Phys. Med. Biol. **45** (2000) 1741–1754. Printed in the UK PII: S0031-9155(00)10329-X

Monte Carlo dose computation for IMRT optimization*

W Laub, M Alber, M Birkner and F Nüsslin

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

Received 13 December 1999, in final form 28 March 2000

Abstract. A method which combines the accuracy of Monte Carlo dose calculation with a finite size pencil-beam based intensity modulation optimization is presented. The pencil-beam algorithm is employed to compute the fluence element updates for a converging sequence of Monte Carlo dose distributions. The combination is shown to improve results over the pencil-beam based optimization in a lung tumour case and a head and neck case. Inhomogeneity effects like a broader penumbra and dose build-up regions can be compensated for by intensity modulation.

1. Introduction

The method of Monte Carlo (MC) dose calculation can potentially offer some advantages over conventional dose computation in complex geometries both in the patient and the linear accelerator head, which are often found in IMRT cases. Whilst it is undisputed that MC algorithms have the potential to model electron transport more accurately in inhomogeneous media, they also afford more elegant ways to deal with the complex scatter patterns of treatment machines and beam modifying equipment. On the other hand, where with standard treatment techniques one cannot avoid the dose loss due to lateral electron transport at low-density surfaces, intensity modulation (IM) can compensate for this by adjustments to the primary fluence. Hence it appears desirable not only to provide for MC dose verification where indicated (Nahum 1997, Mohan 1997), but also for inclusion of the MC dose calculation in the optimization process. For this to be achieved, the first step must be to conceive an algorithm which uses an MC dose computation engine in the iterative optimization routine of an IMRT algorithm.

The MC computation times for photon treatment prevent a simple replacement of the dose computation engine of an IM algorithm, so that several attempts have been made to devise an optimization algorithm for MC dose computation (Jeraj and Keall 1999, Bogner *et al* 1999). In this paper we take the opposite approach to demonstrate how an existing IM algorithm can incorporate MC to full accuracy as an additional stage of optimization.

The method is essentially independent of the first IM optimization stage but for two requirements. Firstly, the objective function must be convex (at least on a sufficiently large set around the global minimum). Secondly, the relative error of the conventional dose computation which is used for updates to the fluence elements must be smaller than some case-dependent constant at all points in the patient. If these requirements are met, the algorithm is shown to converge to the global minimum obtained by the MC dose computation (as opposed to that generated by the conventional dose computation). This minimum does not depend on the way

* The first and second authors contributed equally to this work.

0031-9155/00/071741+14$30.00

the fluence updates are obtained. The IM/MC optimization algorithm is outlined in section 2. In section 3 we give a brief description of the framework in which the MC code EGS4 was implemented. Finally two cases of target volumes adjacent to low-density surfaces are treated in section 4: one lung tumour, and one tumour close to the sinuses and nasal cavities.

2. The IM/MC algorithm

Radiotherapy optimization takes place in fluence space yet with objectives defined in dose space. Every algorithm which invokes derivatives of the objective function with respect to some fluence entity has to perform two distinct dose calculations: one for setting the current dose and one for computing the derivative. We define

$$0 < f(D) = f(T\Phi) \tag{1}$$

$$\partial_\Phi f(D)|_{\Phi^0} = \partial_D f(D)|_{D^0 = T\Phi^0} T \tag{2}$$

$$\partial_{\phi_i} f(D)|_{\Phi^0} = \partial_{d_j} f(D)|_{d_j^0 = T_{ji}\phi_i^0} T_{ji} \tag{3}$$

with $f(D)$ being the objective function depending on the dose distribution D and $\Phi(\alpha, \beta, u, v)$ being the fluence impinging from angle $\Omega = (\alpha, \beta)$ and field coordinate (u, v) and T the linear absorbed dose per unit fluence operator. In the last line we give the discrete version of this relation, where j runs over all points of the dose computation grid of the patient and i enumerates the fluence elements ϕ_i at coordinates (u_i, v_i). The variation and the derivative respectively on the right-hand side are taken with respect to the dose distribution D and the vector of doses (d_j) on grid points j. Notice that we use tensor notation throughout, i.e. for summations over paired indices j on one side of the equation the symbol $\sum_j$ is omitted.

