Physics in Medicine & Biology

PAPER • **OPEN ACCESS**

Prioritized efficiency optimization for intensity modulated proton therapy

To cite this article: Birgit S Müller and Jan J Wilkens 2016 *Phys. Med. Biol.* **61** 8249

OPEN ACCESS
IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology
Phys. Med. Biol. **61** (2016) 8249–8265 doi:10.1088/0031-9155/61/23/8249

Prioritized efficiency optimization for intensity modulated proton therapy

Birgit S Müller[1,2] **and Jan J Wilkens**[1,2]

[1] Department of Radiation Oncology, Technical University of Munich, Klinikum rechts der Isar, Ismaninger Straße 22, 81675 Munich, Germany
[2] Physik-Department, Technical University of Munich, James-Franck-Straße 1, 85748 Garching, Germany

E-mail: birgit.mueller@tum.de

Received 14 April 2016, revised 15 September 2016
Accepted for publication 27 September 2016
Published 3 November 2016

Abstract

A high dosimetric quality and short treatment time are major goals in radiotherapy planning. Intensity modulated proton therapy (IMPT) plans obtain dose distributions of great conformity but often result in long delivery times which are typically not incorporated into the optimization process. We present an algorithm to optimize delivery efficiency of IMPT plans while maintaining plan quality, and study the potential trade-offs of these interdependent objectives. The algorithm is based on prioritized optimization, a stepwise approach to implemented objectives. First the quality of the plan is optimized. The second step of the prioritized efficiency optimization (PrEfOpt) routine offers four alternatives for reducing delivery time: minimization of the total spot weight sum (A), maximization of the lowest spot intensity of each energy layer (B), elimination of low-weighted spots (C) or energy layers (D). The trade-off between dosimetric quality (step I) and treatment time (step II) is controlled during the optimization by option-dependent parameters. PrEfOpt was applied to a clinical patient case, and plans for different trade-offs were calculated. Delivery times were simulated for two virtual facilities with constant and variable proton current, i.e. independent and dependent on the optimized spot weight distributions. Delivery times decreased without major degradation of plan quality; absolute time reductions varied with the applied method and facility type. Minimizing the total spot weight sum (A) reduced times by 28% for a similar plan quality at a constant current (changes of minimum dose in the target $<1\%$). For a variable proton current, eliminating low-weighted spots (C) led to remarkably faster delivery (16%). The implementation of an efficiency-

optimization step into the optimization process can yield reduced delivery times with similar plan qualities. A potential clinical application of PrEfOpt is the generation of multiple plans with different trade-offs for a multicriteria optimization setting. Then, the planner can select the preferred compromise between treatment time and quality for each individual patient.

Keywords: prioritized optimization, IMPT treatment planning, delivery efficiency

(Some figures may appear in colour only in the online journal)

1. Introduction

Intensity modulated radiotherapy treatment planning involves optimization of multiple criteria. Dosimetric goals defined in physicians' prescriptions, such as the coverage of the planning target volume (PTV) with a prescribed dose and dose limits to organs at risk (OARs), are often conflicting and require compromises. In addition to those dosimetric goals, short treatment times are a major concern in radiotherapy for various reasons: faster plan deliveries increase patient comfort, cost-effectiveness and the number of patients that can be handled by the clinic. Shorter irradiation times potentially reduce intrafractional dose uncertainties, due to patient movement and organ deformation, and may further have a positive impact on biological response (Paganetti 2005, Bewes *et al* 2008, Suzuki *et al* 2011, Mittauer *et al* 2013, van de Water *et al* 2015). During the last few years faster irradiation techniques have been developed, volumetric arc therapy being an example of a modality that is being increasingly used due to its short treatment times (Otto 2008).

In intensity modulated radiotherapy with photons (IMXT), as well as protons (IMPT), the degree of modulation often determines the quality of the plan and consequently the irradiation time (RT time). The aim of greater efficiency is correlated with plan quality, such that the major challenge when increasing delivery efficiency is not to compromise dosimetric quality. Treatment planning faces a trade-off problem, not only with respect to conflicting dosimetric goals but also between plan quality and delivery efficiency (Craft *et al* 2007, Bortfeld and Webb 2009, Mittauer *et al* 2013, Wilkie *et al* 2013).

Multicriteria optimization (MCO) is a method well suited to investigate the optimization of interdependent objectives. MCO treatment planning is based on the creation of a database of optimal plans, the so-called Pareto front. It allows the user to search interactively by surface navigation for the plan with the best trade-off (Craft *et al* 2005). Craft *et al* utilized this approach to study the complexity of IMXT plans as a measure of efficiency versus quality (Craft *et al* 2007).

Previous studies on correlations between delivery efficiency and quality demonstrated that RT times of IMXT and IMPT plans can often be decreased without compromising plan quality (Coselmon *et al* 2005, Craft *et al* 2007, Kang *et al* 2008, Mittauer *et al* 2013, Cao *et al* 2014, van de Water *et al* 2015). Sometimes plans are optimized to an unnecessary complexity (Bortfeld and Webb 2009). The degeneracy of the solution space allows one to generate similar dose distributions using different fluence maps or spot patterns (Bortfeld and Webb 2009, Cao *et al* 2014).

Different methods for decreasing IMPT delivery times have already been published (Kang *et al* 2008, Cao *et al* 2014, van de Water *et al* 2015) and are partly realized in clinical practice. IMPT RT times can be reduced at different stages of the planning process: prior to optimization by selecting a large spot grid, via post-processing methods, e.g. by eliminating selective spots,

or as part of the optimization routine. The initial spot raster, given by the lateral spot distance Δx and Δy and the steps in depth Δz, defines the number of available spots. Larger grids typically reduce application times but concurrently decrease the number of degrees of freedom for the optimization. A trade-off strategy, which keeps the number of spots at a minimum level but still high enough to achieve adequate dosimetric qualities, is required (Hillbrand and Georg 2010). Non-uniform depth scanning decreases the number of required energies, in particular for deep-seated tumors (Kang *et al* 2008). Recent publications have demonstrated remarkable reductions of IMPT times by eliminating energy layers, especially for facilities with large energy switch times (Cao *et al* 2014, van de Water *et al* 2015).

IMPT delivery times strongly depend on the individual facility and its characteristics. Proton currents are frequently assumed to be constant or are not specifically addressed. The characteristic of variable currents (dependent on the optimized spot weight distribution) is crucial for some clinics: treatment currents determined by the lowest spot weight per energy layer often result in low currents and prolong treatment times.

