Automated prioritised 3D dose-based MLC segment generation for step-and-shoot IMRT

Schipaanboord, B.W.K.; Breedveld, S.; Rossi, L.; Keijzer, Marleen; Heijmen, Ben

DOI
10.1088/1361-6560/ab1df9

Publication date
2019

Document Version
Accepted author manuscript

Published in
Physics in Medicine and Biology

Citation (APA)
Schipaanboord, B. W. K., Breedveld, S., Rossi, L., Keijzer, M., & Heijmen, B. (2019). Automated prioritised 3D dose-based MLC segment generation for step-and-shoot IMRT. *Physics in Medicine and Biology*, *64*(16), 1-14. Article 165013. https://doi.org/10.1088/1361-6560/ab1df9

Automated prioritised 3D dose-based MLC segment generation for step-and-shoot IMRT

B.W.K. Schipaanboord[1], S. Breedveld[1], L. Rossi[1], M. Keijzer[2], B. Heijmen[1]

[1] Erasmus MC - Cancer Institute, University Medical Center Rotterdam, Department of Radiotherapy, the Netherlands

[2] Delft Institute of Applied Mathematics, Delft University of Technology, the Netherlands

E-mail: b.schipaanboord@erasmusmc.nl

Abstract. Segmentation can degrade a high-quality dose distribution obtained by fluence map optimisation (FMO). A novel algorithm is proposed for generation of MLC segments to deliver an FMO plan with step-and-shoot IMRT while minimising quality loss. All beams are considered simultaneously while generating MLC segments for reproducing the 3-dimensional FMO dose distribution. Segment generation is only steered by the 3D FMO dose distribution, i.e. underlying FMO fluence profiles are not considered. The algorithm features prioritised generation of segments, focusing on accurate reproduction of clinical objectives with the highest priorities. The performance of the segmentation algorithm was evaluated for 20 prostate patients, 15 head-and-neck patients, and 12 liver patients. FMO dose distributions were generated by automated multi-criteria treatment planning (Pareto-optimal plans) and subsequently segmented using the proposed method. Various segmentation strategies were investigated regarding prioritisation of objectives and limitation of the number of segments. Segmented plans were dosimetrically similar to FMO plans and for all patients a clinically acceptable segmented plan could be generated. Substantial differences between FMO and segmented fluence profiles were observed. Avoidance of the usual reconstruction of 2D FMO fluence profiles for segment generation, and instead simultaneously generating segments for all beams to directly reproduce the 3D FMO dose distribution is a likely explanation for the obtained results. For the strategies of limiting the number of segments large reductions in number of segments were observed with minimal impact on plan quality.

Keywords: step-and-shoot IMRT, treatment plan optimisation, inverse IMRT planning, prioritised MLC segmentation, column generation

1. Introduction

In intensity modulated radiation therapy (IMRT) treatment planning, the optimisation problem may be split into a fluence map optimisation (FMO) phase and a segmentation phase to convert the optimised fluences into multi-leaf collimator (MLC) segments. An advantage of this approach is that the FMO problem can be modelled as a convex

multi-criterial optimisation problem (Breedveld et al., 2019) with a guaranteed globally optimal solution in optimisation. As treatment machine limitations (e.g. limitations of the MLC) are not fully accounted for in FMO, the deliverable plan resulting from the segmentation phase may have a quality loss compared to the FMO plan. Additionally, decisions on multi-criterial trade-offs, as made during FMO, are generally not explicitly taken into account in the segmentation phase (Salari and Unkelbach, 2013).

One approach for segmentation in static step-and-shoot IMRT is to separately segment the beam fluence profiles for the involved beams into sets of deliverable segments by stratifying fluences into discrete intensity levels and subsequently generate feasible segments for each beam that match the optimised fluence profile (Xia and Verhey, 1998, Süss et al., 2007). In general, the more segments are included the better the fluence profile can be replicated. To restrict the treatment delivery time, pre-defined trade-offs between plan quality and treatment time can be used to restrict the number of intensity levels and number of segments (Craft et al., 2007). To the best of our knowledge, published MLC segmentation approaches for static step-and-shoot IMRT plans are all based on independent segmentation of the 2-dimensional fluence profiles of all beams (Long et al., 2016, Gören and Taşkin, 2015, Luan et al., 2006, Süss et al., 2007, Sun and Xia, 2004). Sequencing the fluences for each beam separately excludes mutual dosimetric compensation of imperfect segmentations of the 2-dimensional beam fluence profiles to optimally reproduce the initial 3-dimensional FMO dose distribution.

Extensive research has been done to improve MLC segmentation with non-discretised intensity levels and leaf positions (Long et al., 2016), to investigate segmentation efficiency under various MLC constraints (Gören and Taşkin, 2015), to explore regularization in the dose domain before segmentation (Nguyen et al., 2015) and to minimise beam-on-time (Crooks et al. 2002, Boland et al. 2004, Ahuja and Hamacher 2005). However, none of the published methods explicitly account for differences in objective priorities during segmentation. Consequently, discrepancies between FMO fluence and sequenced fluence may potentially lead to dose deviations in the PTV and OARs with uncontrolled balances, i.e. without explicitly considering the clinical priorities.

In contrast to FMO followed by segmentation, Direct Aperture Optimisation (DAO) has been proposed to directly generate MLC segments (Shepard et al., 2002, Romeijn et al., 2005, Men et al., 2007). DAO operates under the "What you see is what you get" principle, meaning that at every stage of the optimisation process the treatment plan is directly feasible for delivery and no segmentation phase (with possible loss in plan quality) is needed. However, including the non-convex modelling of the (physical) constraints of the collimator and treatment device leads to a non-convex optimisation problem. The column generation (CG) approach has been proposed as heuristic in the field of DAO to solve the optimisation problem in radiotherapy (Men et al., 2007, Carlsson, 2008, Cassioli and Unkelbach, 2013). Research on CG approaches for DAO includes investigations on convergence (Carlsson and Forsgren, 2014), generation of segments under various MLC constraints (Men et al., 2007) and inclusion of pre-defined

multi-criterial trade-offs (Salari and Unkelbach, 2013).

In our centre we have given preference to plan optimisation using FMO followed by segmentation, because of the guaranteed Pareto and global optimality of the FMO plans that we generate with *Erasmus-iCycle*, an algorithm for automated a priori Multi-Criterial treatment plan Optimisation (MCO) (Breedveld et al., 2012). For each patient, a single Pareto-optimal FMO plan is generated with clinically favourable trade-offs, considering all treatment objectives with explicitly assigned priorities. This differs from a posteriori MCO (e.g. Craft and Richter 2013, Bokrantz and Miettinen 2015) in which, for each patient, a set of Pareto-optimal plans is generated with automated planning, while selection of a clinically favourable plan is performed by a user.

In this study, a prioritised dose-based MLC segment generation method is proposed which minimises 3-dimensional plan quality loss compared to the FMO plan by placing extra consideration on high priority clinical objectives. To reconstruct the 3-dimensional FMO dose, a CG approach was implemented that simultaneously optimises the beam segments for all treatment beams, rather than replicating the 2-dimensional fluences separately as is done in other published MLC segmentation methods. Segment generation was only steered by the 3D FMO dose distribution, i.e. underlying FMO fluence profiles were not considered. In this sense, the term segmentation as applied in this paper has a slightly different meaning from that in the literature. The CG approach was chosen because of its intuitive mechanism of generating segments and proven effectiveness in IMRT treatment planning (Carlsson, 2008, Romeijn et al., 2005, Salari and Unkelbach, 2013).

The overall goal was to achieve high quality radiotherapy treatment plans by combining global optimality of the FMO plan with a fast prioritised "DAO-like" segmentation. The segmentation was tailored to individual patients by using the prioritised approach in combination with personalised clinical objectives obtained from the FMO plan. The proposed technique was developed in the context of CyberKnife robotic radiotherapy, but can be applied for segmentation of any static step-and-shoot IMRT or stereotactic body radiation therapy (SBRT) plan. Possibilities for keeping calculation and delivery times low were included in the investigations. The segmentation performance was evaluated for prostate, head-and-neck and liver tumours.

2. Materials & Methods

This section starts with briefly describing the applied FMO (section 2.1). Next, CG is introduced in section 2.2, while the proposed segmentation with CG is described in section 2.3. Prioritised steering on personalised objectives is described in section 2.4, which includes approaches to minimise the number of segments. Finally, plan evaluation criteria and details on our computational study are presented in sections 2.5 and 2.6, respectively.

Table 1. Overview of the clinical cases and FMO plans.

Treatment site	Cases	Treatment unit	Beams		Prescribed dose (Gy)	Fractions
Prostate	20	CyberKnife	25	non-coplanar	38	4
Head-and-neck	15	Conventional linac	9	coplanar	46	23
Liver	12	CyberKnife	25	non-coplanar	60	3

2.1. Patients, FMO treatment plans and dose calculation model

An overview of the clinical cases and FMO plans used to evaluate the performance of the MLC segmentation is presented in table 1. All FMO plans were generated with fully automated multi-criterial optimisation as implemented in Erasmus-iCycle (Breedveld et al., 2012, 2017). A pencil-beam approach was used to describe dose delivered to the patient, i.e. $d = Ax$ with d the vector containing the patient's voxel doses, A the dose deposition matrix, and x the pencil beam weights (see also Breedveld et al., 2006, 2017, Nguyen et al., 2015, Zhu et al., 2012). For CyberKnife plans the beamlet and segmentation resolutions were defined at 800 mm from the source, while this was 1000 mm for conventional linac plans. For all plans a beam energy of 6 MV was used. For the prostate treatments, FMO was performed using pencil beams with a 5 x 5 mm^2 beamlet resolution, while a 5 x 10 mm^2 beamlet resolution was used for liver and head-and-neck cancer. These FMO resolutions have shown to provide a good balance between plan quality and computational efficiency. Since MLC segmentation of a FMO dose distribution can result in degradation of plan quality, segmentation on a higher resolution than the FMO resolution can compensate for potential degradation in plan quality. We modelled the CyberKnife InCise2 MLC and performed all segmentations (including the conventional linac plans) for this MLC, which has 2 banks of 26 leaves with a leaf thickness of 3.85 mm defined at 800 mm SAD. Since our dose engine is limited to integer values of resolution only, the segmentation was performed on a resolution of a 1 x 4 mm^2.