Whereas for dose computation it is sufficient to know the scalar products $T_{ji}\phi_i$ of the rows of T_{ji} and the fluence vector ϕ_i, for the derivative the entire matrix of operator T_{ji} has to be known, which is usually more expensive to compute to the same accuracy. Several optimization algorithms exploit the fact that these dose computations do not necessarily have to be done by the same dose calculation engine: a faster yet less accurate one may be selected for the derivatives (Preiser *et al* 1997, Wang *et al* 1995, Xing *et al* 1998)†. We propose to use a PB algorithm for computing the derivatives $\partial_{\phi_i} f$ which are used to update the fluence profiles in the IM optimization algorithm. This method suffers from the setback that convergence to the global minimum can be uncertain and under adverse conditions impossible. The convergence of gradient algorithms depends on the quality of the information in the derivative computation. If this information is corrupted to a degree that the direction of the next step is no longer 'downhill', the algorithm can become trapped in a cycle rather than converge to the minimum, although this cycle may correspond to an acceptable solution.

Please notice that this issue also pertains to an algorithm which does use MC to compute the entire operator T_{ji}. Since MC is a stochastic method, any result has to be considered an approximation to the 'true' result up to some statistical error. The following requirement for the accuracy of the dose computation to ensure convergence is equally valid for a full MC optimization. Notice that the computation of a pencil beam dose to be stored as a column in T_{ji} demands disproportionately more particle histories to attain the same statistical variance than if the row sums were computed together.

To save MC computation time, we propose to start the algorithm with a fluence Φ^0 obtained from an initial optimization step with, say, pencil-beam (PB) dose computation. If the PB dose stood up to the standards of MC, a rerun of the optimization routine with $D^0 = T_{\mathrm{PB}}\Phi^0$ replaced

† These IM algorithms usually utilize a combination of PB for the dose computation and primary fluence for the computation of derivatives.

with $D^0_{\mathrm{MC}} = T_{\mathrm{MC}}\Phi^0$ would not yield a different result. In practice, this will not be the case and the algorithm will deliver an update to the fluence Φ^1 as results from the PB derivative calculations. In general, convergence of the IM/MC algorithm can only be guaranteed if the PB derivative provides sufficiently accurate information. This is formalized in the following theorem whose proof is given in the appendix.

For a differentiable and convex objective function a sequence of objective function values obtained by some convergent algorithm decreases monotonically, i.e.

$$f(D^{k+1}_{\mathrm{MC}}) < f(D^k_{\mathrm{MC}}) \qquad \text{for any } k > 0 \tag{4}$$

if the error in the derivative due to the approximations in T_{PB} is bounded as:

$$|(\partial_D f(D)|_{D^k}(T_{\mathrm{MC}} - T_{\mathrm{PB}}))_i| \leqslant \tfrac{1}{2}|(\partial_D f(D)|_{D^k} T_{\mathrm{PB}})_i| \qquad \text{for all } i \tag{5}$$

at each iteration k. This result is independent of the kind of algorithm used as long as it employs at most first derivatives.

Notice that this condition applies to all columns separately. If this condition is violated, the algorithm can come to a halt before reaching the minimum. This convergence barrier depends strongly on the largest elements in the dose-derivative vector $\partial_{d_j} f$ and thus on the problem†. For IM/MC optimization to be faithful, it follows that the difference $|(T_{\mathrm{MC}} - T_{\mathrm{PB}})_{ji}| < \epsilon (T_{\mathrm{PB}})_{ji}$ has to be bounded with some $\epsilon > 0$, which makes the use of a finite-size PB algorithm rather than the primary fluence approximation indispensable. In practice, it is sufficient that condition equation (5) holds when the termination criterion of the optimization routine is met. This will be given if the positive and negative entries in $\partial_{d_j} f$ differ by no more than a factor of roughly $1/\epsilon$ on average, i.e. quite substantially. Although present, the most dominant error will not be the statistical variance of the MC dose distribution, and hence this cannot influence the result unduly. In the appendix, this issue is discussed further, and also a necessary condition for algorithm convergence is given.