With respect to delivery efficiency, certain spot distributions may be more favorable than others for different centers. In this context it should also be mentioned that proton facilities face a lowest spot limit given by the monitor chamber which can lead to undeliverable spot patterns. Post-processing interactions such as manual elimination of the 'responsible spots' can solve the issue but may degrade plan quality (Zhu *et al* 2010, Howard *et al* 2014). By incorporating the limitations of the monitor unit into the optimization, deliverable spot distributions are assured while plan quality is maintained (Cao *et al* 2013).

Here we present an optimization routine for IMPT planning which integrates delivery efficiency into a prioritized optimization scheme in order to reduce treatment times and meanwhile control plan quality. Prioritized treatment planning was first suggested by Wilkens *et al* (2007) for IMXT plan optimization as a stepwise approach to implementation of optimization goals. It translated the hierarchy of objectives into a mathematical optimization routine. The prioritized efficiency-optimization routine presented here ('PrEfOpt') optimizes the plan quality first and allows treatment time to be reduced via various alternative methods in the final step of the routine.

We demonstrate the feasibility and potential of the efficiency-optimization method by applying it to a clinical astrocytoma case. In order to compare the efficacy of the implemented method and its dependence on proton current properties two generic facility types were implemented, i.e. with constant and variable proton currents.

2. Material and methods

2.1. The prioritized efficiency-optimization routine

The prioritized efficiency-optimization algorithm (PrEfOpt) prescribes a stepwise approach to implementation of optimization objectives. Each step is realized by a corresponding objective function, while the order of steps refers to the clinical relevance of the corresponding goal. PrEfOpt consists of two steps (see figure 1; for mathematical formulations see section 2.2). Step I comprises two consecutive sub-runs to optimize the dose distribution according to the clinical prescription: first (step Ia) the homogeneous coverage of the PTV is optimized; second (step Ib) the mean dose to selected structures is reduced by considering prior achievements, introduced as constraints. For both sub-runs maximum dose constraints are specified prior to step Ia, and these are maintained throughout the whole optimization routine including the following step II.

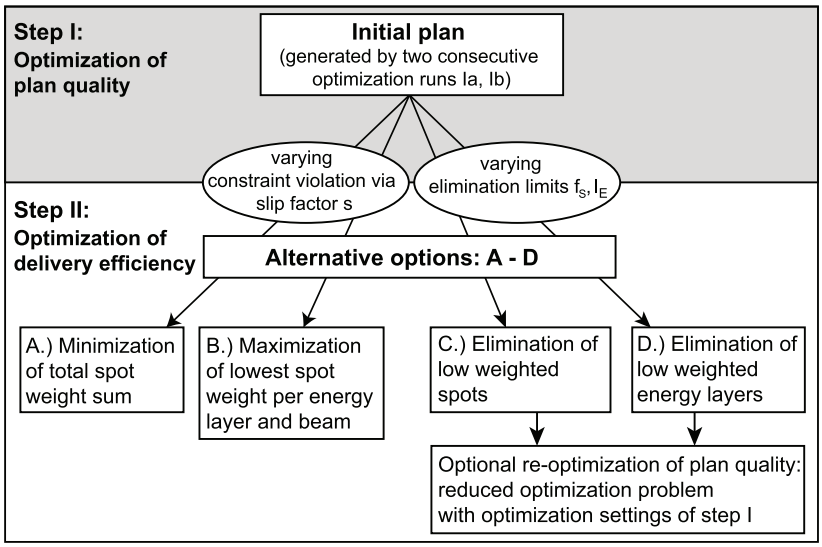

Figure 1. Optimization routine of PrEfOpt: plan quality is obtained in step I according to Wilkens *et al* (2007); step II offers four alternative methods for reducing the treatment time. The trade-off between plan quality and treatment time, i.e. between steps I and II, is controlled by two types of 'trade-off parameters': the slip factor and the elimination limits.

The treatment time is optimized in step II by one of four alternative methods. Method A reduces the sum over all spot intensities. Method B maximizes the lowest spot intensity per energy and beam. For both methods prior dosimetric achievements are turned into hard constraints which may be violated within a small specified range given by the so called slip-factor s. Method C and D reduce the spot pattern by eliminating spots and energy layers, respectively. The number of eliminated spots (C) is defined by a variable limiting factor f_S (here applied to the median spot weight). Instead of comparing each spot with the median spot weight the criterion may be adapted as one wishes, e.g. to the average PTV dose contribution of each spot. The number of eliminated energy layers per beam (method D) is defined by the parameter l_E: the l_E lowest weighted spot layers are deleted. For both methods, C and D, a successive re-optimization run of the plan quality using the remaining spots/layers and otherwise identical parameters as in step I is optional.

2.2. Mathematical formulation of PrEfOpt

Each step of PrEfOpt comprises an objective function to optimize the vector $\underline{\omega}$ of n spot weights ω_j which multiplied by the influence matrix D_{ij} yields the dose $D_i(\underline{\omega})$ to the voxel i. OAR and PTV describe the set of voxels of the corresponding structure volumes (total number of voxels: N_{OAR} and N_{PTV}). Each spot weight ω_j belongs to one beam B (number of beams: N_B) and one energy layer.

Let us introduce the following mathematical definitions:

The energy of the kth energy layer of the mth beam, with the number of energy slices M_m in the mth beam:

$$E_{m,k} \text{ with } k = \{1, ..., M_m\} \tag{1}$$

and the ensemble of spots j in this energy layer:

$$L_{m,k} = \{ j \in \{1, ..., n\} \,|\, j \text{ belongs to } k\text{th energy layer of the } m\text{th beam} \} \tag{2}$$

All optimization runs are subject to:

$$\omega_j \geqslant 0 \quad \forall\ j \in \{1, ..., n\} \tag{3}$$

and defined hard maximum dose constraints $D^{\max}$ to OARs:

$$D_i(\underline{\omega}) \leqslant D_i^{\max} \quad \forall\ i \in \text{OAR} \tag{4}$$

Step I: Optimization of homogeneous tumor coverage with the prescribed dose D^{prescr} by the standard quadratic objective function (Ia) and reduction of the mean dose D^{mean} to selective OARs (Ib) while the prior optimized PTV coverage is maintained by turning the achievements of step Ia into hard constraints. The introduction of a so-called slip-factor $\tilde{s} \geqslant 0$ allows for small deteriorations to enlarge the solution space. For the discussed case $\tilde{s} = 0.1$.

