registration accuracy is described in Sec. II B.

II.A.1. Step 1: Initialization of the featurelet grid

The image region to be registered is covered by a regular 3D grid of equally sized rectangular 3D regions in the reference image. Figure 3 shows the exhale and inhale CTs of a thorax RCCT data set, with the exhale/inhale CT as reference/target image for deformable registration, respectively.

The featurelet size is a parameter of the algorithm which has to be chosen according to the typical extent of distinct anatomical features (e.g., bronchial/vascular branches in the case of lung). In principle, a smaller featurelet size (and thus larger number of featurelets) allows a better resolution of the deformation field in terms of the registration of finer details. However, the robustness of the registration is deteriorated if the featurelet size is chosen too small. A systematic investigation of the optimal featurelet size will be presented in the Appendix for the example of lung registration.

Denoting the reference image by A and the voxel intensity at position $\boldsymbol{p}_a$ in the reference image by $A(\boldsymbol{p}_a)$, the ith reference featurelet $\mathcal{A}_i$ can be formally described as

$$\mathcal{A}_i(\boldsymbol{p}_a) = \begin{cases} A(\boldsymbol{p}_a) & \forall \boldsymbol{p}_a \in \Omega_{\text{rect}}(\boldsymbol{X}_{a,i}) \\ 0 & \text{elsewhere} \end{cases}, \tag{1}$$

where $\Omega_{\text{rect}}(\boldsymbol{X}_{a,i})$ denotes a rectangular region as the domain of the ith featurelet, centered around $\boldsymbol{X}_{a,i}$.

II.A.2. Step 2: Local rigid registration of featurelets

It is assumed that the images are rigidly preregistered so that the center of the mass shift and the bulk rotation vanish. The basic task of the algorithm is to individually register each reference featurelet to the target image B, i.e., to find a transformation T_i between $\mathcal{A}_i$ and the corresponding target featurelet,

$$\mathcal{B}_i(\boldsymbol{p}_b) = \begin{cases} B(\boldsymbol{p}_b) & \forall \boldsymbol{p}_b \in T_i(\Omega_{\text{rect}}(\boldsymbol{X}_{a,i})) \\ 0 & \text{elsewhere} \end{cases}, \tag{2}$$

which maximizes the similarity between $\mathcal{A}_i$ and $\mathcal{B}_i$. For reasons of algorithmic efficiency, it is desirable to restrict the search space of possible transformations to *translations*, i.e., $T_i(\Omega_{\text{rect}}(\boldsymbol{X}_{a,i})) = \Omega_{\text{rect}}(\boldsymbol{X}_{a,i} + \boldsymbol{\delta X}_i) \equiv \Omega_{\text{rect}}(\boldsymbol{X}_{b,i})$, with $\boldsymbol{\delta X}_i$ denoting a 3D-shift vector to be determined for each featurelet. This simplification is valid if the deformations/local rotations present in the data set are small over the domains Ω of the featurelets.

Several similarity measure functions such as the sum of squared intensity differences (SSD), the correlation coefficient (CC), and mutual information (MI) have been proposed for quantification of image similarity.[2] SSD is the simplest

and fastest intramodality similarity measure; however, it is not robust against variations of absolute intensity differences for equivalent structures (e.g., lung density changes during respiratory cycle).

For intramodality registration (especially CT-CT), CC is a more robust similarity measure. The correlation coefficient of a reference featurelet $\mathcal{A}_i$ and its corresponding (shifted) target featurelet $\mathcal{B}_i$ reads (featurelet index i dropped)

$$\mathrm{CC}(\mathcal{A},\mathcal{B}) = \frac{\Sigma_{\boldsymbol{p}_a \in \Omega_{\mathrm{rect}}(\boldsymbol{X}_a)} (\mathcal{A}(\boldsymbol{p}_a) - \bar{\mathcal{A}}) \cdot (\mathcal{B}(T(\boldsymbol{p}_a)) - \bar{\mathcal{B}})}{\sqrt{\Sigma_{\boldsymbol{p}_a \in \Omega_{\mathrm{rect}}(\boldsymbol{X}_a)} (\mathcal{A}(\boldsymbol{p}_a) - \bar{\mathcal{A}})^2 \cdot \Sigma_{\boldsymbol{p}_a \in \Omega_{\mathrm{rect}}(\boldsymbol{X}_a)} (\mathcal{B}(T(\boldsymbol{p}_a)) - \bar{\mathcal{B}})^2}}. \quad (3)$$

Here, $T(\cdot)$ is the translation operator, and $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ are the average voxel values in the domains of the reference and (shifted) target featurelet, respectively.

Unlike CC, the entropy-based mutual information approach[24–26] can be applied both to intra- and intermodal registration problems. In our algorithm, we used the normalized mutual information[27] (NMI) to quantify similarity of featurelets regions,

$$\mathrm{NMI}(\mathcal{A},\mathcal{B}) = \frac{H(\mathcal{A}) + H(\mathcal{B})}{H(\mathcal{A},\mathcal{B})}. \quad (4)$$

$H(\mathcal{A})$ and $H(\mathcal{B})$ denote the marginal entropies of the voxel intensity values inside of the featurelet domains, and $H(\mathcal{A},\mathcal{B})$ is the corresponding joint entropy. For the purpose of calculating image entropies, histograms of intensity values need to be calculated, where an appropriate number of bins has to be chosen. In our implementation, we follow a study of Filev *et al.*[28] about the related problem of 2D-subregion registration, where a number of $32/32 \times 32$ bins was found to be most appropriate for calculation of the marginal/joint entropy, respectively.

We implemented both CC and NMI as similarity measures: NMI is required for intermodality registration, and the computationally less expensive CC was implemented to assess its performance for intramodality registration. For the purpose of determining the optimal featurelet shift, the respective similarity measure was maximized within a rectangular local search region, allowing 3D shifts $\boldsymbol{\delta X}$ with a step size given by the voxel size in each direction. The dimensions of the search region are parameters of the algorithm, and should be chosen large enough to capture the maximal displacement in the region to be registered. As the similarity measure function can have more than one maximum in the search region, we implemented the optimization essentially as an exhaustive search. To speed up the calculation, an adaptive multigrid approach was used: in a first step the similarity measure was evaluated on a coarse search grid (average in-plane step size: 2 voxels per dimension, i.e., undersampling by factor of 4), thereby roughly determining the positions of candidates for maxima. In a second step, the overall maximum was determined by an exhaustive search on a refined search grid around the candidates with the step size given by the voxel size.

II.A.3. Step 3: Assessment of local registration quality

The configuration of the similarity measure in the search region and especially around the optimum provides important information about the registration quality of the individual featurelets. In essence, three cases can be distinguished:

- Case I: Ideally, the local image information of a featurelet is sufficient to allow unambiguous registration to a corresponding target image subregion. Figures 2(a)–2(c) depict a typical example for this situation, where the optimal position of the target featurelet is characterized by a unique, distinct optimum of the similarity measure.
  However, for a part of the featurelets this is not the case. Two main types of problems can be identified:
- Case II: The local image information does not allow an unique registration of a featurelet, which is expressed as a degenerate optimum of the similarity measure function. A typical example is shown in Figs. 2(d)–2(f).
- Case III: The featurelet covers a region with low or no image contrast; thus, the similarity measure function shows low similarity values for all possible shifts and no clear optimum. Figures 2(g)–2(i) show an example for this case.

In the latter cases the correct position of the target featurelet cannot be determined based on image contrast alone, but additional information from adjacent featurelets has to be used. This is implemented in terms of a subsequent relaxation step (see Sec. II A 4) for which the featurelets were categorized according to the three cases described above by the following criteria:

- *Similarity measure values:* If no local optimum of the similarity measure can be found, or if the value of the global optimum inside the search region is below a predefined threshold value ($\mathrm{CC}_{\mathrm{thresh}}$ or $\mathrm{NMI}_{\mathrm{thresh}}$), the respective featurelet is assumed to have bad registration quality [case III above, Fig. 2(i)].

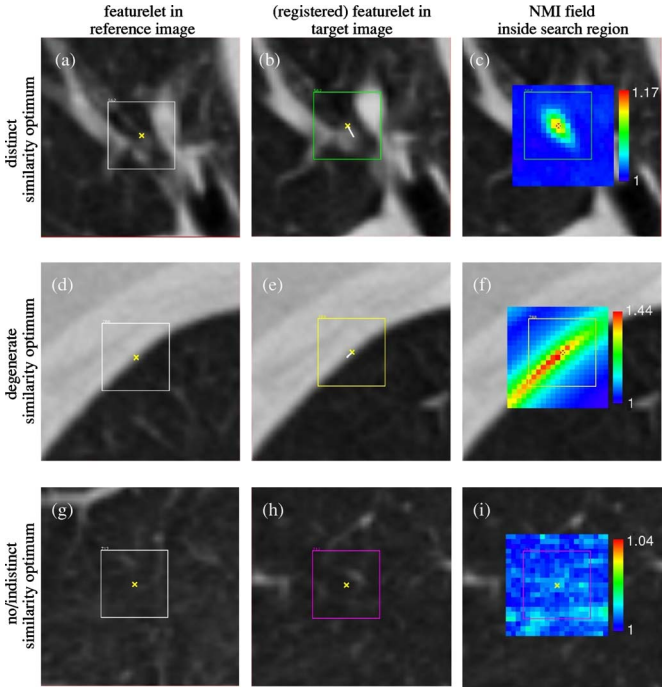

FIG. 2. Examples for the three typical situations encountered in local rigid registration of featurelet regions (2D views of 3D featurelets and search regions): distinct (a)–(c), degenerate (d)–(f), or no/indistinct (g)–(i) similarity optimum (case I/II/III; see Sec. II A 3). First column: featurelet position in the reference image; second column: (registered) featurelet in the target image with the (projected) 3D-shift vector represented as line; third column: local similarity (here: normalized mutual information, NMI) for all possible shifts of the featurelet inside of search region.

- *Eigenvalues of Hessian matrix at optimum:* If a maximum above the similarity measure threshold can be found, a criterion assessing the grade of its distinctiveness is needed. This criterion uses the curvature matrix of the maximum. The eigenvalues of this curvature matrix are a measure for the curvature of the similarity measure function in the three orthogonal principal directions. At a true maximum, these eigenvalues are negative. Thus, if all three eigenvalues are below a predefined threshold EV_{thresh}, the registered featurelet is assumed to have good registration quality [case I above, Fig. 2(c)]. Otherwise, the featurelet is categorized to have a degenerate optimum as the registration in the direction of the lowest curvature is not reliable [case II above, Fig. 2(f)].

CC_{thresh} (or NMI_{thresh}) and EV_{thresh} are parameters of the algorithm that need to be determined separately for each similarity measure, and their appropriate values also depend on the image modalities to be registered. As the parameter EV_{thresh} controls the classification between case I- and case II featurelets, its value has to be chosen appropriately to the typical "feature content" in the respective image region to be registered. Thus, its optimal value will also depend on the organ site to be registered, where class-solutions for head-and-neck, thorax, etc. should be obtainable. A systematic investigation of the optimal thresholds will be presented in the Appendix on the example of lung registration.

II.A.4. Step 4: Relaxation

Due to inevitable local registration ambiguities or misregistrations, a deformable registration ansatz solely based on local image information produces unphysical deformation fields which locally violate injectivity and continuity requirements. This has to be corrected for by imposing additional assumptions which assure physically meaningful deformation fields. In our implementation this is realized by moving the respective featurelets (cases II and III) to positions minimizing the overall deformation energy of the featurelet grid. For case II featurelets the additional constraint to stay within the degenerate optimum is imposed. Elsewhere, registration accuracy is ensured by keeping the positions of the correctly registered featurelets (case I).

Numerically, minimization of the deformation energy was implemented as a relaxation algorithm: Assigning a virtual mass m_i to each featurelet, the dynamics of the ith-target featurelet as represented by its center position $\boldsymbol{X}_{b,i}$ is governed by

$$m_i\ddot{\boldsymbol{X}}_{b,i} = -D\cdot\dot{\boldsymbol{X}}_{b,i} + \boldsymbol{F}^{[i]}_{\text{springs}} + \boldsymbol{F}^{[i]}_{\text{image}}. \tag{5}$$

Here, D is a damping factor, and the dynamics is driven by two types of forces, which are typically referred to as "internal" and "external" force.[3] The internal forces give rise to the deformation energy of the featurelet grid and are represented by $\boldsymbol{F}^{[i]}_{\text{springs}}$ for each featurelet i, which is the resulting force imposed by virtual springs connected to the center of each 3D neighbor (in total: 26),

$$\boldsymbol{F}^{[i]}_{\text{springs}} = k\cdot\sum_{l\in\{\text{neighbors}\}}(|\boldsymbol{X}_{b,i}-\boldsymbol{X}_{b,l}| - L_0^{[i,l]}), \tag{6}$$

where k is the spring constant, and $L_0^{[i,l]}=|\boldsymbol{X}_{a,i}-\boldsymbol{X}_{a,l}|$ the baseline distance of the center positions in the regular featurelet grid in the reference image.

The external force $\boldsymbol{F}^{[i]}_{\text{image}}$ introduces constraints originating from image information, thereby ensuring a "fit" to the locally available data. The image force for the ith-target featurelet in our model is given by the gradient of the similarity measure S, Eq. (3) and (4), at the respective position $\boldsymbol{X}_{b,i}$, i.e.,

$$\boldsymbol{F}^{[i]}_{\text{image}} = \lambda_i\cdot\nabla S(\mathcal{A}_i,\mathcal{B}_i), \tag{7}$$

with λ_i acting as a balance of the spring forces. A Savitzky-Golay smoothing filter[29] (polynomial degree $M=2$, window size $n_L=n_R=2$) was used for the gradient calculations to improve numerical stability and convergence.

A second-order Runge-Kutta method[29] was used to calculate the evolution of Eq. (5). Due to the damping term, the asymptotic equilibrium with $\dot{\boldsymbol{X}}_{b,i}=0=\ddot{\boldsymbol{X}}_{b,i}$ for all featurelets represents the state of minimal deformation energy. The parameters used were $k=0.9$ and $D=1.4$; $m=1$ for the featurelets assessed as cases II and III, and $m=10^7(\gg 1)$ for the well-registered featurelets (case I), thereby effectively fixing these. The image force, Eq. (7), is used only for the featurelets with degenerate similarity optimum (case II), for which the multipliers are initialized as $\lambda_i=|\boldsymbol{F}^{[i]}_{\text{springs}}|/\overline{\nabla S}$ with $\overline{\nabla S}$ as

the mean image gradient in the neighborhood of the initial position, thereby having $\boldsymbol{F}^{[i]}_{\text{springs}}$ and $\boldsymbol{F}^{[i]}_{\text{image}}$ numerically in the same order.

II.A.5. Step 5: Calculation of global deformation field

The shift vectors $\boldsymbol{\delta X}_i$ of the target featurelets Eq. (2) relative to the corresponding reference featurelets Eq. (1) represent the displacement field sampled on the regular grid of featurelet centers in the reference image. The deformation field at any position can be obtained by B-spline interpolation between the $\boldsymbol{\delta X}_i$, thereby ensuring continuity or continuous derivatives if required. For the results presented in this study, a B-spline of order 1, which is analogous to trilinear interpolation, was used.

II.B. Validation studies

The registration quality was assessed visually by red-green overlay of the registered data sets. Additionally, a more quantitative evaluation of registration accuracy is desirable. We implemented two methods to estimate registration residuals on the example of 4D-lung deformable registration as described in the following.

II.B.1. Virtual thorax phantom

For real patient data sets the true deformation field against which to compare the algorithm is not accessible. Instead, a known deformation field could be applied to a given image data set to generate an artificial deformed image. Using these two images as input for a deformable registration algorithm to be evaluated, the known underlying deformation field should be retrieved.

Guerrero *et al.*[12] and Wang *et al.*[17] used thin-plate splines with a limited number of fiducial points taken from real patient data to generate a deformation field, while Crum *et al.*[19] derived a deformation field from a biomechanical model of the organ site using finite-element calculations. For this study, we used an independent, free-form deformable registration algorithm described in Zhang *et al.*[18] and Lu *et al.*[4] for creation of a virtual thorax phantom. The deformation field was calculated by using this algorithm for registration of the exhale and inhale state CTs of a lung RCCT patient data set. Notice that the registration accuracy of the algorithm is irrelevant for the purpose of creation of the virtual phantom, provided that the deformation field roughly represented the physiological deformations in the original lung dataset.

The virtual phantom for the evaluation studies encompassed an artificially deformed exhale state CT and the original inhale state CT with known underlying deformation field between these.

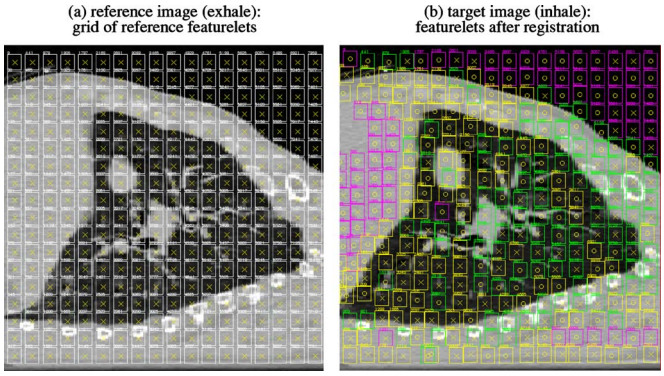

FIG. 3. Grid of featurelets in sagittal view (a) before registration and (b) after local registration and relaxation. The featurelet centers are marked by crosses [if in plane] or open circles [if they lie outside the shown sagittal plane; featurelets that moved completely out-of-plane are omitted in (b)]. For each featurelet (b) indicates the result of automatic assessment of registration quality (cf. Fig. 2) as used for the relaxation step (Sec. II A 4).

II.B.2. Landmark-based evaluation on RCCT patient test cases

Organs like the liver and lung show a large number of internal anatomical structures, which facilitates a landmark-based evaluation approach of registration accuracy.[9,22,23,30]

We evaluated registration accuracy of the presented algorithm using lung RCCT data sets. The positions of N_{AL} vascular or bronchial bifurcations as anatomical landmarks (AL) were marked by a physician in the inhalation and exhalation CTs, and the corresponding displacements $\boldsymbol{\delta p}_k^{[\text{actual}]}$ of the $k=1,\ldots,N_{\text{AL}}$ landmarks were determined and compared to the displacements $\boldsymbol{\delta p}_k^{[\text{predicted}]}$ calculated by deformable registration. The *registration residuals*

$$\boldsymbol{R}_k = \boldsymbol{\delta p}_k^{[\text{predicted}]} - \boldsymbol{\delta p}_k^{[\text{actual}]}, \tag{8}$$

are a direct measure for the alignment accuracy of the algorithm. For a comprehensive overview we calculated mean and standard deviation for the 3D residuals $|\boldsymbol{R}|$ and for each vector component separately, and determined the maximal 3D residuals.

III. RESULTS

III.A. Patient test cases

III.A.1. Intramodality registration (CT-CT)

The algorithm was tested on five RCCT patient data sets of the thorax region. These were acquired on a Siemens Somatom Open (Siemens Medical Solutions, Erlangen, Germany) in spiral RCCT mode[31] and re-sorted into ten respiratory phases with a slice thickness of 3 mm and in-slice voxel size of 1 mm.

One of these patient data sets was used for generation of a virtual thorax phantom (Sec. II B 1); the results will be presented below in Sec. III B. A quantitative evaluation of registration accuracy using anatomical landmarks (Sec. II B 2) was based on the other four data sets; see Sec. III C for results.

As an example, Fig. 3 shows the grid of featurelets before

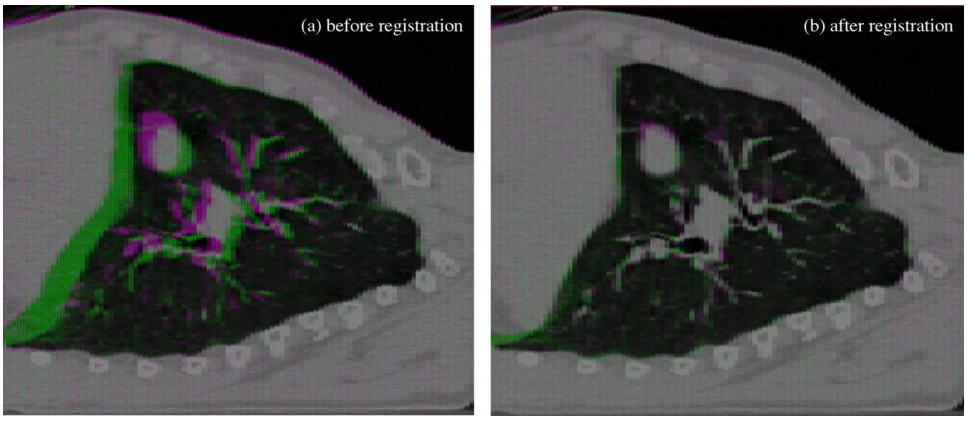

FIG. 4. The data set of Fig. 3, shown as (a) overlay of exhale and inhale CT before registration, and (b) after registration (inhale CT deformed to exhale CT using the deformation field).

and after exhale-to-inhale registration for patient 1. The size of the featurelets was chosen to be $15\times15\times5$ voxels $=14.6\times14.6\times15$ mm, resulting in ~10 400 featurelets for the whole CT as a region of interest for registration. The size of the rectangular local search region for individual featurelet registration was $9\times9\times6$ voxels. NMI was used as a similarity measure in this example. The parameters used for the assessment of local registration quality (Sec. II A 3) were $NMI_{thresh}=1.05$ and $EV_{thresh}=-0.01$.

As indicated in Fig. 3(b), featurelets with a distinct NMI-optimum (case I in Sec. II A 3; in this example ~29% of all featurelets) can be typically found in areas with a high density of clearly visible anatomical structures, while, e.g., featurelets at the diaphragm show mostly degenerate similarity optima as expected (case II; ~37%). The latter featurelets thus needed correction of their shifts in the relaxation step (Sec. II A 4). In areas with low image contrast such as air in lung or deep inside tissue areas, the featurelet shift could not be determined using local image information (case III; ~34%); consequently such featurelets were shifted to positions corresponding to minimal deformation energy in the relaxation step. The final registration result is visualized in Fig. 4. Except for a noticeable misregistration around the posterior part of the diaphragm, which is characterized by non-negligible rotations/local deformations on a scale smaller than the featurelet size used, the overall registration quality appears satisfactory with all visible major image features registered accurately.

A quantitative evaluation is presented in Secs. III B and III C. For more details about the dependence of the fraction of case I/II/III featurelets on the registration parameters, we refer to Table III in the Appendix.

III.A.2. Intermodality registration (CT-CBCT, CT-MRI)

To demonstrate intermodality performance, we applied the algorithm—using NMI as a similarity measure—to head-and-neck test cases.

In Fig. 5, a pretreatment cone-beam CT (CBCT; Elekta Synergy XVI, Elekta Ltd., Crawley, UK) was registered to a planning (helical) CT as reference image. Prior to registration, the image data were resampled to have equal resolution (voxel size: $1.1\times1.1\times2.0$ mm). A featurelet size of $15\times15\times5$ voxels was chosen, resulting in ~5200 featurelets covering the whole data set. The parameters used for registration were $9\times9\times4$ voxels as search region size, $NMI_{thresh}=1.05$, and $EV_{thresh}=-0.01$. A comparison of Figs. 5(c) and 5(d) shows that—even in the presence of dental metal artifacts (mainly in the reference CT) and strong shading/cupping artifacts in the CBCT—a substantial reduction of the residuals could be gained through deformable registration. This is especially obvious in the area of the nasal cavities and the upper spine.

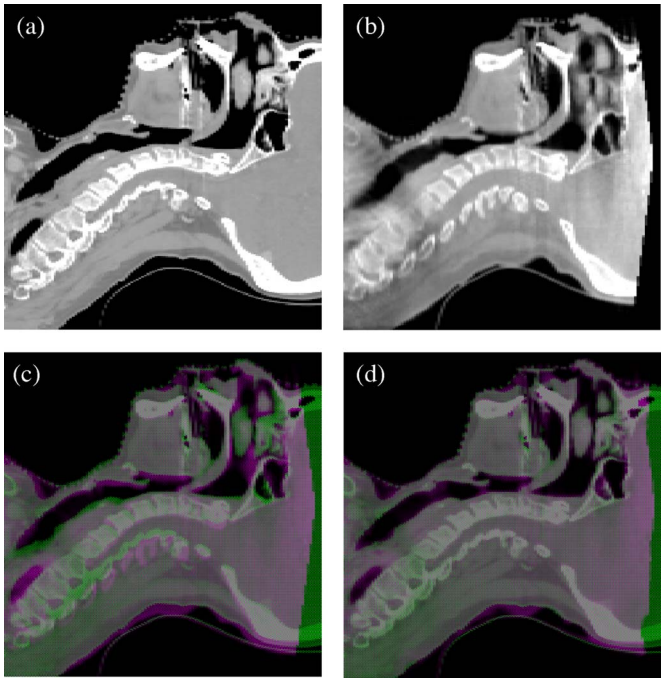

FIG. 5. Exemplary CT-CBCT head-and-neck registration: (a) CT as reference and (b) CBCT as target image; (c)/(d) overlays before/after registration.

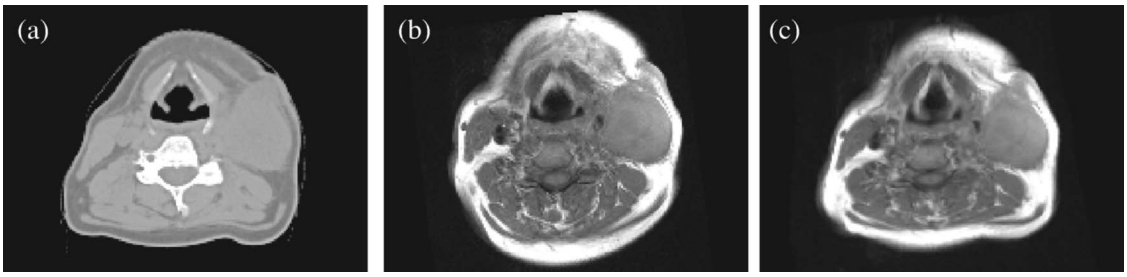

FIG. 6. Exemplary CT-MRI head-and-neck registration: (a) CT as reference and (b)/(c) MRI as target image before/after registration.

As a second example, Fig. 6 shows a CT-MRI registration. The voxel size of both input data sets was $1.3\times1.3\times3.0$ mm. A number of ~1800 featurelets of dimension $21\times21\times7$ voxels each was used for registration with the following parameters: search region size $10\times10\times8$ voxels; $NMI_{thresh}=1.11$, and $EV_{thresh}=-0.01$. The result of the registration is depicted in Fig. 6.

General conclusions about the performance of the presented algorithm for CT-MRI registration problems are planned to be subject to more quantitative future research on a larger number of data sets.