The pseudocode of the IM/MC algorithm reads:

1. Compute T_{PB} with a finite-size pencil-beam algorithm and solve optimization problem defined by minimizing $f(D)$ to yield Φ^0. Store $(T_{\mathrm{PB}})_{ji}$ to compute derivatives in step 4.
2. MC loop counter $k = 0$.
3. Compute $D^0 = T_{\mathrm{MC}}\Phi^0$.
4. On the basis of the MC dose D^k, find fluence Φ^{k+1} by using T_{PB}.
5. Update MC dose: $D^{k+1} = D^k + T_{\mathrm{MC}}(\Phi^{k+1} - \Phi^k)$.
6. Has the algorithm converged? If no, $k = k + 1$, goto 4.

The updates of the fluence elements in step 5 can also be negative, in which case the particle would be given negative statistical weight. The optimization routine in step 4 prevents smaller than minimal total fluence ϕ^k_i by a penalty term. Convergence can be checked by comparing the objective function value of the intermediate MC/PB dose and the corresponding MC dose. If the difference between both objective function values satisfies the termination criterion of the optimization routine in step 4, the IM/MC loop terminates. The statistical error of the MC dose distribution does not affect convergence since due to the laws of error propagation it would be in the order of 10^{-4} for the derivatives, and 10^{-5} (relative) for the objective function, i.e. well below the convergence threshold.

† Large entries in this vector are connected either to underdosage in the target volume or overdosage in organs at risk, so that a failure of convergence is more likely to occur if strongly conflicting objectives have to be met. This implies that a failure of convergence will result either in fluence spikes or dips around the outline of the target volume or organs at risk.

The quintessential advantage of this scheme is that MC computation time is not wasted on the intermediate updates of the dose distribution, but is only used to generate a converging sequence of verification dose distributions. This can be captured in the figure of MC efficiency

$$K = \frac{\text{no of MC events in first run}}{\text{no of all MC events}}. \tag{6}$$

Typically, the MC efficiency lies around $0.95 \ldots 0.8$, showing that the IM/MC optimization takes at most 20 per cent more time than merely MC verification†.

This high number is in part due to the effect that organs at risk can be exposed to a higher dose when computed by the MC algorithm which leads to readjustments of the respective constraints. This affects mainly highly sensitive structures such as the eye or the lungs, where electron scatter is usually underestimated by PB algorithms. The IM algorithm in use here employs constrained biological optimization which does not necessitate user-adjusted weights; thus, once a set of constraints has been decided upon, the algorithm can be left unattended during the MC computations. The objective for the target volumes was to minimize clonogenic cell survival under quadratic maximum dose constraints. The optimization was constrained by maximum total organ dose as computed by the Kutcher–Burman DVH reduction technique (Kutcher and Burman 1989) for the organs at risk; the lung model included a dose-volume restriction based on a local dose-response curve for pneumonitis. The optimization is performed with a conjugate gradient algorithm.

The pencil beam algorithm was based on Monte Carlo computed finite-sized kernels. It included a 1D inhomogeneity correction and accounted for lateral electron transport without density correction. This algorithm and the finite-size PB dose computation used are described elsewhere (Alber and Nüsslin 1999, Alber *et al* 2000). Virtually all physical objective functions published so far are convex and hence eligible for an IM/MC optimization.

3. Implementation of EGS4

The MC code EGS4 is well established in radiotherapy physics (Nelson *et al* 1985, Ma *et al* 1996, Nahum 1995). The accelerator head model chosen for this implementation has been shown to perform well (Fippel 1999). It includes a finite sized source as a simplified model of head scatter. The energy spectrum used was for an Elekta SL 20 linac at 6 MV. The inclusion of scatter from the beam modifiers (MLC, compensators) is possible because the leaf trajectories or compensator profiles are known before entering each IM/MC iteration, such as to optimize on the basis of the actually achievable dose. However, the inclusion of beam modifier models is beyond the scope of this paper and will be dealt with in respect of compensators in a forthcoming publication.

The fluence matrices were arranged to form a hexagonal grid of conical finite size pencil beams with isocentre diameters of 5 mm and 3 mm in the stereotactical case. The minimum fluence was set to a fixed percentage of the maximum fluence of the same field to model compensators of 6 cm thickness. The unit fluence was normalized such that it was roughly equivalent to one monitor unit (MU) total fluence (not per fraction). For each fluence element i, a number of particles proportional to the fluence ϕ_i with an even distribution across the hexagon were simulated. The proportionality constant was chosen to be 160 particles per unit fluence cm^{-2} resulting in a standard deviation of roughly one per cent in the target volume.