2.2. Column generation

CG is generally used to solve large-scale problems. The large-scale optimisation problem is denoted as the *Master Problem* (MP). Instead of solving the MP directly, the MP is solved by iteratively solving a restricted version of the problem denoted as the *Restricted Master Problem* (RMP). The RMP only includes a subset of the original decision-variables (i.e. the beamlet intensities x). During each iteration of the CG method, the RMP is solved and the solution is projected onto the MP. The projection on the MP can be used to identify the next promising subset of decision variables, which will be added to the RMP in the subsequent CG iteration. This identification step is called the *Pricing Problem* (PP). If no new decision-variables can be identified, the MP is solved to optimality. For a detailed description of CG in RT, see Romeijn et al., 2005, Men et al., 2007, Carlsson and Forsgren, 2014.

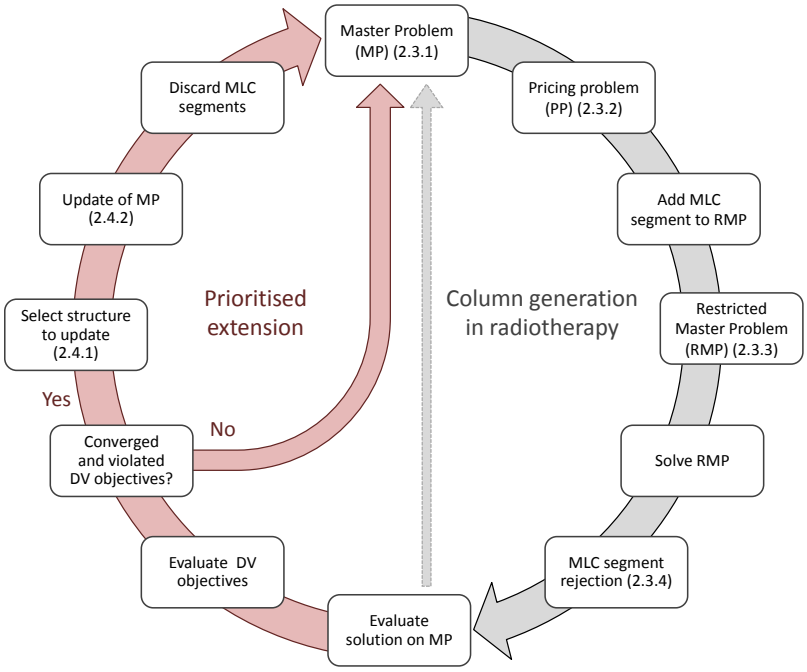

Figure 1. Schematic representation of the column generation (CG) approach for prioritised MLC segmentation as a combination of the grey and red workflows. The numbers between brackets refer to corresponding paragraphs.

2.3. Problem definition and segment generation

The CG workflow, as introduced by Romeijn et al. (2005) for DAO in radiotherapy, was in this study used to segment FMO dose distributions by iteratively identifying promising MLC segments, see figure 1. The CG workflow is denoted in grey, while the proposed prioritised workflow consists of the grey workflow, followed by red.

The MP for plan segmentation is formulated in section 2.3.1. In each iteration the most promising segment is identified by solving the PP (section 2.3.2) and then added to the RMP (section 2.3.3). The RMP only contains the segments identified so far and it is solved to optimality to determine the intensities of the segments. At the end of each iteration, segments for which the intensity falls below the minimum required Monitor Units (MU) are removed (section 2.3.4). Then, in the next iteration the PP is again solved to identify the next promising segment.

For the proposed prioritised MLC segmentation, the CG workflow is incorporated into an adaptive framework (section 2.4). If a segmented solution converges to a dose distribution that does not comply with one or more of the DV criteria, segmentation is re-started with extra emphasis on high priority objectives.

2.3.1. Master Problem With x denoting the fluence vector, our MP is formulated by:

$$\begin{aligned} &\text{minimise} && f(x) + \omega p(x) && \text{(MP)} \quad (1) \\ &\text{subject to} && x \geq 0 \end{aligned}$$

$x \geq 0$ ensures non-negative fluences and $f(x)$ and $p(x)$ are given by:

$$f(x) = \sum_{v \in V} \|\eta_v{}^{1/2}(Ax - d_v^F)\|_2^2 \tag{2}$$

$$p(x) = \frac{1}{N_{PTV}} \sum_{j=1}^{N_{PTV}} e^{-\alpha(A_j x - D^P)} \tag{3}$$

The quadratic term $f(x)$ is the main driving force of the MP, similar to objective functions commonly used in treatment planning (Breedveld et al., 2006, Carlsson and Forsgren, 2014). The vector η_v contains voxel-dependent weight factors for each volume $v \in V$, which play a key role in the MP adaptation part of the algorithm described in section 2.4. d_v^F defines the FMO dose distribution. $f(x)$ can be rewritten in the canonical form to increase computational efficiency (Breedveld et al., 2006).

The term $p(x)$ with weight ω in equation (1) puts extra emphasis on attaining an adequate PTV coverage by penalising under-dosage of the PTV, for which the Logarithmic Tumour Control Probability (LTCP) is used (equation (3), as proposed by Alber and Reemtsen (2007)). D^P denotes the prescribed dose for the PTV, N_{PTV} the number of sampled PTV voxels, and $A_j x$ the dose delivered to PTV voxel j. α is a constant related to cell survival (Alber and Reemtsen, 2007), which was tuned to achieve adequate coverage. An α equal to 0.90, 0.82 and 0.40 was used for prostate, head-and-neck and liver respectively.

2.3.2. Pricing problem and feasible segment generation For the identification of the most promising feasible segment, an approach similar to Romeijn et al. 2005 has been implemented, in which the gradient from the MP is projected onto the beamlet grid for each of the beams. Beamlets with a negative gradient are favourable for inclusion into the next segment, as these indicate the most effective descent direction for the MP. These individual beamlets are grouped together into feasible MLC segments by constructing a layered graph (per beam direction) for possible combinations of adjacent negative beamlets. Mechanical restrictions of the MLC device are taken into account during construction of the graph. The following segment restrictions, similar to the restrictions of the CyberKnife InCise2 MLC, were enforced: the MLC segment contains only one contiguous opening, a minimum number of 2 leaf pairs open per segment (7.7 mm in total), a minimum opening size of 7.6 mm in the direction of the leaves, and interdigitation is allowed. Given that our dose engine operates on integer values of resolution, a field size restriction of two leaf pairs open per segment (8 mm in total) and a minimum opening size of 8 mm in the direction of the leaves were used instead. The graph is subsequently solved using a shortest path algorithm. The pricing problem was implemented in C++ using Boost Graph Libraries (v1.58) and solved using a Bellman-Ford shortest path algorithm. A post-processing step was implemented to guarantee feasible segments. Feasible segments are generated for all beam directions in parallel, but only the most promising segment (the one with the largest sum of negative contributions) is selected and added to the RMP.

2.3.3. Restricted Master Problem The restricted version of the Master Problem is formulated in equation (4), where the quadratic (2) and LTCP (3) terms are now given by equations (5) and (6). Variable x_r denotes segment intensities. In each iteration, a promising segment is identified and added to the RMP; a new column is added to the optimisation problem. The RMP is subsequently solved to optimality to determine the intensities x_r for included MLC segments. For this we use our in-house developed solver (Breedveld et al., 2017), specifically designed and tuned for solving radiotherapy optimisation problems, but any non-linear solver could in principle be used.

$$\begin{aligned} &\text{minimise} && f_r(x_r) + \omega p_r(x_r) && \text{(RMP)} \quad (4) \\ &\text{subject to} && x_r \geq 0 \end{aligned}$$

Where the quadratic and LTCP terms are now given by:

$$f_r(x_r) = \sum_{v \in V} \|{\eta_v}^{1/2}(A_r x_r - d_v^F)\|_2^2 \tag{5}$$

$$p_r(x_r) = \frac{1}{N_{PTV}} \sum_{j=1}^{N_{PTV}} e^{-\alpha(A_{rj}x_r - D^P)} \tag{6}$$

2.3.4. Segment rejection For treatment delivery a minimum MU/segment is imposed because dose delivery for MU below this threshold may be inaccurate. Also, segments added to the RMP in an early stage of the segmentation process can decline in relevance due to addition of newer segments. We have chosen not to enforce the minimum MU/segment constraint while solving the RMP, in order to maintain the ability to identify and remove segments for which the contribution to the solution diminishes. When the intensity of a segment drops below the minimum the segment is removed from the RMP. Additionally, after segment removal, the intensities of the remaining segments are re-optimised and it is again verified whether they fulfil the minimum MU constraint. Discarding redundant columns (segments) from the RMP reduces the size of the problem which improves the computational efficiency of solving the RMP. For the hypo-fractionated SBRT plans (prostate and liver) in this study a minimum MU/segment per fraction of 5 was used and for the conventionally fractionated plans the minimum MU/segment per fraction was 3.

2.4. Prioritised MLC segmentation

The performance regarding the posed (personalised) objectives is tracked during segmentation (section 2.4.1). If a segmented solution converges towards a solution that does not comply with one or more of the objectives, the MP is updated in an attempt to better reflect the requested trade-offs (section 2.4.2), taking into account the clinical priorities.