Step Ia:

$$F^{\text{Ia}}(\underline{\omega}^{\text{Ia}}) = \sum_{i \in \text{PTV}} (D_i(\underline{\omega}^{\text{Ia}}) - D^{\text{prescr}})^2 = \min \tag{5}$$

Step Ib:

$$F^{\text{Ib}}(\underline{\omega}^{\text{Ib}}) = \frac{1}{N_{\text{OAR}}} \sum_{i \in \text{OAR}} D_i(\underline{\omega}^{\text{Ib}}) = \min \tag{6}$$

subject to:

$$F^{\text{Ia}}(\underline{\omega}^{\text{Ib}}) = \sum_{i \in \text{PTV}} (D_i(\underline{\omega}^{\text{Ib}}) - D^{\text{prescr}})^2 \leqslant F^{\text{Ia}}(\underline{\omega}^{\text{Ia}})(1 + \tilde{s}) \tag{7}$$

Step II: Optimization of delivery efficiency by one of four alternative methods (A–D).

Objective functions of cases A and B are optimized with respect to previous achievements. A slip factor $s \geqslant 0$ is introduced to control the trade-off between the plan quality (step I) and treatment time.

Case A: Minimization of the sum over all spot weights:

$$F^{\text{II}}(\underline{\omega}^{\text{II}}) = \sum_{j=1}^{n} \omega_j^{\text{II}} = \min \tag{8}$$

subject to:

$$\sum_{i \in \text{PTV}} (D_i(\underline{\omega}^{\text{II}}) - D^{\text{prescr}})^2 \leqslant F^{\text{Ia}}(\underline{\omega}^{\text{Ia}})(1 + s)^2 \tag{9}$$

$$\frac{1}{N_{\text{OAR}}} \sum_{i \in \text{OAR}} D_i(\underline{\omega}^{\text{II}}) \leqslant F^{\text{Ib}}(\underline{\omega}^{\text{Ib}}) \tag{10}$$

Note that the slip factor is only applied to the PTV term, not to the achieved D^{mean} of the OAR.

Case B: Maximization of the lowest spot weight $\breve{\omega}_{m,k} = \min(\omega_j | \, j \in L_{m,k})$.

The available number of spot is reduced beforehand by excluding all spots which had zero weight after step Ib:

$$\omega_j^{\text{II}} = 0 \qquad \forall \quad j \quad \text{with} \quad \omega_j^{\text{Ib}} = 0 \tag{11}$$

$$F^{\mathrm{II}}(\breve{\underline{\omega}}) = \sum_{B} \sum_{k=1}^{M_m} \breve{\omega}_{m,k} = \max \tag{12}$$

subject to equations (9) and (10) and:

$$\omega_j^{\mathrm{II}} > \breve{\omega}_{m,k} \qquad \forall\ j \in L_{m,k} \tag{13}$$

Cases C and D reduce the optimization problem by a defined number of spots (limiting factor f_S) (C) or energy layers (limit l_E) (D).

Case C: Elimination of spot weights which are lower than the specified threshold given by the median spot weight $\tilde{\omega}_j^{\mathrm{Ib}}$ and the variable factor f_S:

$$\omega_j^{\mathrm{II}} = 0 \qquad \forall \quad j \quad \text{with} \quad \omega_j^{\mathrm{II}} \leqslant f_S \cdot \tilde{\omega}_j^{\mathrm{Ib}} \quad \text{with } f_S \geqslant 0 \tag{14}$$

Case D: Elimination of the l_E lowest weighted energy layers of each beam:

$$\begin{aligned} &\omega_j^{\mathrm{II}} = 0 \quad \forall\ j \in \{l_E \text{ lowest sums of spot weights with } j \in L_{m,k}\} \\ &\text{with } l_E \in \mathbb{N} \end{aligned} \tag{15}$$

Case C and D: Optional re-optimization of the resulting optimization problem with the reduced number of spots and the optimization parameter used in step I.

2.3. Treatment time implementation

The algorithm for and the potential of each optimization method are demonstrated based on time calculations with two generic facility types. The implemented virtual facility types are based on simplified assumptions and present examples of different current specifications, which primarily serve to demonstrate potential differences in the results derived using the chosen efficiency-optimization method.

Treatment times t_{RT} were calculated as the sum over energy switch times Δt_E, the time for the adjustment of each spot position $\Delta t_A = 0.1$ ms (adapted from van de Water *et al* (2015)) and the actual beam-on time Δt_S. The minimum spot weight was defined as $w_{\mathrm{feasible}} = 1$, referring to 10^6 particles. All spot weights $\omega_j \leqslant w_{\mathrm{feasible}}$ were set to zero after optimization. The default energy switch time was $\Delta t_E = 1$ s.

The simulated facilities differed in their proton current specifications.

Facility type 1: constant proton current $I_{\mathrm{c}} = 0.5$ nA, independent of the optimized spot weights.

Facility type 2: variable proton current $I_{\mathrm{v}}(\breve{\omega}_{m,k})$ in the range of $[I_{\min}, I_{\max}]$, dependent on the lowest spot intensity of each energy layer and beam $\breve{\omega}_{m,k}$. $I_{\mathrm{v}}(\breve{\omega}_{m,k})$ was implemented by a linear function between a lower and upper weight limit, $\omega_{ll} = \omega_{\mathrm{feasible}}$ and $\omega_{ul} = 30$:

$$\begin{aligned} &\breve{\omega}_{m,k} = \omega_{ll} \quad \rightarrow \quad I_{\mathrm{v}} = I_{\min} = 0.1 \text{ nA} \\ &\breve{\omega}_{m,k} \geqslant \omega_{ul} \quad \rightarrow \quad I_{\mathrm{v}} = I_{\max} = 2.0 \text{ nA} \\ &\omega_{ll} \leqslant \breve{\omega}_{m,k} \leqslant \omega_{ul} \quad \rightarrow \quad I_{\mathrm{v}} = \frac{I_{\max} - I_{\min}}{\omega_{ul} - \omega_{ll}} \cdot \breve{\omega}_{m,k} + \frac{I_{\min}\omega_{ul} - I_{\max}\omega_{ll}}{\omega_{ul} - \omega_{ll}} \end{aligned} \tag{16}$$

The efficacy of each method was evaluated by relative time savings Δt_{RT}:

$$\Delta t_{\mathrm{RT}} = \frac{t_{\mathrm{RT}}(\omega_j^{\mathrm{Ib}}) - t_{\mathrm{RT}}(\omega_j^{\mathrm{II}})}{t_{\mathrm{RT}}(\omega_j^{\mathrm{Ib}})} \tag{17}$$

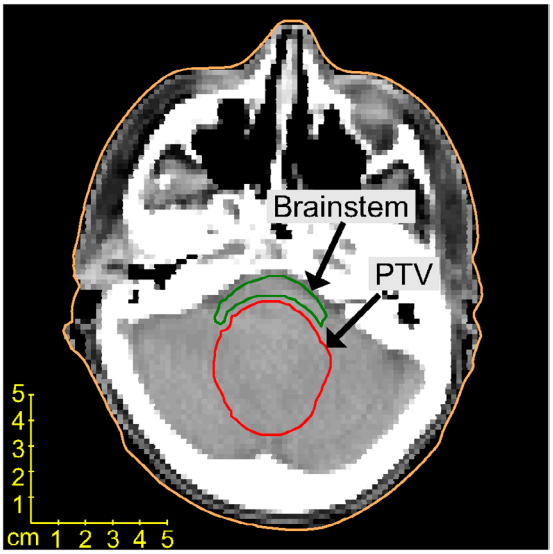

Figure 2. Astrocytoma patient geometry.