In the patient geometry, the physical properties were attributed according to tissue type as defined by contours or density. CT numbers (in Hounsfield units) were converted into electron

† According to a tentative estimate, the MC efficiency of a straightforward approach which uses MC throughout the optimization would be around 0.3 to as low as 0.1 for constrained optimization.

densities. Thresholds of the CT numbers were set to distinguish a small number of different media or tissues. For example, simulations in the head distinguished the media brain, soft tissue, bone and air. Voxels outside the brain contour with CT numbers between −1000 and −80 were attributed to air, CT numbers between −79 and +100 were attributed to soft tissue and CT numbers larger than +100 were attributed to bone. The simulation of a lung tumour defined in an Alderson phantom distinguished the materials air, Aldersonlung, Aldersonmuscle and bone as defined in ICRU 37. The CT cube was transformed such that the outermost contour is circumscribed by a grid of $128 \times 128 \times z$ cubic voxels. This procedure results in voxel sizes of about $(3\ \text{mm})^3$ for thorax cases, and less than $(2\ \text{mm})^3$ for head and neck cases.

The electron and photon cut-offs ECUT and PCUT were set to 611 keV and 100 keV, the electron and photon production thresholds AE and AP were set to 521 keV and 10 keV. The value of the parameter ESTEPE (maximum fractional energy a charged particle can lose per step) was set to 4%. These parameters were found by Lovelock *et al* (1995) to be those of the fastest trial. Computations were performed on a 533 MHz DEC Alpha at a speed of about 10^7 histories per hour, which gives total computation times of 13 h and 6 h respectively in the two cases presented in this paper. The lung case was finished in 11 iterations of the IM/MC loop, whereas the optical nerve tumour took 23 iterations.

4. Results

A comparison between a PB algorithm and EGS4 in homogeneous media reveals rather more of the quality of the PB algorithm than of the advantages of IM/MC optimization. The examples were chosen to demonstrate the viability of the method. For treatment sites without low-density interfaces the effects would be much less pronounced, so that an IM/MC algorithm would converge very quickly.

4.1. Lung tumour

In most practical lung cases, the PTV contour lies partially in lung tissue. The dose gradient in these regions is significantly shallower than in water; this implies that with respect to dose homogeneity at the PTV outline, an IM/MC algorithm would have to compensate, and to effect a sharp dose fall-off, penumbra sharpening effects could occur.

The lung tumour (PTV I) in this case was defined in an Alderson phantom and was planned with elective nodal irradiation to the mediastinum (PTV II) (figure 1). Five coplanar photon beams of 6 MV were arranged at (hand-optimized) angles of 0, 170, 195, 220 and 330 degrees gantry angle. The resolution of the dose computation grid was $(3\ \text{mm})^3$, the pencil-beam resolution was 5 mm at isocentre distance. The pencil beams were arranged on a hexagonal grid. The optimization employed biological constraints defined as total organ doses of which only pneumonitis and myelitis as well as the quadratic maximum-dose constraint on the mediastium remained active, i.e. dose limiting.

The PB optimization yielded a homogeneous total volume dose to the PTV I of 77.4 Gy, which was diminished to 73.1 Gy according to the EGS4 verification computation. The subsequent IM/MC optimization improved the result to 75.4 Gy. The constraints were not markedly violated according to the MC verification. The full equivalent dose to the target volume could not be restored since active constraints were present which would have been violated by a further increase in dose. This also rules out a simple rescaling of the dose which would have violated the dose limiting constraints of the organs at risk.

Figure 2 shows the dose–volume histograms (DVHs) of the PTV I for the PB dose, the MC verified dose, and the IM/MC optimized dose. The lung DVH shows an increase at low