2.4.1. Prioritised objectives for segmentation For the three investigated tumour sites, the tumour and OAR objectives with assigned priorities as used for the segmentation of FMO plans are presented in table 2a. They are in line with the clinical treatment planning protocols at Erasmus MC. For each patient, the personalised goal values for the OAR objectives are obtained from the Pareto-optimal FMO plan generated with Erasmus-iCycle to obtain trade-offs in OAR sparing during the segmentation like they were made during the automated multi-criterial FMO. During segmentation there is no need to obtain a PTV coverage higher than requested in the clinical protocol, even if it is obtained during FMO. Therefore, to generate maximum space for OAR sparing during segmentation, the tumour objectives are enforced as provided in the treatment protocol.

When during a segmentation all objective functions have converged and one or more of the objective functions have not reached the goal value, an update of the MP is performed, i.e. the MP is modified to put more emphasis on the objective function with an unattained goal with the highest priority (table 2a) in a subsequent segmentation run. When multiple objectives with equal priority are not fulfilled, the structure with the largest deviation from the desired objective value is selected.

2.4.2. Updates of the MP during segmentation Two mechanisms are used simultaneously for updates of the MP. The first approach is adjustment of the individual voxel weights η_v within the quadratic objective function (equation 2). Increasing the weight will magnify the difference between the attained dose in the segmented solution and the reference dose for that voxel, as obtained in the FMO plan. This will put more emphasis on that particular voxel for attaining its reference dose. The second option is to adjust the reference dose d_v^F within the quadratic objective function. By adjusting the reference dose for a voxel the difference in dose will be increased, thereby increasing the contribution to the MP objective function, but this option also favours deviations from the original FMO plan.

For each update, only the voxels in the selected structure that do not comply with the criteria contribute to adjustment of the MP cost function. A maximum of three MP updates was enforced to limit calculation time and to remain close to the FMO solution. The values of ω, η_v and d_v^F with the updates are presented in section 3.1. If the update of the objective function yields a plan of inferior plan quality than before the update the segmentation falls back on the previous plan.

Segmentations are eventually terminated when the MP objective function has converged with all clinical objectives (table 2a) met, or if the maximum number of MP resets (three) has been reached. The convergence criterion was defined as the objective value being within 10% of its current value over the last 10 iterations for segment additions. This criterion was relaxed to 12.5% when all clinical objectives were met. As a result, fewer segments were included when an adequate plan has already been achieved.

Table 2. Personalised objectives for prioritised segmentation of a prostate, head-and-neck or liver FMO plan with assigned priorities (Pr.); for each patient, the goal values for the OAR objectives were the plan parameters in the corresponding FMO dose distributions, while the tumour objectives were always enforced as stated in the treatment planning protocol (a) and hard constraints to evaluate clinical acceptability of plans (b).

(a)

Prostate			**Head-and-neck**			**Liver**		
Tumour objective		Pr.	Tumour objectives		Pr.	Tumour objectives		Pr.
PTV	$V_{D>38Gy} > 95\%$	1	PTV	$V_{D>43.7Gy} > 98\%$	1	PTV	$V_{D>60Gy} > 95\%$	1
				$D_{1cc} < 49.2\ Gy$	2		$D_{1cc} < 75\ Gy$	2
Personalised objectives		Pr.	Personalised objectives		Pr.	Personalised objectives		Pr.
Rectum	D_{1cc}	2	Spinal cord	D_{1cc}	3	Liver - GTV	$V_{D>15Gy}$	3
Bladder	D_{1cc}	3	Brainstem	D_{1cc}	3	Duodenum	D_{1cc}	4
Urethra	$D_{5\%}$	4	Parotid L/R	D_{mean}	4	Small bowel	D_{1cc}	4
Urethra	$D_{10\%}$	4	SMG L/R	D_{mean}	4	Stomach	D_{5cc}	4
Urethra	$D_{50\%}$	4	Larynx	D_{mean}	5	Spinal cord	D_{1cc}	5
			Cochlea L/R	D_{mean}	6	Oesophagus	D_{1cc}	5
			Oesophagus	D_{mean}	6	Kidney L/R	$V_{D>15Gy}$	6
			Oral cavity	D_{mean}	6			

(b)

Prostate			**Head-and-neck**			**Liver**		
Clinical constraints			Clinical constraints			Clinical constraints		
PTV	$V_{D>38Gy} > 95\%$		PTV	$V_{D>43.7Gy} > 98\%$		PTV	$V_{D>60Gy} > 95\%$	
Rectum	$D_{1cc} < 32.3\ Gy$			$D_{1cc} < 49.2\ Gy$			$D_{1cc} < 75\ Gy$	
Bladder	$D_{1cc} < 38\ Gy$		Spinal cord	$D_{1cc} < 50\ Gy$		Liver - GTV	$V_{D<15Gy} \geq 700\ cc$	
			Brainstem	$D_{1cc} < 60\ Gy$		Duodenum	$D_{1cc} < 30\ Gy$	
						Small bowel	$D_{1cc} < 30\ Gy$	
						Stomach	$D_{5cc} < 22.5\ Gy$	
						Spinal cord	$D_{1cc} < 18\ Gy$	
						Oesophagus	$D_{1cc} < 27\ Gy$	
						Kidney L/R	$V_{D>15Gy} \leq 33\%$	

2.4.3. Segment reduction An important contributor to treatment delivery time is the number of segments. To investigate possibilities for active steering on the number of segments, we have implemented and evaluated six segmentation approaches: three prioritised segmentation (PS) methods and three non-prioritised (noPS) methods:

(i) *PS_full*: Full prioritised segmentation as described in sections 2.4.1 and 2.4.2.

(ii) *PS_remove*: PS_full, followed by stepwise removal of segments. For every removal step, the segments are ranked based on their relative contribution to the PTV mean dose. Subsequently, the segment with the lowest contribution is removed and

the intensities x_r of the remaining segments (equation (4)) are re-optimised. This process is continued until a tumour objective is violated.

(iii) *PS_terminate*: Start prioritised segmentation like in PS_full, but terminate as soon as all tumour objectives are met after the last update of the MP.

(iv) *noPS_full*: Start segmentation like in PS_full, but do not update the MP, i.e. the prioritised list of personalised objectives is not used during segmentation.

(v) *noPS_remove*: noPS_full, followed by stepwise removal of segments (see ii for details).

(vi) *noPS_terminate*: Start segmentation like in noPS_full, but terminate as soon as all tumour objectives are met.

2.5. Plan evaluation criteria

Plans segmented with the various approaches were mutually compared and compared with FMO. The analyses focused on clinical acceptability, dosimetric quality, number of segments, MU and segmentation time. Criteria for clinical acceptability are summarised in table 2b. Dose-volume parameters and the *Conformation Number (CN)* as proposed by van 't Riet et al. (1997) were used to quantify dosimetric quality. Additionally, visual inspections of the dose distributions were performed.

2.6. Computation times

Segmentations were performed on a dual CPU system, consisting of 2 octocore Intel Xeon E5-2690 CPUs, running at 2.90 GHz and with 128 GB of memory. For the various segmentation approaches, calculation times were recorded.

3. Results

Segmentation parameters found to be suitable for prioritised segmentation are presented in section 3.1. Prior to presenting the overall performance results for the segmentation approaches in section 3.3, one head-and-neck case is discussed in detail in section 3.2.

3.1. Prioritised segmentation parameters

A weight ω of 10^3 for the LTCP term of the objective function (1) was found to work adequately for all tumour sites and was kept fixed throughout the investigations. As mentioned in section 2.3.1, the contribution of the LTCP term diminished when an adequate PTV coverage was attained. The value of 10^3 provided an appropriate trade-off between steering on sufficient PTV coverage (when necessary) and reconstruction of the FMO dose distribution provided. The voxel weights η_v of the quadratic part of the objective function (2) were all set to 1 at the start of the segmentation. For updates of the MP cost function, the weights for selected voxels were increased from 1 to 5, to 10 and to 15 for subsequent updates. For OAR objectives, the voxel reference doses

d_v^F, initially obtained from the FMO dose distribution were decreased by 0.33 Gy at the same time.

3.2. Segmentation performance - example patient

Figure 2 shows for an example patient axial and sagittal slices through the PTV for the FMO plan (a) and the segmented PS_full plan (b). FMO and segmented dose distributions were similar, though small deviations were noticeable. As required by the clinical protocol (table 2b), in both plans more than 98% of the PTV was covered by at least 95% of the prescribed dose (yellow isodose line in figure 2, 43.7 Gy=95%). Isodose lines of the segmented plan were somewhat smoother, which can be explained by the difference in fluence modulation. In FMO, the fluence is modulated per beamlet, although some form of regularisation (smoothing) is applied. In segmented plans, collections of connecting beamlets are irradiated with equal intensity, which generally results in smoother fluence profiles. In principle, the higher modulation in FMO could have resulted in a more conformal plan. However, in this case a minor increase in CN was observed in the segmented plan; CN=0.77 for FMO and 0.79 for the segmented plan. FMO and segmented fluence profiles for all beams are shown in figure 3. Even though FMO and segmented dose distributions were similar, substantial differences in fluence profiles per beam were observed.

DVHs for the FMO plan and all segmented plans are presented in figure 4. For all three PS methods the goal values for both PTV objectives were attained (table 2), with PTV coverages of 98.9%, 98.8% and 98.8% and PTV D_{1cc} of 49.0 Gy, 49.1 Gy and 49.2 Gy for PS_full, PS_remove and PS_terminate respectively. In comparison, in none of the noPS plans the PTV D_{1cc} goal was achieved (PTV D_{1cc} of 50.1 Gy). Since the PTV D_{1cc} was already violated for noPS_full, no segments could be removed without violating one of the tumour objectives in the reduction step and therefore all noPS plans were equal. Figure 5 shows for the example HN case the PTV objective functions ($V_{D>43.7Gy}$ and $V_{D>49.2Gy}$) as a function of iteration number during the segmentation process. In order to meet for both functions the goal values three MP updates for the PTV were necessary, to place more emphasis on crucial voxels that contributed to PTV overdose.