III.B. Virtual thorax phantom

Voxelwise evaluation of registration residuals was performed based on a virtual thorax phantom as described in Sec. II B 1. Mean and standard deviation of the (*a priori*) displacements of this phantom were 2.9 ± 2.8 mm. Using the (inhale deformed to) exhale CT as reference and the (original) inhale CT as the target image for deformable registration, the 3D residuals could be reduced to 1.1 ± 1.2 mm for a typical choice of registration parameters (featurelet size $15\times15\times5$ voxels; size of search region $15\times15\times8$ voxels; $NMI_{thresh}=1.05$; $EV_{thresh}=-0.01$). A comprehensive overview of the results in terms of histograms for the displacement field and the 3D residuals is provided by Fig. 7, which illustrates the substantial reduction of residuals by the presented algorithm. Similarly to the findings in Sec. III A 1 (see also Fig. 4), the voxels with residuals >4 mm (~2% of all lung voxels) are predominantly located around the diaphragm.

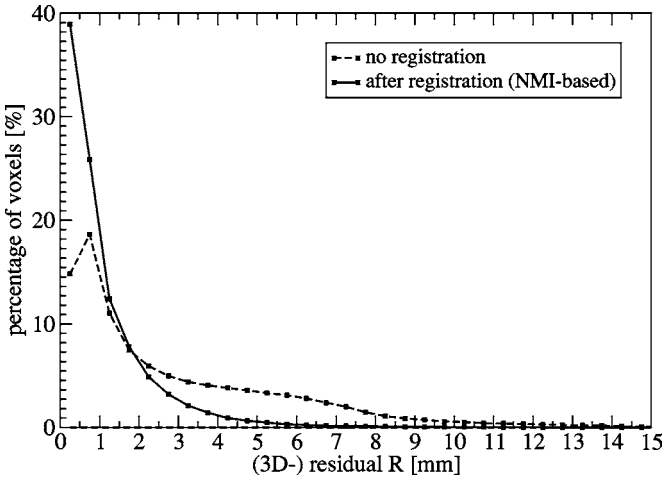

FIG. 7. Distribution of 3D residuals based on deformable registration of the virtual thorax phantom.

Due to the detailed statistical information about registration accuracy, a virtual phantom is especially useful for determination of optimal values for the different registration parameters. For a systematic investigation we refer to the Appendix.

III.C. Landmark-based evaluation

Landmark-based evaluation (Sec. II B 2) was performed based on lung RCCT data sets of four patients (voxel size: $1\times1\times3$ mm), which had $N_{AL}=11-15$ homologous points marked inside a single lung lobe. The magnitude of breathing motion was determined by the displacements $\boldsymbol{\delta p}_k^{[actual]}$ of the $k=1,\ldots,N_{AL}$ landmarks between the inhale and exhale CTs. Averaging over all 55 landmarks of all patients, this resulted in 0.1 ± 1.7 mm in right-left (RL), -1.1 ± 1.7 mm in anterior-posterior (AP), and -7.0 ± 5.6 mm in inferior-superior (IS) direction, with 7.8 ± 5.1 mm for the 3D motion. After deformable registration, the corresponding registration residuals $\boldsymbol{R}_k$, Eq. (8), were determined. This resulted in -0.3 ± 0.8 mm in RL, 0.0 ± 0.9 mm in AP, and 0.1 ± 1.5 mm in IS direction averaged over all patients, with 1.6 ± 1.0 mm for the 3D residuals. Table I summarizes the results for the different patients. Obviously, for all three directions the residuals are considerably smaller than the breathing motion itself and in the order of the voxel size or smaller.

Following a plot visualization introduced by Coselmon *et al.*,[9] Fig. 8 gives a detailed overview of the observed RL/AP/IS shifts and the corresponding inhale-exhale registration residuals for all landmarks of all patients. As to be expected from Table I, most of the landmarks show small residuals in the order of the voxel size in the respective direction or smaller, with maximal residuals of −2.1/2.9/4.2 mm in the RL/AP/IS direction. As an additional finding, increased residuals are not systematically associated with larger shifts of landmarks (i.e., breathing motion itself).

To avoid a possibly biased evaluation of registration accuracy, the physician marking the anatomical points was asked to distribute the landmarks regularly over the lung volume. Thus, for each of the datapoints of the plots in Figs.

TABLE I. Mean and standard deviation of the inhale-exhale registration residuals $\boldsymbol{R}_k$, Eq. (8), and the breathing motion (displacements $\boldsymbol{\delta p}_k^{[\text{actual}]}$) according to the $k=1,\ldots,N_{AL}$ anatomical landmarks between the inhale and exhale state CTs. The final row gives mean and standard deviation taken over all landmarks of all patients.

		Inhale-exhale registration residuals of landmarks (mm) *[w/o registration, i.e., breathing motion (mm)]*			
Patient	N_{AL}	RL	AP	IS	3D
1	15	-0.3 ± 0.8	-0.1 ± 0.7	-0.4 ± 1.1	1.3 ± 0.9; max: 3.6
		(−0.3±0.9)	*(−1.3±1.5)*	*(−7.8±4.8)*	*(8.2±4.6; max: 15.4)*
2	11	0.2 ± 0.7	0.3 ± 1.2	0.0 ± 1.2	1.5 ± 1.0; max: 3.6
		(−0.6±0.9)	*(−1.7±1.1)*	*(−2.7±2.8)*	*(4.2±1.5; max: 6.1)*
3	14	-0.1 ± 0.9	-0.5 ± 0.8	0.5 ± 1.4	1.8 ± 0.7; max: 2.9
		(1.2±2.7)	*(−0.1±2.4)*	*(−9.4±6.0)*	*(10.4±5.5; max: 21.2)*
4	15	-0.7 ± 0.6	0.2 ± 0.7	0.2 ± 1.9	1.8 ± 1.3; max: 4.6
		(0.1±1.1)	*(−1.3±1.3)*	*(−7.2±6.1)*	*(7.8±5.7; max: 21.3)*
Average ($\Sigma N_{AL}=55$)		-0.3 ± 0.8	0.0 ± 0.9	0.1 ± 1.5	1.6 ± 1.0; max: 4.6
		(0.1±1.7)	*(−1.1±1.7)*	*(−7.0±5.6)*	*(7.8±5.1; max: 21.3)*

8(a)–8(c), the abscissa values were chosen to represent the relative RL/AP/IS positions in the considered lung lobe for the respective patient as defined by

$$\frac{p_{k,z}-\frac{1}{2}(Z_{\min}^{\text{lung}}+Z_{\max}^{\text{lung}})}{|Z_{\min}^{\text{lung}}-Z_{\max}^{\text{lung}}|} \tag{9}$$

for the relative IS position (and correspondingly for RL and AP). Here, $p_{k,z}$ is the IS coordinate of the kth landmark as manifest in the reference CT of the respective patient, and $Z_{\min}^{\text{lung}}(Z_{\max}^{\text{lung}})$ is the minimum (maximum) IS extent of the lung lobe in the reference CT. From the plots in Fig. 8, the positions of the landmarks used for evaluation were distributed fairly regularly, and no systematic association of the size of the residuals with the relative landmark position exists.

III.D. Calculation time

The calculation times of the test cases presented in Sec. III A are summarized in Table II. The step of local rigid registration of the featurelets (Sec. II A 2) was the most time-consuming part of the algorithm. However, this part can be parallelized, and thus a major speedup can be expected by the growing availability of multiprocessor/multicore computer hardware. To assess the efficiency of parallelization, the calculation time has been determined when using a varying number of CPUs (1/2/8 processors) on a multiprocessor PC; see Table II. This shows that parallelization of the local rigid registration step allows efficient speedup, which scales almost linearly with the number of available CPUs. Using all eight CPUs of the PC, the overall calculation time ranged from 14 s for the CT-CBCT test case to 38 s for the CT-CT test case.

For more details on the influence of the registration parameters on the calculation time, we refer to the Appendix.

IV. DISCUSSION

"Featurelet"-based deformable registration, i.e., the division of global deformable registration into rigid registration of a multitude of small image subvolumes ("featurelets") proves to be an efficient concept both for intra- and intermodal registration. The division into subproblems allows parallelization in a straightforward manner, which puts "online deformable registration" with calculation times well below 1 min on recent multiprocessor/-core computer hardware in reach.

The presented algorithm has been applied to exemplary patient test cases such as inhale-exhale registration of thorax CTs as well as CT-CBCT and CT-MRI registration of head-and-neck data sets, which resulted in visually satisfactory registration results. A rigorous quantitative evaluation for lung registration based on a virtual thorax phantom and manually placed landmarks in four patient data sets consistently resulted in a sub-voxel-size 3D alignment accuracy of 1–2 mm in terms of the mean and standard deviation of the 3D residuals (voxel dimensions: $1\times1\times3$ mm), as demanded for deformable registration applications in the field of radiotherapy.

Deformable registration algorithms have to consider two sources of information: image information and physical properties. Often, *a priori* information about tissue properties[32] and tissue boundary conditions[33] is not available or subject to uncertainties. It is the advantage of the presented featurelet-based algorithm that it allows model-independent registration in image regions with sufficient local image information. However, in regions with no or insufficient image information, it provides physically reasonable interpolation by means of minimization of elastic energy.

The basic premise of the featurelet concept is that the deformations present in a data set can be locally approximated by rigid registrations. Thus, featurelet-based registration in the form of the presented algorithm has its limits for organs and image regions with non-negligible local deformations, which deteriorate the possibility of local rigid (translatory) featurelet registration. In the case of inhale-exhale registration of the lung, we observed such problems mostly

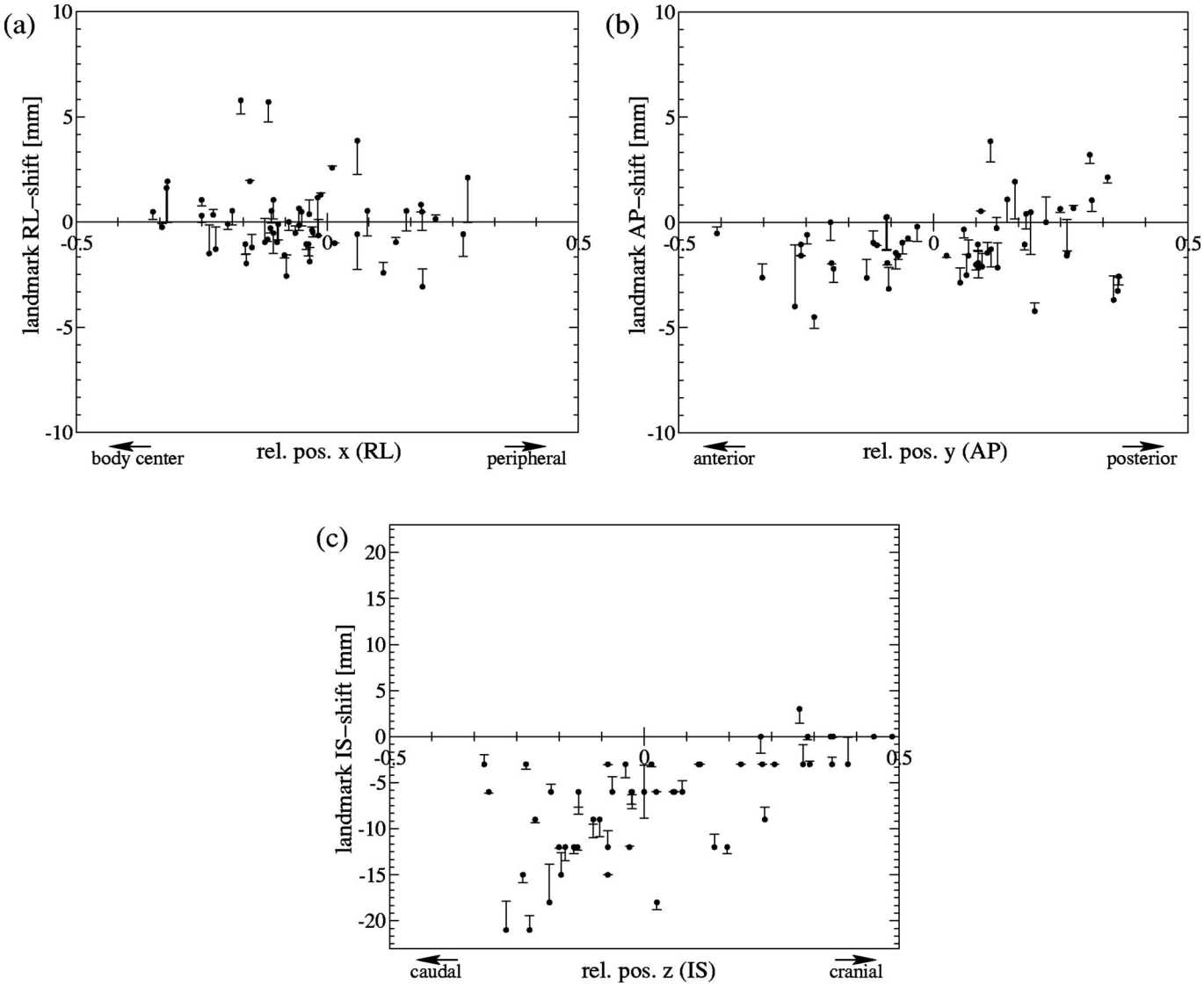

FIG. 8. Overview plot of the results of inhale-exhale registration for the anatomical landmarks of the four patients: The ordinate values of the dot markers give the displacements of the landmarks, while the (signed) lengths of the lines represent the values of the corresponding registration residuals, Eq. (8). The abscissa values are the relative positions of the landmarks, Eq. (9).

in the area of the diaphragm/diaphragm-chest wall interface (see, e.g., Fig. 4). Another phenomenon is the sliding of organ surfaces relative to adjacent structures, which is the case for the liver or lung relative to the chest wall. Thus, an efficient detection of featurelets in such "problematic" regions and modified algorithmic handling of local registration of these featurelets should be the subject of future developments to further improve the presented algorithm. Here, possible directions of research are featurelet transformations beyond translatory rigid registration, as well as hierarchical[34]

TABLE II. Calculation times for deformable registration of the test cases presented in Sec. III A using a dual-quadcore PC (2 Intel Xeon X5355 processors @ 2.66 GHz, 8 CPU cores altogether). The calculation times for local rigid registration (Sec. II A 2) and relaxation of the featurelet grid (Sec. II A 4) are specified separately. Additionally, the calculation time of the local rigid registration step has been determined separately using 1, 2, and 8 processors.

Test case	Registered region (voxels)	Rigid registration (s) (1/2/8 CPUs)	Relaxation (s)	Overall (s) (1/2/8 CPUs)
Thorax (CT-CT)	360×270×120	171/89/24	13	185/103/38
H&N (CT-CBCT)	225×225×115	64/33/9	4	69/38/14
H&N (CT-MRI)	378×210×70	120/62/16	2	122/64/19

TABLE III. Registration residuals inside of lung for NMI- or CC-based registration with different parameters used. μ_R and σ_R denote the mean and standard deviation of the 3D residuals, and $R_{X\%}$ is the X-percentile of the distribution of residuals ($\sim 7.8 \times 10^5$ voxels used for comparison). The last three columns show the number of featurelets used for registration of the right lung, the fraction of case I/II/III featurelets (see Sec. II A 3 and Fig. 2), and the calculation times (rigid registration/relaxation/overall) as determined on a dual-dualcore PC (2 Intel Xeon 5150 processors @ 2.66 GHz, 4 CPU cores altogether, 2 CPUs used for the calculations).

		Featurelet size XY/Z (voxel)	Search radius XY/Z (voxel)	NMI_{thresh} or CC_{thresh} [ ]	EV_{thresh} [ ]	$\mu_R \pm \sigma_R$ (mm)	$R_{95\%}/R_{99\%}/R_{99.9\%}$ (mm)	Number of featurelets	Featurelets case I/II/III (%)	Calculation time[a] (s)
	A			1.02		0.99 ± 0.94	2.86/4.39/7.00		44.1/55.4/0.5	14.5/3.6/19.2
	B			1.05	−0.02	0.97 ± 0.95	2.85/4.46/7.08		44.1/53.5/2.4	14.2/3.6/18.9
	C	11/5	9/4	1.15		0.97 ± 0.96	2.95/4.55/6.62	2677	44.1/44.9/11.0	14.1/3.2/18.5
	D			1.25		1.14 ± 1.08	3.36/4.91/6.81		38.6/30.2/31.2	14.3/2.5/17.9
	E				−0.005	1.06 ± 1.10	3.28/5.46/8.33		85.3/12.3/2.4	14.3/1.5/16.9
	F			1.05	−0.01	0.98 ± 1.00	2.99/4.88/7.82		71.3/26.3/2.4	14.4/2.2/17.8
	G				−0.04	1.09 ± 0.93	2.91/4.36/6.87		11.8/85.8/2.4	14.2/4.7/20.1
	H				−0.005	1.02 ± 1.01	3.06/4.88/7.53		85.2/12.0/2.8	13.7/1.2/16.0
N M I	I	15/5	9/4	1.05	−0.01	1.00 ± 0.96	2.97/4.58/6.71	1666	72.5/24.7/2.8	13.7/1.7/16.4
	J				−0.02	1.04 ± 0.99	3.05/4.57/6.70		45.7/51.5/2.8	13.7/2.7/17.5
	K				−0.04	1.20 ± 0.98	3.15/4.61/6.71		14.2/83.0/2.8	13.8/7.2/22.2
	L	19/9	9/4	1.05	−0.01	1.18 ± 1.13	3.52/5.19/7.34	713	69.8/26.6/3.5	15.9/1.5/18.4
	M	9/3	9/4	1.05	−0.03	0.97 ± 0.97	2.88/4.72/7.27	5755	19.9/78.1/2.0	19.2/7.0/27.5
	N		6/3			0.96 ± 0.89	2.80/4.25/6.11		73.5/23.6/2.9	8.4/1.4/11.0
	O	15/5	12/5	1.05	−0.01	1.06 ± 1.07	3.26/5.14/7.87	1666	71.7/25.1/3.2	27.2/1.9/30.2
	P		14/6			1.10 ± 1.19	3.42/6.06/9.38		71.4/25.6/2.9	40.2/2.1/43.5
	Q		15/8			1.10 ± 1.22	3.44/6.04/10.1		71.2/25.6/3.2	51.5/2.2/54.7
	R	9/3			−0.04	1.01 ± 1.00	2.97/4.66/7.61	5755	34.1/64.5/1.4	11.9/7.3/20.5
C C	S	11/5	9/4	0.35	−0.02	1.08 ± 0.98	2.98/4.63/7.40	2677	44.1/53.8/2.1	9.8/2.8/13.8
	T	15/5			−0.01	1.23 ± 1.13	3.46/5.24/7.49	1666	54.6/43.1/2.3	9.9/2.0/13.0
	U	19/9			−0.005	1.32 ± 1.12	3.57/5.08/7.34	713	59.6/37.6/2.8	12.1/1.6/14.8
		Input displacement field				2.92 ± 2.81	8.51/12.30/14.80		–	

[a]Notice that the calculation times of registration experiments with different featurelet sizes are not directly comparable due to the different size of the registered volumes (=number of featurelets × featurelet size).

or multigrid approaches that subdivide the domain used for calculation of the similarity measure to subfeaturelet regions.

V. CONCLUSIONS

Applications in "4D-radiotherapy" such as (4D) online adaptive planning, interactive 4D contouring on RCCT data sets, or image-guided (4D) treatment dose accumulation for proper treatment evaluation require near real-time deformable registration. This demand is addressed by the presented deformable registration algorithm, which is based on the principle of multiple fast local registrations of small featurelet regions and a subsequent novel correction scheme for mismatched featurelets. As shown by a number of intra- and intermodality image registration examples, its speed and parallelizability open up possibilities for "online deformable registration" with sufficient accuracy for radiotherapy applications.

In its present implementation the algorithm uses rectangular featurelet regions of finite fixed size and thus shows limitations in image regions with large local deformations. Future research will focus on overcoming such limitations, e.g., by the inclusion of models for specific applications. It is a unique advantage of featurelet-based registration that its principle does not depend on a model, so that maximum portability and versatility for use with different organ sites and image modalities is guaranteed.

ACKNOWLEDGMENTS

The authors would like to acknowledge D. Thorwarth (Section f. Biomedical Physics, University Hospital Tübingen) and V. Scheel (Department of Radiology, University Hospital Tübingen) for providing test cases for CT-MRI registration. The authors also wish to thank G. Mageras (Department of Medical Physics, Memorial Sloan-Kettering Cancer Center) for fruitful discussions about deformable registration. This work has been supported in part by Deutsche Krebshilfe e.V. Grant No. 106280, DFG Grant No. AL 877/1-2, and a grant by Siemens Medical Solutions, Forchheim, Germany.

APPENDIX: INFLUENCE OF MATCH PARAMETERS ON REGISTRATION ACCURACY AND CALCULATION TIME

Due to the comprehensive statistical information about registration accuracy that can be derived from voxelwise

evaluation, a virtual phantom is especially useful for parameter studies seeking to determine optimal parameter combinations of a deformable registration algorithm.

Using the (inhale deformed to) exhale CT as the reference and the (original) inhale CT as the target image for the featurelet-based registration algorithm, repeated registration experiments were performed under variation of the following parameters: featurelet size (the *X*- and *Y*-size was chosen to be equal for all experiments), size of rectangular search region (*X*- and *Y*-sizes chosen equal), NMI_{thresh} or CC_{thresh}, and EV_{thresh} (Sec. II A 3). The results for NMI- and CC-based registration are presented in Table III. For these studies, the registration accuracy was determined in terms of 3D residuals for each voxel inside the right lung of the virtual phantom and is summarized as mean μ_R and standard deviation σ_R over all lung voxels, as well as the 95%-, 99%-, and 99.9% percentiles. Additionally, Table III specifies the fraction of case I/II/III-featurelets (Sec. II A 3) and the calculation time for each parameter set.

Experiments A–Q in Table III used NMI as a similarity measure. In experiments A–D the influence of the parameter NMI_{thresh} on the registration accuracy was investigated. It turns out that the latter is fairly insensitive to this parameter as long as it does not exceed a value of ~1.15, which obviously is the typical NMI "noise level" for this registration problem: Generally, if the similarity measure threshold is set too high, unnecessarily many featurelets are erroneously classified as "case III" featurelets ["noisy match;" compare Sec. II A 3 and Fig. 2(i)], thereby possibly ignoring available local image information in the subsequent relaxation step. To be on the safe side, we chose $NMI_{thresh}=1.05$ for the following experiments.

In experiments E–G (H–K) the influence of the parameter EV_{thresh} was investigated. Again, the registration accuracy turns out to be fairly insensitive to this parameter within a certain range. However, to obtain optimal registration quality a slightly different value of EV_{thresh} should be chosen for different featurelet sizes: For a featurelet size of $11\times11\times5$ voxels/$15\times15\times5$ voxels the optimal value for EV_{thresh} is ~−0.02/−0.01, respectively.

Concerning the influence of the featurelet size on the registration accuracy, a comparison of experiments B, I, L, and M shows that for large featurelets the residuals increase: The validity of the approximation that deformations are small over the domain of a featurelet is compromised for too large featurelets, which leads to misregistrations. Note that the calculation time increases both for very small and large featurelet sizes. Thus, parameter set I (featurelet size $15\times15\times5$ voxels) appears as a good compromise of registration accuracy and calculation time.

Experiments N–Q show that the registration accuracy moderately depends on the size of the search region (due to possible multiple local optima of the similarity measure function, if similar structures exist in the vicinity of a anatomical feature).

Based on this, we used a featurelet size of $15\times15\times5$ voxels, $NMI_{thresh}=1.05$, and $EV_{thresh}=-0.01$ as parameters for the CT-CT and CT-CBCT registrations of patient data presented in Sec. III.

Similar parameter studies were performed for CC as a similarity measure. The results are shown as experiments R–U in Table III, which already represent parameter sets with optimal EV_{thresh} for a given featurelet size. A value of ~0.35 was found to be a reasonable value for CC_{thresh} in all experiments. According to these experiments, smaller featurelet sizes are favorable for CC-based registration (parameter sets R, S). With respect to calculation time, CC-based registration is typically 20%–30% faster compared to NMI-based registration. Notice, however, that the applicability of CC as a similarity measure is restricted to intramodality registration problems.

[a]Electronic mail: matthias.soehn@med.uni-tuebingen.de

[1]D. Sarrut, "Deformable registration for image-guided radiation therapy," Z. Med. Phys. **16**, 285–297 (2006).
[2]D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," Phys. Med. Biol. **46**, R1–R45 (2001).
[3]T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: A survey," Med. Image Anal. **1**, 91–108 (1996).
[4]W. Lu, M.-L. Chen, G. H. Olivera, K. J. Ruchala, and T. R. Mackie, "Fast free-form deformable registration via calculus of variations," Phys. Med. Biol. **49**, 3067–3087 (2004).
[5]B. Zitová and J. Flusser, "Image registration methods: A survey," Image Vis. Comput. **21**, 977–1000 (2003).
[6]D. Yan, D. A. Jaffray, and J. W. Wong, "A model to accumulate fractionated dose in a deforming organ," Int. J. Radiat. Oncol. Biol. Phys. **44**, 665–675 (1999).
[7]B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif. Intell. **17**, 185–203 (1981).
[8]D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," IEEE Trans. Med. Imaging **18**, 712–721 (1999).
[9]M. M. Coselmon, J. M. Balter, D. L. McShan, and M. L. Kessler, "Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines," Med. Phys. **31**, 2942–2948 (2004).
[10]V. Pekar, E. Gladilin, and K. Rohr, "An adaptive irregular grid approach for 3D deformable image registration," Phys. Med. Biol. **51**, 361–377 (2006).
[11]R. Shekhar, P. Lei, C. R. Castro-Pareja, W. L. Plishker, and W. D'Souza, "Automatic segmentation of phase-correlated CT scans through nonrigid image registration using geometrically regularized free-form deformation," Med. Phys. **34**, 3054–3066 (2007).
[12]T. Guerrero, G. Zhang, T.-C. Huang, and K.-P. Lin, "Intrathoracic tumour motion estimation from CT imaging using the 3D optical flow method," Phys. Med. Biol. **49**, 4147–4161 (2004).
[13]J. P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," Med. Image Anal. **2**, 243–260 (1998).
[14]G. E. Christensen, B. Carlson, K. S. C. Chao, P. Yin, P. W. Grigsby, K. Nguyen, J. F. Dempsey, F. A. Lerma, K. T. Bae, M. W. Vannier, and J. F. Williamson, "Image-based dose planning of intracavitary brachytherapy: Registration of serial-imaging studies using deformable anatomic templates," Int. J. Radiat. Oncol. Biol. Phys. **51**, 227–243 (2001).
[15]M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi, "Large deformation three-dimensional image registration in image-guided radiation therapy," Phys. Med. Biol. **50**, 5869–5892 (2005).
[16]W. Lu, G. H. Olivera, Q. Chen, K. J. Ruchala, J. Haimerl, S. L. Meeks, K. M. Langen, and P. A. Kupelian, "Deformable registration of the planning image (kVCT) and the daily images (MVCT) for adaptive radiation therapy," Phys. Med. Biol. **51**, 4357–4374 (2006).
[17]H. Wang, L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, D.