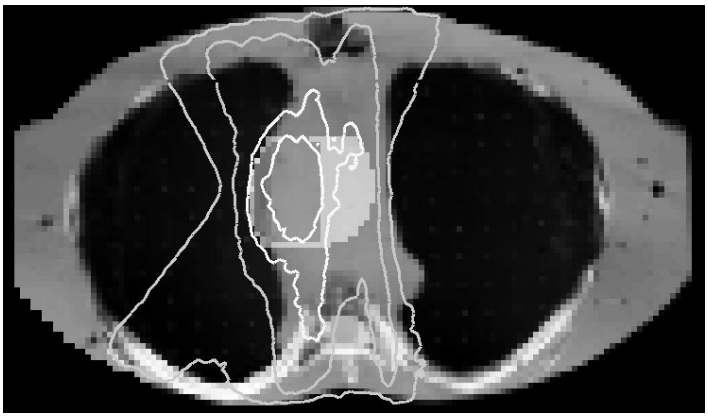

Figure 1. The MC verification of the PB dose distribution of the lung tumour in the lowest slice of the (dark) PTV I which can be seen adjacent to the (brighter) PTV II. The isolines shown are 95, 80, 50 and 20% of the prescription dose of 70 Gy. The 95% isodose line does not lie completely outside the PTV I. Since there are no counteracting constraints, this can be amended by the IM/MC optimization.

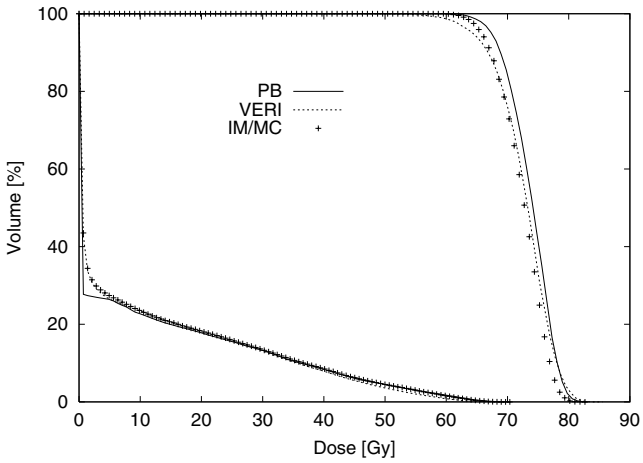

Figure 2. The DVH for the lung case shows a loss of more than 5 Gy to small volumes of the PTV I according to the MC verification (broken curve) of the PB optimized plan (full curve). This loss is partially compensated for by the IM/MC optimization (crosses). The other DVH shown is for the ipsilateral lung with a difference of about ten per cent volume at low dose levels due to the increased penumbra and electron transport in lung.

dose, which is due to electron transport and penumbra broadening. All other DVHs are very similar and therefore omitted.

The fluence profiles in figures 3 and 5 belong to the beams from 0 and 220 degrees after IM/MC optimization. The difference from the PB profiles is displayed in figures 4 and 6. The effects of both penumbra sharpening and scatter compensation can be seen at the cranial and caudal boundary of the PTV I. The same is not observed along the other boundaries since these compensations would cause overdosage to the PTV II. The high peak in front of the profile in figure 6 compensates for the underdosage displayed in figure 1.

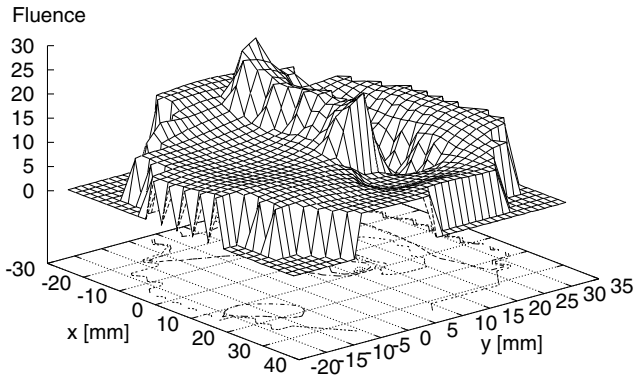

Figure 3. The fluence profile of the 0° field of the lung case after IM/MC optimization. The field is defined on a plane at 500 mm source distance. The fluence is given in arbitrary units. The broad region in the centre shows peaks of penumbra sharpening. The dip which is partially hidden corresponds to the spinal cord.