2.4. Patient case and treatment planning parameters

The algorithm was implemented in a research planning system for 3D spot scanning protons (Lomax 1999, Nill *et al* 2004) in the computational environment of CERR (MATLAB) (Deasy *et al* 2003, Schell and Wilkens 2010). All runs were optimized with the commercial toolbox Mosek (www.mosek.com).

PrEfOpt was applied to the case of an astrocytoma patient (figure 2) previously treated with photons at our institute. We selected two opposed fields of 90° and 270° which were simultaneously optimized. The spot spacing within the PTV (plus a margin of 0.5 cm) was $\Delta x = \Delta y = 0.4$ cm and $\Delta z = 1.2 \cdot w_{80}$ (width w at the 80% intensity level of the Bragg peak). Available energies ranged from 50 MeV to 250 MeV in steps of 1 MeV. The beam featured a constant lateral width of $\mathrm{FWHM} = 0.5 \cdot \Delta x$ at the patient surface (FWHM, full width half maximum of a Gaussian profile).

The prescribed dose was 60 Gy (30×2 Gy) to the PTV. Dosimetric quality was analyzed by various dose–volume histogram (DVH) indicators of the PTV and the brainstem, referring to the brainstem excluding the PTV plus 2 mm (figure 2). Minimum and maximum doses were evaluated as $D_{\min}(1\ \mathrm{cm}^3)$ and $D_{\max}(1\ \mathrm{cm}^3)$ referring to the dose received by 1 cm^3 of the corresponding volume. Due to the formulation of the algorithm, the maximum dose to OARs was not increased in step II. In order to compare the potential of each method we determined the plan of the shortest treatment time of each facility type which fulfilled the dosimetric criteria of the initial plan up to a maximum change of $\Delta D = \pm 1\%$ and $\pm 2\%$ of $D_{\min}(1\ \mathrm{cm}^3)$(PTV) and $D_{\max}(1\ \mathrm{cm}^3)$(PTV).

Case-dependent trade-off parameters allowed us to control and steer the compromise between plan quality (in terms of PTV coverage) and treatment time for each optimization method. The utilized parameter values were: $s \in [0.1, ..., 4.8]$, $f_S \in [0.01, ..., 2]$ and $l_E \in [1, .., 12]$. The relative time savings Δt_{RT} were calculated for all methods and facility combinations.

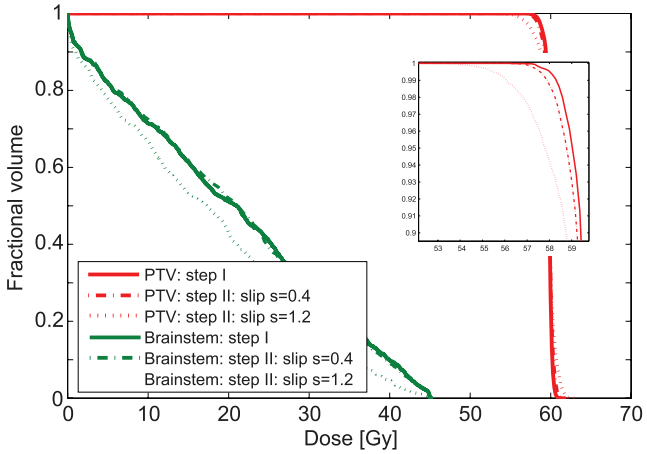

Figure 3. Dose–volume histogram of the initial plan after step I (solid line) and two plans after efficiency optimization (step II) the plan via 'minimization of the total spot weight sum' (method A) with slip $s = 0.4$ (dashed line) and $s = 1.2$ (dotted line).

3. Results

3.1. Analysis of treatment times and trade-offs for the astrocytoma patient

3.1.1. Illustration of the slip-factor concept. Figure 3 depicts the slip-factor concept for the minimization of total spot weight sum with $s = 0.4$ and $s = 1.2$. The larger slip factor permitted greater deviations from the initially achieved PTV coverage (step I), which enabled shorter treatment times. The calculated RT times for constant current were reduced from 75.7 s to 56.4 s for $s = 0.4$ and to 48.2 s for $s = 1.2$. Compromises were found in slightly decreased minimum doses: $D_{\mathrm{min}}(1\ \mathrm{cm}^3)(\mathrm{initial}) = 58.5$ Gy, $D_{\mathrm{min}}(1\ \mathrm{cm}^3)(s = 0.4) = 58.1$ Gy and $D_{\mathrm{min}}(1\ \mathrm{cm}^3)(s = 1.2) = 57.1$ Gy. The 100% coverage of the PTV with the 95% isodose was reduced to 99.9% for $s = 0.4$ and 98.7% for $s = 1.2$. Due to the defined hard constraints on D_{max} for all runs and on the achieved D_{mean} of step Ib (see equations (4) and (10)), brainstem doses of both efficiency-optimized plans were equally as good or lower than in the initial plan.

3.1.2. Trade-off analysis for different optimization method–facility combinations. We calculated plans of different trade-offs for each facility–method combination by varying the corresponding parameters s, f_S and l_E (for the utilized values see section 2.4 and for results see table 1). The correlations between treatment times and plan quality are visualized in trade-off plots (figures 4 and 5). Each plan is represented by three evaluated DVH criteria.

3.1.3. Facility with a constant proton current. Minimizing the total spot weight sum (figure 4, left) resulted in remarkable time savings, indicated by a gap in treatment times at similar doses between step I and the first plan of step II ($s = 0.4$) (see also figure 3). The least time was saved via 'maximization of the lowest spot weight' (designed for facility type 2) and spot elimination (C) without re-optimization. The efficacy of the latter was improved by re-optimizing the problem after spot elimination (figure 4, center) which enabled the elimination of a larger number of spots. Dose distributions derived by eliminating energy layers were qualitatively not comparable to the initial plan and required subsequent re-optimization of plan quality

Table 1. Saved time for deviation criteria of $\Delta D = \pm 1\%$ and $\pm 2\%$ to the DVH points of the initial plan (step I): $D_{\min}(1\ \mathrm{cm}^3)$ and $D_{\max}(1\ \mathrm{cm}^3)$ of the PTV are within 99%, 98% and 101%, 102% of the initially obtained DVH points, respectively.