A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," Phys. Med. Biol. **50**, 2887–2905 (2005).
[18]T. Zhang, Y. Chi, E. Meldolesi, and D. Yan, "Automatic delineation of on-line head-and-neck computed tomography images: Toward on-line adaptive radiotherapy," Int. J. Radiat. Oncol. Biol. Phys. **68**, 522–530 (2007).
[19]W. R. Crum, C. Tanner, and D. J. Hawkes, "Anisotropic multi-scale fluid registration: Evaluation in magnetic resonance breast imaging," Phys. Med. Biol. **50**, 5153–5174 (2005).
[20]A. Pevsner, B. Davis, S. Joshi, A. Hertanto, J. Mechalakos, E. Jorke, K. Rosenzweig, S. Nehmeh, Y. E. Erdi, J. L. Humm, S. Larson, C. C. Ling, and G. S. Mageras, "Evaluation of an automated deformable image matching method for quantifying lung motion in respiration-correlated CT images," Med. Phys. **33**, 369–376 (2006).
[21]Y. H. Lau, M. Braun, and B. F. Hutton, "Non-rigid image registration using a median-filtered coarse-to-fine displacement field and a symmetric correlation ratio," Phys. Med. Biol. **46**, 1297–1319 (2001).
[22]M. Söhn, M. Birkner, D. Yan, and M. Alber, "Automatic deformable registration of the lung for 4D respiratory correlated CT (RCCT) datasets," Radiother. Oncol. **76** (Suppl.), S59 (2005).
[23]U. Malsch, C. Thieke, P. E. Huber, and R. Bendl, "An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy," Phys. Med. Biol. **51**, 4789–4806 (2006).
[24]P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," Int. J. Comput. Vis. **24**, 137–154 (1997).
[25]F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," IEEE Trans. Med. Imaging **16**, 187–198 (1997).
[26]J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual information based registration of medical images: A survey," IEEE Trans. Med. Imaging **22**, 986–1004 (2003).
[27]C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," Pattern Recognit. **32**, 71–86 (1999).
[28]P. Filev, L. Hadjiiski, B. Sahiner, H.-P. Chan, and M. A. Helvie, "Comparison of similarity measures for the task of template matching of masses on serial mammograms," Med. Phys. **32**, 515–529 (2005).
[29]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[30]K. M. Brock, J. M. Balter, L. A. Dawson, M. L. Kessler, and C. R. Meyer, "Automated generation of a four-dimensional model of the liver using warping and mutual information," Med. Phys. **30**, 1128–1133 (2003).
[31]E. C. Ford, G. S. Mageras, E. Yorke, and C. C. Ling, "Respiration-correlated spiral CT: A method of measuring respiratory induced anatomic motion for radiation treatment planning," Med. Phys. **30**, 88–97 (2003).
[32]Y. Chi, J. Liang, and D. Yan, "A material sensitivity study on the accuracy of deformable organ registration using linear biomechanical models," Med. Phys. **33**, 421–433 (2006).
[33]J. Liang and D. Yan, "Reducing uncertainties in volumetric image based deformable organ registration," Med. Phys. **30**, 2116–2122 (2003).
[34]B. Likar and F. Pernuš, "A hierarchical approach to elastic registration based on mutual information," Image Vis. Comput. **19**, 33–44 (2001).

Appendix F

IMRT optimization in a quasi-periodically deforming patient model

accepted for publication in the
International Journal of Radiation Oncology, Biology and Physics

IMRT optimization in a quasi-periodically deforming patient model

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

Martin Weinmann
Department of Radiotherapy, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany

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

Purpose: The common method of IMRT-planning in the presence of geometric uncertainties uses a static CT dataset for dose calculation as substitute for dose accumulation in the moving tissue. We present the implementation of a probability-based 4D-planning approach which explicitly optimizes the accumulated dose-to-moving tissue estimated using the patient's probability density function (pdf) of respiratory motion. This is termed "optimization in tissue-eye-view" (TEV).

Materials and Methods: The method incorporates 4D-Monte Carlo (MC) dose calculation in multiple geometries of a respiratory correlated CT dataset. The instance doses are weighted according to the breathing pdf and accumulated in a common reference geometry which involves dose warping based on deformable registration. The algorithm produces deliverable MLC segments and was tested on an example lung cancer patient dataset with large target excursion. Accumulated doses of the moving target and OARs of this plan (4Dp) were compared to corresponding margin-based static IMRT-plans for freebreathing (FBp) and gated (GATINGp) treatment as well as target tracking (TRACKp).

Results: TRACKp provided best target coverage. Both 4Dp and GATINGp gave similar results for target coverage and lung dose, with significantly better target coverage than for FBp.

Conclusions: The presented 4D-planning concept offers an alternative to gating by providing the optimal dose for freebreathing IMRT treatment. While the focus of this study was 4D-lung planning, the approach can be generally applied for IMRT optimization in randomly deforming patient models.

Keywords: 4D-radiotherapy, IMRT, deformable geometries, deformable image registration, 4D-Monte Carlo dose calculation

I. INTRODUCTION

The currently most common planning approach employs a single CT scan to geometrically define the gross-tumor volume (GTV) and organs at risk (OARs) to determine the dose distribution. However, the existence of inevitable geometrical uncertainties such as rigid and non-rigid interfractional setup errors or intrafractional organ motion causes a discrepancy between the dose delivered to the moving tissue and computations performed on the static geometry. The common method to account for this is the extension of the clinical target volume (CTV) by a margin, and to ensure coverage of the resulting larger planning target volume (PTV) (ICRU report 50[1]). With regard to internal organ motion (eg. breathing) the ICRU report 62[2] additionally introduced the concept of the internal target volume (ITV) as hull of possible tumor positions relative to bony structures. By using the PTV as surrogate merely for planning, CTV coverage is sought to be guaranteed in the presence of uncertainties.

A large PTV margin, however, goes at the expense of irradiating additional healthy tissue. Moreover, it may happen that the PTV overlaps with adjacent OARs. In this case an ill-defined tradeoff decision between PTV coverage and OAR sparing has to be made. Finally, the very position uncertainty of the target can have an impact on the density distribution, rendering the dose calculation itself uncertain. A pragmatic and efficient way for margin reduction is the use of image guidance (IGRT), optionally combined with adaptive concepts to individualize treatment margins[3]. However, remaining intrafractional uncertainties of random nature (eg. prostate) or periodic nature (freebreathing treatment of lung) necessitate non-vanishing margins.

The problems encountered in PTV-based planning are a consequence of planning on a static patient geometry, which is a rather incomplete model of reality in the presence of large geometrical deviations. The actually relevant quantity is the *accumulated dose* in the moving/deforming CTV/OAR-structures. The accumulated dose changes physically – due to changes in the density distribution – as well as physiologically, as tissue voxels move away from the planning position.

As first step from the static 'beam-eye-view' (BEV) to this 'tissue-eye-view' (TEV), it was realized already 1987 by Leong[4] that random errors lead to blurring of the dose distribution as seen in TEV. For random positional errors, this can be approximately modeled

by dose-convolution[5], fluence-convolution[6] or isocentre-sampling[7]. These methods model only the effect of rigid body motion on dose. Ultimately the dose accumulated by the moving tissue needs to be estimated, which requires information about local displacement vectors between different geometries. Dose distributions of different geometries are then 'warped' and accumulated to a reference geometry. Such TEV-based dose evaluation has first been proposed by Yan *et al.* 1999[8]. The combination with the use of Monte Carlo (MC) dose calculation represents a powerful tool for evaluation of '4D-doses' especially for regions with large tissue density heterogeneities like lung[9,10].

The possibility to account explicitly for accumulated dose in deforming geometries enables '4D-planning'. As at the time of planning the exact individual motion pattern during the treatment course is unknown, such approaches have to make assumptions about the probability density function (pdf) of motion. The basic idea of pdf-based TEV-optimization then is to use the *expectation value* of dose as planning surrogate.
Such an approach was first implemented for real patient data in terms of an adaptive framework by Birkner *et al.*[11] for prostate, bladder and rectum. The dose distribution itself was calculated only in the planning geometry, which is a good approximation for the pelvic region. However, for head-and-neck (H&N), lung cases etc. with large inhomogeneities and tissue-air interfaces this is not a valid approximation. McShan *et al.*[12] presented an implementation of pdf-based TEV-optimization called MIGA (multiple instance geometry approximation), where dose is explicitly calculated for each geometry instance. This was presented on a H&N case for rigid shifts. Trofimov *et al.*[13] applied a method similar to MIGA for planning of a lung and liver case, where organ deformations as obtained from deformable registration were considered. Zhang *et al.*[14] presented 4D inverse lung planning using multiple geometries of a respiratory-correlated CT-dataset for breathing-pdf-weighted dose accumulation based on deformable registration. Differently from MIGA, dose calculation was performed on a single geometry (average density CT) for speed-up reasons.

The approaches mentioned above presented non-segmented plans using pencil beam (PB) or convolution superposition (CVSP) algorithms for beamlet dose calculation. In this study we present an implementation of pdf-based TEV-optimization with MC-based dose calculation in multiple geometries, resulting in a fully segmented freebreathing '4D-plan'. For dynamic MC the calculation time does not scale with the number of geometry instances[10]. This facilitates 4D-planning with MC accuracy in time frames comparable to conventional static planning.

In the following, we apply the algorithm to the problem of freebreathing lung treatment for a clinical test case with large breathing excursion. The results of 4D-planning in terms of accumulated dose to tumor and lung are compared to margin-based static IMRT-plans for freebreathing and gated treatment as well as tracking.

II. MATERIALS AND METHODS

A. Concepts

1. IMRT dose calculation in a static geometry

Given a fluence distribution Φ in terms of the fluence weights $\{\phi\}$ of n beamlets or segments as discrete fluence elements, the dose D_i to voxel i in the fixed treatment room coordinate system is given by the weighted sum of contributions from all fluence elements as expressed by the dose operator T:

$$D_i = \sum_{j=1}^{n} T_{ij}\phi_j \qquad (1)$$

T_{*j} holds the *static* dose distribution of fluence element j, which is determined based on the density information of a single planning geometry. Therefore, T is the *static* dose operator.

2. Representation of the patient's dynamics by multiple deformable geometries

For lung patients, the respiratory-correlated CT (RCCT) technique allows to reconstruct CT geometries of different respiratory phases which provides a straightforward way to represent the dynamics of patient individual breathing motion in terms of multiple phase CTs. For this study the RCCT dataset was aquired using a Siemens Somaton Open in spiral RCCT-mode. The scanner software allows separate, amplitude-based reconstruction of phases in the inhale ('In') and exhale ('Ex') branch of the breathing curve for user-defined respiratory levels in the range $0 \ldots 100$ (0=full expiration, 100=full inspiration).

TEV-optimization uses information about motion of all tissue elements within the relevant structures to enable calculation of dose-to-moving-tissue. This requires deformable registration providing displacement fields that relate image voxels of different RCCT-geometries. For this study, a featurelet-based deformable registration algorithm[15] was used.

3. Probabilistic description of breathing motion

As in fig. 1a, significant irregularities in the breathing curve can be observed over the time of a treatment fraction, which renders deterministic modelling of respiratory motion for calculation of dose-to-moving-tissue

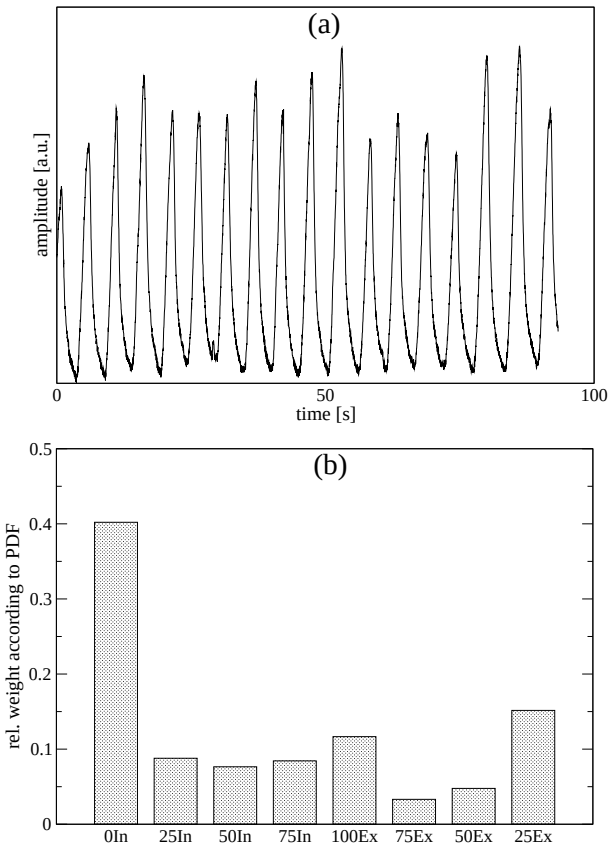

FIG. 1: (a) Breathing curve of the example patient during aquisition of the RCCT. (b) Relative time spend in the different breathing phases of the RCCT dataset as derived from the pdf of the breathing curve in (a).

infeasible. However, a probabilistic description of breathing in terms of the probability density function (pdf) of the breathing curve provides information about the relative time spent in each breathing phase *on average*. Thus each phase CT can be associated with a weight factor expressing the relative importance of the respective instance geometry to represent a 'snapshot' of the underlying dynamic process (fig. 1b).

4. IMRT dose calculation in multiple deformable geometries

The 4D-planning concept presented here optimizes the *expected* dose accumulated by the moving tissue elements.

For this, the dynamics of the patient geometry is discretized into N instances (sect. II A 2). Each of these instances is associated with an instance weight w_k ($w_k < 1$, $\sum_{k=1}^{N} w_k = 1$; sect. II A 3).
The instance geometries are used for determining the (static) dose operators T^k and thereby the static dose distributions D^k (eq. 1). Chosing one of the instance geometries ($k = 1$, say) as reference geometry, dose accumulation is realized by *warping* the instance doses to the reference geometry using the displacement vectors between corresponding tissue elements (sect. II A 2). The accumulated dose $\langle D \rangle$ can then be expressed as weighted average:

$$\langle D \rangle = \sum_{k=1}^{N} w_k \cdot \tilde{D}^k = \sum_{k=1}^{N} w_k \cdot (\mathcal{W}_k \circ D^k) \qquad (2)$$

Here $\tilde{D}^k$ is the k-th instance dose as warped to the reference geometry, which can be formalized by introduction of a *warping operator* $\mathcal{W}_k$ (with $\mathcal{W}_1 = \mathrm{Id}$ as identity operator).
This is illustrated in fig. 2 for the example of a beamlet dose distribution: Fig. 2a/b show (static) dose distributions D^{exhale} and D^{inhale} in the treatment room coordinate system. The effect of the warping operator on D^{inhale}, i.e. $\tilde{D}^{\mathrm{inhale}}$, is shown in fig. 2c: This is the dose deposited in the tissue voxels in the inhale geometry as in fig. 2b, but represented in the reference coordinate system of the exhale geometry. Finally, fig. 2d shows the accumulated dose $\langle D \rangle$.

For the purposes of IMRT optimization, eq. 2 can be rewritten in terms of the fluence weights by expressing the instance doses D^k through eq. 1:

$$\langle D_i \rangle = \sum_{k=1}^{N} w_k \cdot \sum_{j=1}^{n} (\mathcal{W}_k \circ T^k)_{ij} \phi_j = \sum_{j=1}^{n} \langle T_{ij} \rangle \phi_j \qquad (3)$$

is the accumulated dose in voxel i of the reference geometry with

$$\langle T_{ij} \rangle = \sum_{k=1}^{N} w_k \cdot (\mathcal{W}_k \circ T^k)_{ij}. \qquad (4)$$

$\langle T \rangle$ is the motion-averaged version of the static dose operator T (fig. 2d).

a. Beamlet precalculation For beamlet-based IMRT-optimization of the fluence distribution, $\langle T_{ij} \rangle$ can be precalculated by separate warping of all beamlets according to eq. 4. This allows efficient update of the accumulated dose distribution $\langle D \rangle$ according to eq. 3 during iterative optimization.

b. 4D-MC dose calculation For MC algorithms calculation time depends on the overall number of simulated particle histories n_{hist} which is split according to the instance weights w_k among the dose calculations for the multiple instances, i.e. $n_{hist,k} = w_k \cdot n_{hist}$. Thus MC allows efficient calculation of the accumulated dose essentially independent of the number of instances.

5. Constrained optimization based on isoeffects

The isoeffect framework is used for the optimization costfunctions to quantify dose-response characteristics of different tissues[16,17]. The quantities used for the purpose

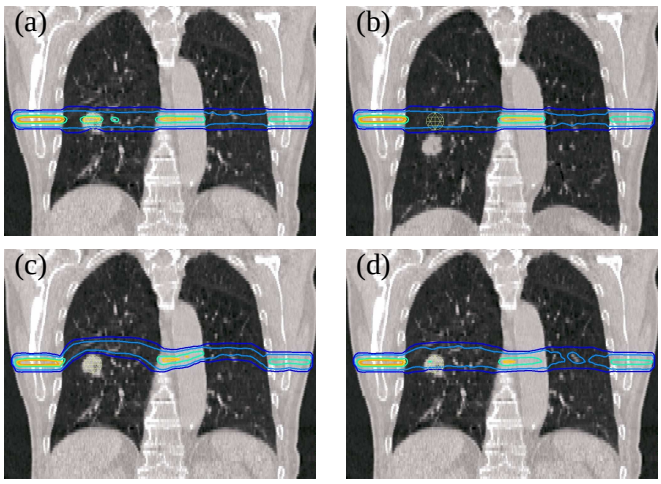

FIG. 2: Beamlet warping: (a)/(b) beamlet doses calculated in the exhale/inhale CT. (c) beamlet dose of inhale CT warped to exhale reference geometry; (d) beamlet dose as accumulated using all 8 CTs.

of this study are:

- *poisson cell-kill EUD* (equivalent uniform dose) to describe tumor control[18]
- *normal tissue parallel complication model*, quantifying the *mean damage* to an organ caused by irradiation[16,17]
- *maximum dose*, controlled by a physical overdose constraint in terms of the root mean square (rms) overdosage above a threshold dose.

In order to ensure the existence of a solution for a given set of prescriptions, we implement IMRT-optimization in terms of a *constrained optimization* problem[19], where local tumor cell survival is minimized (i.e. tumor control is maximized) while isoeffects to critical structures are constrained to maximum tolerable values.

B. Implementation

The methods described in this study were implemented into the clinical IMRT-planning software Hyperion, which allows constrained, EUD-based biological optimization based on MC-dose distributions[20]. In short, Hyperion uses a two-stage ansatz of IMRT-optimization: The first stage optimizes the fluence map based on a discretization into beamlets, where beamlet dose calculation is performed using a dedicated pencil-beam algorithm with 3D-density corrections[21]. For the second stage, the fluence distribution is segmented, and henceforth both segment-weights and -shapes are iteratively optimized based on MC-calculated segment doses[22]. For the latter the fast MC-code XVMC is used[23,24].

C. Patient test case

We tested the presented algorithm for IMRT-planning of a moving lung tumor, where the intrafractional quasi-periodic breathing motion was explicitly incorporated into optimization.
Planning was based on a RCCT dataset with eight CTs (0/25/50/75In and 100/75/50/25Ex, sect. II A 2; voxel size: $1 \times 1 \times 3$mm). The corresponding instance weight factors are given in fig. 1b. The tumor had a diameter of 2.4cm and was located posteriorily in the lower right lung (see figs. 2, 3). The breathing excursion was 2.9cm, mainly in cranio-caudal (CC) direction.

To allow comparison of the 4D-planning method to other approaches, the following four plans were created:

a. 4D-plan (4Dp): In this approach, the expected dose to the CTV was optimized by explicit dose-calculation in all eight breathing phases and pdf-based dose accumulation as described in sect. II A 4. The reference geometry was the exhale CT.

b. static freebreathing PTV-plan (FBp): The PTV of this plan was the ITV which incorporated the CTV-positions in all eight RCCT phases. Dose-calculation was based on the *average density CT* (fig. 3a) which was constructed by weighted averaging for each voxel the Hounsfield-values of all RCCT phases (weight factors w_k).

c. static exhale-gating PTV-plan (GATINGp): The PTV of this plan was given by an ITV that covered the CTV-positions in the RCCT-phases 0In, 25In and 25Ex. The latter two phases were considered to incorporate inevitable geometric uncertainties of the exhale position during the gating window. Dose-calculation was based on the *exhale-average density CT* resulting from weighted averaging of the 0In/25In/25Ex-CTs (fig. 3d).

d. tracking plan (TRACKp): To simulate target tracking, an optimal static IMRT-plan was determined separately for each of the instance geometries, with the target for each plan given by the CTV in the respective phase CT. The instance dose distributions of the resulting eight plans were retrospectively warped and accumulated to the reference phase. This idealized 'tracking plan' is based on the assumption that continuous 'live'-tracking without residuals is possible.

For all plans, the availability of imaging devices for daily online, target-based setup correction is assumed, and residual setup errors are neglected as simplification. In particular, no additional setup margin were applied.

The prescription and other planning parameters were chosen equally for each planning method to guarantee comparability. The prescription was as follows:

- *target:* 55.0Gy in 10 fractions to the respective target (prescribed as poisson cell-kill EUD with $\alpha = 0.5$); overdosage constraint: max. 2Gy rms overdosage above 55Gy.
- *right (ipsilateral) lung (volume without target):* parallel contraint: max. 5% mean damage ($k = 3$,

d_0 = 20Gy); overdosage constraint: max. 0.05Gy rms overdosage above 20Gy for voxels with $\geq$ 2cm distance to target.

- *left (contralateral) lung:* parallel constraint: max. 1% mean damage.
- *unspecified tissue inside of skin contour:* overdosage constraint: max. 0.1Gy rms overdosage above 20Gy.

Eleven beam angles were chosen initially for each plan (20/155/175/195/ 215/235/270/295/310/325/345°), but the optimizer automatically switched off unfavourable beams with low contribution to the plan. The parameters for the sequencer were chosen such as to perform conversion of the beamlet-based fluence maps into segments with high fidelity.

Other planning parameters were: $3\times3\times3\text{mm}^3$ voxel size for dose calculation; $4\times2\text{mm}^2$ beamlet size; 6MV photon beams; $\sim 3.5\cdot10^6$ simulated histories/segment (corresponding to 3% MC-accuracy/segment) assuming Elekta BeamModulator with 4mm leaf width as MLC hardware.

III. RESULTS

A. Dose distributions

The planning geometry and MC-dose distributions for the different plans are shown in fig. 3, the DVHs are plotted as fig. 4 and table I summarizes the resulting isoeffects and EUDs. For the two PTV-based plans two types of dose distributions are specified: the static dose calculated in the average CT as used for planning and the accumulated dose to allow for comparison.

Following the isotoxic prescription, all plans show very similar OAR doses, while differences mainly reflect in the target dose (fig. 4, tab. I). None of the plans fully met the presciption of 55.0Gy EUD due to the strict OAR constraints, however, significant differences were found. The best coverage was found for TRACKp (53.0Gy accumulated CTV-EUD) as could be expected for this idealized planning approach. In contrast, the lowest target dose of all plans was found for FBp with 45.1Gy static EUD to the PTV. For this planning method the PTV encompassed a relatively large region and thus could not be covered well for the given constraints. Recalculation of this plan on the different RCCT-phases showed that the accumulated EUD to the CTV was actually higher by about 2Gy, however, significant target overdosage and dose inhomogeniety was found for the CTV. A similar difference between static PTV-dose and accumulated CTV-dose was found for GATINGp (tab. I). This renders the static PTV-dose as poor predictor of the actual dose-to-moving-tumor. Due to its smaller PTV region, GATINGp shows a significantly better target coverage (50.9Gy accumulated CTV-EUD) than FBp. A similar target coverage of 50.2Gy was found for 4Dp. Compared to FBp this is a significant improvement without exceeding the respective overdosage constraint, which also reflects in a better and steeper DVH of 4Dp.

Fig. 3 gives an insight into how the approaches redistribute dose to arrive at different target coverage under isotoxic conditions. For the two PTV-based plans (FBp and GATINGp) the static dose distribution conforms to the respective PTV-region as prescribed (figs. 3a,d). As the gating-PTV is significantly smaller than the FB-PTV, less volume of lung as main dose-limiting organ is irradiated, facilitating higher target doses as compared to the FBp. Breathing motion causes lung tissue to move in and out of the irradiated PTV-region, while only the moving CTV stays inside. This leads to considerable dose blurring for the accumulated dose, manifest as spreadout of the low-dose region in CC-direction while the higher doses are confined to a smaller region around the CTV (figs. 3b,e). The latter can be partly contributed to the fact that significant dose-buildup can happen only inside the high tissue density of the moving CTV, thus dose 'follows' the tumor.
Compared to FBp, 4Dp conforms the accumulated dose significantly better to the CTV (fig. 3c). While the dose-blurring effect in CC-direction is reduced, thereby sparing tissue in the ITV-region, a larger volume anteriorily and posteriorily of the CTV receives doses between 10-30Gy, effectively rendering 4Dp isotoxic to FBp (tab. I).

B. Fluence distributions

Fig. 5 shows fluence distributions from a number of beams for FBp and 4Dp. While FBp has locally increased fluence mainly at the field edges (penumbra sharpening), 4Dp typically concentrates fluence in areas with highest probability of CTV occupancy (the breathing phases around exhale, see fig. 1b) as seen in the respective BEV projection. This can be interpreted as fluence redistribution between breathing phases, where underdosage during phases that are occupied only short in time (phases around inhale) is compensated for by increased fluence in regions of high CTV occupancy.

The approaches resulted in deliverable plans with 9-11 beams and ∼60-100 segments. We computed the total delivered radiation energy $E_{rad,rel}$ (product of monitor units and segment area, summed up for all segments) relative to TRACKp, which had the minimal value. While 4Dp is similarly efficient to TRACKp in terms of total energy ($E_{rad,rel}$ = 1.02), both static PTV-based plans resulted in $E_{rad,rel} = 1.06$.