3.3. Segmentation performance - all patients

Figures 6, 7 and 8 show population mean DVHs for the three patient groups. DVHs per individual patient can be found in the supplementary materials. For prostate and liver the mean DVHs for the six segmentation approaches were very close to those of the FMO plans. For the OARs this also held for head-and-neck cancer, but for the three noPS segmentation approaches the mean $V_{D>49.2Gy}$ for the PTV exceeded the clinical dose constraint (table 2b), see inset of figure 7. Figure 9 shows the number of clinically acceptable plans, which for head-and-neck indeed shows that for each of the three noPS approaches only 1 out of 15 plans was clinically acceptable.

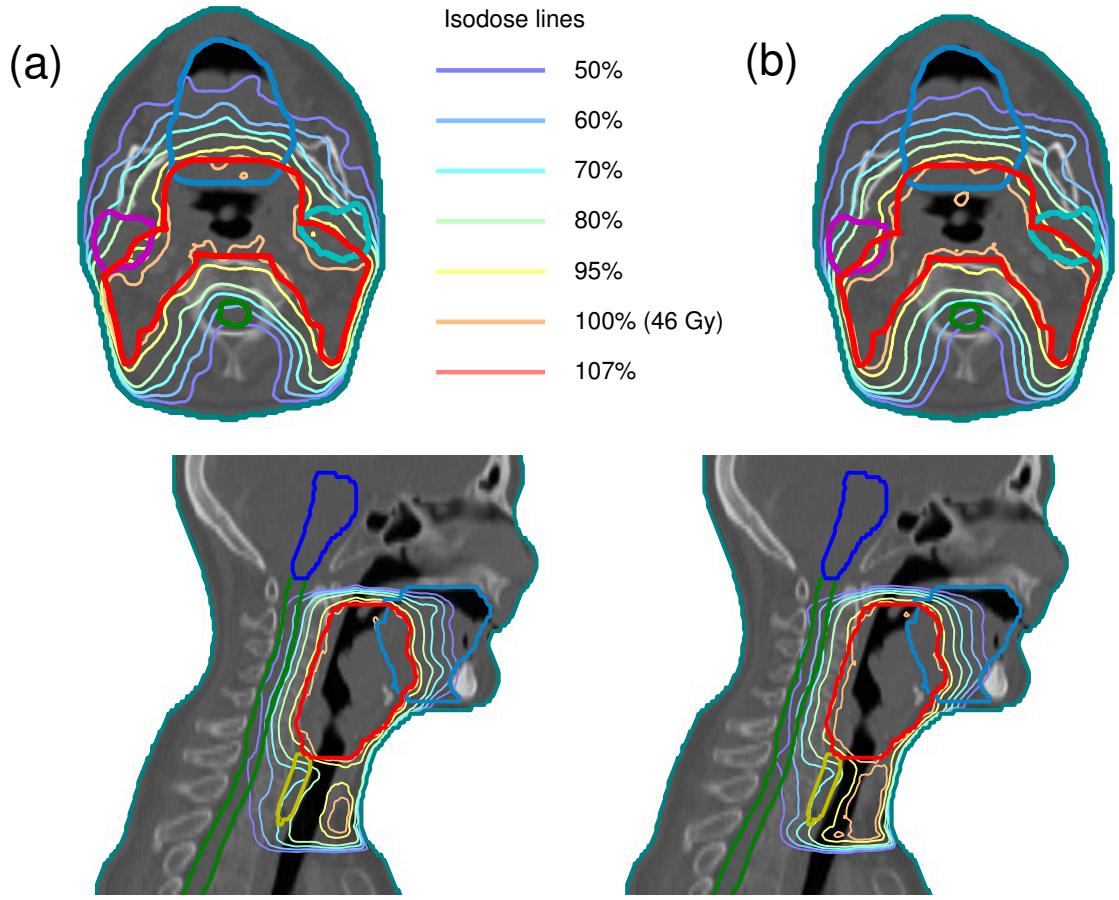

Figure 2. Similar FMO (a) and PS_full segmented (b) dose distributions for the example HN patient discussed in section 3.2. Depicted structures: PTV (red), spinal cord (green), brainstem (blue), parotid L/R (cyan/purple), oesophagus (yellow) and oral cavity (light blue).

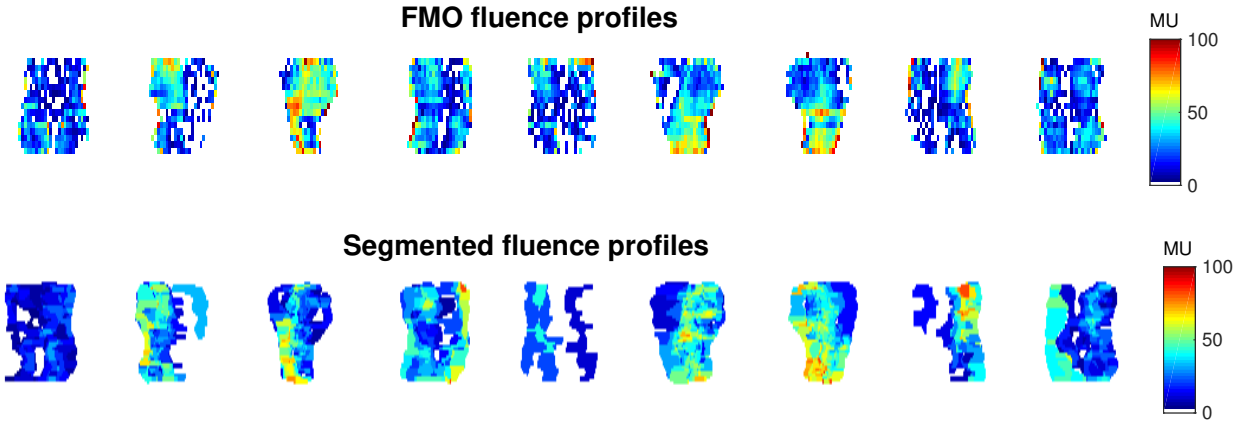

Figure 3. Fluence profiles to realise the FMO and segmented plans for the example patient discussed in section 3.2, plotted per beam. Even though the FMO and segmented dose distributions are similar (figure 2), substantial differences between FMO and segmented fluence profiles per beam were observed.

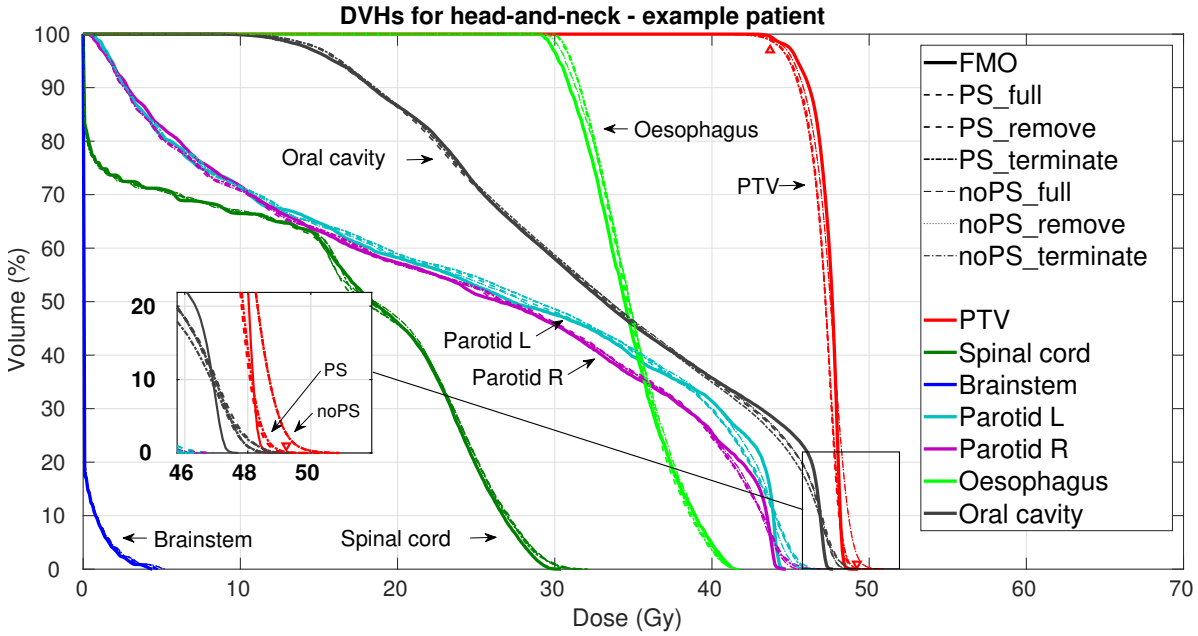

Figure 4. DVHs for FMO and segmented plans for the example head-and-neck case presented in section 3.2. Tumour objectives are denoted with red triangles. Inset: close-up of the DVHs around maximum dose. All three PS plans met both tumour objectives compared to none of the noPS plans (none of them fulfilled the PTV D_{1cc} objective).