				Spot elimination (C)		Energy layer elimination (D)	
Evaluation DVH criteria (%)	Proton current	Minimization of overall spot weight sum (A) (%)	Maximization of lowest spot weight per energy (B) (%)	w/o reopt. (%)	with reopt. (%)	w/o reopt. (%)	with reopt. (%)
$\Delta D = \pm 1$	Constant	28	1	2	14	0	32
	Variable	27	15	16	21	0	26
$\Delta D = \pm 2$	Constant	34	1	3	17	3	41
	Variable	32	15	20	30	2	31

Note: Efficiency determining parameters of the derived plans, such as the number of spots and energy layers, are provided in table A1 in the appendix. w/o reopt., without reoptimization.

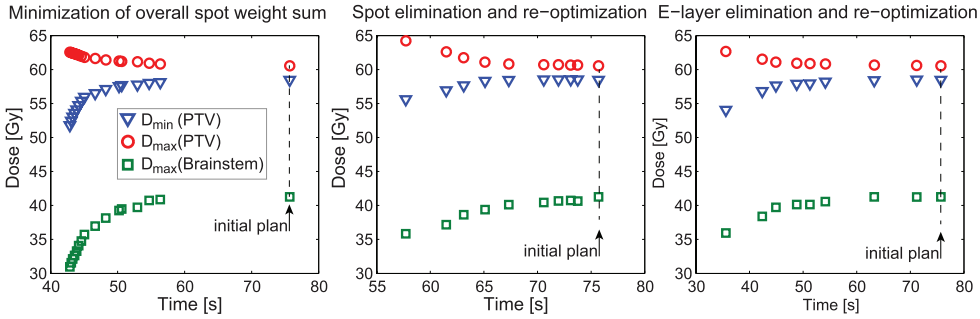

Figure 4. Trade-off between plan quality and delivery time for a constant treatment current $I_c = 0.5$ nA. Each subplot illustrates the results optimized using the optional method given in the label above each subfigure. Plans are represented by three DVH points. The plan results of step I are marked as 'initial plans'.

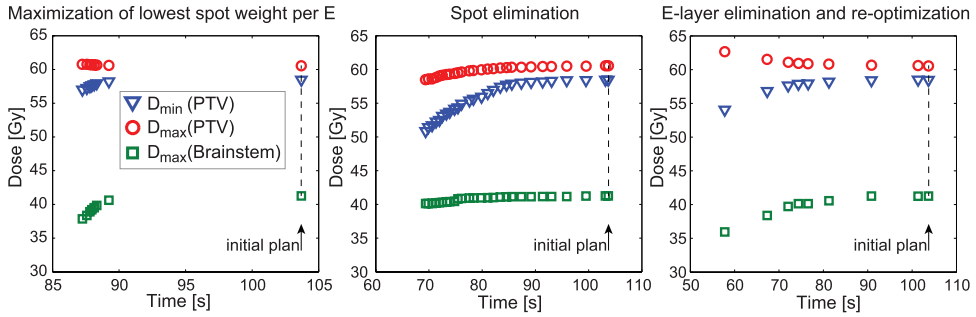

Figure 5. Trade-off between plan quality and delivery time for a variable treatment current I_v. Each subplot illustrates the results optimized by the optional methods given in the text above each subfigure. Plans are represented by three DVH points. The plan results of step I are marked as 'initial plans'.

(figure 4, right). The hereby derived time reductions achieved the shortest RT times compared with all other methods. The presented trade-off plots showed a trend of diverging $D_{\min}(1\ \mathrm{cm}^3)$ and $D_{\max}(1\ \mathrm{cm}^3)$ of the PTV and decreasing $D_{\max}(1\ \mathrm{cm}^3)$ of the brainstem (implemented constraints in equations (4) and (10)) with shorter treatment times.

3.1.4. Facility with a variable proton current. By maximizing the lowest spot weight of each energy layer (figure 5, left) times were reduced by, for example, $\Delta t_{\mathrm{RT}}(I_{\mathrm{v}}) = 14.8\%$ at a constant coverage of 100% and a decrease in $\Delta D_{\min}(1\ \mathrm{cm}^3) < 0.5\%$. Similar results were achieved by deleting low-weighted spots (C) (figure 5, center). A larger number of spots was eliminated with consecutive re-optimization, which led to shorter RT times than without re-optimization. Greater time reductions were obtained by method D with re-optimization (figure 5, right) and by method A. The latter, and the re-optimization of method C, did not continuously reduce RT times with the number of eliminated spots/reduced spot weight sum but indicated that fewer spots or energy layers, as well as smaller sums of spot intensities, do not necessarily result in shorter treatment times.

3.2. Varying energy switch times

In order to investigate the impact of energy switch times, delivery times were additionally calculated with $\Delta t_E = 5$ s (see table A3, appendix). Different relative time reductions were derived with a varying impact dependent on the applied optimization method. Compared with the results for $\Delta t_E = 1$ s (see tables 1 and A3 in the appendix) the relative decrease of delivery times was larger for the 'elimination of low-weighted energy layers': delivery times were reduced by $\Delta t_{\mathrm{RT}}(I_{\mathrm{c}}) = 37\%$ and $\Delta t_{\mathrm{RT}}(I_{\mathrm{c}}) = 47\%$ for dose changes of $\Delta D < 1\%$ and $\Delta D < 2\%$, respectively. Contrary results were found for all other presented methods (A–C).

3.3. Varying geometry: prostate case

PrEfOpt was further applied to the case of a patient with prostate cancer, with the main differences being a larger target volume and higher required energies. Beam angles were fixed at 90° and 270°. For the facility with a constant current, the methods showed similar trends, as suggested by the astrocytoma patient (see table A3 in the appendix): the shortest RT times were obtained by eliminating low-weighted energy layers (D), for example $\Delta t_{\mathrm{RT}}(I_{\mathrm{c}}) = 24\%$ (figure 6, left), for dose changes of $\Delta D_{\min/\max}(1\ \mathrm{cm}^3)(\mathrm{PTV}) < 1\%$ (for plan details see table A2 in the appendix). RT times achieved by minimizing the overall spot weight sum (A) did not strictly monotonically decrease with quality degradation, as observed for the astrocytoma patient (figure 4, left), caused by variations in the number of energy layers. For the facility with a variable current, method A only achieved time reductions for compromises larger than 1% of the minimum PTV dose. The shortest RT times were obtained by the elimination of low-weighted spots without re-optimization (C) with $\Delta t_{\mathrm{RT}}(I_{\mathrm{v}}) = 37\%$ (figure 6, right) for $\Delta D_{\min/\max}(1\ \mathrm{cm}^3)(\mathrm{PTV}) < 1\%$. A larger number of spots was eliminated with consecutive re-optimization but this led to smaller time savings than without re-optimization (for plan details see table A2 in the appendix).