C. Numerical performance

To quantify the numerical overhead of the 4D-planning method as compared to conventional static margin-based planning, we found an increase of about 35% in calcula-

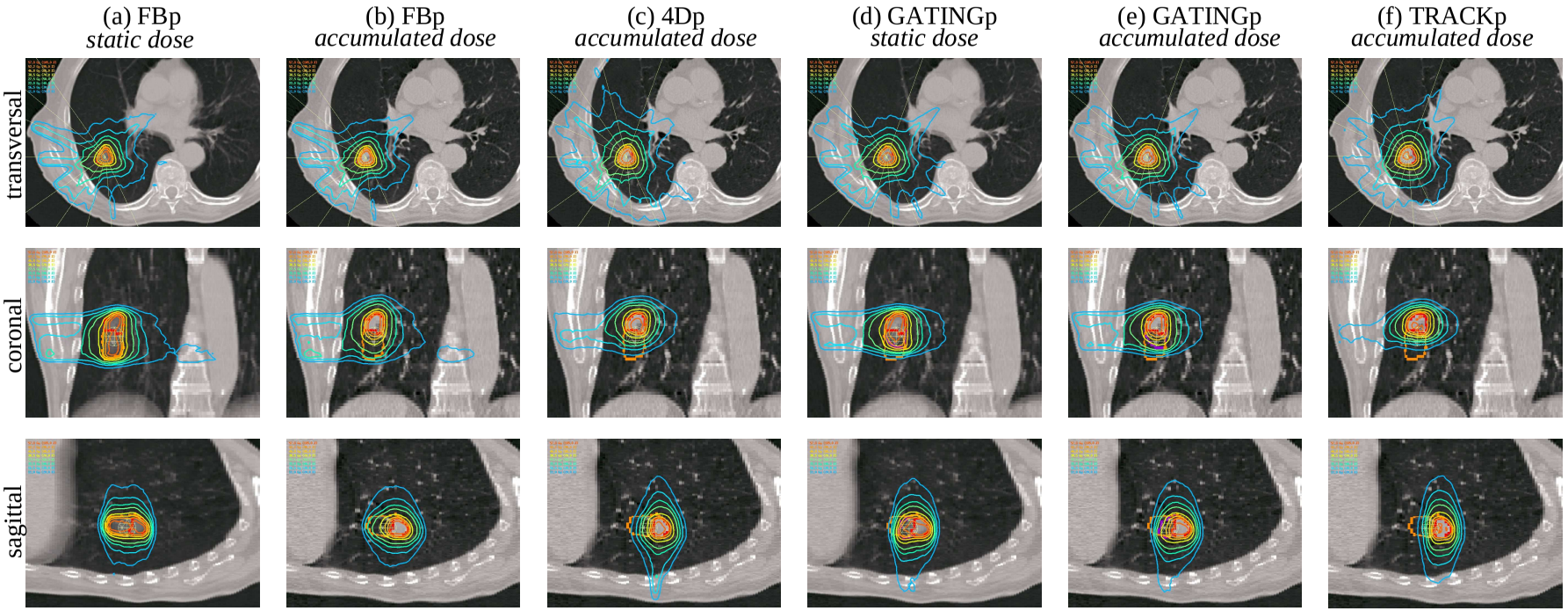

FIG. 3: MC-dose distributions for the different plans. [isodose lines: 57.8/52.2/49.5/44.0/38.5/22.0/16.5/11.0Gy; contours: PTV (orange), gating-PTV (purple; only in (d),(e)) and CTV_{exhale} (red)].

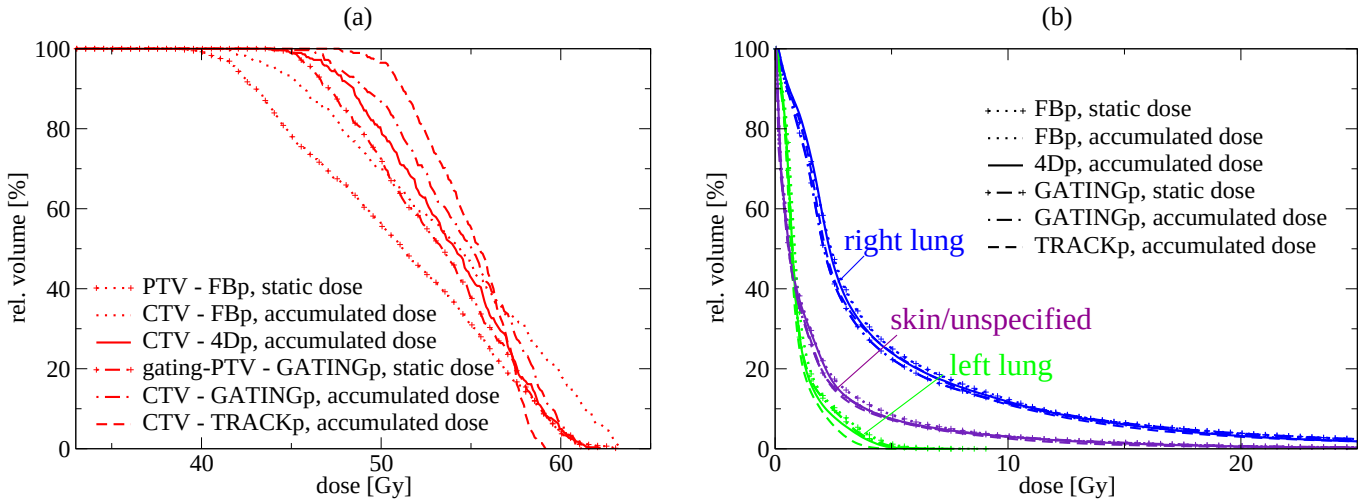

FIG. 4: DVHs of the dose distributions shown in fig. 3: (a) target DVHs; (b) DVHs of the OARs.

	objective	*constraints*				
	target	target	right lung		left lung	skin/unspecified
	poisson EUD [Gy]	*rms overdosage >55Gy [Gy]*	*parallel, mean damage [%]*	*rms overdosage >20Gy* [Gy]*	*parallel, mean damage [%]*	*rms overdosage >20Gy [Gy]*
prescription	*55.0*	*2.0*	*5.0*	*0.05*	*1.0*	*0.10*
FBp						
–static	45.1	1.9	5.1	0.05	0.1	0.10
–accumulated	47.4	3.2	5.0	0.18	0.1	0.10
4Dp						
–accumulated	50.2	2.0	5.0	0.05	0.1	0.10
GATINGp						
–static	49.3	1.9	5.0	0.05	0.1	0.10
–accumulated	50.9	2.3	5.0	0.11	0.1	0.10
TRACKp						
–accumulated	53.0	1.6	4.7	0.02	0.0	0.04

TABLE I: Prescribed and resulting isoeffects/EUDs for the different plans. For FBp and GATINGp, the values denoted by 'static' refer to the static dose distributions calculated on the respective average CTs as used for optimization (target: freebreathing-PTV or gating-PTV, respectively). All other values refer to the accumulated dose distribution (target: CTV). [*applied only to voxels with ≥2cm distance to the target]

tion time when comparing 4Dp with FBp. The main cause for the prolonged calculation time is the initialization of the beamlets with the pencil-beam algorithm, which in case of the 4D-approach involves explicit beamlet dose-calculations in all instance geometries and warping to the reference geometry.

IV. DISCUSSION AND CONCLUSIONS

The 4D-planning concept of optimization in 'tissue-eye-view' (TEV) for freebreathing lung treatment facilitates direct IMRT-optimization of accumulated dose-to-moving-tissue by explicit optimization in multiple instance geometries. The concept formulates treatment planning as pdf-based probabilistic optimization problem[11–14,25,26] (and references therein), using the patient-individual breathing-pdf to describe quasiperiodic respiratory uncertainties. Deformable registration as central tool is used to warp and accumulate the separately calculated and pdf-weighted instance doses to a common reference geometry. In this way, the beamlet/segment-doses are locally convolved with the local respiratory uncertainties. IMRT-optimization is then based on these 'blurred' beamlet/segment-doses which represent the expectation value of the accumulated dose-to-moving-tissue. It is the strength of this approach that it removes the two main paradoxes of conventional PTV-

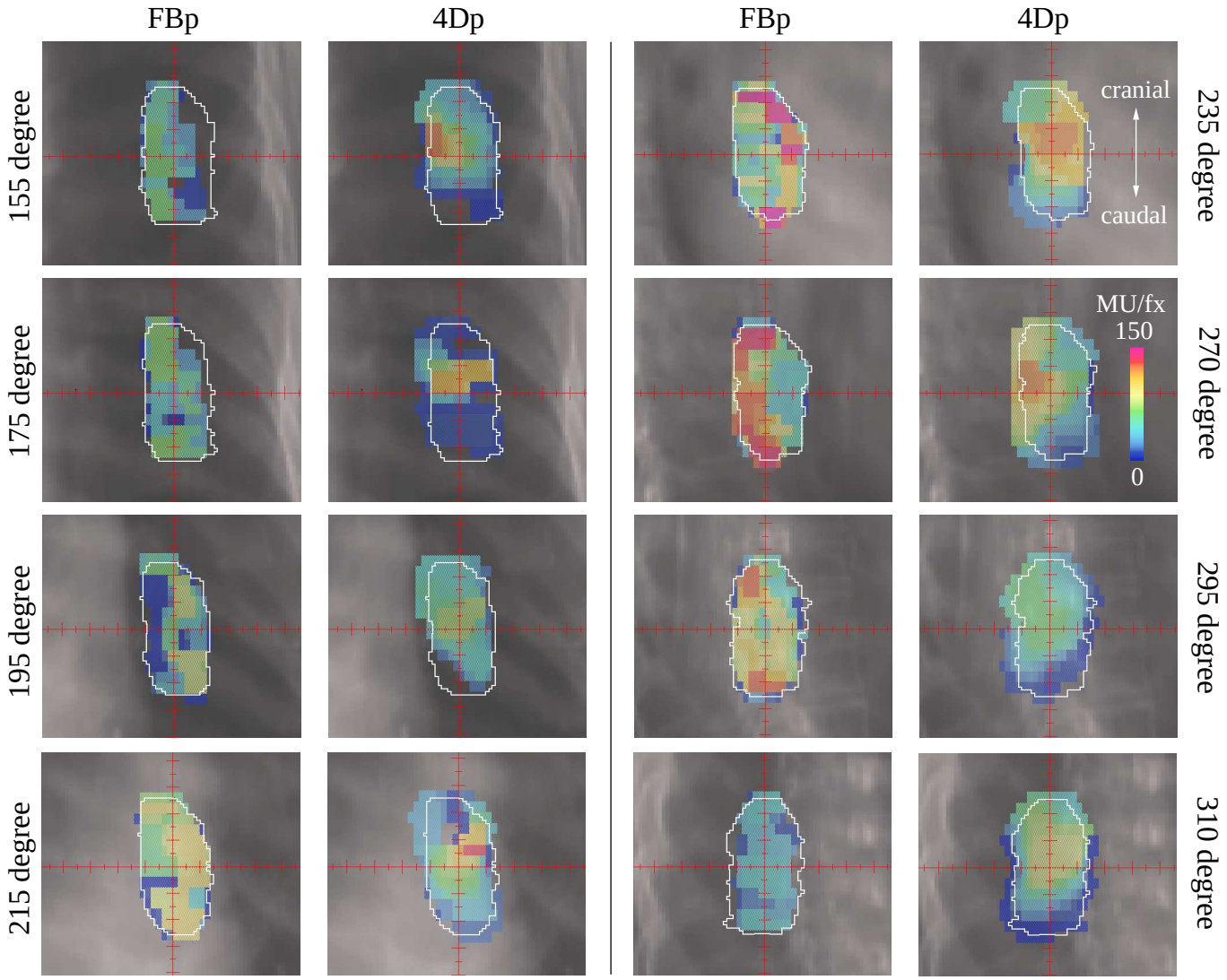

FIG. 5: Beam-eye-view of the fluence distributions of eight beam directions for the freebreathing PTV-plan and 4D-plan (white contours: freebreathing PTV).

based planning that arise as a consequence of planning on a static patient geometry: (1) The question which density grid to use for dose-calculations — this is a problem of static PTV-based planning especially in the presence of large geometric changes, where the dose distribution itself dynamically changes due to movement. (2) The problem of PTV-OAR overlaps — this leads to multiple voxel associations and thus an ill-defined optimization problem. These problems do not exist for 4D-TEV optimization, where prescription and dose scoring can be realized based on the non-overlapping, physiological structures of CTV and OARs.

The presented implementation of 4D-TEV optimization incorporates concurrent 4D-MC dose calculation in the multiple phase CTs of a RCCT-dataset to assure accurate estimation of dose-to-moving-tissue with calculation times essentially independent of the number of instances N. This facilitates calculation of fully-segmented 4D-plans in clinically realistic time frames. The use of other dose calculation methods is conceivable, for which, however, calculation time scales with N. Addressing this, Zhang *et al.* 2008[14] presented an accelerated, approximate 4D-dose-calculation scheme for 4D-inverse-planning.

Planning assumed daily volumetric imaging and setup to the daily mean target position for all presented plans. Based on a challenging patient example with large breathing excursion, we found that freebreathing treatment with the 4D-approach performed similarly well to gated treatment both with respect to accumulated target- and OAR-doses. Thus it provides a less demanding alternative to gating in day-to-day delivery without sharing the disadvantage of a reduced duty cycle and thereby prolonged treatment time. Idealized target tracking performed superior to all other approaches. This finding might be specific to the example patient who

showed an unusually large breathing excursion of 2.9cm. A future planning study with more patients should help to identify the influence of tumor excursion, size as well as relative position on the relative performance of the approaches. Recently, Zhang *et al.* 2008[14] reported comparable performance of a related 4D inverse planning approach and target tracking based on four patients.

The present study focused on incorporation of quasiperiodic intrafractional uncertainties in terms of respiratory deformation. Other uncertainties, that were not explicitly considered yet, are discussed in the following.
Generally, a major source of error in lung treatment are patient setup errors. Several studies found significant daily variation of the mean tumor position[27] and corresponding internal correlates[28]. Concerning systematic errors, the safest way of incorporation into planning is a setup margin around the (moving) CTV. As this is contrary to the original intention of 4D-TEV planning, future development of alternative methods is desirable. In general, it is advisable to minimize systematic errors by online image-guided setup to the daily mean target position[29,30]. Random setup errors can be approximately incorporated by combining the presented 4D-method with isocentre-sampling methods[7], or by explicitly optimizing in additional multiple shifted instances[12].
An important issue is the *robustness* of the 4D-TEV approach to changes of the breathing pattern during the course of treatment. As for all pdf-based 4D-planning approaches, optimization is performed in the frequency domain of motion and consequently such methods are insensitive to changes in the breathing curve itself (i.e. the exact spatial positions of tissue elements over time) as long as the motion is a *stationary process* that is described by the same pdf as assumed during planning. However, potential differences in the breathing pattern between planning and treatment expressed by a different pdf translate to uncertainties in the accumulated dose. Quantification of the sensitivity of the presented 4D-planning method to such uncertainties can be performed in a straightforward way by variation of the instance weights according to the expected uncertainties and is subject to future research. Generally, it is desirable to explicitly incorporate measures of robustness into the planning process[25], which has been recently proposed for breathing-pdf uncertainties[26].
Another aspect of breathing pattern uncertainties are changes in the motion amplitude. As these can lead to systematic errors, image-based monitoring of tumour excursion and the use of an appropriate margin around the CTV might be advisable. Recent studies suggest that the extent of tumor motion itself as seen in an RCCT seems to be stable for a majority of patients[27,28,31], but more studies in this direction are desirable with respect to requirements of 4D-planning. Moreover, a combination of 4D-planning with 4D-adaptive concepts along the lines of corresponding margin-based concepts[29] could help to identify patients who could benefit from replanning.
Registration errors of the deformable registration algorithm are another source of possible systematic errors in calculation of accumulated dose. While the algorithm used for this study has been shown to have residuals $\leq$ 3mm (voxelsize used for dose calculation) for more than 95% of all voxels in lung[15], generally a careful visual evaluation of the registration quality for all phases is advisable to prevent registration errors especially around the tumour. As registration errors in different phases are spatially uncorrelated, dosimetric errors decrease with the number of instances used for dose accumulation.

To conclude, 4D-optimization with deliverable segments and realistic doses was presented. The method generates robust fluence distributions and unsophisticated plans for rapid delivery, thereby providing an alternative to gating. Apart from its application to lung, it has potential for other entities like H&N, breast or liver.

* Corresponding author: Matthias Söhn, Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany; Tel: ++49/(0)7071/2986061; Fax: ++49/(0)7071/295920; Email: Matthias.Soehn@med.uni-tuebingen.de

[1] "Prescribing, recording and reporting photon beam therapy", ICRU Report 50, International Commission on Radiation Units and Measurements, Bethesda, MD, 1993.

[2] "Prescribing, recording and reporting photon beam therapy (supplement to ICRU Report 50)", ICRU Report 62, International Commission on Radiation Units and Measurements, Bethesda, MD, 1999.

[3] D. Yan, F. Vicini, J. Wong and A. Martinez, "Adaptive radiotherapy", Phys. Med. Biol. **42**(1), 123–132 (1997).

[4] J. Leong, "Implementation of random positioning error in computerized radiation treatment planning systems as a result of fractionation", Phys. Med. Biol. **32**(3), 327–334 (1987).

[5] T. Craig, J. Battista and J. V. Dyk, "Limitations of a convolution method for modeling geometric uncertainties in radiation therapy. I. The effect of shift invariance", Med. Phys. **30**(8), 2001–2011 (2003).

[6] I. J. Chetty, M. Rosu, N. Tyagi, L. H. Marsh, D. L. McShan, J. M. Balter, B. A. Fraass and R. K. T. Haken, "A fluence map convolution method to account for respiratory motion in three-dimensional dose calculations of the liver: A Monte-Carlo study", Med. Phys. **30**(7), 1776–1780 (2003).

[7] S. A. Naqvi and W. D. D'Souza, "A stochastic convolution/superposition method with isocenter sampling to evaluate intrafraction motion effects in IMRT", Med. Phys. **32**(4), 1156–1163 (2005).

[8] D. Yan, D. A. Jaffray and J. W. Wong, "A model to accumulate fractionated dose in a deforming organ", Int. J. Radiat. Oncol. Biol. Phys. **44**(3), 665–675 (1999).

[9] H. Paganetti, H. Jiang, J. A. Adams, G. T. Chen and E. Rietzel, "Monte Carlo simulations with time-dependent geometries to investigate effects of organ motion with high

temporal resolution", Int. J. Radiat. Oncol. Biol. Phys. **60**(3), 942–950 (2004).

[10] P. J. Keall, J. V. Siebers, S. Joshi and R. Mohan, "Monte Carlo as a four-dimensional radiotherapy treatment-planning tool to account for respiratory motion", Phys. Med. Biol. **49**(16), 3639–3648 (2004).

[11] M. 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**(10), 2822–2831 (2003).

[12] 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)", Med. Phys. **32**(5), 1510–1521 (2005).

[13] A. 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**(12), 2779–2798 (2005).

[14] P. 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).

[15] M. Söhn, M. Birkner, Y. Chi, J. Wang, D. Yan, B. Berger and M. Alber, "Model-Independent, Multi-Modality Deformable Image Registration by Local Matching of Anatomical Features and Minimization of Elastic Energy", Med. Phys. **35**(3), 866–878 (2008).

[16] M. Alber and F. Nüsslin, "An objective function for radiation treatment optimization based on local biological measures", Phys. Med. Biol. **44**(2), 470–493 (1999).

[17] M. Alber, "Normal tissue dose-effect models in biological dose optimisation", Z. Med. Phys. **18**(2), 102–110 (2008).

[18] A. Niemierko, "Reporting and analyzing dose distributions: A concept of equivalent uniform dose", Med. Phys. **24**(1), 103–110 (1997).

[19] M. Alber and R. Reemtsen, "Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method", Optimization Methods and Software **22**(3), 391–411 (2007).

[20] M. Alber, M. Birkner, W. Laub and F. Nüsslin, "Hyperion: an integrated IMRT planning tool", in *Proc. XIIIth Int. Conf. on the Use of Computers in Radiation Therapy*, edited by W. Schlegel and T. Bortfeld. (Heidelberg: Springer, 2000), pp. 46–48.

[21] U. Jeleń and M. Alber, "A finite size pencil beam algorithm for IMRT dose optimization: density corrections", Phys. Med. Biol. **52**(3), 617–633 (2007).

[22] M. Alber and F. Nüsslin, "Optimization of intensity modulated radiotherapy under constraints for static and dynamic MLC delivery", Phys. Med. Biol. **46**(12), 3229–3239 (2001).

[23] M. Fippel, "Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm", Med. Phys. **26**(8), 1466–1475 (1999).

[24] M. Sikora, O. Dohm and M. Alber, "A virtual photon source model of an Elekta linear accelerator with integrated mini MLC for Monte Carlo based IMRT dose calculation", Phys. Med. Biol. **52**(15), 4449–4463 (2007).

[25] J. Unkelbach and U. Oelfke, "Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions", Phys. Med. Biol. **49**(17), 4005–4029 (2004).

[26] T. C. Y. Chan, T. Bortfeld and J. N. Tsitsiklis, "A robust approach to IMRT optimization", Phys. Med. Biol. **51**(10), 2567–2583 (2006).

[27] J.-J. Sonke, J. Lebesque and M. van Herk, "Variability of four-dimensional computed tomography patient models", Int. J. Radiat. Oncol. Biol. Phys. **70**(2), 590–598 (2008).

[28] G. Hugo, C. Vargas, J. Liang, L. Kestin, J. W. Wong and D. Yan, "Changes in the respiratory pattern during radiotherapy for cancer of the lung", Radiother. Oncol. **78**(3), 326–331 (2006).

[29] G. D. Hugo, D. Yan and J. Liang, "Population and patient-specific target margins for 4D adaptive radiotherapy to account for intra- and inter-fraction variation in lung tumour position", Phys. Med. Biol. **52**(1), 257–274 (2007).

[30] I. S. Grills, G. Hugo, L. L. Kestin, A. P. Galerani, K. K. Chao, J. Wloch and D. Yan, "Image-Guided Radiotherapy via Daily Online Cone-Beam CT Substantially Reduces Margin Requirements for Stereotactic Lung Radiotherapy", Int. J. Radiat. Oncol. Biol. Phys. **70**(4), 1045–1056 (2008).

[31] M. Guckenberger, J. Wilbert, J. Meyer, K. Baier, A. Richter and M. Flentje, "Is a single respiratory correlated 4D-CT study sufficient for evaluation of breathing motion?", Int. J. Radiat. Oncol. Biol. Phys. **67**(5), 1352–1359 (2007).

Appendix G

A finite size pencil beam for IMRT dose optimization

published in
Physics in Medicine and Biology 2005; **50**(8): 1747–1766

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **50** (2005) 1747–1766 doi:10.1088/0031-9155/50/8/009

A finite size pencil beam for IMRT dose optimization

U Jeleń[1,2], M Söhn[2] and M Alber[2]

[1] Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, Al Mickiewicza 30, 30-059 Krakow, Poland
[2] Section for Biomedical Physics, Clinic for Radiooncology, University of Tübingen, Hoppe-Seyler-Str 3, D-72076 Tübingen, Germany

E-mail: Urszula.Jelen@med.uni-tuebingen.de

Received 15 November 2004, in final form 7 February 2005
Published 6 April 2005
Online at stacks.iop.org/PMB/50/1747

Abstract

Dose optimization for intensity modulated radiotherapy (IMRT) using small field elements (beamlets) requires the computation of a large number of very small, often only virtual fields of typically a few mm to 1 cm in size. The primary requirements for a suitable dose computation algorithm are (1) speed and (2) proper consideration of the penumbra of the fields which are composed of these beamlets. Here, a finite size pencil beam (fsPB) algorithm is proposed which was specifically designed for the purpose of beamlet-based IMRT. The algorithm employs an analytical function for the cross-profiles of the beamlets which is based on the assumption of self-consistency, i.e. the requirement that an arbitrary superposition of abutting beamlets should add up to a homogeneous field. The depth dependence is stored in tables derived from Monte Carlo computed dose distributions. It is demonstrated that the algorithm produces accurately the output factors and cross-profiles of typical multi-leaf-shaped segments. Due to the accurate penumbra model, the dose distribution features physically feasible gradients at any stage of the iterative optimization, which eliminates the problem of large discrepancies in normal tissue dose due to misaligned gradients between optimized and recomputed treatment plans.

1. Introduction

Dose computation for radiotherapy photon beams has always had to find the balance between available computation power, accuracy requirements and the dosimetric paradigm. Each dose calculation approach inevitably employs approximations that limit the range of its applicability. In particular, new, complex treatment techniques such as 3D conformal radiotherapy or intensity modulated radiotherapy (IMRT) require high spatial accuracy. Commonly, dose computation methods have been developed towards broad beam applications. In this paper,

we present an algorithm which is designed for the specific requirements of beamlet-based dose optimization for IMRT.

One approach to IMRT optimization starts with the discretization of the radiation fields into small elements, called beamlets. A beamlet corresponds to a complex fluence distribution: all particles that pass through a small rectangular opening in the field in a given plane. It becomes apparent that in the presence of collimator scatter, the fluence distribution of a beamlet depends on the position of all collimators, usually multi-leaf collimators (MLCs) and jaws. Consequentially, the interpretation of a beamlet as a 'unit fluence element' is to some extent over-simplified, and in any case rather unwieldy for direct dose computation. This is also true for Monte Carlo dose computation algorithms, which could in principle be used to compute the fluence phase space for arbitrary fluence elements. However, these methods also need to be commissioned by broad beam measurements, because small field measurements are confounded by dosimetric problems. Thus, even Monte Carlo methods are limited in their accuracy for calculation of beamlet dose distributions. Moreover, given that the optimum size of a beamlet for MLC delivery is about 2 mm times the leaf width (Bortfeld *et al* 2000), it would not only be very difficult and approximate to compute the dose distribution from physical principles, but also exceedingly time consuming, even in the age of affordable supercomputer power in the shape of PC clusters. At this resolution, an average five field prostate plan amounts to 2500 beamlets, while head and neck treatment plans can easily exceed 15 000 beamlets.

One of the marks of IMRT is that there can be a beam edge anywhere in the field, in particular if the projections of organs of risks and the target overlap. Many MLC segments are used to irradiate the target volume. As a consequence, the dose, especially in the mostly shielded organs at risk, stems from out-of-field penumbra dose (leakage, scatter) to a large degree, rather than primary fluence. Here, it is absolutely essential to use the correct measure of the beam penumbra during the dose optimization to facilitate the accurate placement of dose (or rather exposure time) gradients. Failure to do so leads to overdose of critical structures which goes unnoticed during the optimization process (Schwarz *et al* 2003).

Therefore, we formulate the requirements for a dose computation algorithm for beamlet-based IMRT as:

- it does not require accurate fluence modelling, e.g. can be commissioned from broad beam measurements;
- it predicts the proper penumbra anywhere in the field;
- it allows order of magnitude 10^4 beamlet dose computations in a reasonable time.

The essential idea in the following is that the dose distribution of a beamlet can be interpreted as the difference between two broad beams with an incremental change of, say, the position of one leaf. Conversely, any composition of equally weighted, abutting beamlets should approximate a broad beam dose distribution. This property of *self-consistency* is central to the development in section 2. A similar line of argument was followed in appendix B of the paper by De Gersem *et al* (2001). There, a self-consistent analytic cross-profile is used to construct the approximate incremental difference between broad beam dose distributions. In section 3, we investigate the properties of this method with respect to the composition of broad beams from beamlet dose distributions. Results of the fitting procedure used to define the beamlet cross-profile shape on the basis of broad beams are presented and performance of the algorithm is tested in a water phantom. Next, the accuracy of the method is compared to Monte Carlo in selected examples of field shapes typical for IMRT. In section 4, the presented method is discussed in the context of IMRT dose optimization.