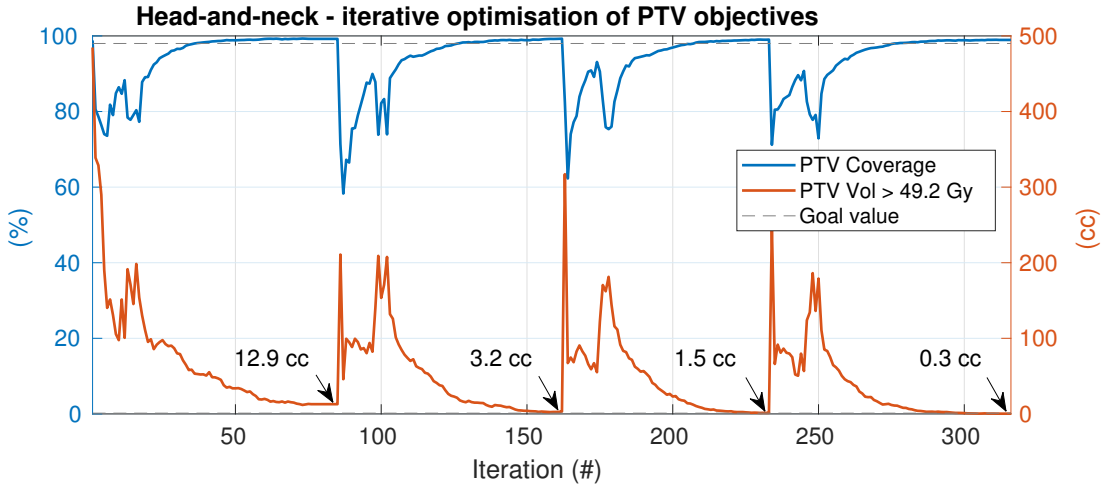

Figure 5. For the example head-and-neck patient discussed in section 3.2, tumour objective functions' values during the PS_full prioritised segmentation plotted against iteration number. In order to meet the goal values (table 2a), three updates of the objective function were necessary. Updates were performed at iteration 85 (PTV $V_{D>49.2Gy} = 12.9$ cc), at iteration 162 (PTV $V_{D>49.2Gy} = 3.2$ cc) and at iteration 233 (PTV $V_{D>49.2Gy} = 1.5$ cc). At the end of PS_full, PTV $V_{D>49.2Gy} = 0.3$ cc, which is within the goal value (< 1 cc, table 2a).

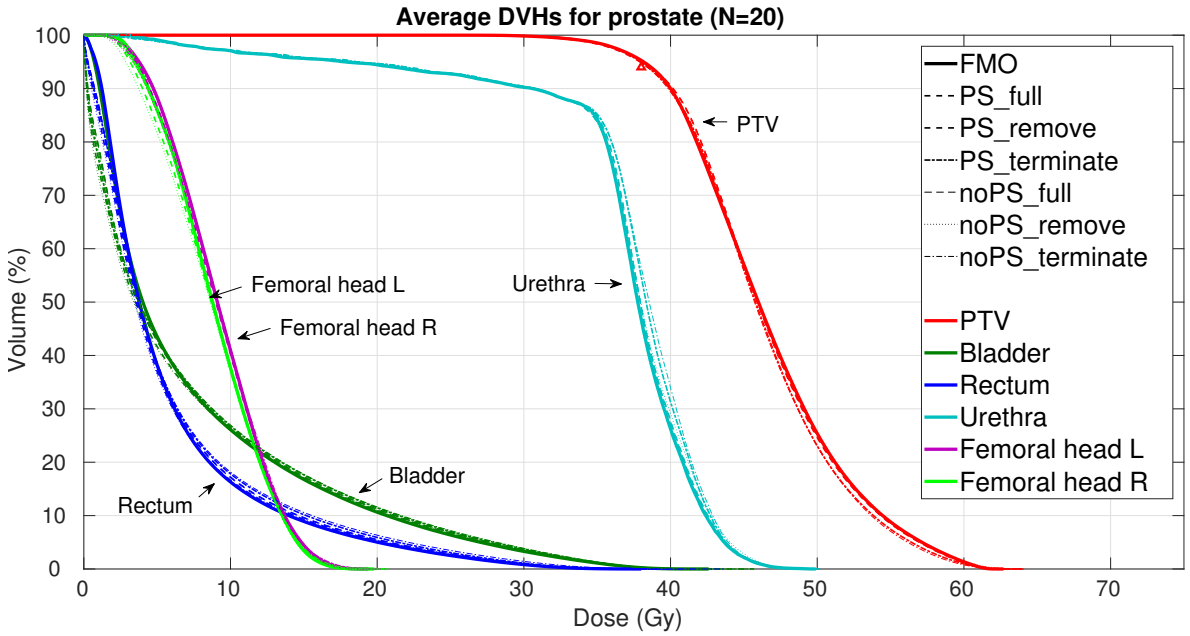

Figure 6. Average DVHs for the prostate cases (N=20) for the FMO plans (thick solid lines) and the segmented plans (dashed). The PTV coverage objective is denoted with a red triangle.

Figures 10, 11 and 12 show details on the performance of the six segmentation approaches for the three investigated tumour sites. The subfigure "PTV D1cc" of figure 11 shows that acceptability issues with noPS plans for head-and-neck cancer were indeed indeed related to too large PTV volumes receiving high dose. Among the three PS approaches, differences between achieved dosimetric parameters for head-and-neck were clinically irrelevant (figure 11). Overall the best segmentation approach for this group seems PS_terminate as it has the lowest # segments (median: 75), the lowest # iterations (median: 334) and the lowest MU (median per fraction: 979 MU). Also for prostate and liver cancer, PS_terminate is often a good choice, mainly because of the relatively low # segments, low # iterations and low total MU.

Another interesting observation are the differences in deviation observed per segmentation method over various objectives with different priorities (table 2a). For example, for prostate cancer the smallest deviations were observed for the most important OAR (i.e. rectum). With decreasing importance (increasing priority number) the deviations in dose from the FMO plan increased (compare objective subplots in figure 10). No similar trend was observed for the noPS approaches, which indicates that the proposed extension of the segmentation technique is able to reduce dose deviations in a prioritised manner.

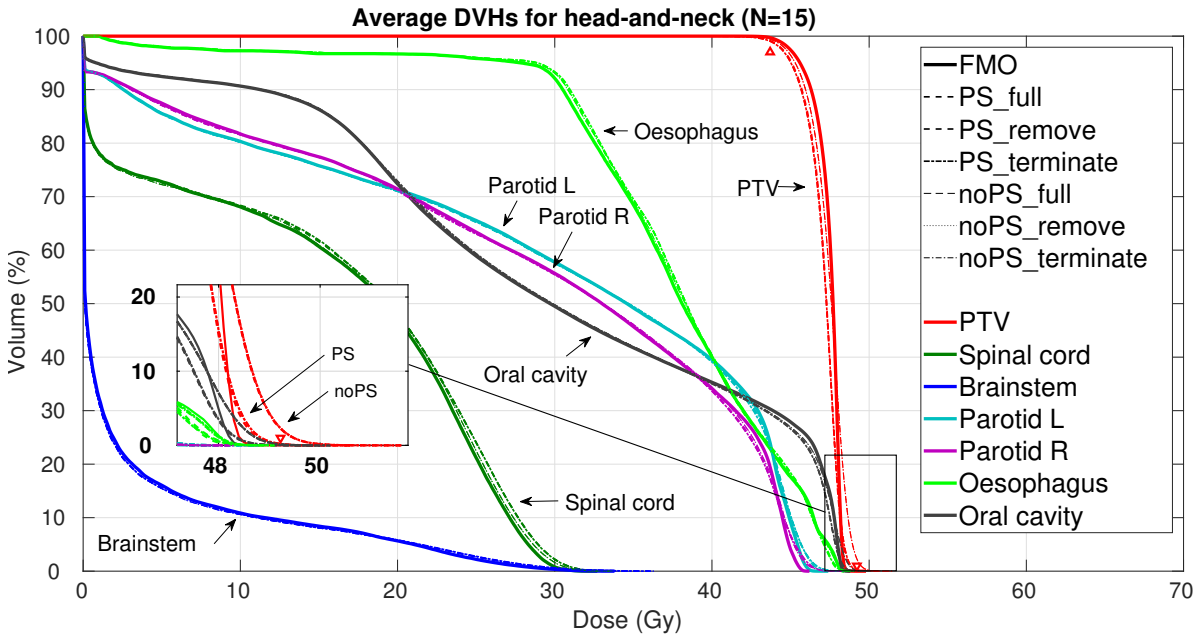

Figure 7. Average DVHs for the head-and-neck cases (N=15) for the FMO plans (thick solid lines) and the segmented plans (dashed). Inset: close-up of the DVHs around maximum dose. PTV objectives are denoted with red triangles.

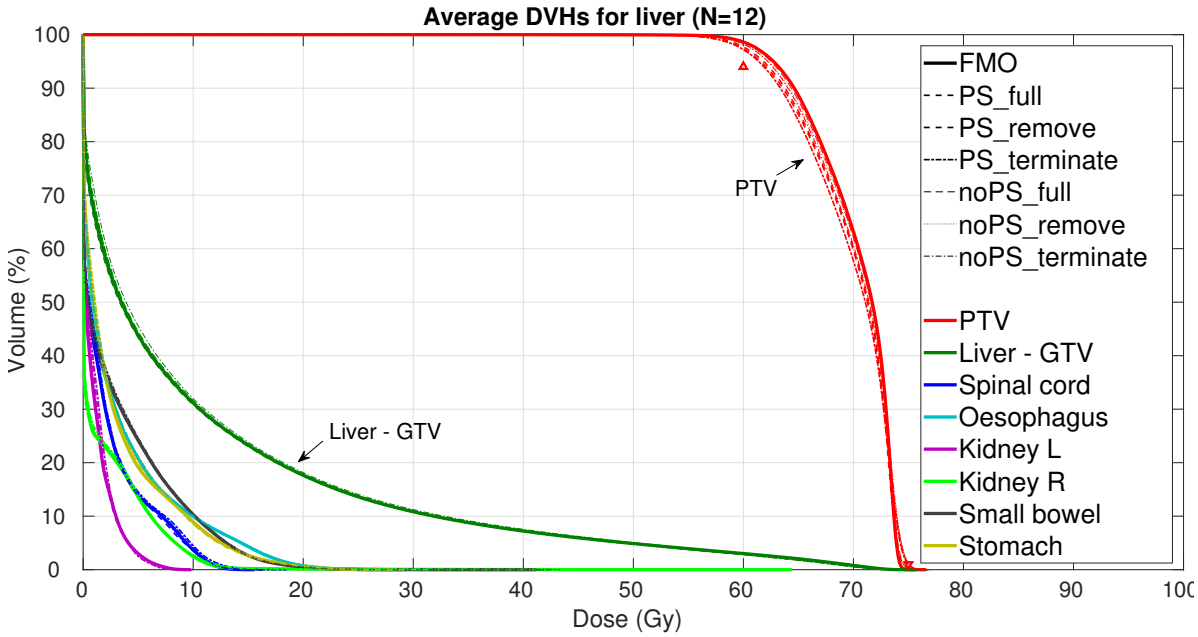

Figure 8. Average DVHs for the liver cases (N=12) for the FMO plans (thick solid lines) and the segmented plans (dashed). PTV objectives are denoted with red triangles.