4. Discussion

We presented the treatment planning algorithm PrEfOpt, a two-step routine, to optimize IMPT plans dosimetrically and to reduce treatment times while controlling plan quality. PrEfOpt

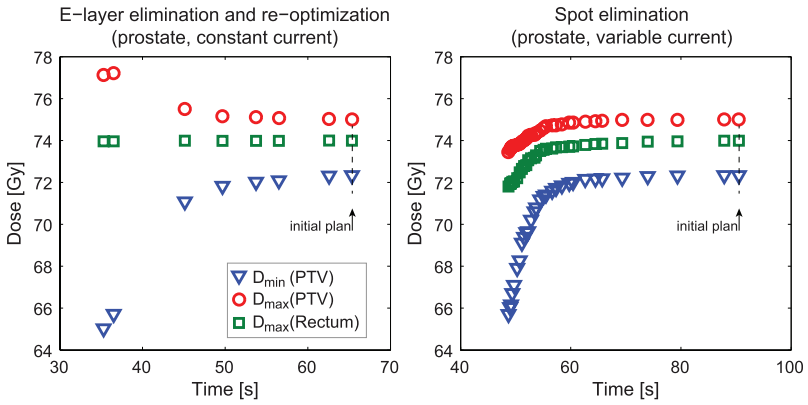

Figure 6. Trade-off between quality and treatment times for a prostate cancer case. Presented are the method–facility combinations which achieved the largest time reductions: 'energy layer elimination with re-optimization' for a constant current (left) and 'spot elimination' of low-weighted spots without re-optimization for a variable current (right).

offers different methods for increasing delivery efficiency which were applied to two patients, one with an astrocytoma and one with prostate cancer. For both constant and variable current, reductions in irradiation time were achievable without major compromises in plan quality. A generally valid identification of 'the most effective' optimization method is not possible. Delivery times strongly vary with the implemented currents and energy switch times (see section 2.3). As the initial RT times of facility 1 and facility 2 differed, absolute times are not comparable. The facility simulations served to analyze the potential of the methods for each facility type separately.

To compare the efficacy of the methods we selected the plans with the shortest application times which still fulfilled certain DVH criteria (see table 1). As the number of calculated plans was limited and varied between methods, the proximity of the chosen plan to the exclusion criteria, and thus corresponding qualities, differed slightly. We determined the shortest feasible RT time and the 'first plan' violating the criteria. The 'true' feasible irradiation time lies somewhere in between.

The derived RT times indicated methods A and D to be clearly superior for a constant current for both patients. While these methods also gave the largest time savings for the facility with a variable current for the astrocytoma case, contrary results were found for the prostate cancer case. The minimization of the overall spot weight sum, as well as re-optimizations of quality after spot or energy layer eliminations (methods C and D) do not necessarily lead to shorter RT times for variable currents: (re-)appearing low spot intensities which consequently reduce the current may result in even longer delivery times. Method B was designed to specifically increase the variable current, which, similar to the pure elimination of low-weighted spots, considerably decreased corresponding RT times. The latter is a simple post-processing step that cannot be considered as an actual optimization method. As it allows one to reduce RT times for I_v with comparably little effort it is considered to be a valuable alternative to more complex efficiency optimizations.

With respect to methods A and B, the formulations decreased the sum of spot weights and increased the variable currents, respectively, as intended in all cases. A simultaneous influence on distinctive efficiency-determining parameters, such as the number of spots and energy layers or the overall weight sum in case B, is, however, not 'prohibited' and may work

against the intended shorter RT times. The introduction of additional constraints could solve this situation. It should be noted though that a large number of constraints could influence the feasibility of the optimization problem.

Besides the dependence on treatment currents, the selection of the most suitable time reduction method depends on the energy switch times. Obviously the larger the energy switch time, the more impact is given to the 'energy layer elimination' method. The great potential for reducing energy layers to give a decrease in treatment times was published recently (Cao *et al* 2014, van de Water *et al* 2015). van de Water *et al* (2015) presented an iterative optimization routine to reduce energy layers (and spots) in robust IMPT treatment planning in a multicriterion optimization system. Based on prioritized optimization the algorithm assured the plan quality by introducing dose constraints. The published time savings for energy switches of 1 s and 5 s are consistent with our results (method D with re-optimization; see table A3 in the appendix). Both studies concluded that due to the degenaracy of IMPT plans, times can be reduced without affecting dosimetric plan quality. Our findings underline this conclusion and complement the research by discussing different facility types for which different efficiency-optimization methods may have greater potential.

We assessed the plan quality by evaluating DVH points of the dose distributions. In IMPT range uncertainties are of great concern; these are reduced by robust planning or incorporated by robust optimization (Unkelbach *et al* 2007, 2009, Pflugfelder *et al* 2008). We did not perform any robustness analysis here. Each method may have a different impact on how spot weights are (re-)distributed. More pronounced weighting of high proton energies (e.g. by method A) could occur. Since the biological effect of proton radiation increases at the distal edge of the Bragg Peak, these changes in spot patterns may cause undesired, crucial dosimetric consequences in patient treatment, and thus require specific considerations; similarly some methods may derive more homogeneous spot patterns over all energies, which could be beneficial in terms of robustness. To see that weights were not shifted between both beams, such that the majority of the dose is only delivered by one of the beams after efficiency optimization, we evaluated the ratio between the total spot weight sum of both beams; it remained mostly constant with slight variations between cases and methods (see tables A1 and A2 in the appendix). In clinical settings the maintenance of plan robustness has to be assured similar to dosimetric quality.

Further limitation of the treatment planning procedure might be found in the objective functions of step I. The minimization of the mean dose (step Ib) may not lead to the most optimal plan quality for every case. Step I is merely an example, and can obviously be replaced by different objective functions.

In step II the slip factor was exclusively applied to the results of step Ia, i.e. the PTV coverage. Thus doses to OARs were kept at a constant level or even improved. The introduction of a slip factor on OARs may lead to greater time savings in some cases, and could be considered as an option to modify the PrEfOpt routine.

Provided that a plan has been efficiency optimized (by any method), a method to decide on the best trade-off is required. What degree of dosimetric degradation is still acceptable, and is it acceptable to influence the plan quality at all with the intention of just saving time? These are difficult questions, partly involving ethical issues, which should be considered in this context.