2. Methods

The common concept of pencil beam (PB) algorithms uses a *pencil kernel* which can be understood as a convolution of some unit photon fluence of a monodirectional beam with the energy spread distribution in water. The potential of convolution-based models employing point kernels for dose calculation in radiotherapy was exploited by different investigators (Boyer and Mok 1984, Mackie *et al* 1985, Mohan *et al* 1986, Ahnesjö *et al* 1987) but the long computation time of the full 3D convolution has led to the development of group of algorithms based on 2D convolution of pencil beam kernels (Mohan and Chui *et al* 1987, Ahnesjö *et al* 1992). The pencil kernels are usually obtained by means of Monte Carlo calculations (Mohan *et al* 1986, Mackie *et al* 1988), either directly as pencil kernels or as a superposition of point kernels. Application of kernels obtained in this way is complicated as it requires knowledge about the energy spectrum and primary fluence distribution and the electron contamination of the modelled photon beam. In order to avoid these difficulties, methods have been developed to derive kernels directly from broad beam measurements by deconvolution (Chui and Mohan 1988) or differentiation of parameters like the product of output factor and tissue-phantom ratio (Bortfeld *et al* 1993), scatter-to-primary ratio (Ceberg *et al* 1996) or the scatter dose (Storchi and Woudstra 1996, Storchi *et al* 1999).

The pencil kernel may derive from a finitely sized unit fluence (as opposed to an infinitesimally narrow ray) under consideration of the particular geometry of the multileaf collimator used to shape the IMRT fields. This approach was studied by Bourland and Chaney (1992) and improved by Ostapiak *et al* (1997). Both works assume that the radiation beam can be geometrically divided into identical finite size pencil beams that allow to reconstruct the dose distribution of the full beam by superposition. These finite size pencil beams are self-consistent in the sense that they can be used to construct arbitrary field shapes, or conversely, do not depend on the broad beam from which they were derived. Finite size pencil beam kernels are generated in these models by Monte Carlo simulations. The following method constructs an analytical form of a finite size pencil beam kernel (fPBK) from broad beam dose distribution with an emphasis on self-consistency in order to compute beamlet dose distributions. Thus, it follows the same concept as a finite size pencil beam algorithm.

The main assumption of the model is that a self-consistent analytical function, with parameters determined by a fitting procedure from cross-profiles of broad beam dose distributions calculated with Monte Carlo, can be used to describe the shape of the fPBK. Depth dependence of the dose is in turn reproduced by means of a 'family' of depth curves used to scale the amplitude of the profiles. These depth curves are also obtained from Monte Carlo calculations for different field sizes. Due to self-consistency of the fPBK, the dose distribution of a broad beam can be composed of a sum of abutting beamlets. By choosing the size of the beamlets appropriately and small enough, arbitrary MLC field shapes can be created.

2.1. Self-consistent cross-profile

In order to construct the analytical form of the pencil beam kernel, let $p(x)$ be a dose cross-profile of a semi-infinite field with field edge at $x = 0$ (figure 1). Assume that $p(x)$ can be approximated for $x < 0$ by a sum of exponentials:

$$p(x) = \frac{1}{2} \sum_{i=1}^{n} w_i \exp(u_i x) \tag{1}$$

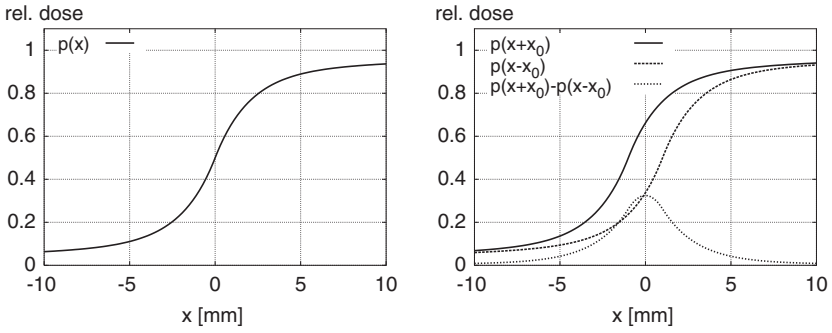

Figure 1. Cross-profile $p(x)$ of semi-infinite field with edge at $x = 0$ and pencil beam cross-profile of width $2x_0$ as the difference between two semi-infinite fields shifted by $2x_0$.

with weight factors $w_i > 0$, $\sum_{i=1}^{n} w_i = 1$ and slopes $u_i > 0$. Requiring self-consistency, the profile for $x \geqslant 0$ reads

$$p(x) = 1 - \frac{1}{2}\sum_{i=1}^{n} w_i \exp(-u_i x) \tag{2}$$

so that two abutting semi-infinite fields of this shape add up to a constant dose distribution.

We define a one-dimensional finite size pencil beam cross-profile of width $2x_0$ as the difference between two semi-infinite fields shifted by $2x_0$ (figure 1). If the profile is centred about $x = 0$, it derives from $p(x + x_0) - p(x - x_0)$ as

$$P(x, \vec{w}, \vec{u}, x_0) = \begin{cases} \sum_{i=1}^{n} w_i \sinh(u_i x_0) \exp(u_i x) & \text{for} \quad x < -x_0 \\ 1 - \sum_{i=1}^{n} w_i \exp(-u_i x_0) \cosh(u_i x) & \text{for} \quad -x_0 \leqslant x \leqslant x_0 \\ \sum_{i=1}^{n} w_i \sinh(u_i x_0) \exp(-u_i x) & \text{for} \quad x_0 < x. \end{cases} \tag{3}$$

Note that with this definition, the profile is normalized to

$$\int_{-\infty}^{\infty} P(x, \vec{w}, \vec{u}, x_0)\, \mathrm{d}x = 2x_0. \tag{4}$$

In practice, it is often sufficient to truncate the sum in (3) after the second element, see figure 3. Then, one exponential $f(x, u_2, x_0)$ would model the primary penumbra which results from the collimation of primary fluence (source size, shape of leaf ends) and the electron spread at the field edge, while the other $f(x, u_1, x_0)$, with a much shallower tail-off, would model off-axis head scatter and phantom scatter distributions. We define

$$f(x, u_i, x_0) = P(x, 1, u_i, x_0). \tag{5}$$

In two dimensions, the dose distribution $F(x, y, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0)$ of a pencil beam of width $(2x_0, 2y_0)$ is defined as the product of two independent one-dimensional profiles with equivalent weights w_i [3]:

$$\begin{aligned} F(x, y, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0) &= w_1 f_x(x, u_{1x}, x_0) f_y(y, u_{1y}, y_0) \\ &\quad + (1 - w_1) f_x(x, u_{2x}, x_0) f_y(y, u_{2y}, y_0) \end{aligned} \tag{6}$$

where f_x, f_y denote components as described above in the x and y directions respectively (parameters u_1 and u_2 along the x and y axes may differ slightly due to linac construction

[3] The requirement of equivalent weights is convention rather than necessity.

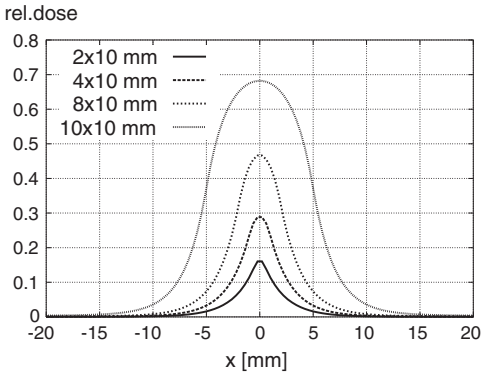

Figure 2. Typical pencil beam cross-profiles for widths $2x_0 = 1, 2, 4$ and 10 mm.

details e.g. distance between source and collimator jaws). Again, it can be shown that the profile is normalized to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x, y, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0)\,\mathrm{d}x\,\mathrm{d}y = 4x_0y_0. \tag{7}$$

Some examples for this function are shown in figure 2.

The parameters $u_{1x}, u_{1y}, u_{2x}, u_{2y}, w_1$ depend on phantom depth and distance-to-source which is a consequence of geometrical broadening and secondary photon production, i.e. phantom scatter, which will primarily affect w_1 and u_1. This is taken into account in the definition of the full dose distribution of a beamlet below.

2.2. Single beamlet dose distribution

The full dose distribution of a beamlet derives from a product of a scaling factor $A(d, \theta)$ (dependent on the depth d and offset θ from the central beam axis) and the function $F(x, y, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0)$ (equation (6)). Let $D(\vec{r})$ be the dose distribution of a beamlet. Let $\vec{r}$ and $\vec{a}$ originate in the source point and let $\vec{a}$ be a unit vector in the direction of the beamlet. $\vec{r_a}$ denotes the projection of $\vec{r}$ onto $\vec{a}$. Then

$$D(\vec{r}) = F\left(\frac{x}{r_a}, \frac{y}{r_a}, \vec{w}(d), \vec{u}_x(d), \vec{u}_y(d), x_0, y_0\right) \cdot A(d_{\mathrm{rad}}, \theta) \cdot \left(\frac{1}{r_a}\right)^2 \tag{8}$$

where x, y are the projections of $\vec{r} - \vec{r_a}$ onto the unit vector of the x-axis and y-axis of the plane perpendicular to the beamlet direction, d is the geometrical depth of the point $\vec{r_a}$ below the patient surface, $A(d_{\mathrm{rad}}, \theta)$ is the scaling factor for profiles as a function of the radiological depth and off-axis distance, defined along raylines with offset θ to the central beam axis, d_{rad} is the radiological depth of the point $\vec{r_a}$, θ is the angle between the central axis of the beam and the beamlet and r_a is the length of $\vec{r_a}$ in units of the source-to-isocentre distance.

With this definition, the geometric broadening of the penumbra due to the divergent beam geometry is factored into the first two variables, while a (usually much smaller) depth-dependent broadening can be taken into account directly by the tabulated steepness parameters u_1 and u_2. Effects such as off-axis beam softening or central depression are absorbed in the angular variability of the $A(d, \theta)$ curves. Note that $A(d, 0)(1/r_a)^2$ for large fields tends to the central axis depth–dose curve as the value of the function $F(x, y, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0)$ becomes close to unity.

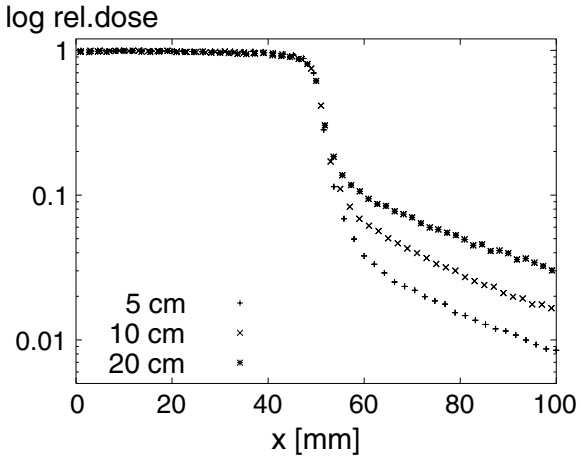

Figure 3. Beam penumbra of a 10 × 10 cm^2 6 MV field in depths of 5, 10 and 20 cm. All profiles are scaled in the x direction and dose to map onto each other. It can be seen that the steepness of primary and secondary penumbra can be assumed to be constant with depth and that both components can be described well by an exponential function. The relative weight of both components changes with depth due to phantom scatter.

The factors $A(d, \theta)$ facilitate also the introduction of longitudinal density corrections by using the concept of radiological depth. For a given geometrical depth d, where the dose calculation is performed, the density weighted depth along the ray line is calculated and used to read the value of the tabulated $A(d, \theta)$ factor.

Due to the divergent geometry of the radiation beam, abutting beamlets calculated according to (8) are self-consistent only on a spherical surface centred on the radiation source point. Usually, the dose distribution of a broad beam is measured in a plane, but for sufficiently small beamlets this fact can be neglected for commissioning the pencil beam kernel.

2.3. Beam commissioning

The beam commissioning procedure for the presented algorithm determines tables of parameters used in (8): $w_1(d)$, $u_{ix}(d)$, $u_{iy}(d)$ and $A(d, \theta)$ and is a two-stage process. Starting from Monte Carlo computations of various fields in a water phantom with a source–surface distance of 90 cm, in a first step, the steepness parameters u_{ix}, u_{iy} and weights w_1 are determined depending on field size and depth. These parameters are determined for a variety of field sizes to show that they depend only very weakly on field size (otherwise the assumption of self-consistency would not be warranted). The Monte Carlo code XVMC was used for all simulations with a virtual source model (Fippel 1999, Fippel *et al* 1999, Fippel *et al* 2003) that was commissioned for an Elekta Precise linear accelerator (Elekta, Crawley, UK).

Cross-profiles acquired at various depths are geometrically rescaled to isocentre distance (see figure 3). Note that due to the self-consistency, it is possible to set the width of the analytical cross-profile $2x_0, 2y_0$ to the width of the master profile at isocentre distance. The function given in (6), rewritten to account for $y = 0$ for x profile and vice versa, is fitted with a least-squares Levenberg–Marquardt algorithm (Press *et al* 1992) to the rescaled Monte Carlo cross-profiles. The auxiliary factor a accounts for depth-dependent changes of the profile amplitude.

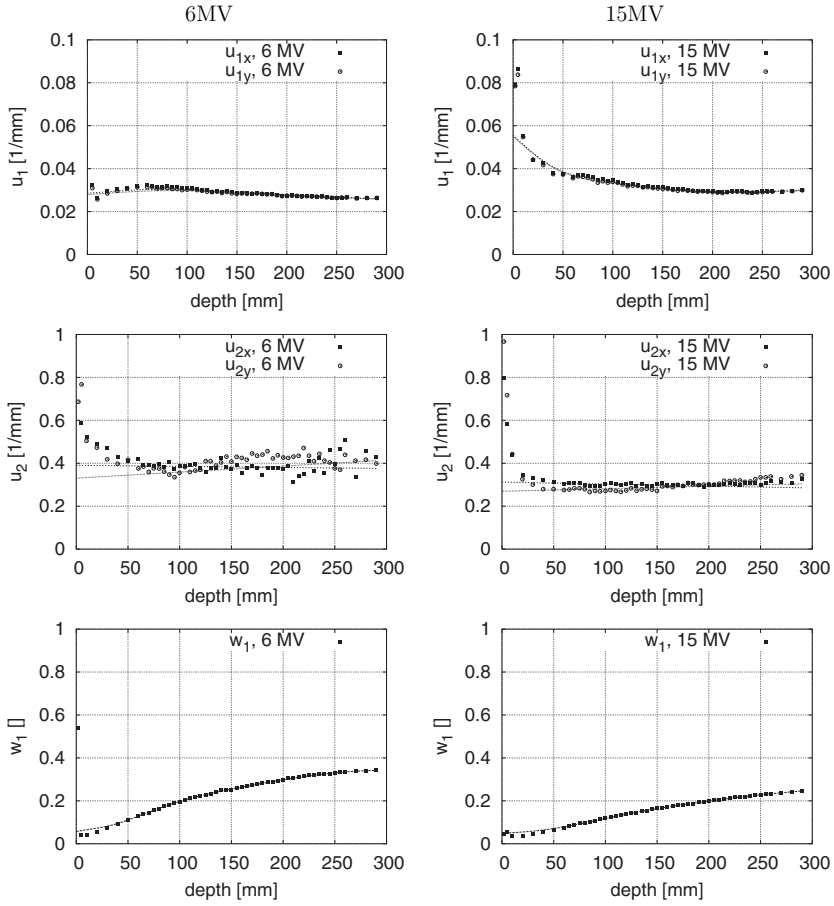

Figure 4. Commissioned values of u_1, u_2 and w_1 for the 6 MV and 15 MV photon beams for 10×10 cm^2 field (broken curves are result of smoothing and interpolation).

$$a \cdot F(x, 0, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0)$$

$$= \begin{cases} w_1 \sinh(u_{1x}x_0)\exp(u_{1x}x)(1-\exp(-u_{1y}y_0)) & \\ \quad +(1-w_1)\sinh(u_{2x}x_0)\exp(u_{2x}x)(1-\exp(-u_{2y}y_0)) & \text{for} \quad x < -x_0 \\ w_1(1-\cosh(u_{1x}x)\exp(-u_{1x}x_0))(1-\exp(-u_{1y}y_0)) & \\ \quad +(1-w_1)(1-\cosh(u_{2x}x)\exp(-u_{2x}x_0))(1-\exp(-u_{2y}y_0)) & \text{for} \quad -x_0 \leqslant x \leqslant x_0 \\ w_1 \sinh(u_{1x}x_0)\exp(-u_{1x}x)(1-\exp(-u_{1y}y_0)) & \\ \quad +(1-w_1)\sinh(u_{2x}x_0)\exp(-u_{2x}x)(1-\exp(-u_{2y}y_0)) & \text{for} \quad x_0 < x. \end{cases} \tag{9}$$

The values of $u_{1x}, u_{1y}, u_{2x}, u_{2y}$ and w_1 resulting from the fitting procedure, interpolated and smoothed as shown in figure 4 are tabulated for depths from 0 to 300 mm in 2 mm increments for field sizes 4×4, 10×10 and 20×20 cm^2 and stored in look-up tables.

In a second step, the scaling factor $A(d, \theta)$ is computed by

$$\mathrm{A}(d, \theta) = D(R\tan\theta) \cdot \frac{R^2 \cos^{-2}\theta}{F(R\tan\theta, 0, \vec{w}(d/\cos\theta), \vec{u}_x(d/\cos\theta), \vec{u}_y(d/\cos\theta), x_0, y_0)} \tag{10}$$

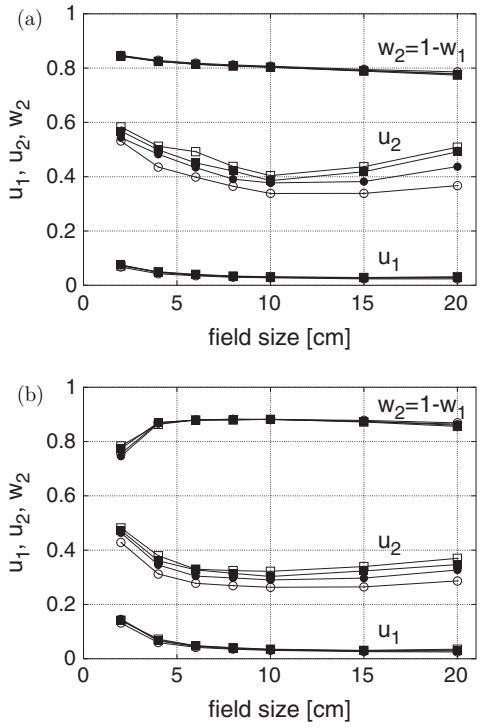

Figure 5. Dependences of the fit parameters on the field size and source–surface distance (parameters defined at depth of 10 cm): (a) for beam energy 6 MV, (b) for beam energy 15 MV. ○—SSD 75 cm, ●—SSD 85 cm, ■—SSD 90 cm, □—SSD 95 cm.

where R is the nominal source distance of the considered Monte Carlo calculated profile $D(x)$ along the central axis, and d is its depth. Note that $(R\tan\theta)$ is a parametrization of the off-axis position x of the profile and $(d/\cos\theta)$ is the depth of a calculated point x along the ray. From this, a table of $A(d,\theta)$ curves can be obtained as a function of offset θ and depth d and stored in the base data file for different field sizes.

3. Results

3.1. Commissioned profile parameters

Input data for the beam commissioning procedure are depth–dose curves and cross-profiles computed with the Monte Carlo code XVMC (Fippel 1999) for a set of square fields (2×2, 4×4, 6×6, 8×8, 10×10, 15×15, 20×20 and 25×25 cm^2) in the standard set-up of source–surface distance (SSD) of 90 cm, source–isocentre distance (SID) of 100 cm for two energies: 6 MV and 15 MV in the homogeneous water phantom.

Some curves of commissioned values for the profile steepness parameters u_1, u_2 and the weight w_1 for photon beam energies 6 MV and 15 MV (field size 10×10 cm^2) are presented in figure 4. The plots demonstrate that, while steepness parameters u_1 and u_2 remain almost constant with depth, the weight w_1 representing the ratio between secondary and primary penumbra component increases, reflecting the increase of phantom scatter.

As the steepness parameters and weights depend only slightly on the field size and the source–surface distance (figure 5) commissioned values of parameters were tabulated only for

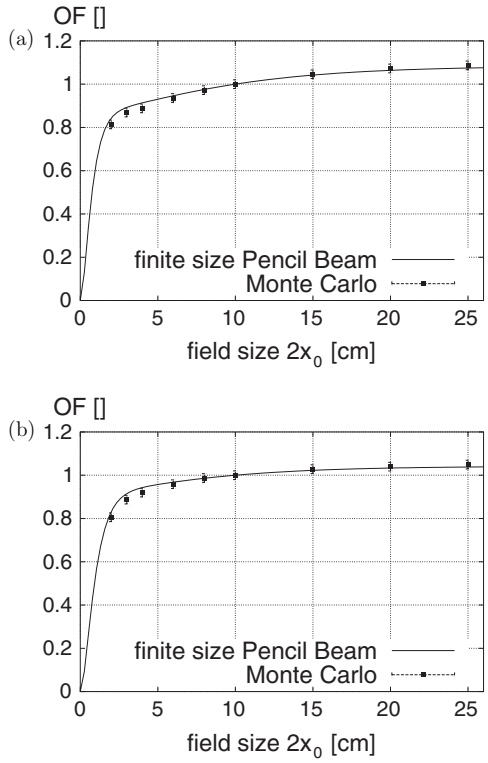

Figure 6. Output factors for square fields for (a) 6 MV, (b) 15 MV photon beams.

4×4, 10×10 and 20×20 cm^2 fields and used to calculate the scaling factor curves $A(d, \theta)$ for various fields.

3.2. Output factors

From (6), it is possible to derive an expression for the output factor of rectangular fields of size $2x_0 \times 2y_0$ centred around the central axis. The output factor is the ratio of the dose at the central axis of a given field relative to a reference field. For the dose at reference depth on the central axis we obtain

$$\begin{aligned} \mathrm{OF}\ (x_0, y_0) &= \frac{F(0, 0, \vec{w}, \vec{u}_x, \vec{u}_y, x_0, y_0)}{F(0, 0, \vec{w}, \vec{u}_x, \vec{u}_y, x_r, y_r)} \\ &= \frac{w_1(1 - \exp(-u_{1x}x_0))(1 - \exp(-u_{1y}y_0))+}{w_1(1 - \exp(-u_{1x}x_r))(1 - \exp(-u_{1y}y_r))+} \\ &\quad \frac{(1 - w_1)(1 - \exp(-u_{2x}x_0))(1 - \exp(-u_{2y}y_0))}{(1 - w_1)(1 - \exp(-u_{2x}x_r))(1 - \exp(-u_{2y}y_r))} \end{aligned} \tag{11}$$

where $2x_r, 2y_r$ is the field size of the reference field. Figure 6 displays a remarkable agreement between output factors for a set of square fields computed by Monte Carlo and calculated from (11) using parameters obtained from fitting of a cross-profile of a 10×10 cm^2 field at depth 10 cm. Monte Carlo computed and finite size pencil beam (fsPB) calculated output factors are in accordance within 2% for fields from 2×2 cm^2 to 25×25 cm^2.

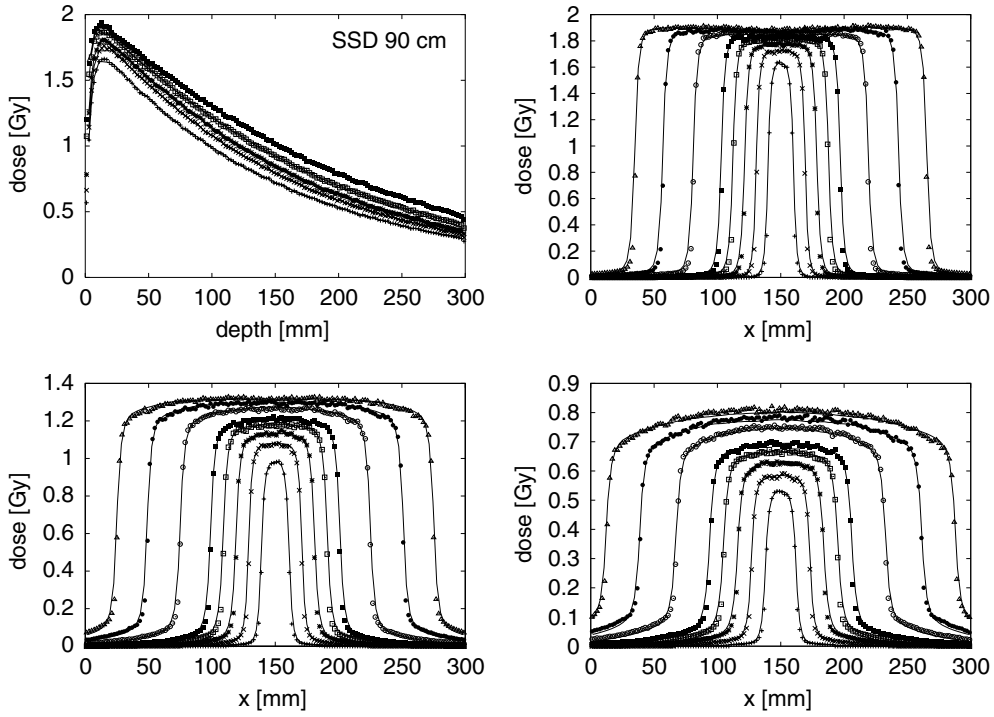

Figure 7. Monte Carlo (symbols) and fsPB (lines) calculated depth–dose curves for 2 × 2, 4 × 4, 6 × 6, 10 × 10 and 20 × 20 cm^2 fields and cross-profiles at 2, 10 and 20 cm depth for 2 × 2, 4 × 4, 6 × 6, 8 × 8, 10 × 10, 15 × 15, 20 × 20 and 25 × 25 cm^2 fields (6 MV, SSD 90 cm).

3.3. Phantom tests

3.3.1. Square fields. Tests of the accuracy of the algorithm were performed in a virtual water phantom by comparison with Monte Carlo. Cross-profiles and depth–dose curves computed with the finite size pencil beam algorithm were compared to corresponding Monte Carlo calculation results for the following field sizes: 2 × 2 cm^2, 4 × 4 cm^2, 6 × 6 cm^2, 8 × 8 cm^2, 10 × 10 cm^2, 15 × 15 cm^2, 20 × 20 cm^2 and 25 × 25 cm^2. Plots for source-surface distance of 90 cm are presented in figures 7 and 9 for photon beam energies 6 MV and 15 MV. The algorithm was tested additionally for different SSD distances: 75, 85 and 95 cm. Depth–dose curves for these cases are presented in figures 8 and 10 for the field sizes 2 × 2 cm^2, 4 × 4 cm^2, 6 × 6 cm^2, 10 × 10 cm^2, 15 × 15 cm^2 and 20 × 20 cm^2.