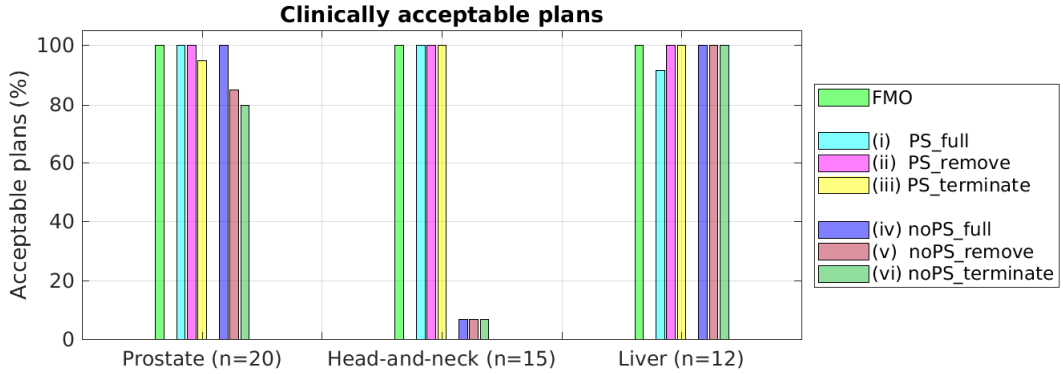

Figure 9. Percentage of clinically acceptable plans. A plan was considered acceptable if for all constraints the obtained values were within 0.25 Gy or 0.25% of imposed values (table 2b). Plans segmented using the prioritised methods (PS) outperformed the non-prioritised methods (noPS). Especially for head-and-neck where more emphasis was needed on crucial voxels, in 14 out of the 15 cases, in order to meet the PTV D_{1cc} constraint.

4. Discussion

MLC segmentation approaches were investigated that aim at accurate reconstruction of optimised FMO distributions, while complying with the mechanical limitations of the treatment device. To reconstruct 3-dimensional FMO dose distributions, a column generation approach was implemented that simultaneously optimised the beam segments for all treatment beams, while ignoring the underlying FMO fluence profiles.

For OARs the proposed segmentation method uses a convex quadratic objective function to minimise the voxel-wise differences between the intended FMO dose and the segmented dose. During prioritised segmentation, this objective function may iteratively be adapted to maximally reproduce OAR dose parameters, while considering the clinical priorities. In initial attempts, we tried to also use for the PTV only a quadratic cost function, similar to the OARs. However, achieving clinically acceptable PTV coverages and maximum doses often failed. Therefore, the quadratic function was supplemented with an LTCP term. Also for FMO plan generation, the LTCP cost function is often used for obtaining adequate PTV dose (Alber and Reemtsen, 2007, Breedveld et al., 2012, 2017).

In our study, segmented plans were in good agreement with the FMO plans (section 3.3), and the number of segments and MU of the generated plans are in line with our clinical experience. Quantitative comparison with published segmentation methods is difficult due to large variations in clinical cases and plans, and in applied segmentation objectives, parameters and quality measures. Also, in published studies the number of evaluated cases is generally relatively low (typically 2 to 10 compared to 47 in this study). However, qualitative comparisons indicate a superior FMO plan reproduction with the proposed approach. Explanations for this could be i) the direct reconstruction of the 3D FMO dose distribution with total ignorance of obtained FMO fluences, and simultaneous

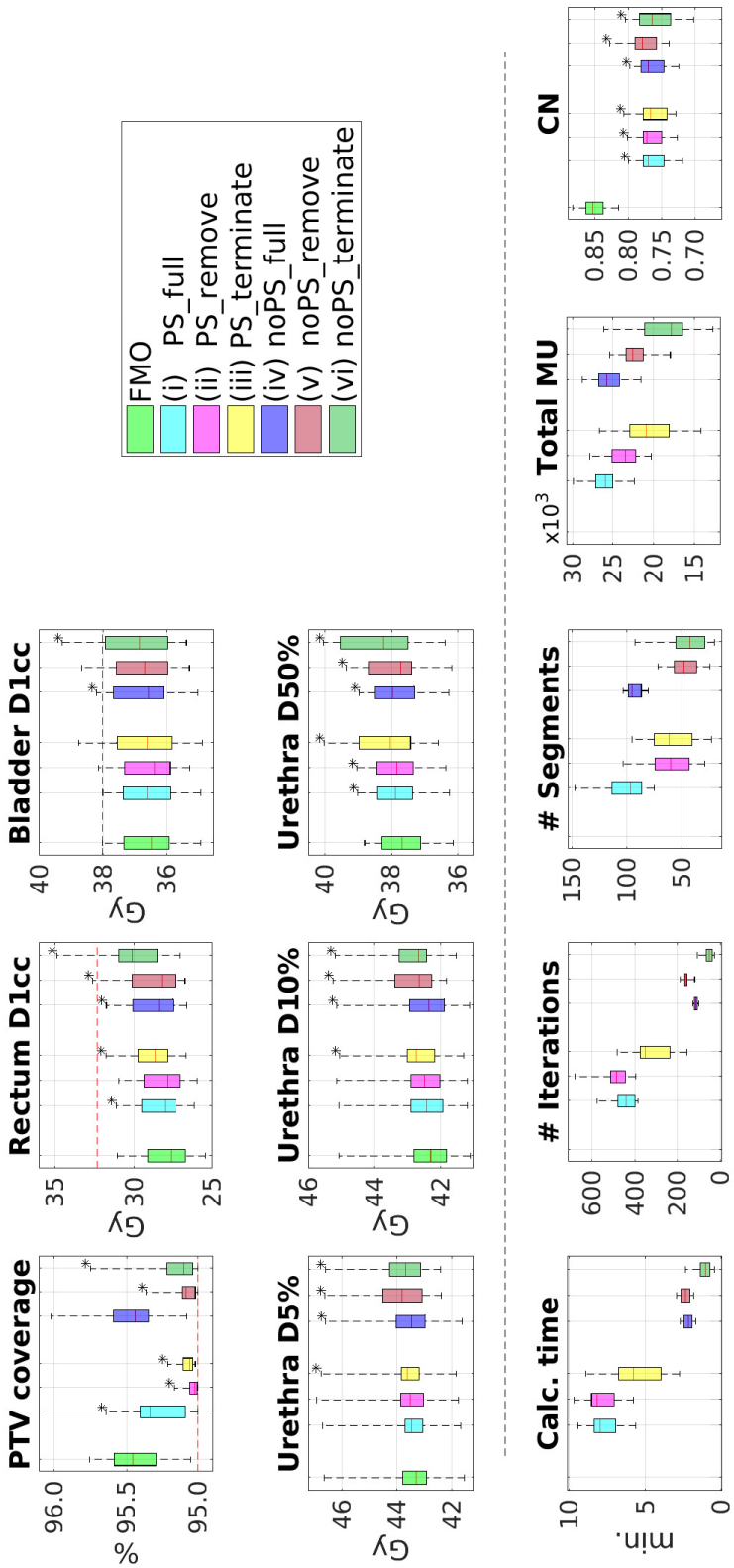

Figure 10. Comparison of treatment plan characteristics for FMO and the six segmentation approaches for prostate cancer. The whiskers in the boxplots denote the range of the data, i.e. minimum and maximum observed values and the coloured bars the 25-75% percentile range with the median value depicted with a horizontal line. Horizontal red dashed lines denote clinical constraints as summarised in table 2b. A plan was considered acceptable if for all constraints the obtained values were within 0.25 Gy or 0.25% of imposed values. MU = Monitor Unit, CN = Conformation Number, an asterisk (*) denotes a significant ($p < 0.05$) difference compared to FMO (Wilcoxon Signed Rank test).