The question of whether anyone should (have to) make a crucial decision on the trade-off between efficiency and quality was discussed earlier by Bortfeld and Webb (2009). We agree completely with their consideration that it would mostly be worth to wait a minute longer for a better plan (Bortfeld and Webb 2009). There may, however, be patients for whom the benefit of shorter treatment times is larger than the impact of minor compromises in the calculated

dose distributions. Further, our work and prior publications indicated that large time savings do not necessarily mean a reduction in quality (Coselmon *et al* 2005, Craft *et al* 2007, Kang *et al* 2008, Mittauer *et al* 2013, Cao *et al* 2014, van de Water *et al* 2015).

One method for selecting the treatment plan with the best trade-off is by execution of an automatic script which selects the plan with the shortest time that still fulfills specific DVH criteria. This method will always pick the 'worst' plan within the defined interval of acceptable DVH values, and may not always result in the same choice being made as would be selected by a person. Method A for a constant current (figure 4, left) presents a large decrease in RT time between the initial and the efficiency-optimized plan with the best quality, and comparably small time improvements were only achieved with further plan degradation. The 'first plan' would most likely be the 'trade-off choice' made by any person.

Moreover, automated plan selection exclusively takes DVH points into account rather than comparing the actual dose distributions. The best trade-off choice may vary with the derived dose distributions and the patient's background, i.e. the intention of the treatment and the physical condition of the patient.

The creation of a plan database similar to MCO treatment planning could be a useful application for realizing efficiency improving algorithms in clinical practice. PrEfOpt generates plans with different trade-offs which could serve to fill the database. By providing a slider on the planning interface which represents the treatment time along with sliders to control dosimetric objectives, the trade-off between time and quality could be steered interactively by the user. MCO would give control of the the trade-off to the planner or/and responsible physician rather than being hard-coded in a planning system.

5. Conclusion

Our work has demonstrated the potential of efficiency optimization in IMPT planning via different optimization methods. By prioritizing objectives, PrEfOPt achieved remarkable time reductions while maintaining the previously obtained dosimetric quality. Inevitable trade-offs between quality and delivery time were detected, indicating a limit to the achievement of reasonable time reductions. Even though the application of any type of efficiency optimization does not generally guarantee to decrease RT times, efficiency optimization is still considered to be useful in clinical practice. If a plan cannot be improved with regards to RT time with a similar plan quality the initial plan may still be selected for treatment. Generally the implementation of a routine after plan quality optimization to increase delivery efficiency could be of great potential for clinical treatment planning.

In order to decide on the best trade-off plan, PrEfOpt could serve to calculate plans with different trade-offs to fill a MCO database. The user could then navigate to the preferred trade-off between plan quality and delivery time.

Acknowledgments

Supported by DFG Cluster of Excellence: Munich-Centre for Advanced Photonics.

Appendix

Table A1. Efficiency-determining quantities of plans derived by the different PrEfOpt methods compared with the initial plan properties for the case of an patient with astrocytoma.

						Method C				Method D		
	Initial plan	Method A		Method B		w/o re-opt.		with re-opt.		w/o re-opt.	with re-opt.	
Evaluation criteria ΔD		±1%	±2%	±1%	±2%	±1%	±2%	±1%	±2%	±2%	±1%	±2%
Trade-off parameter s, f_s, l_E	/	0.5	0.9	0.5	0.8	0.2	0.3	0.8	1.0	1	9	11
RT time 1 (I_c)	75.7	54.7	50.2	75.0	74.8	74.1	73.4	65.1	63.1	73.4	51.3	44.9
RT time 2 (I_v)	103.7	76.0	71.0	88.3	87.8	87.4	83.4	76.9	72.7	101.3	76.6	72.0
Number of E-layers[a]	47	35	32	47	47	47	47	46	46	45	29	24
Number of spots	1914	1200	1076	1914	1914	1759	1694	1018	834	1894	1361	1238
Sum over $\underline{w}$	3.0×10^4	2.4×10^4	2.3×10^4	2.8×10^4	2.7×10^4	3.0×10^4	2.9×10^4	2.8×10^4	2.7×10^4	3.0×10^4	2.7×10^4	2.7×10^4
Average $\breve{\omega}_{m,k}$	2.7	1.7	2.4	3.4	3.4	3.6	3.9	5.8	6.3	2.6	2.1	1.8
Average variance $(\omega_{m,k})$	145.5	205.5	222.7	130.2	133.2	145.2	145.8	259.8	335.2	143.7	241.3	234.7
Ratio of the sum over all spot weights of beam 1 to the sum over all spot weights of beam 2	0.9	1.0	1.0	0.9	0.9	0.9	0.9	1.0	0.9	0.9	1.4	1.4

Note: The table presents the plan results of the shortest RT times for both facility types (constant current (RT time 1) and variable current (RT time 2)) with maximal dose changes of 1% and 2% in $D_{\min}(1\ \mathrm{cm}^3)$ and $D_{\max}(1\ \mathrm{cm}^3)$ of the PTV. The prescribed dose to the PTV was 60 Gy (30 x 2 Gy) with the initial plan values of $D_{\min}(1\ \mathrm{cm}^3) = 58.5$ Gy and $D_{\max}(1\ \mathrm{cm}^3) = 60.6$ Gy and 100% coverage of the PTV with 95% of the prescribed dose.

[a] The number of energy layers exclusively includes layers which contain at least one non-zero spot weight.

Table A2. Efficiency-determining quantities of plans derived by the different PrEfOpt methods compared with the initial plan properties in the case of a patient with prostate cancer.

						Method C				Method D	
	Initial plan	Method A		Method B		w/o re-opt.		with re-opt.		with re-opt.	
Evaluation criteria ΔD		±1%	±2%	±1%	±2%	±1%	±2%	±1%	±2%	±1%	±2%
Trade-off parameter s, f_s, l_E	/	0.2	0.5	0.8	1.2	0.2	0.3	0.9	1.0	5	7
RT time 1 (I_c)	65.3	55.1	50.5	62.7	62.2	63.9	63.5	59.0	58.1	49.7	45.1
RT time 2 (I_v)	90.5	107.0	65.7	78.0	77.8	57.3	54.5	58.7	52.3	74.7	71.2
Number of E-layers[a]	36	31	29	35	35	36	36	36	36	25	22
Number of spots	1017	813	590	855	805	893	858	503	454	750	663
Sum over $\underline{\omega}$	6.0×10^4	5.0×10^4	4.9×10^4	6.0×10^4	6.0×10^4	5.9×10^4	5.9×10^4	5.6×10^4	5.5×10^4	5.4×10^4	5.1×10^4
Average $\breve{\omega}_{m,k}$	16.7	2.5	9.4	29.3	31.6	22.8	25.2	28.0	34.1	11.9	12.4
Average variance $(\omega_{m,k})$	4518.8	4999.5	6208.3	6670.2	7354.7	4551.0	4537.5	6336.4	5942.7	5971.9	6312.5
Ratio of the sum over all spot weights of beam 1 to the sum over all spot weights of beam 2	1.1	1.3	1.2	1.3	1.4	1.1	1.1	1.1	1.1	1.1	1.3

[a] The number of energy layers exclusively includes layers which contain at least one non-zero spot weight.