The accuracy of the dose computation with the proposed algorithm, in comparison to the Monte Carlo results, was evaluated using formulae and tolerances proposed by Venselaar *et al* (2001). According to this concept, the ratio of dose deviation between Monte Carlo and finite size pencil beam computation results and local Monte Carlo calculated dose is used as a measure of discrepancy.

$$\delta = \frac{D_{\mathrm{PB}} - D_{\mathrm{MC}}}{D_{\mathrm{MC}}} \cdot 100\%. \tag{12}$$

An exception from this formula is recommended by the cited work and other investigators (Van Dyk *et al* 1993) in the low dose regions outside the geometrical field outline. Due to inherent uncertainties present both in the dose measurements and calculations for those regions, it is considered to be more useful to relate the deviation to the dose at a point at the

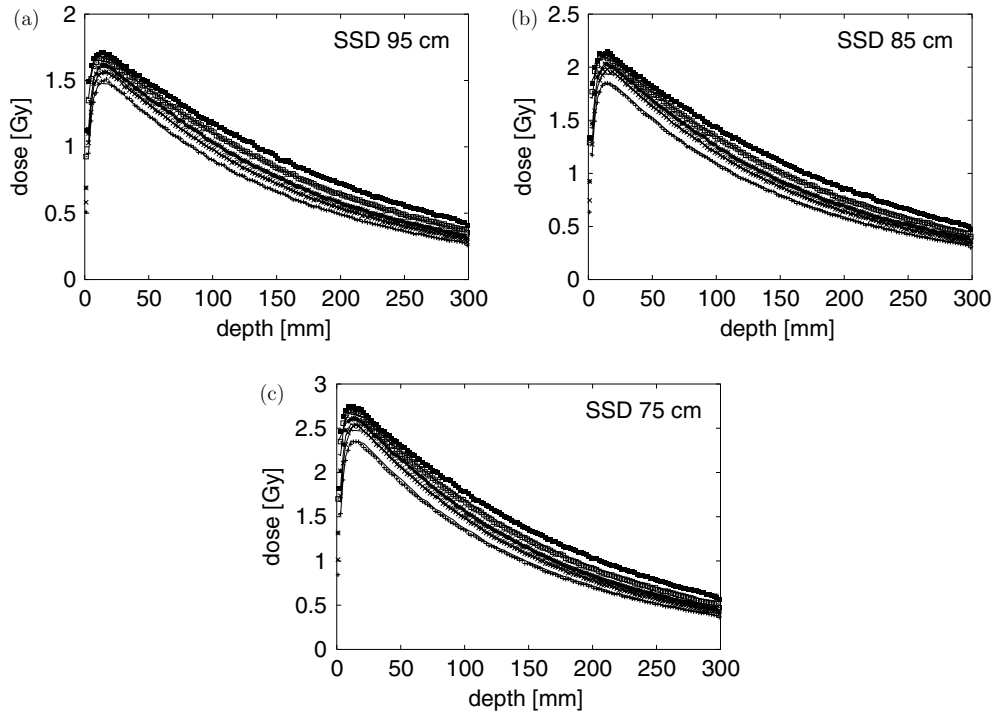

Figure 8. Monte Carlo (symbols) and fsPB (lines) calculated depth–dose curves for 2×2, 4×4, 6×6, 10×10 and 20×20 cm^2 fields for 6 MV with (a) SSD 95 cm, (b) SSD 85 cm (c) SSD 75 cm.

same depth on the central axis. For the calculations performed to commission and validate finite size pencil beam algorithm, the accuracy of Monte Carlo was 1% of the value on the central axis (Fippel *et al* 1999), resulting in the equal absolute error everywhere in the field, but increase of the relative error in low dose regions. Therefore, the fitting procedure assigns weights to fitted points and allows larger deviations in those regions.

For simple geometries (homogeneous phantom, different SSD values, square or rectangular fields, oblique incidence allowed), the criteria proposed in the work of Venselaar *et al* (2001), based on Van Dyk *et al* (1993) and Fraass *et al* (1998) and experience with modern treatment planning systems (Venselaar and Welleweerd 2001), are the following:

- for points on the central beam axis beyond the $d_{\max}$: $\delta = 2\%$
- for points in the build-up region: $\delta = 10\%$
- for points beyond the $d_{\max}$ not on the central axis: $\delta = 3\%$
- for points outside the geometrical beam edges: $\delta = 3\%$ when related to the dose value on the central axis ($\delta = 30\%$ when related to the local dose value).

For penumbra regions, due to large dose gradients, accuracy criteria are expressed in millimetres as the acceptable shift of isodose lines placement, and are 2 mm for the simple geometry described above.

The overall depth–dose curve (DDC) accuracy for SSD 90 cm, calculated as the mean deviation of dose values for all points, is within 1% for all field sizes considered and for both energies (figures 7 and 9). Locally the deviation exceeds 1%, mostly for larger fields like 20×20 or 25×25 cm^2, reaching up to 4% in the build-up region where the statistical noise of Monte Carlo leads to suboptimal fitting results for cross-profiles.

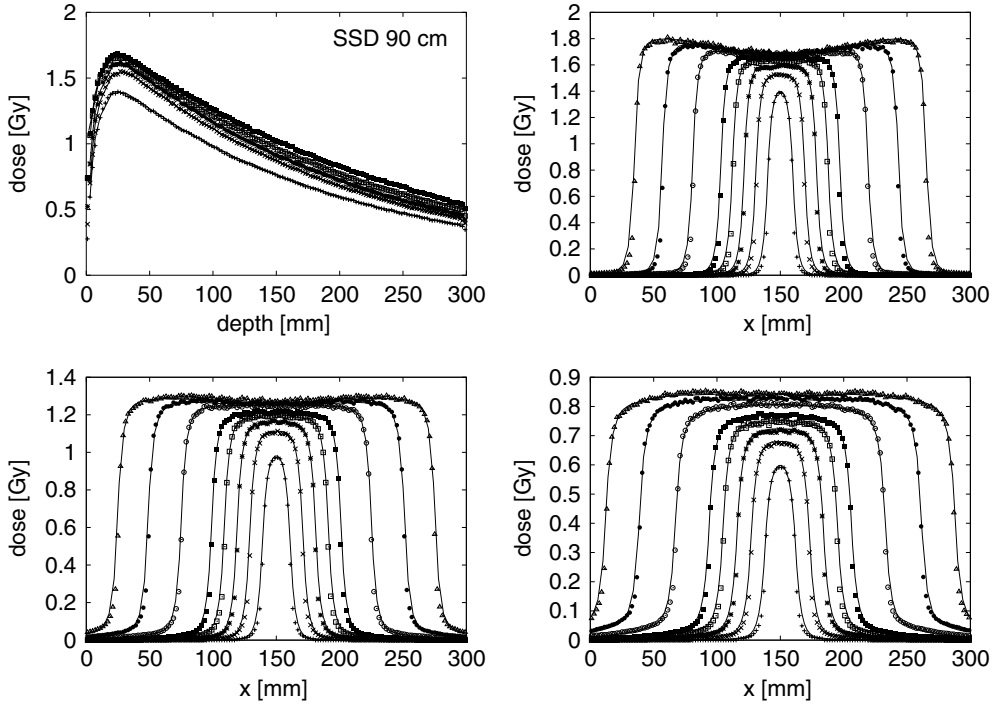

Figure 9. Monte Carlo (symbols) and fsPB (lines) calculated depth–dose curves for 2 × 2, 4 × 4, 6 × 6, 10 × 10 and 20 × 20 cm^2 fields and cross-profiles at 2, 10 and 20 cm for 2 × 2, 4 × 4, 6 × 6, 8 × 8, 10 × 10, 15 × 15, 20 × 20 and 25 × 25 cm^2 fields (15 MV, SSD 90 cm).

The accuracy of depth–dose calculation remains within 1–2% for SSD distances within 75–95 cm for both energies. Only locally, in the build-up region and at larger depths (>20 cm) for SSD 75 cm, deviations around 4% were observed and overall accuracy for smallest fields at SSD 75 cm for energy 15 MV was constantly worse—in the order of 4%.

For both investigated beam energies, dose values in the plateau of cross-profiles and outside the geometrical beam edges, calculated with finite size pencil beam and Monte Carlo methods, agree within 2% for smallest fields and for fields greater than 15 × 15 cm^2 at large depths. Even better agreement of about 1% was observed for the remaining cases. At small depths, the distortion of profiles by central depression is reproduced accurately (plots for depth 2 cm presented in figure 7 and 9).

In the penumbra regions the accuracy expressed as the magnitude of the shift of the 50% isodose line reveals misalignment of around 0.5 mm for all evaluated field sizes, but reaches up to 3.5 mm for edges of penumbra (10% and 90% isodoses), especially for large fields.

3.3.2. Irregular fields. The results of dose computation with the finite size pencil beam algorithm were checked for some typical segment shapes obtained during IMRT treatment planning of breast and head-and-neck cases. Three examples are shown below:

(a) an elongated, narrow and slightly off-axis segment, beam energy 6 MV (figure 11)
(b) a small segment (mean radius $\simeq$ 1.5 cm) at large off-axis distance (around 10 cm in both directions), beam energy 6 MV (figure 12)
(c) a large segment of irregular shape not containing the isocentre, beam energy 15 MV (figure 13).

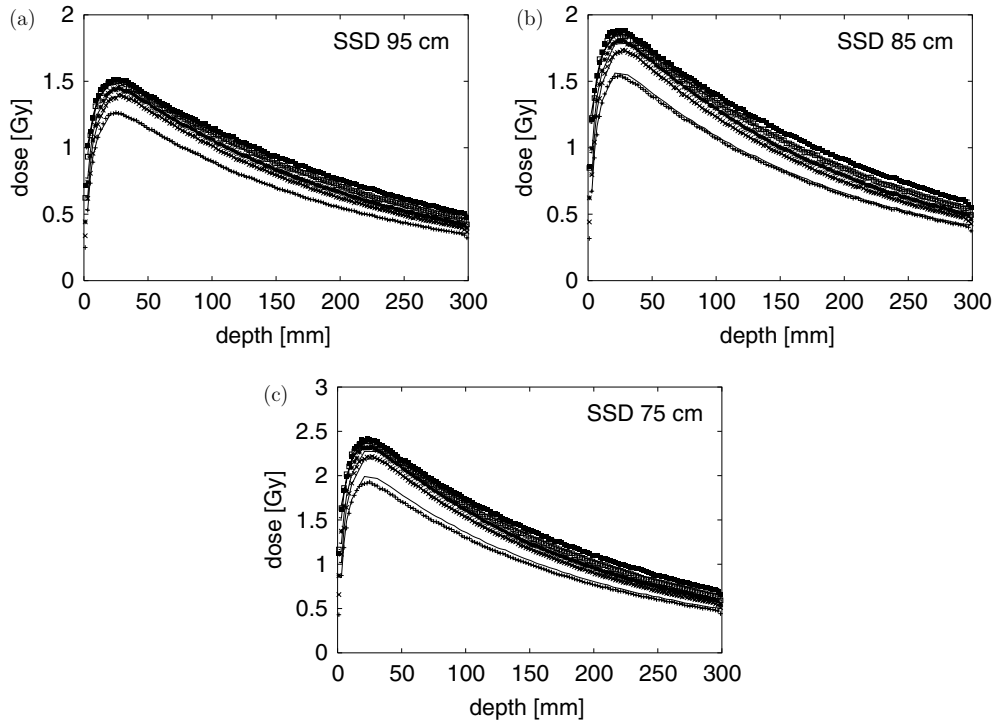

Figure 10. Monte Carlo (symbols) and fsPB (lines) calculated depth–dose curves for 2 × 2, 4 × 4, 6 × 6, 10 × 10 and 20 × 20 cm^2 fields for 15 MV with (a) SSD 95 cm, (b) SSD 85 cm. (c) SSD 75 cm.

For this test, distributions of the dose deposited in the water phantom irradiated with fields of selected shapes were calculated with finite size pencil beam and Monte Carlo. Comparison reveals that the dose distributions calculated with the Monte Carlo method are reproduced with an accuracy of about 2% for cases (a) and (c) except for regions shielded only with multileaf collimator leaves (figures 11 and 13). The algorithm was commissioned for fields created by collimator blocks and MLC leaves and does not model properly transmission through leaves and effects like interleaf leakage. Cross-profiles of fields (a) and (c) reveal that the penumbra region is placed in these cases with spatial accuracy of 1 mm. For the small off-axis segment presented as case (b) fsPB calculation results overestimate the dose, compared to Monte Carlo, by at most 4% and penumbra misalignment is in order of 2–3 mm (figure 12).

3.4. Clinical tests

The proposed algorithm was tested for two clinical cases: head-and-neck, using a beam of energy 6 MV and a prostate case calculated for a beam of energy 15 MV. For both cases, dose distributions of a single conformal beam (for the prostate case at angle 0° and for the head-and-neck case at 70°) were calculated using either the fsPB algorithm or Monte Carlo. The results of the calculations are shown in figures 14 and 15.

Dose distributions in the selected planes were evaluated using the concept of χ-factor (Bakai *et al* 2003) being a refinement of γ-evaluation method proposed by Low *et al* (1998). For all clinical tests the following acceptance citeria, acknowledged as a goal for dose calculation (ICRU 1987), were used: $\Delta d = 2$ mm of misalignment and $\Delta D = 2\%$ dose

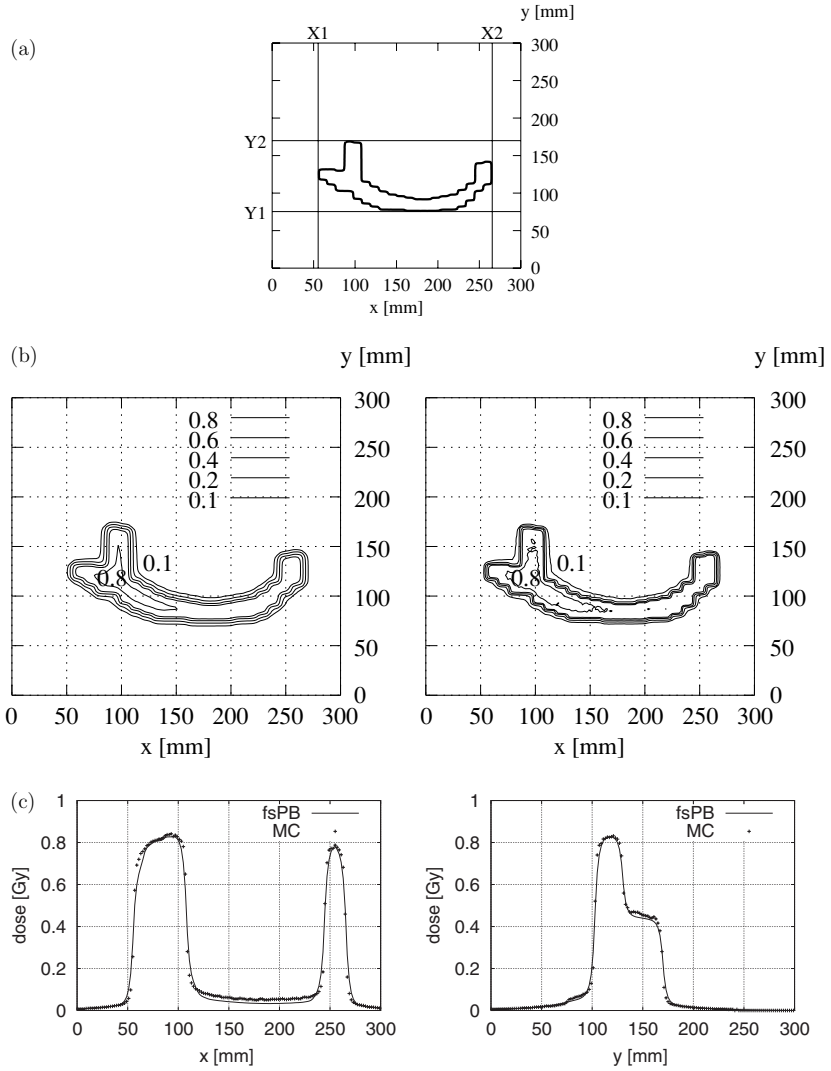

Figure 11. Results of dose computation for an exemplary IMRT segment (photon beam energy 6 MV): (a) segment localization and shape (jaw positions marked as X1, X2, Y1 and Y2), (b) finite size pencil beam (left) and Monte Carlo (right) computed dose distributions at 15 cm depth, (c) profiles at depth 15 cm for $y = 120$ mm and for $x = 85$ mm—fsPB (lines) and MC (symbols).

difference (relative to the local value of the Monte Carlo calculated dose). Especially in the prostate case, a fundamental difference between Monte Carlo and pencil beam algorithm can be made out in the persistent deviation along the field edges. This discrepancy occurs in regions shielded by the MLC leaves only, where the leaf transmission is not modelled properly and is in order of 2% of the central axis dose. For both cases, regions expanding form the isodose line 1% outwards, where the discrepancy exceeds chosen acceptance criteria are present, but the deviation, related to the dose on the central axis remains for those regions within 0.5%.

The third clinical example is a nine beam head-and-neck plan performed using simultaneous integrated boost technique with three targets planned to 50, 60 and 70 Gy with

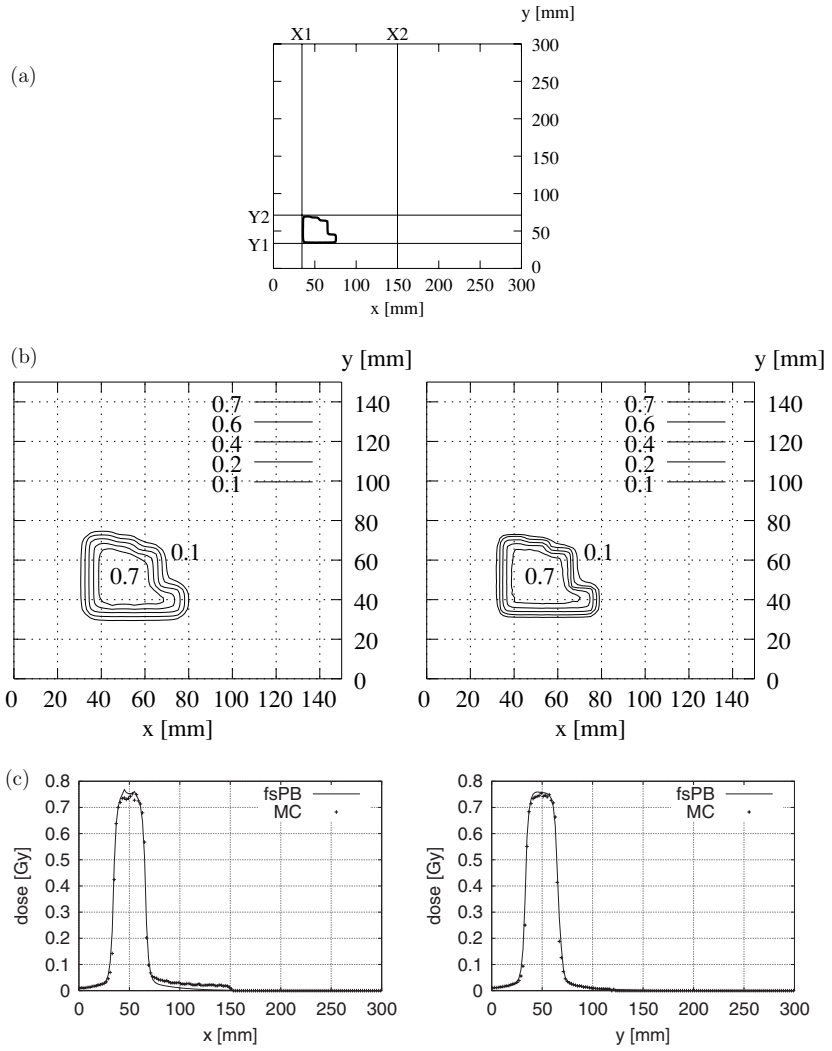

Figure 12. Results of dose computation for an exemplary IMRT segment (photon beam energy 6 MV): (a) segment localization and shape (jaw positions marked as X1, X2, Y1 and Y2), (b) finite size pencil beam (left) and Monte Carlo (right) computed dose distributions at 15 cm depth, (c) profiles at depth 15 cm for $y = 55$ mm and for $x = 55$ mm—fsPB (lines) and MC (symbols).

the constraint of 35 Gy equivalent uniform dose for the spinal cord. The presented plan, computed on a dose grid of $(3.2\ \text{mm})^3$, with photon beam energy of 6 MV consists of 233 segments to be delivered with an Elekta MLC. The dose distribution computed with Monte Carlo and the proposed finite size pencil beam algorithm in an exemplary slice and the corresponding χ-factor distribution are presented in figure 16 and reveal satisfactory agreement, especially in the large gradient area around the spinal cord.

4. Discussion

The algorithm presented here was designed for the purposes of beamlet-based IMRT optimization. It is sufficiently fast (approximately 4000 beamlets/min for a head and neck

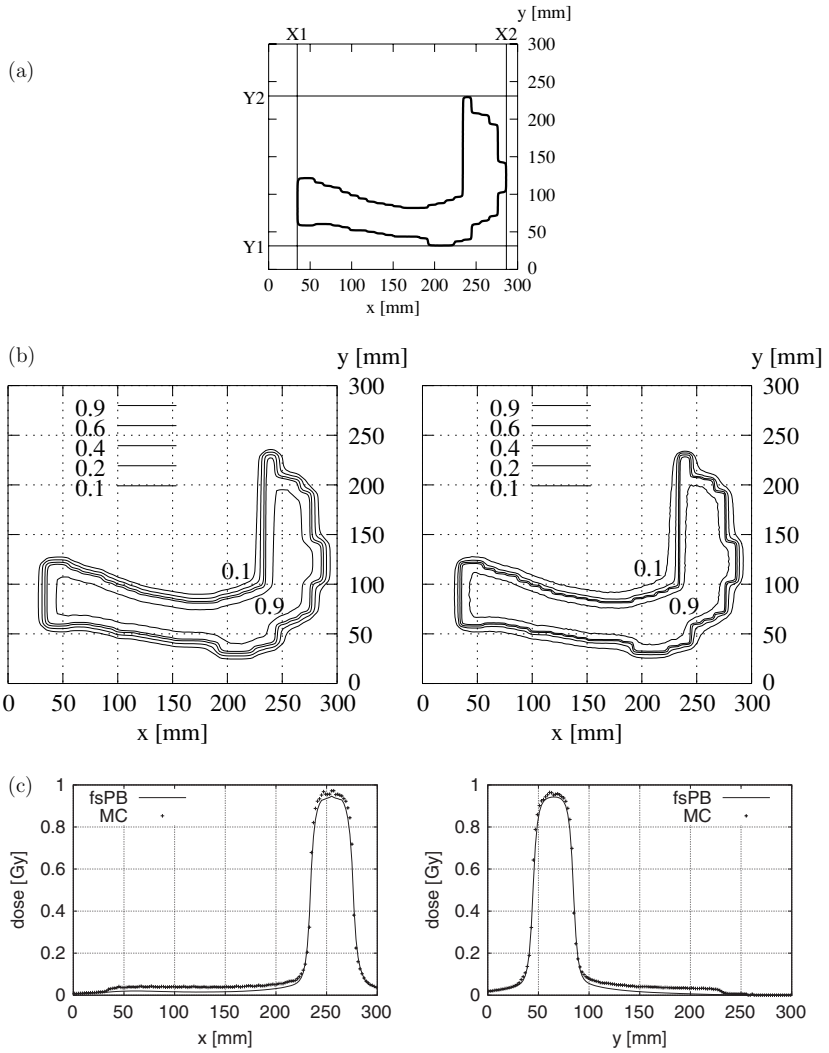

Figure 13. Results of dose computation for an exemplary IMRT segment (photon beam energy 15 MV): (a) segment localization and shape (jaw positions marked as X1, X2, Y1 and Y2), (b) finite size pencil beam (left) and Monte Carlo (right) computed dose distributions at 15 cm depth, (c) profiles at depth 15 cm for $y = 150$ mm and for $x = 150$ mm—fsPB (lines) and MC (symbols).

case on a $(3\ \text{mm})^3$ dose grid on a 2.66 GHz Intel Xeon PC) in order to pre-compute and store the beamlet dose distributions prior to optimization. Due to its construction, the dose gradients within modulated fields are always very close to deliverable dose gradients. As a consequence, the intermediate dose distributions during iterative optimization never become unrealistic, which is an important factor for fast convergence (Laub *et al* 2000, Jeraj *et al* 2002) and also guarantees that a re-computation with a high-precision dose algorithm like Monte Carlo does not show gross discrepancies, especially in the dose gradients around organs at risk.

A second objective in the algorithm design was that commissioning should be possible with broad beam measurements to avoid systematic measurement errors for small field sizes.

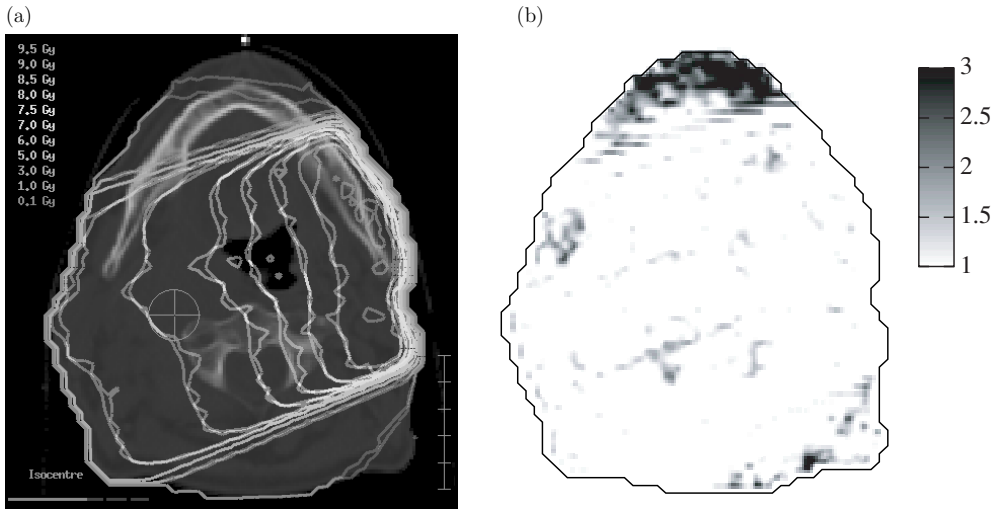

Figure 14. (a) Dose distributions obtained with Monte Carlo and finite size pencil beam algorithm (overlapped) for head-and-neck case, beam energy 6 MV, (b) corresponding χ-factor distribution ($\Delta d = 2$ mm, $\Delta D = 2\%$ relative to the local value of the Monte Carlo calculated dose, full line—patient contour, values >3 are not resolved).