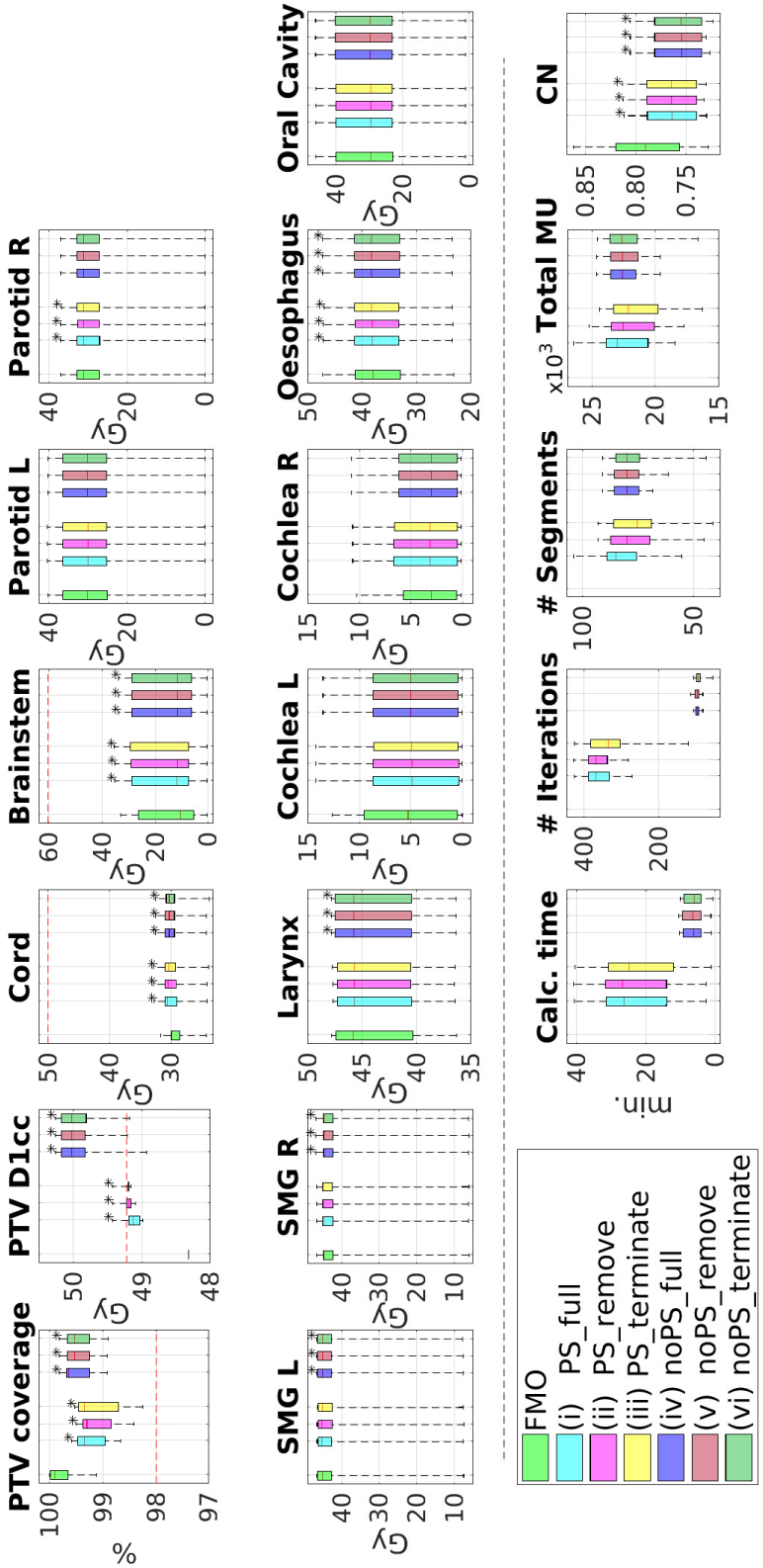

Figure 11. Comparison of treatment plan characteristics for FMO and the six segmentation approaches for head-and-neck cancer. The whiskers in the boxplots denote the range of the data, i.e. minimum and maximum observed values and the coloured bars the 25-75% percentile range with the median value depicted with a horizontal line. Horizontal red dashed lines denote clinical constraints as summarised in table 2b. A plan was considered acceptable if for all constraints the obtained values were within 0.25 Gy or 0.25% of imposed values. MU = Monitor Unit, CN = Conformation Number, an asterisk (*) denotes a significant ($p < 0.05$) difference compared to FMO (Wilcoxon Signed Rank test).

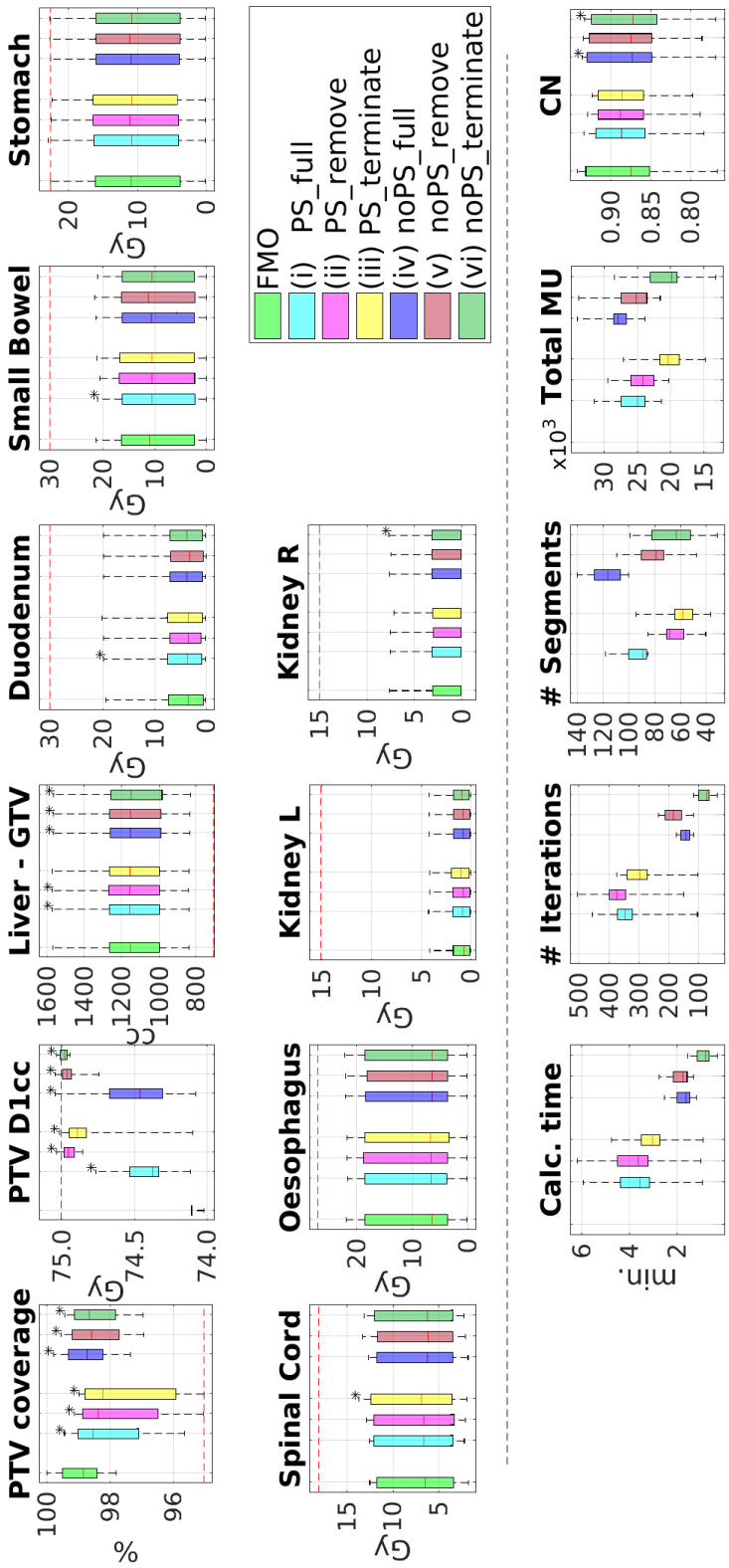

Figure 12. Comparison of treatment plan characteristics for FMO and the six segmentation approaches for liver cancer. The whiskers in the boxplots denote the range of the data, i.e. minimum and maximum observed values and the coloured bars the 25-75% percentile range with the median value depicted with a horizontal line. Horizontal red dashed lines denote clinical constraints as summarised in table 2b. A plan was considered acceptable if for all constraints the obtained values were within 0.25 Gy or 0.25% of imposed values. MU = Monitor Unit, CN = Conformation Number, an asterisk (*) denotes a significant ($p < 0.05$) difference compared to FMO (Wilcoxon Signed Rank test).

segment generation for all beam directions, instead of the generally observed focus on reconstruction of separate fluence profiles, and ii) the prioritised approach with an explicit drive to avoid plan quality losses for the highest clinical priorities. With this approach, segmentation is fully focused on maintaining the quality of the FMO plan with minimal limitations in selecting optimal segments and preservation of the clinical trade-offs. We have indeed observed substantial differences between FMO- and segmented fluence profiles for similar dose distributions (see figures 2 and 3 for an example patient). Possibly, the featured large freedom in beam segment selection has contributed to the high quality of the reconstructed plans with clinically acceptable numbers of segments (see figures 10, 11 and 12). Nguyen et al. 2015 investigated dose domain regularisation for MLC segmentation and observed that the segmented fluence profiles with and without regularisation could be substantially different. A direct comparison with published or commercially available segmentation approaches would be very interesting. For such studies it would be important to eliminate potential bias, e.g. originating from difference in dose calculation models, segmentation objectives or evaluation measures.