Note: The table presents the plan results of the shortest RT times for both facility types (constant current (RT time 1) and variable current (RT time 2)) with maximal dose changes of 1% and 2% in $D_{min}(1\ cm^3)$ and $D_{max}(1\ cm^3)$ of the PTV. The prescribed dose to the PTV was 74 Gy (37×2 Gy) with the initial plan values of $D_{min}(1\ cm^3) = 72.3$ Gy and $D_{max}(1\ cm^3) = 75.0$ Gy and 100% coverage of the PTV with 95% of the prescribed dose.

Table A3. Comparison of time savings (%) for energy switch times of $\Delta t_E = 1s$ and $\Delta t_E = 5s$ for the astrocytoma (ASTRO) and prostate (PRO) patients of selective optimization methods.

	Eval. criteria ΔD	Proton current	Method A		Method B		Method C w/o. re-opt.		Method C with re-opt.		Method D w/o. re-opt.		Method D with re-opt.	
Δt_E			1 s (%)	5 s (%)	1 s (%)	5 s (%)	1 s (%)	5 s (%)	1 s (%)	5 s (%)	1 s (%)	5 s (%)	1 s (%)	5 s (%)
ASTRO	±1	Const.	28	26	1	0	2	1	14	6	0	0	32	37
		Var.	27	26	15	5	16	6	21	9	0	0	26	27
	±2%	Const.	34	32	1	0	3	1	17	6	3	4	41	47
		Var.	32	32	15	5	20	7	30	12	2	4	31	36
PRO	±1%	Const.	16	14	4	3	2	1	11	3	0	0	24	29
		Var.	0	1	14	7	37	14	35	14	0	0	17	26
	±2%	Const.	23	20	5	3	3	1	11	3	0	0	31	36
		Var.	27	23	14	7	40	15	42	16	0	0	21	32

References

Bewes J M, Suchowerska N, Jackson M, Zhang M and McKenzie D R 2008 The radiobiological effect of intra-fraction dose-rate modulation in intensity modulated radiation therapy (IMRT) *Phys. Med. Biol.* **53** 3567–78

Bortfeld T and Webb S 2009 Single-Arc IMRT? *Phys. Med. Biol.* **54** N9–20

Cao W, Lim G, Liao L, Li Y, Jiang S, Li X, Li H, Suzuki K, Zhu X, Gomez D and Zhang X 2014 Proton energy optimization and reduction for intensity-modulated proton therapy *Phys. Med. Biol.* **59** 6341–54

Cao W, Lim G, Li X, Li Y, Zhu X R and Zhang X 2013 Incorporating deliverable monitor unit constraints into spot intensity optimization in intensity-modulated proton therapy treatment planning *Phys. Med. Biol.* **58** 5113–25

Coselmon M M, Moran J M, Radawski J D and Fraass B A 2005 Improving IMRT delivery efficiency using intensity limits during inverse planning *Med. Phys.* **32** 1234–45

Craft D, Halabi T and Bortfeld T 2005 Exploration of tradeoffs in intensity-modulated radiotherapy *Phys. Med. Biol.* **50** 5857–68

Craft D, Süss P and Bortfeld T 2007 The tradeoff between treatment plan quality and required number of monitor units in intensity-modulated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **67** 1596–605

Deasy J O, Blanco A I and Clark V H 2003 CERR: a computational environment for radiotherapy research *Med. Phys.* **30** 979–85

Hillbrand M and Georg D 2010 Assessing a set of optimal user interface parameters for intensity-modulated proton therapy planning *J. Appl. Clin. Med. Phys.* **11** 3219

Howard M, Beltran C, Mayo C S and Herman M G 2014 Effects of minimum monitor unit threshold on spot scanning proton plan quality *Med. Phys.* **41** 091703

Kang J H, Wilkens J J and Oelfke U 2008 Non-uniform depth scanning for proton therapy systems employing active energy variation *Phys. Med. Biol.* **53** N149–55

Lomax A J 1999 Intensity modulation methods for proton radiotherapy *Phys. Med. Biol.* **44** 185–205

Mittauer K, Lu B, Yan G, Kahler D, Gopal A, Amdur R and Liu C 2013 A study of IMRT planning parameters on planning efficiency, delivery efficiency, and plan quality *Med. Phys.* **40** 061704

Nill S, Bortfeld T and Oelfke U 2004 Inverse planning of intensity modulated proton therapy *Z. Med. Phys.* **14** 35–40

Otto K 2008 Volumetric modulated arc therapy: IMRT in a single arc? *Med. Phys.* **35** 310–17

Paganetti H 2005 Changes in tumor cell response due to prolonged dose delivery times in fractionated radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **63** 892–900

Pflugfelder D, Wilkens J J and Oelfke U 2008 Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy *Phys. Med. Biol.* **53** 1689–700

Schell S and Wilkens J J 2010 Advanced treatment planning methods for efficient radiation therapy with laser accelerated proton and ion beams *Med. Phys.* **37** 5330–40

Suzuki K, Gillin M T, Sahoo N, Zhu X R, Lee A K and Lippy D 2011 Quantitative analysis of beam delivery parameters and treatment process time for proton beam therapy *Med. Phys.* **38** 4329–37

Unkelbach J, Bortfeld T, Martin B C and Soukup M 2009 Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning *Med. Phys.* **36** 149–63

Unkelbach J, Chan T C Y and Bortfeld T 2007 Accounting for range uncertainties in the optimization of intensity modulated proton therapy *Phys. Med. Biol.* **52** 2755–73

van de Water S, Kooy H M, Heijmen B and Hoogeman M 2015 Shortening delivery times of intensity modulated proton therapy by reducing proton energy layers during treatment plan optimization *Int. J. Radiat. Oncol. Biol. Phys.* **92** 460–8

Wilkens J J, Alaly J R, Zakarian K, Thorstad W L and Deasy J O 2007 IMRT treatment planning based on prioritizing prescription goals *Phys. Med. Biol.* **52** 1675–92

Wilkie J R, Matuszak M M, Feng M, Moran J M and Fraass B A 2013 Use of plan quality degradation to evaluate tradeoffs in delivery efficiency and clinical plan metrics arising from IMRT optimizer and sequencer compromises *Med. Phys.* **40** 071708

Zhu X R, Sahoo N, Zhang X, Robertson D, Li H, Choi S, Lee A K and Gillin M T 2010 Intensity modulated proton therapy treatment planning using single-field optimization: the impact of monitor unit constraints on plan quality *Med. Phys.* **37** 1210–19