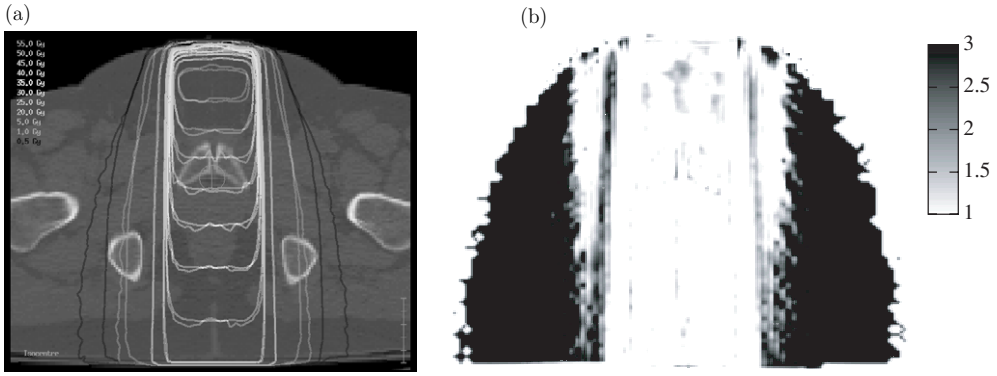

Figure 15. (a) Dose distributions obtained with Monte Carlo and finite size pencil beam algorithm (overlapped) for prostate case, beam energy 15 MV, (b) Corresponding χ-factor distribution ($\Delta d = 2$ mm, $\Delta D = 2\%$ relative to the local value of the Monte Carlo calculated dose, values >3 are not resolved).

As was shown with the computation of output factors for small fields, this objective could be achieved without compromising accuracy for smallest fields. This is of particular importance for MLC-based IMRT delivery.

As for any specialized dose computation algorithm, simplifying assumptions had to be made. This is mostly the assumption of self-consistency and field size and SSD independence of the penumbra component that originates from accelerator head and phantom scatter. The problems with small fields and short SSD at 15 MV stem from this approximation, but seem acceptable. For more standardized conditions, the accuracy of the fsPB algorithm is sufficient for all practical purposes. Its obvious shortcomings are the lack of lateral density correction and the inclusion of MLC leaf shape effects like tongue and groove and inter-leaf scatter.

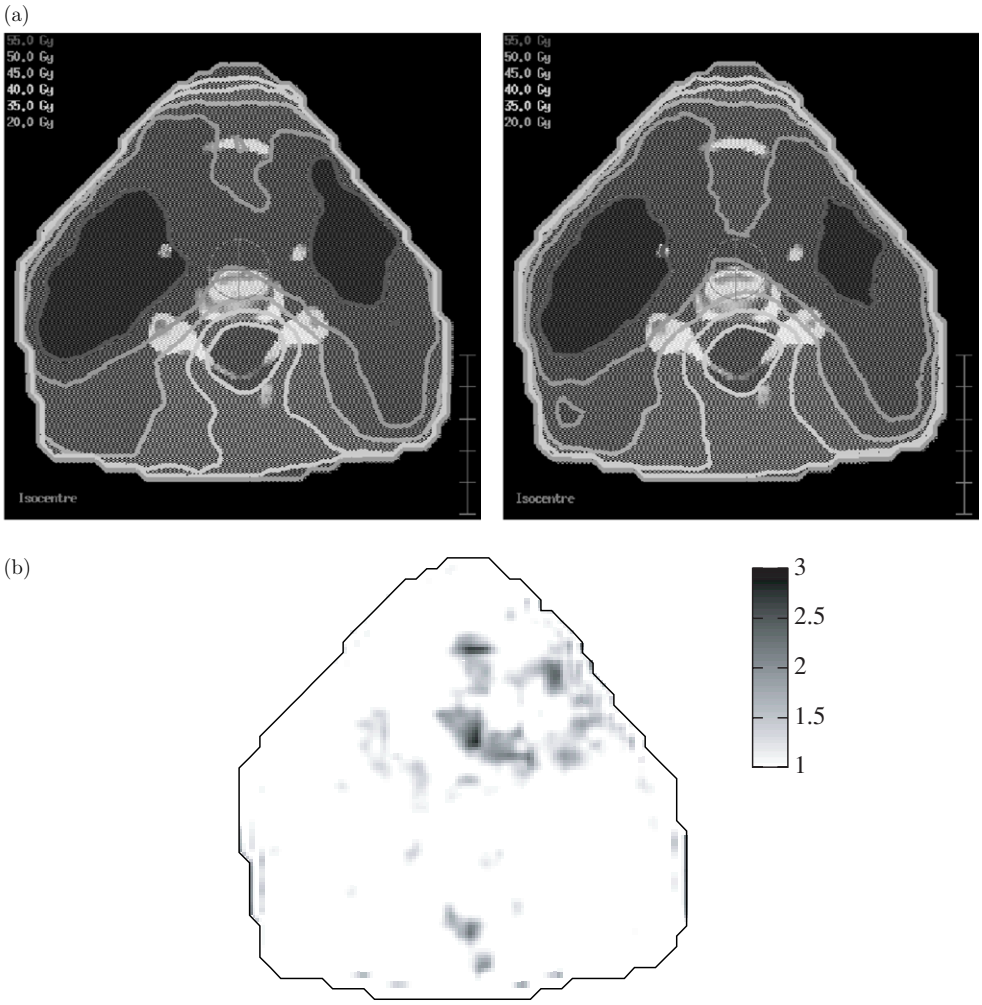

Figure 16. (a) Dose distributions obtained with finite size pencil beam algorithm (left) and Monte Carlo (right) for a nine beam head-and-neck plan, beam energy 6 MV. (b) Corresponding χ-factor distribution ($\Delta d = 2$ mm, $\Delta D = 2\%$ relative to the local value of the Monte Carlo calculated dose, full line—patient contour, values >3 are not resolved).

While it is not clear how the latter can be included in this framework, the former is the subject of current research.

The algorithm presented here has been in clinical use for IMRT planning since 2001 in combination with the Monte Carlo code XVMC as a second stage verification computation. The concept of pre-computation and storage of beamlet dose distributions as well as the close proximity between optimized (with fsPB) and final dose (with Monte Carlo) is a crucial element in accelerating IMRT routine planning.

Acknowledgments

This work was supported by DFG grant no 33/7-2 and Marie Curie Fellowship HPMT-CT-2001-00221-03.

References

Ahnesjö A, Andreo P and Brahme A 1987 Calculation and application of point spread functions for treatment planning with high energy photon beams *Acta Oncol.* **26** 49–56

Ahnesjö A and Asparadakis M M 1999 Dose calculations for external photon beams in radiotherapy *Phys. Med. Biol.* **44** R99–155

Ahnesjö A, Saxner M and Trepp A 1992 A pencil beam model for photon dose calculations *Phys. Med. Biol.* **36** 973–85

Bakai A, Alber M and Nüsslin F 2003 A revision of the γ-evaluation concept for the comparison of dose distributions *Phys. Med. Biol.* **48** 3543–53

Bortfeld T, Oelfke U and Nill S 2000 What is the optimum leaf width of a multileaf collimator? *Med. Phys.* **27** 2494–502

Bortfeld T, Schlegel W and Rhein B 1993 Decomposition of pencil beam kernels for fast dose calculation in three-dimensional treatment planning *Med. Phys.* **20** 311–18

Bourland J D and Chaney E L 1992 A finite-size pencil beam model for photon dose calculations in three dimensions *Med. Phys.* **19** 1401–12

Boyer A L and Mok E C 1984 A photon dose distribution model employing convolution calculations *Med. Phys.* **12** 169–77

Ceberg C P, Bjärngard B E and Zhu T C 1996 Experimental determination of the dose kernel in high-energy x-ray beams *Med. Phys.* **23** 505–11

Chui C-S and Mohan R 1988 Extraction of pencil beam kernels by the deconvolution method *Med. Phys.* **15** 138–44

De Gersem W, Claus F, De Wagter C, Van Duyse B and De Neve W 2001 Leaf position optimization for step-and-shoot IMRT *Int. J. Radiat. Oncol. Biol. Phys.* **51** 1371–88

Fippel M 1999 Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm *Med. Phys.* **26** 1466–75

Fippel M, Haryanto F, Dohm O, Nuẅsslin F and Kriesen S 2003 A virtual photon energy fluence model for Monte Carlo dose calculation *Med. Phys.* **30** 301–11

Fippel M, Laub W, Huber B and Nüsslin F 1999 Experimental investigation of a fast Monte Carlo photon beam dose calculation algorithm *Phys. Med. Biol.* **44** 3039–54

Fraass B, Doppke K and Hunt M *et al* 1998 American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning *Med. Phys.* **25** 1773–829

ICRU 1987 Use of Computers in External Beam Radiotherapy Procedures with High-energy Photons and Electrons *ICRU Publication 42* (Bethesda, MD: International Commission on Radiation Units and Measurements)

Jeraj R, Keall P J and Siebers J V 2002 The effect of dose calculation accuracy on inverse treatment planning *Phys. Med. Biol.* **47** 391–407

Laub W, Alber M, Birkner M and Nüsslin F 2000 Monte Carlo dose computation for IMRT optimization *Phys. Med. Biol.* **45** 1741–54

Low D A, Harms W B, Mutic S and Purdy J A 1998 A technique for the quantitative evaluation of dose calculation *Med. Phys.* **25** 656–61

Mackie T R, Bielajew A F, Rogers D W O and Battista J J 1988 Generation of photon energy deposition kernels using the EGS Monte Carlo code *Phys. Med. Biol.* **33** 1–20

Mackie T R, Scrimger J W and Battista J J 1985 A convolution method of calculating dose for 15-MV x rays *Med. Phys.* **12** 188–96

Mohan R and Chui C 1987 Use of fast Fourier transforms in calculating dose distributions for irregularly shaped fields for three-dimensional treatment planning *Med. Phys.* **14** 70–7

Mohan R, Chui C and Lidofsky L 1986 Differential pencil beam dose computation model for photons *Med. Phys.* **13** 64–73

Ostapiak Z, Zhu Y and Van Dyk J 1997 Refinements of the finite-size pencil beam model of three-dimensional photon dose calculation *Med. Phys.* **24** 743–50

Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in C. The Art of Scientific Computing* 2nd edn (New York: Cambridge University Press) pp 683–7

Schwarz M, Bos L J, Mijnheer B J, Lebesque J V and Damen E M F 2003 Importance of accurate dose calculations outside segment edges in intensity modulated radiotherapy treatment planning *Radiother. Oncol.* **69** 305–14

Storchi P and Woudstra E 1996 Calculation of the absorbed dose distribution due to irregularly shaped photon beams using pencil beam kernels derived from basic beam data *Phys. Med. Biol.* **41** 637–56

Storchi P R M, van Battum L J and Woudstra E 1999 Calculation of pencil beam kernel from measured photon beam data *Phys. Med. Biol.* **44** 2917–28

Van Dyk J, Barnett R B, Cygler J E and Shragge P C 1993 Commissioning and quality assurance for treatment planning computers *Int. J. Radiat. Oncol. Biol. Phys.* **26** 261–73

Venselaar J and Welleweerd H 2001 Application of a test package in an intercomparison of the performance of treatment planning systems used in a clinical setting *Radiother. Oncol.* **60** 203–13

Venselaar J, Welleweerd H and Mijnheer B 2001 Tolerances for the accuracy of photon beam dose calculations of treatment planning systems *Radiother. Oncol.* **60** 191–201

Appendix H

A study of the robustness of IMPT and IMRT for prostate against organ movement

accepted for publication in the
International Journal of Radiation Oncology, Biology and Physics

A study of the robustness of IMPT and IMRT for prostate against organ movement

Martin Soukup, Ph.D. [a,b] Matthias Söhn, Dipl. Phys. [a]
Di Yan, D.Sc. [c] Jian Liang, Ph.D. [c] Markus Alber, Ph.D. [a]

[a] *Section for Biomedical Physics, University Hospital for Radiation Oncology, Tübingen, Germany*

[b] *CMS GmbH, Freiburg, Germany*

[c] *Department of Radiation Oncology, William Beaumont Hospital, Royal Oak, MI*

Corresponding author: Martin Soukup, Ph.D., Section for Biomedical Physics, University Hospital for Radiation Oncology, Hoppe-Seyler-Str. 3, 72076 Tübingen, Germany; Tel: ++49 7071 2986062; Fax: ++49 7071 295920; E-mail: Martin.Soukup@cmsrtp.com

Running title: Robustness of IMPT and IMRT for prostate against organ movement

Acknowledgements: This work was supported in part by Deutsche Forschungsgemeinschaft, Grant No. 33/8-2.

This work was presented in part at the 27th Annual Congress of The European Society for Therapeutic Radiology and Oncology (ESTR0), Göteborg, Sweden, September 14-18, 2008

Conflict of Interest Notification: Martin Soukup is employed by CMS GmbH, Freiburg, Germany since 2009.

Abstract

Purpose: Intensity modulated radiotherapy with photons (IMRT) and protons (IMPT) produces dose distributions that have high conformality to the PTV and sufficient sparing of the OAR if calculated on a single static CT. For prostate patients, organ movement with related changes of the density distribution in the irradiated volume occurs during the treatment course. We evaluated the sensitivity of IMPT and IMRT plans to organ movement.

Methods and Materials: IMPT and IMRT treatment plans were evaluated for 4 patients with an average of 16 CT data sets per patient. The treatment plans were recalculated on all treatment CTs, the dose was accumulated in the reference geometry by means of a deformable registration algorithm. Accurate dose calculation

methods were applied for both IMPT and IMRT.

Results: With IMPT unacceptably low total doses in the GTV were observed for patients with gas in the rectum in the planning CT. To achieve similar total EUD and EUD spread as with IMRT, two methods were crucial for IMPT - a rectal gas water-equivalent density overwrite in the original planning CT and initial beam weight setting to achieve a homogeneous dose distribution over the whole PTV for each field separately. An improvement of the total EUD for prostate and rectum wall was observed also for IMRT with the water-equivalent density overwrite of the rectal cavities.

Conclusions: The sensitivities of IMPT and IMRT to organ movement are of the same order if applying appropriate planning strategies. The latter is especially crucial for IMPT.

Key words: Intensity modulated radiation therapy, Proton therapy, Organ motion, Prostate cancer, Monte Carlo

1 Introduction

The method with the largest potential in proton therapy is IMPT [1] as it offers the largest degree of freedom by optimizing the energy fluence of narrow proton beams. However, the more heterogeneous the dose distributions of each field in the PTV is, the more carefull one has to be regarding the robustness of the treatment plan [2].

In order to decide whether the rather costly proton therapy is beneficial for prostate patients compared to modern photon irradiation techniques, several treatment planning comparisons were performed ([3], [4], [5], [6], [7], [8], [9]). Similar dose distributions were usually found in the high dose region, for the low and intermediate dose levels, protons were superior.

However, proton beam ranges are also very sensitive to the material crossed, and by uncertainties about the material the Bragg peak can be shifted or in the presence of lateral heterogeneities distorted (e.g. [10], [11]). To diminish the influence of heterogeneities caused by a rectal gas filling and to increase the volume of the rectum, proton therapy is often performed with a rectal balloon filled with water (e.g. [12]).

Zhang et al. [7] have shown the impact of changing patient geometries for prostate treatment with passive scattering proton therapy and IMRT. They compared the mean dose values for 10 patients resulting from 8 CTs per patient and reported not worse dose distribution changes for protons as for IMRT.

However, due to the high conformality and heterogenous dose distributions of the fields, IMPT can be expected to be potentially more sensitive to organ movements than the passive scattering technique. In this study we therefore tried to estimate the influence of

organ movement on the accumulated dose distribution during the irradiation course for IMRT and IMPT.

2 Methods and Materials

IMPT and IMRT treatment plans were evaluated for 4 patients from the William Beaumont Hospital (Royal Oak, USA) with an average of 16 CT data sets per patient. The treatment plans were calculated on one CT data set and recalculated on the CTs taken during the irradiation course (treatment CTs). As we wanted to see the influence of the choice of the initial CT, the procedure was repeated two to three times for each patient for the first three CT geometries with the same planning contours and prescription. Three of the four patients had residuals of the contrast agent in the bladder and/or rectum in the first CT. For these patients, the treatment planning was performed on the second and third CT only.

Following the clinical IMRT routine in Tübingen, the CTV, rectum and bladder were created as a hull of the contours from the first three CTs based on [13]. The PTV results from the expansion of the CTV by a 7 mm margin. For IMRT 9 equidistant fields (0,40,80,120,160,200,240,280,320 degrees) and for IMPT two opposing fields from 90 and 270 degrees were applied. The same PTV and OAR prescription was used for IMRT and IMPT. A dose grid of $3x3x3mm^3$ was defined. The treatment planning system for both IMRT and IMPT was Hyperion [14]. The dose prescription was based on the EUD concept with the Poisson cell kill model for the PTV and serial constraints for the OAR. The goal for PTV was 78 Gy ($\alpha = 0.4$), the constraints were for the bladder an EUD of 60 Gy (k=8) and for rectum 65 Gy (k=12; entire rectum volume, representing the late effects like rectal bleeding [15]) and 48 Gy (k=4; volume not overlapping with PTV, representing the acute reactions). Additionally, quadratic overdosage constraints were applied for the PTV and healthy tissue. The constraints can not be exceeded, thus the optimization algorithm was trying to find the best PTV coverage within the given toxicity levels. Further details about the optimization concept can be found in [16], [17], [18]. To estimate the influence of gas volumes in rectum, another set of treatment plans was created with rectal gas water-equivalent density overwrite in the original planning CT.

All CTs were matched to the bony structures around the prostate. Thus daily patient setup corrections (shifts, rotations) to reach the position of the planning CT were simulated. The dose was recalculated on each matched CT. The approach of correction to bony structures is applied routinely for all IMRT prostate patients in Tübingen. With this approach, the material changes seen by the proton beam can be expected to be smaller than for soft tissue based patient setup schemes. The evaluation of the robustness for target based patient setup is out of the scope of this paper and should be subject of further investigation.

To evaluate the total dose to organs correctly, the dose distributions from all CTs were

warped and accumulated in the reference geometry with a deformable organ registration algorithm [19] for the prostate, bladder and the rectum wall. From the accumulated dose distributions the accumulated EUDs were evaluated for the entire organs.

Due to the high degeneracy of the optimization problem in IMPT, many solutions fulfill the optimization constraints. Thus, the final "optimum" beam weights depend very much on the initial conditions. Two methods were used to define the initial conditions. In the first method, beams from all fields were optimized at the same time starting with a constant beam weight. In the second method, two optimizations stages were introduced - in the first stage, beams were optimized for each field separately with relaxed skin (unspecified tissue within the skin contours) constraints and in the second stage all preoptimized beams from both fields were optimized again at the same time. The second method is thus similar to the approach of Paul Scherer Institut (Villingen, Switzerland) where the beams are assigned with initial beam weights leading to a homogeneous spread out Bragg peak [1], [20].

The initial beam size is $\sigma_x = \sigma_y = 3$ mm, the beams were placed on a raster of 4 mm x 4 mm on the spot plane and 4 mm water equivalent Bragg peak spacing in depth. The dose was calculated with a pencil beam algorithm [21] that was also implemented in the commercially available treatment planning system CMS XiO (St. Louis, USA). As we were interested in effects due to changing heterogeneities, a high level beam decomposition was chosen (121 subbeams per beam; corresponds to approximately 1.5 mm subbeam distance) and because of the higher initial proton energies also the large angle scatter correction was used. For these settings, we previously reported a systematic error of less than 1.5% if compared with the Monte Carlo code VMCpro [22] for prostate cases [23]. In case gas was found in the rectum in the original CT, an additional treatment plan was performed with the gas cavity overwritten with water-equivalent density.

IMRT was calculated for an Elekta linear accelerator with Elekta Beam Modulator as MLC for the nominal energy of 15 MV. The Monte Carlo algorithm XVMC [24] with the virtual source model [25] was used for the treatment plan segment optimization. The number of primary particle histories was set to achieve a statistical uncertainty of 2.5% per field segment.

3 Results

Robustness of treatment plans was considered for different approaches (planning CT rectal gas density with and without overwriting to water-equivalent density, different setting of initial beam weights). The dose was recalculated on each treatment CT geometry, warped to the first geometry with the deformable organ registration algorithm, accumulated and finally the EUDs were evaluated. Concurrently, the dose was recalculated on each treatment geometry, the EUDs were evaluated separately and the mean EUDs of all configurations were calculated (See Table 1). As a measure of robustness we consider

Table 1
Accumulated EUD after recalculation in the treatment CTs for patients with largest amount of rectal gas in the planning CT (Patient A with planning CT nr 0, patient B with CT nr 1, patient C with CT nr 2). 2nd column, the mean EUD over all treatment CTs and its standard deviation. 'W' marks the rectal gas water density overwrite in the original planning CT, 'Sep' marks the IMPT optimization method with field dependent initial spot weights.

	A0 EUD [Gy]		B1 EUD [Gy]		C2 EUD [Gy]	
	accum.	mean	accum.	mean	accum.	mean
Prostate $\alpha = 0.4$						
IMRT	78.4	78.3 ± 0.2	78.1	78.0 ± 0.3	76.5	76.6 ± 0.8
IMRT W	78.7	78.7 ± 0.1	78.1	78.0 ± 0.3	77.8	77.8 ± 0.6
IMPT	68.4	64.1 ± 11.3	52.8	51.5 ± 10.9	71.5	65.9 ± 12.4
IMPT Sep	74.4	70.4 ± 8.6	67.7	65.8 ± 7.7	73.3	67.9 ± 10.6
IMPT W	77.9	77.8 ± 0.9	77.9	77.6 ± 1.8	77.5	77.3 ± 1.2
IMPT W Sep	78.5	78.5 ± 0.1	78.2	78.1 ± 0.2	77.4	77.4 ± 1.0
Bladder						
IMRT	65.1	66.6 ± 3.1	61.0	61.9 ± 3.2	59.7	61.0 ± 4.0
IMRT W	65.6	67.1 ± 3.2	61.0	61.9 ± 3.2	60.3	61.8 ± 4.2
IMPT	63.0	65.5 ± 3.2	60.2	61.6 ± 4.4	58.6	60.6 ± 4.5
IMPT Sep	63.8	66.2 ± 3.1	60.2	61.5 ± 3.3	58.1	60.2 ± 4.5
IMPT W	63.1	65.4 ± 3.4	59.9	61.4 ± 4.1	58.6	60.6 ± 4.7
IMPT W Sep	63.8	66.2 ± 3.3	60.1	61.5 ± 3.4	58.4	60.4 ± 4.6
Rectum wall late						
IMRT	66.2	66.4 ± 1.0	67.0	67.8 ± 1.8	66.3	66.3 ± 1.9
IMRT W	63.5	63.6 ± 1.2	64.3	65.0 ± 1.9	65.1	65.1 ± 2.7
IMPT	57.4	58.3 ± 3.5	63.3	64.0 ± 2.6	53.1	56.3 ± 9.3
IMPT Sep	57.6	58.3 ± 2.8	63.3	64.1 ± 1.5	55.7	58.2 ± 7.3
IMPT W	65.3	66.4 ± 1.8	65.1	66.3 ± 2.7	62.1	64.4 ± 6.5
IMPT W Sep	63.7	64.6 ± 1.1	64.9	66.0 ± 1.3	62.6	64.3 ± 3.9
Rectum wall acute						
IMRT	54.2	53.7 ± 1.8	54.1	55.0 ± 2.3	55.3	54.4 ± 2.6
IMRT W	51.1	50.5 ± 1.7	52.0	52.7 ± 2.3	52.6	51.8 ± 3.3
IMPT	39.7	39.9 ± 4.0	47.3	48.3 ± 3.2	37.8	39.8 ± 9.6
IMPT Sep	40.2	40.4 ± 3.6	48.2	49.3 ± 2.6	38.9	40.6 ± 8.5
IMPT W	47.5	47.9 ± 2.4	50.4	51.8 ± 3.4	45.9	47.1 ± 7.3
IMPT W Sep	46.5	46.8 ± 2.1	50.3	51.7 ± 2.4	46.1	47.1 ± 5.6

the difference between the planned EUD and the applied EUD (both accumulated and evaluated separately) and its variability.

For IMPT the sensitivity to organ motion was strongly dependent on the method of optimization (initial beam weights) and on the gas cavities in the rectum. Starting the optimization from same beam weights lead to two patched fields accordingly to published data (e.g. [20]). This method resulted in a higher spread of the EUD values in the treatment CTs in prostate and rectum wall (Table 1). The accumulated EUD to the prostate was up to about 25 Gy (plan B1 in Table 1) lower than the prescribed value. More robust plans could be achieved with preoptimizing the beam weights for each field separately, but still the difference to the prescribed value can be up to about 10 Gy (plan B1). Best results for IMPT in terms of robustness provided a combination of the rectal gas water-equivalent density overwrite in the original planning CT and preoptimizing the fields separately (Figures 1, 2). Under these conditions, comparable results as with IMRT were achieved for the EUD in the prostate and the serial constraints with high coefficients (higher dose parts of the organs) for the bladder and the rectum wall (Table 2). Starting the optimization from the same beam weights and using the rectal gas water-equivalent

Table 2
Accumulated EUD for all plans with rectal gas water density overwrite in the original planning CT for IMRT and IMPT. IMPT is with field dependent initial spot weights. 'planned' denotes the EUD accumulated on the three organ contours used for the hull definition on the original plan, 'realized' denotes the accumulated EUD on the treatment CTs from the fourth CT onward.

	A0	A1	A2	B1	B2	C1	C2	D1	D2
Prostate $\alpha = 0.4$	[Gy]	[Gy]	[Gy]	[Gy]	[Gy]	[Gy]	[Gy]	[Gy]	[Gy]
IMRT W planned	78.7	78.6	78.7	78.0	78.0	77.5	77.5	75.4	75.6
IMRT W realized	78.7	78.4	78.5	78.1	78.4	77.4	77.8	74.6	74.9
IMPT W Sep planned	78.7	78.7	78.6	78.1	78.1	77.7	77.6	76.0	75.9
IMPT W Sep realized	78.5	78.6	78.5	78.2	78.1	77.6	77.4	75.2	75.1
Bladder									
IMRT W planned	61.3	61.4	61.4	58.2	58.1	60.2	60.4	59.0	59.0
IMRT W realized	65.6	65.3	65.6	61.0	61.0	59.9	60.3	60.6	60.6
IMPT W Sep planned	59.8	59.7	59.5	56.9	57.0	59.1	59.0	58.0	57.9
IMPT W Sep realized	63.8	63.4	63.2	60.1	60.2	58.7	58.4	59.6	59.5
Rectum wall late									
IMRT W planned	64.4	64.3	64.3	63.2	63.6	66.5	66.7	61.2	61.3
IMRT W realized	63.5	63.3	63.3	64.3	65.0	64.8	65.1	65.0	65.1
IMPT W Sep planned	64.4	64.2	63.9	63.0	62.9	65.6	65.6	60.2	60.3
IMPT W Sep realized	63.7	63.6	63.2	64.9	64.5	62.8	62.6	64.2	64.1
Rectum wall acute									
IMRT W planned	51.7	51.6	51.6	51.5	51.7	53.8	53.9	51.0	51.3
IMRT W realized	51.1	51.0	50.9	52.0	52.3	52.6	52.6	55.4	55.6
IMPT W Sep planned	46.9	46.7	46.4	48.8	48.9	48.9	48.8	45.7	45.7
IMPT W Sep realized	46.5	46.4	46.0	50.3	49.9	46.3	46.1	52.1	52.0

density overwrite in the original planning CT was less robust than with the separate field preoptimizing. Although it lead to similar accumulated EUDs for the prostate and rectum wall, the spread of the EUDs in the treatment CTs was much higher (e.g. the mean EUD standard deviation for prostate increased from 0.1 Gy to 0.9 Gy for plan A0 and for the rectum wall late effects from 1.3 Gy to 2.7 Gy for plan B1).