In this paper, FMO treatment plans were generated with Erasmus-iCycle (Breedveld et al., 2012), which has been successfully implemented in clinical practice for fully automated multi-criterial generation of clinically deliverable plans for head-and-neck, prostate, advanced lung cancer and advanced cervical cancer. For these tumour sites, Erasmus-iCycle is used for FMO plan generation, while the Monaco TPS (Elekta AB, Stockholm, Sweden) is effectively used for segmentation (Voet et al., 2013, 2014, Sharfo et al., 2015, Gala et al., 2017, Heijmen et al., 2018). For automated offline treatment planning calculation time is not crucial, but in other scenarios it could be. Therefore, we have investigated calculation times for the segmentation approaches. When plan quality is the most important aspect, the PS_full is most suitable; fully converged segmentation with the best plan quality. When calculation time is more important, for example in the case of online-adaptive treatment, the PS_terminate could be a more suitable option. It provides a reduction in number of segments compared to the PS_full, with only minimal impact on plan quality. Additionally, the *terminate* and *remove* plans generally have a lower number of MU and so these plans are more efficient to deliver compared to the fully converged plans. A possible drawback of the terminate and remove approaches could be that the personalised objectives obtained from the FMO dose distributions are too challenging to reconstruct under the mechanical limitations. In that case, a plan would be returned with a high number of segments while fulfilling the highest feasible objectives. Regarding the terminate approach, since the segmentation minimises the difference between the FMO dose and segmented dose, terminating the segmentation when the tumour objectives are met does not necessarily mean that none of the remaining objectives are met. The segmentation primarily works on all dose points simultaneously with an extra emphasis on PTV coverage due to the LTCP term, prior to placing extra emphasis on other prioritised objectives. Investigations on further reduction of calculation times using GPU are on-going; preliminary results indicate a

potential reduction in calculation time with a factor of 5.

The prioritised extension of the segmentation has been observed to be effective, but the impact is patient dependent and it increases calculation time. An option for clinical practice would be to always perform a full prioritised segmentation and let the user decide which plan is most appropriate per patient, given that the intermediate (e.g. PS_terminate, noPS_full, noPS_terminate) plans are also available when PS_full is performed.

Interesting opportunities for further improving the proposed MLC segmentation approach would be to integrate published improvements in DAO techniques into this MLC segmentation method. One interesting approach in particular would be to integrate the Aperture Shape Optimization (ASO) algorithm as proposed by Cassioli and Unkelbach (2013), which optimises the shapes of the included segments in between CG iterations. However, this will also increase computation time. Another useful improvement might be to integrate a clinical dose engine in order to account for MLC scatter effects or to include a fuzzy controller to reduce numerical noise on the gradient maps as proposed by Yang et al. (2018).

5. Conclusions

Novel MLC segmentation approaches have been proposed for accurate reconstruction of high-quality FMO dose distributions, while complying with the mechanical limitations of the treatment device. 3-dimensional FMO dose distributions are reconstructed with total ignorance of underlying FMO fluences, and simultaneous segment generation for all beam directions. Due to the proposed prioritised approach, plan reconstruction has an accent on high priority planning objectives. Clinically acceptable segmented dose distributions could be generated for all cases with a plan quality that was in good agreement with the FMO plan and clinically acceptable numbers of segments.

Acknowledgments

The authors want to thank Rebecca Jacobs for her contribution to the segment generation method. We also thank Abdul Wahab Sharfo and Rens van Haveren for their support and for sharing their research data and we thank Warren Kilby (Accuray Inc.) for his helpful comments and insight in the CyberKnife system. This work was in part funded by a research grant from Accuray Inc, Sunnyvale, USA. Erasmus MC Cancer Institute also has a collaboration agreement with Elekta AB, Stockholm, Sweden.

References

Ahuja, R. K. and Hamacher, H. W. (2005). A network flow algorithm to minimize beam-on time for unconstrained multileaf collimator problems in cancer radiation therapy, *Networks* **45**(1): 36–41.

Alber, M. and Reemtsen, R. (2007). Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method, *Optim. Methods Softw.* **22**: 391–411.

Bokrantz, R. and Miettinen, K. (2015). Projections onto the Pareto surface in multicriteria radiation therapy optimization, *Med. Phys.* **42**: 5862.

Boland, N., Hamacher, H. W. and Lenzen, F. (2004). Minimizing beam-on time in cancer radiation treatment using multileaf collimators, *Networks* **43**(4): 226–240.

Breedveld, S., Craft, D., van Haveren, R. and Heijmen, B. (2019). Multi-criteria optimization and decision-making in radiotherapy, *Eur. J. Oper. Res.* **277**(1): 1 – 19.

Breedveld, S., Storchi, P., Keijzer, M. and Heijmen, B. (2006). Fast, multiple optimizations of quadratic dose objective functions in IMRT, *Phys. Med. Biol.* **51**: 3569–3579.

Breedveld, S., Storchi, P., Voet, P. and Heijmen, B. (2012). iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans, *Med. Phys.* **39**: 951–963.

Breedveld, S., van den Berg, B. and Heijmen, B. (2017). An interior-point implementation developed and tuned for radiation therapy treatment planning, *Comput. Optim. Appl.* **68**(2): 209–242.

Carlsson, F. (2008). Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy, *Med. Phys.* **35**(9): 3828–3838.

Carlsson, F. and Forsgren, A. (2014). On column generation approaches for approximate solutions of quadratic programs in intensity-modulated radiation therapy, *Ann. Oper. Res.* **223**(1): 471–481.

Cassioli, A. and Unkelbach, J. (2013). Aperture shape optimization for IMRT treatment planning, *Phys. Med. Biol.* **58**(2): 301.

Craft, D. and Richter, C. (2013). Deliverable navigation for multicriteria step and shoot IMRT treatment planning, *Phys. Med. Biol.* **58**(1): 87–103.

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**: 1596–1605.

Crooks, S. M., McAven, L. F., Robinson, D. F. and Xing, L. (2002). Minimizing delivery time and monitor units in static IMRT by leaf-sequencing, *Phys Med Biol* **47**(17): 3105–3116.

Gala, G. D., Dirkx, M. L. P., Hoekstra, N., Fransen, D., Lanconelli, N., van de Pol, M., Heijmen, B. and Petit, S. F. (2017). Fully automated vmat treatment planning for advanced-stage nsclc patients, *Strahlentherapie und Onkologie*.

Gören, M. and Taşkin, Z. C. (2015). A column generation approach for evaluating delivery efficiencies of collimator technologies in IMRT treatment planning, *Phys. Med. Biol.* **60**(5): 1989–2004.

Heijmen, B., Voet, P., Fransen, D., Penninkhof, J., Milder, M., Akhiat, H., Bonomo, P., Casati, M., Georg, D., Goldner, G., Henry, A., Lilley, J., Lohr, F., Marrazzo, L., Pallotta, S., Pellegrini, R., Seppenwoolde, Y., Simontacchi, G., Steil, V., Stieler, F., Wilson, S. and Breedveld, S. (2018). Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy - An international multi-center validation for prostate cancer, *Radiother Oncol* **128**(2): 343–348.

Long, T., Chen, M., Jiang, S. and Lu, W. (2016). Continuous leaf optimization for IMRT leaf sequencing, *Med. Phys.* **43**(10): 5403.

Luan, S., Wang, C., Chen, D. Z., Hu, X. S., Naqvi, S. A., Wu, X. and Yu, C. X. (2006). An improved MLC segmentation algorithm and software for step-and-shoot IMRT delivery without tongue-and-groove error, *Med. Phys.* **33**(5): 1199–1212.

Men, C., Romeijn, E., Taşkin, C. and Dempsey, J. (2007). An exact approach to direct aperture optimization in IMRT treatment planning, *Phys. Med. Biol.* **52**: 7333–7352.

Nguyen, D., O'Connor, D., Yu, V. Y., Ruan, D., Cao, M., Low, D. A. and Sheng, K. (2015). Dose domain regularization of MLC leaf patterns for highly complex IMRT plans, *Med. Phys.* **42**(4): 1858–1870.

Romeijn, H. E., Ahuja, R. K., Dempsey, J. F. and Kumar, A. (2005). A column generation approach to radiation therapy treatment planning using aperture modulation, *SIAM J. Optim.* **15**: 838–862.

Salari, E. and Unkelbach, J. (2013). A column-generation-based method for multi-criteria direct aperture optimization, *Phys. Med. Biol.* **58**(3): 621–639.

Sharfo, A.-W., Voet, P., Breedveld, S., Mens, J.-W., Hoogeman, M. and Heijmen, B. (2015). Comparison of VMAT and IMRT strategies for cervical cancer patients using automated planning, *Radiother. Oncol.* **114**: 395–401.

Shepard, D. M., Earl, M. A., Li, X. A., Naqvi, S. and Yu, C. (2002). Direct aperture optimization: A turnkey solution for step-and-shoot IMRT, *Med. Phys.* **29**: 1007–1018.

Sun, X. and Xia, P. (2004). A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning, *Med. Phys.* **31**(5): 1158–1165.

Süss, P., Küfer, K.-H. and Thieke, C. (2007). Improved stratification algorithms for step-and-shoot MLC delivery in intensity-modulated radiation therapy, *Phys. Med. Biol.* **52**: 6039–6051.

van 't Riet, A., Mak, A. C., Moerland, M. A., Elders, L. H. and van der Zee, W. (1997). A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate, *Int. J. Radiat. Oncol. Biol. Phys.* **37**(3): 731–736.

Voet, P., Dirkx, M., Breedveld, S., Al-Mamgani, A., Incrocci, L. and Heijmen, B. (2014). Fully automated VMAT plan generation for prostate cancer patients, *Int. J. Radiat. Oncol. Biol. Phys.* **88**: 1175–1179.

Voet, P., Dirkx, M., Breedveld, S., Fransen, D., Levendag, P. and Heijmen, B. (2013). Towards fully automated multi-criterial plan generation: a prospective clinical study, *Int. J. Radiat. Oncol. Biol. Phys.* **85**: 866–872.

Xia, P. and Verhey, L. J. (1998). Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments, *Med. Phys.* **25**(8): 1424–1434.

Yang, J., Gui, Z., Zhang, L. and Zhang, P. (2018). Aperture generation based on threshold segmentation for intensity modulated radiotherapy treatment planning, *Med. Phys.* **45**(4): 1758–1770.

Zhu, L., Niu, T., Choi, K. and Xing, L. (2012). Total-variation regularization based inverse planning for intensity modulated arc therapy, *Technol. Cancer Res. Treat.* **11**(2): 149–162.