An improvement of the accumulated prostate and rectum wall EUD with rectal gas water-equivalent density overwrite in the original planning CT was observed also for IMRT (Table 1). The sensitivity was, however, not as large as for IMPT compared to IMRT. The accumulated EUD for prostate is larger by up to 1.0 Gy and the decrease of EUD (k=12) for the rectum wall between 1.2 Gy and 2.7 Gy. Bladder EUD did not change significantly. An example of the EUD sensitivity to the different optimization methods is presented in Figure 2. In the following, unless explicitly stated otherwise, we present only results for the most robust approaches (i.e. rectal gas water-equivalent density overwrite for IMRT and IMPT and separate field preoptimizing for IMPT).

The DVHs and therefore EUDs showed a systematic difference between the mean values resulting out of averaging the DVH and EUDs without considering the deformable organ registration and accumulated values with deformable organ registration (Table 1, Figure 3). The DVHs with deformable organ registration had less extreme doses on both sides for both IMRT and IMPT. While for the prostate the accumulated EUD of the treatment CTs was equal or higher than the mean EUD by up to 0.3 Gy, the systematic differences were especially large for the OARs. For bladder the accumulated EUD was lower by 1.0 Gy to 2.9 Gy than the mean EUD. The accumulated EUD for the late effects on the

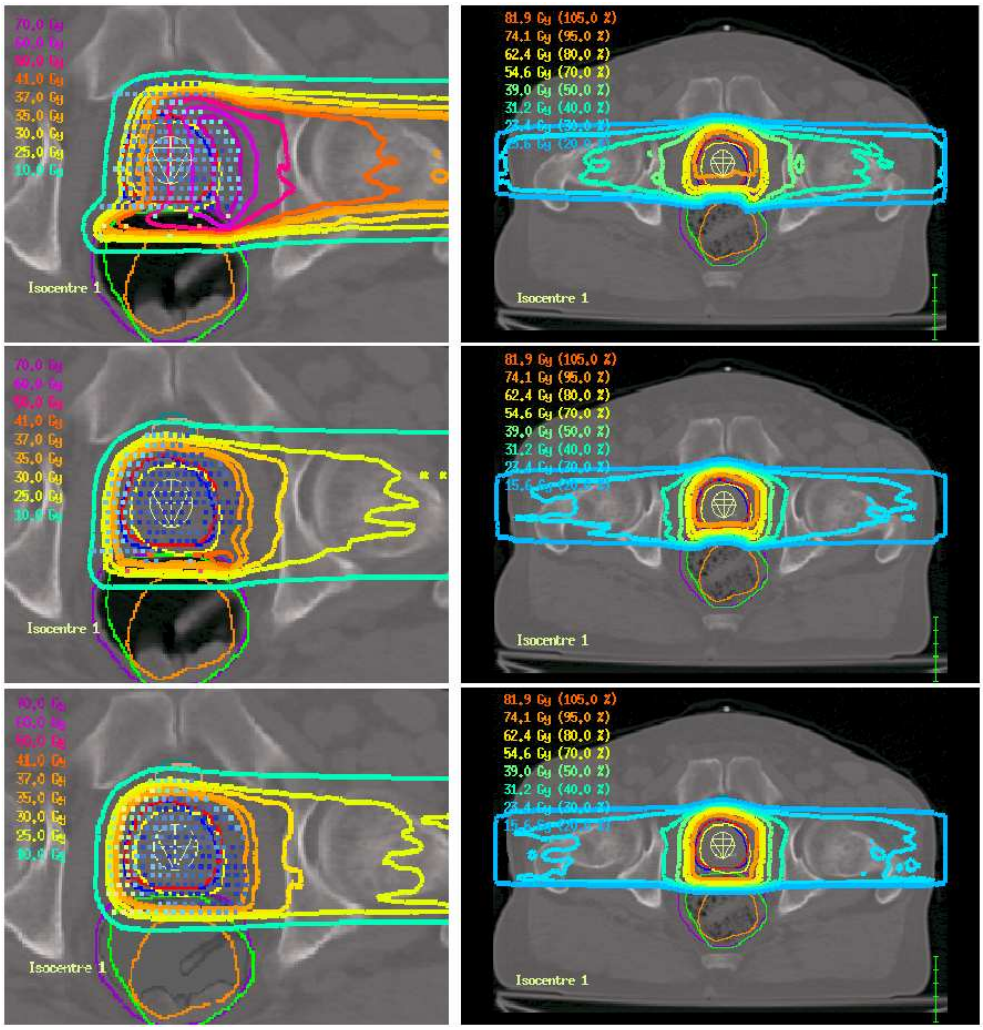

Fig. 1. Left, the dose distributions of the original plan for patient C (planning CT number 2) for one field (Gantry 90^o) obtained with the different IMPT techniques - initial weights constant (top), initial weights corresponding to homogeneous PTV coverage for each field without (IMPT Sep) and with (IMPT W Sep) rectal gas water-equivalent density overwrite in the planning CT (middle, bottom). The colored points represent the Bragg peaks and the optimized beam weight (lowest = blue, highest = red). Right, the corresponding total plan dose distributions after recalculation on another CT (CT number 3).

rectum wall was equal or lower by up to 1.7 Gy. For acute effects on the rectum wall, the difference of accumuluted EUD to the mean EUD reached from 0.8 Gy to -1.6 Gy. Thus, the mean DVH and mean EUD should be used only as a conservative estimate, for the real dose volume histograms and EUD a deformable organ registration algorithm is necessary.

When using different planning CT data sets, the comparison of the accumulated EUDs shows only small differences both for IMRT and IMPT (Table 2). Thus, the choice of planning CT is not critical neither for IMRT nor for IMPT as long as either a CT without rectal gas cavities is chosen or the density of the cavities is overwritten with the water-

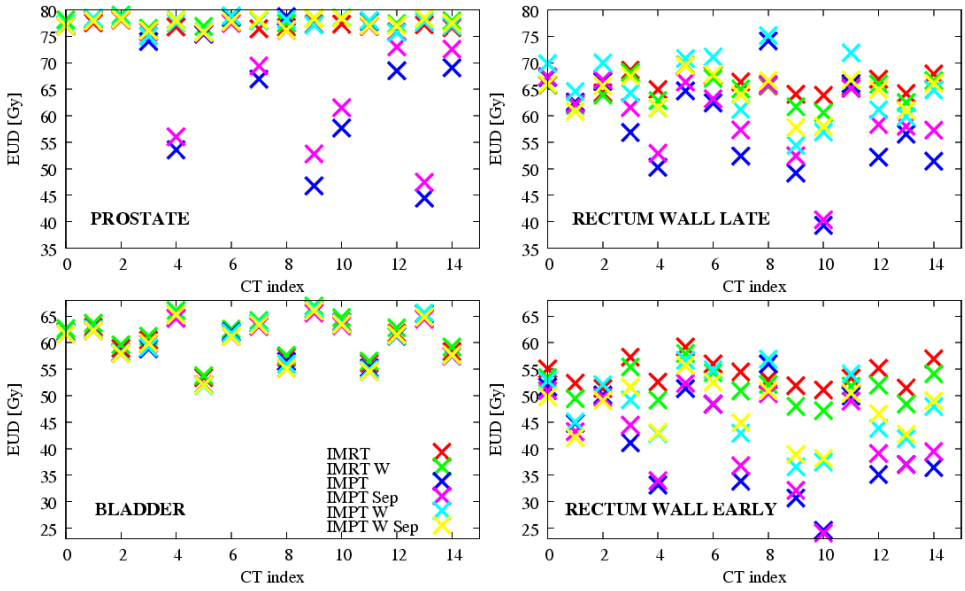

Fig. 2. EUDs for patient C (planning CT number 2) after recalculation on the treatment CTs for different optimization methods ('W' marks the rectal gas water-equivalent density overwrite and 'Sep' the setting of field dependent initial beam weights). The EUDs belonging to the contours of the first three CTs (planning contours) are evaluated on the planning dose distribution.

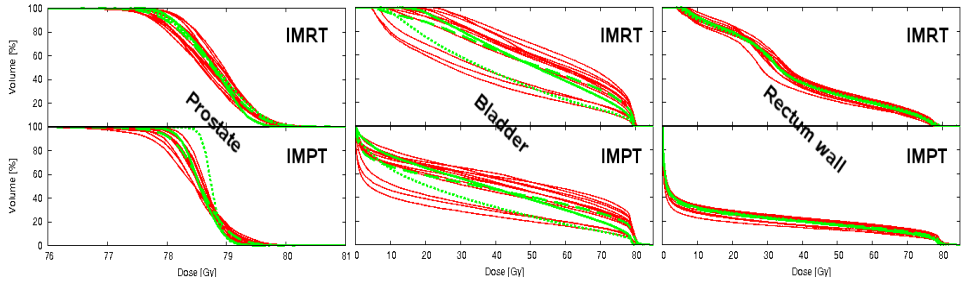

Fig. 3. The DVH comparison for IMRT and IMPT for patient A (planning CT number 0). For both methods, the most robust approaches were used, i.e. rectal gas water-equivalent density overwrite in the original planning CT for the rectal cavities and in case of IMPT also the field dependent initial beam weights (IMRT W, IMPT W Sep). In red, the DVHs after recalculation on the single treatment CTs. In green and dotted, the accumulated DVH of the planning dose distribution on the first three contours. In green and solid, the accumulated DVH on the treatment CTs. In green and dashed, the mean DVH on the treatment CTs resulting from averaging the single DVHs. .

equivalent density.

For patient D, the OAR constraints were too tight to reach the goal of 78 Gy in the PTV because of the large movements of the organs resulting in a big overlap of the PTV and the OAR hulls. An accumulated planning EUD to the prostate of 75.4 Gy (planning CT nr 1) and 75.6 Gy (planning CT number 2) was reached for IMRT and 76.0 Gy and 75.9 Gy for IMPT respectively. Another effect of the big overlap is the heterogeneous dose distribution in the PTV for the original plan. The dose heterogeneity obviously leads to

Table 3
Accumulated EUD averaged over all plans and patients (total of 9). The lower EUD index defines in which CTs the dose was calculated. The upper index marks contour sets which were used for dose evaluation. 'realized' denotes all CTs after the first three CTs. 'planned' denotes the original CT (i.e. the planning CT). E.g. for $EUD^{realized}_{planned}$ the dose was not recalculated for each CT, but the deformed organ contours were used for dose accumulation.

IMRT W	Prostate [Gy] $\alpha = 0.4$	Bladder [Gy] $k = 8$	Rectum wall [Gy] $k = 12$	$k = 4$
$EUD^{planned}_{planned}$	77.6 ± 1.2	59.9 ± 1.4	63.9 ± 1.9	52.0 ± 1.0
$EUD^{realized}_{realized}$	77.4 ± 1.6	62.2 ± 2.5	64.4 ± 0.8	52.6 ± 1.8
$EUD^{realized}_{planned}$	77.4 ± 1.5	62.2 ± 2.5	64.7 ± 0.6	52.9 ± 1.7
$EUD^{planned}_{planned} - EUD^{realized}_{realized}$	0.2 ± 0.4	-2.3 ± 1.7	-0.4 ± 2.2	-0.6 ± 2.2
$EUD^{planned}_{planned} - EUD^{realized}_{planned}$	0.2 ± 0.4	-1.8 ± 1.7	-0.9 ± 1.9	-0.9 ± 2.1
$EUD^{realized}_{realized} - EUD^{realized}_{planned}$	0.0 ± 0.2	0.0 ± 0.2	-0.4 ± 0.4	-0.3 ± 0.2
IMPT W Sep				
$EUD^{planned}_{planned}$	77.7 ± 1.1	58.5 ± 1.1	63.3 ± 2.0	47.4 ± 1.4
$EUD^{realized}_{realized}$	77.4 ± 1.4	60.8 ± 2.1	63.7 ± 0.8	48.4 ± 2.6
$EUD^{realized}_{planned}$	77.6 ± 1.4	60.9 ± 2.2	63.9 ± 0.7	48.5 ± 2.7
$EUD^{planned}_{planned} - EUD^{realized}_{realized}$	0.3 ± 0.3	-2.2 ± 1.8	-0.4 ± 2.6	-1.0 ± 3.3
$EUD^{planned}_{planned} - EUD^{realized}_{planned}$	0.1 ± 0.3	-2.4 ± 1.8	-0.5 ± 2.5	-1.0 ± 3.4
$EUD^{realized}_{realized} - EUD^{realized}_{planned}$	-0.1 ± 0.1	-0.2 ± 0.1	-0.1 ± 0.3	-0.1 ± 0.1

an increase of the sensitivity to organ motions expressed by the largest degradation of the dose in the prostate in the patient collective (Table 2) of about 0.8 Gy for patient D for both techniques. Within our margin concept, the choice is therefore between an increased toxicity in the OAR via relaxing the constraints and increased sensitivity to organ motions.

With 0.8 Gy being the largest degradation, the hull+margin concept of the initial three CTs proved to be very robust for the prostate for both IMPT and IMRT. For the OARs the hull concept was less robust. For patient A, the prescription of 60 Gy for the bladder hull was exceeded in the realized accumulated dose to the bladder by up to 5.6 Gy with IMRT and 3.8 Gy with IMPT. For this patient the first bladder contour was abnormally large reaching about 2 cm further in the cranial direction than most of the following bladder contours. Thus a lot of volume far away from the PTV was included to the bladder hull leading to an underestimation of the EUD on the treatment CTs if compared to the planning EUD. The hull concept for the rectum was conservative enough to lead to accumulated EUDs of rectum wall being only maximally by 0.1 Gy larger than the prescribed 65 Gy to the rectum hull.

By averaging the accumulated EUD over all patients and plans, we could quantify the plan population robustness. For the prostate, the difference between the original treatment plan EUD and the accumulated EUD on the treatment CTs was very low both for IMRT and IMPT - a decrease of 0.2±0.4 Gy, respectively 0.3±0.3 Gy ($EUD^{planned}_{planned} - EUD^{realized}_{realized}$ in Table 3).

The difference between the original treatment plan EUD and the accumulated EUD on the treatment CTs were biggest for the bladder with an increase of 2.3 ± 1.7 Gy for IMRT and 2.2 ± 1.8 Gy for IMPT. The difference between the accumulated EUD on the treatment CTs and the accumulated EUD evaluated on the original planning dose distribution with the contours from the treatment CTs ($EUD^{realized}_{realized} - EUD^{realized}_{planned}$ in Table 3) was very small: 0.0 ± 0.2 Gy, respectively -0.2 ± 0.1 Gy. This indicates that the influence of the changing overall density distribution during the course is marginal for both IMRT and IMPT ($EUD^{realized}_{realized} - EUD^{realized}_{planned}$ in Table 3), if compared to the variation of the bladder contours (size, shape, position) during the irradiation course ($EUD^{planned}_{planned} - EUD^{realized}_{planned}$).

Similarly, but to a lesser extent than for the bladder, the population based accumulated EUD increases also for the rectum wall by 0.4 ± 2.2 Gy (late effects) and 0.6 ± 2.2 Gy (acute effects) for IMRT and 0.4 ± 2.6 Gy and 1.0 ± 3.3 Gy, respectively, for IMPT. Again, the organ form and position changes play a more important role than the overall CT density changes with $EUD^{realized}_{realized} - EUD^{realized}_{planned}$ having negligible values of -0.4 ± 0.4 Gy (late effects) and -0.3 ± 0.2 Gy (acute effects) for IMRT and -0.1 ± 0.3 Gy and -0.1 ± 0.1 Gy respectively for IMPT.

Thus, to estimate the planning EUD reliability, recalculation of the dose on the treatment CTs is not crucial and the evaluation on the planning CT dose distribution with the treatment CT contours is sufficient.

The maximum dose in the OAR is often higher for IMPT than for IMRT. For the same EUD constraint, the sharp gradients in IMPT can result in reduced high dose volumes but with higher maximum dose. In this study, we did not try to tighten the constraints further as the main purpose of this work was to compare the sensitivity of IMPT and IMRT under same conditions.

4 Discussion

The robustness of IMRT and IMPT treatment plans to the inter-fractional organ changes was evaluated for four patients with an average of 16 CT data sets per patient. The robustness of the the treatment plans was very much dependent on the treatment planning approaches.

The rectal gas water-equivalent density overwrite in the planning CT is especially important for the stability of the dose distribution in IMPT. With gaseous cavities in the overlapping parts of PTV and rectum, lower maximal energies are necessary for proton beams crossing this area to reach the distal edge of the PTV. In case of the rectum being filled with more dense material in the following CTs, not enough dose is delivered to these distal parts of the PTV. An additional problem is the dose distortion of beams crossing the lateral heterogeneity interface of gas and tissue leading to a different shape of dose distribution in case this interface disappears. In case there are gas cavities in the following

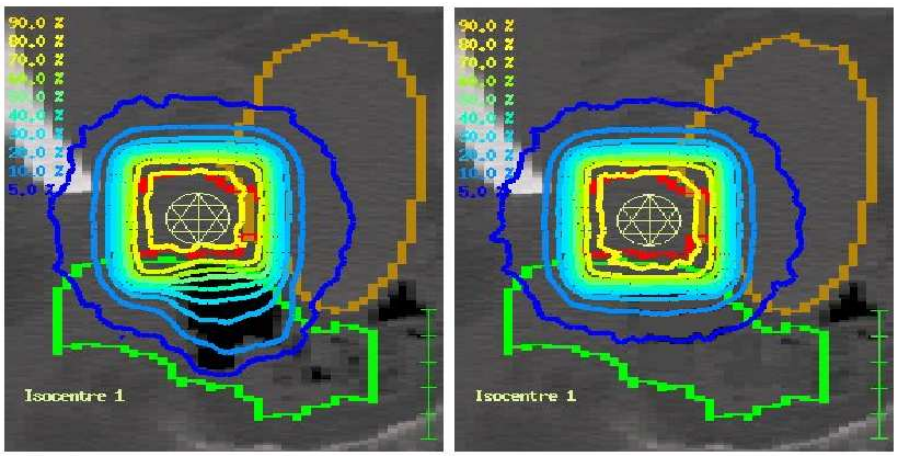

Fig. 4. The isodose distribution from the sagittal view in the isocentric plane for a $5x5cm^2$ photon field (gantry 90^o). Left, the original CT, right a density overwrite of rectal gas filling with water-equivalent density. A problem arises, if the PTV overlaps with the cavity as the optimizer compensates for the dose loss due to lateral scatter.

CTs, the Bragg peaks are shifted into the healthy tissue on the opposing side of the field entrance and cause a dose increase here, fortunately mostly outside of OARs.

The rectal gas water-equivalent density overwrite in the planning CT helps to improve the dose distribution also for IMRT. The reason is different than in case of IMPT. We believe that the improvement is mainly caused by the improved penumbra in water due to preventing the secondary electron disequilibrium at the gas/soft tissue interface. The effect leads to less dose in the overlapping and neighboring part of the PTV and the rectal gas cavities, forcing the optimizer to deliver more dose to the rectum region than in case of water-equivalent density overwriting (Figure 4). Thus, to obtain better robustness of the plan for both IMPT and IMRT, one should prefer performing the plan optimization on a planning CT with rectal cavities artificially overwritten with water instead of the exact original CT with gas cavities. Another possibility of diminishing the influence of the rectal gas cavities is the real filling of rectum with water with help of a rectal balloon (e.g. [12]). A further aspect of the use of rectal balloons is the potential reduction of the organ movements. Thus the use of rectal balloons might lead to an increase of the plan robustness.

The hull+margin concept of the initial three CTs lead to very robust results for the prostate for both IMPT and IMRT. To further increase the dose to prostate without increasing the toxicity and to increase the robustness for OARs, more sophisticated approaches like coverage probability distribution ([26]) or adaptive treatment planning (e.g. [13], [17]) will be necessary.

The solution space is usually more highly degenerate for IMPT and the initial conditions can thus play an important role for the final solution. These solutions can also have different sensitivities (Table 1). An interesting problem in this context is the choice of the cost functions as the degeneracy originates from the particular formulation of the

cost functions. An indirect way of forcing the optimizer to deliver a rather homogeneous dose distribution per field and thus potentially increase the robustness is tightening of the skin constraints (not shown here). However, a too large tightening of the skin constraints leads to a less optimal solution for the OARs and/or the PTV. It appears therefore to be more straightforward to start from a solution with beam weights optimized for robustness (homogeneity) directly instead of trying to optimize for robustness only indirectly via the skin constraints. Another potential way of influencing the homogeneity of dose per field is the use of regularisation techniques such as fluence smoothness penalty functions.

The problem of Hounsfield Unit (HU) uncertainties was not considered in this work. While IMRT is expected to be robust against HU changes in the order of 1% to 3%, such changes can lead to several millimeters shift of the Bragg peak in IMPT. In case the HU error is systematic for the whole patient this can lead to a distortion of the whole dose distribution [27]. Even with daily CTs and repositioning, setup errors will occur to some extent. We did not consider the influence of setup errors.

5 Conclusions

The appearance of large gas cavities in the rectum in the original treatment planning CT is the only geometrical factor significantly degrading the dose distribution in the following CTs for IMPT. This problem can be overcome by overwriting of the rectal gas in the original CT with water-equivalent density.

Another important factor for the sensitivity of the IMPT plan is the choice of the initial proton beam weights. If starting from identical weights, the solution is more sensitive to the geometrical changes with respect to the variance of dose per fraction which can affect biological fractionation effects. In contrast, the IMPT plans become more robust if setting the initial weights such as to obtain a homogeneous coverage of the PTV from each field separately.

Overwriting of the rectal gas in the original CT with water-equivalent density and setting the initial weights of proton beams to obtain a homogeneous coverage of the PTV from each field separately, the sensitivity of IMPT to organ motions is of the same order as with IMRT. The changes of the organ shapes and positions are more important to the accumulated dose in prostate and OAR than the material changes.

The choice of the initial CT for treatment planning was not critical for the results neither for IMRT nor for IMPT.

In terms of estimated probability of late rectal and bladder complications and the prostate coverage in the realized treatment, IMPT was mostly equal or slightly better than IMRT and it is unclear whether the difference is of clinical relevance.

References

[1] Lomax A. Intensity modulation methods for proton radiotherapy. *Phys Med Biol* 1999;44:185-205

[2] Lomax A. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Phys Med Biol* 2008;53:1043-56

[3] Cella L, Lomax A, Miralbell R. Potential role of intensity modulated proton beams in prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys* 2001;49:217-23

[4] Mock U, Bogner J, Georg D, *et al.* Comparative treatment planning on localized prostate carcinoma conformal photon- versus proton-based radiotherapy. *Strahlenther Onkol* 2005;181:448-55

[5] Schneider U, Lomax A, Pemler P, *et al.* The impact of IMRT and proton radiotherapy on secondary cancer incidence. *Strahlenther Onkol* 2006;182:647-52

[6] Trofimov A, Nguyen PL, Coen JJ, *et al.* Radiotherapy treatment of early-stage prostate cancer with IMRT and protons: a treatment planning comparison. *Int J Radiat Oncol Biol Phys* 2007;69:444-53

[7] Zhang X, Dong L, Lee AK, *et al.* Effect of anatomic motion on proton therapy dose distributions in prostate cancer treatment. *Int J Radiat Oncol Biol Phys* 2007;67:620-9

[8] Muzik J, Soukup M, Alber M. Comparison of fixed-beam IMRT, helical tomotherapy, and IMPT for selected cases. *Med Phys* 2008; 35:1580-92

[9] Vargas C, Fryer A, Mahajan C, *et al.* Dose-volume comparison of proton therapy and intensity-modulated radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2008;70:744-51

[10] Goitein M. A technique for calculating the influence of thin inhomogeneities on charged particle beams. *Med Phys* 1978; 5:258-64

[11] Goitein M, Chen G, Ting J, *et al.* Measurements and calculations of the influence of thin inhomogeneities on charged particle beams. *Med Phys* 1978;5:265-73

[12] Slater J, Yonemoto L, Rossi C, *et al.* Conformal proton therapy for prostate carcinoma. *Int J Radiat Oncol Biol Phys* 1998; 42:299-304

[13] Yan D, Lockman D, Brabbins D, *et al.* An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer. *Int J Radiat Oncol Biol Phys* 2000;48:289-302

[14] Alber M, Birkner M, Laub W, *et al.* Hyperion - an integrated IMRT planning tool *13th International Conference on the Use of Computers in Radiation Therapy* (Heidelberg: Springer) 2000;46-8

[15] Söhn M, Yan D, Liang J, *et al.* Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* 2007;67:1066-73

[16] Alber M, Nuesslin F. An objective function for radiation treatment optimization based on local biological measures. *Phys Med Biol* 1999;44:479-93

[17] Birkner M, Yan D, Alber M, *et al.* Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation. *Med Phys.* 2003;40:2822-31

[18] Alber M, Reemtsen R. Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method. *Optim Methods Softw* 2007;22:391-411

[19] Yan D, Jaffray DA, Wong JW. A model to accumulate fractionated dose in a deforming organ. *Int J Radiat Oncol Biol Phys* 1999;44:665-75

[20] Albertini F, Lomax AJ, Hug EB. In regard to Trofimov et al.: Radiotherapy treatment of early-stage prostate cancer with IMRT and protons: a treatment planning comparison (Int J Radiat Oncol Biol Phys 2007;69:444-453). *Int J Radiat Oncol Biol Phys* 2007;69:1333-4

[21] Soukup M, Fippel M and Alber M. A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations. *Phys Med Biol* 2005;50:5089-104

[22] Fippel M and Soukup M. A Monte Carlo dose calculation algorithm for proton therapy. *Med Phys* 2004; 31:2263-73

[23] Soukup M, Alber M. Influence of dose engine accuracy on the optimum dose distribution in intensity-modulated proton therapy treatment plans. *Phys Med Biol* 2007;52:725-40

[24] Fippel M. Fast Monte Carlo dose calculation for photon beams based on VMC electron algorithm. *Med Phys* 1999;26:1466-75

[25] Sikora M, Dohm O, Alber M. A virtual photon source model of an Elekta linear accelerator with integrated mini MLC for Monte Carlo based IMRT dose calculation. *Phys Med Biol* 2007;52:4449-63

[26] Baum C, Alber M, Birkner M, *et al.* Robust treatment planning for intensity modulated radiotherapy of prostate cancer based on coverage probabilities. *Radiother Oncol* 2006;78:27-35

[27] Lomax A. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Phys Med Biol* 2008;53:1027-42