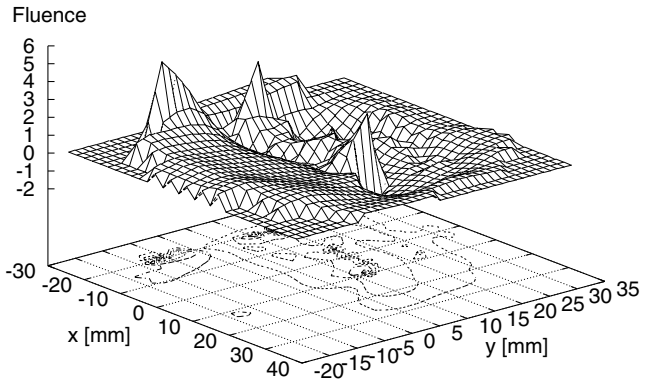

Figure 4. This plot shows the difference between the fluence profile with PB dose computation and IM/MC optimization, figure 3. The fluence has changed markedly at the outline of PTV I, where the algorithm compensates for the broader penumbra due to scatter losses.

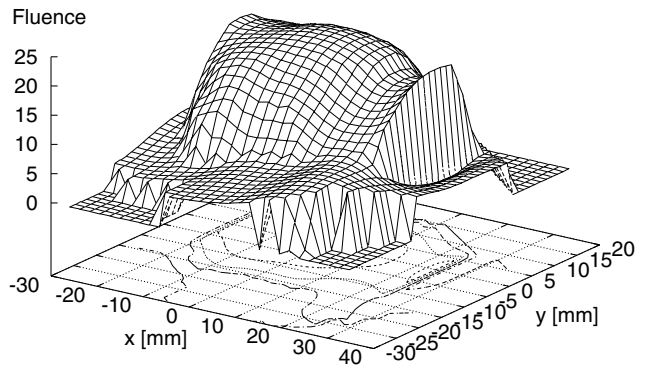

Figure 5. The fluence profile of field 220° field of the lung case after IM/MC optimization. The bulk in the centre corresponds to the PTV I and again penumbra sharpening effects can be seen.

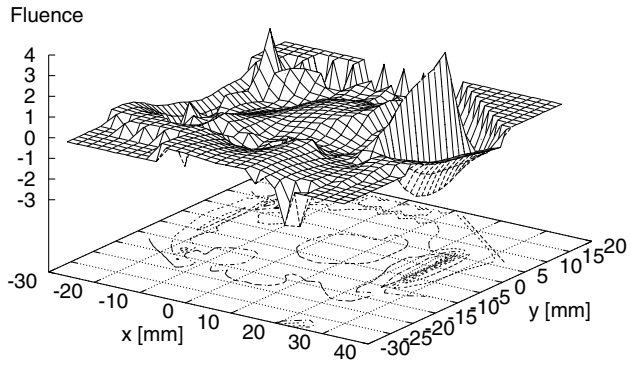

Figure 6. This plot shows the difference between the fluence profile obtained with PB dose computation and IM/MC optimization, figure 5. The pronounced peak on the right corresponds to the underdosed region in the PTV I shown in figure 1.

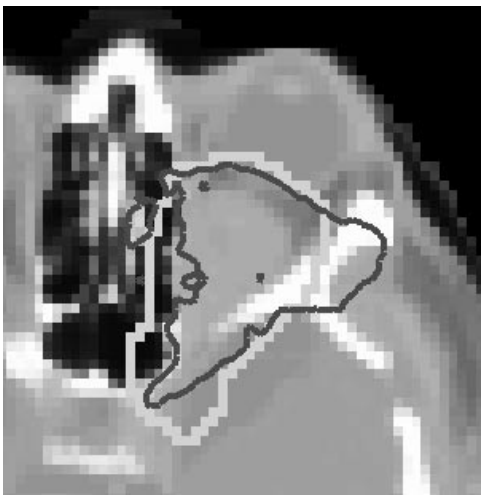

Figure 7. The PTV of the optic nerve tumour can be seen as the bright line, the dark line is the 95% isodose of the result of the PB optimization according to the MC verification computation. The rebuild-up region of the frontal fields extends over about 1 cm into the PTV and makes the 95% isodose recede into the tumour volume. The 95% isodose is normalized to the prescription dose of 70 Gy.

4.2. Orbital tumour

In head and neck treatment cases, often a genuine tumour–air interface can be found which can lead to a dose rebuild-up region in the tumour volume if a beam is chosen at a large angle to this surface. An IM/MC algorithm can take into account these effects.

The case presented here is a tumour of the optical nerve which is adjacent to the sinuses. Four non-coplanar photon beams of 6 MV impinging from directions (gantry, couch): $(-50^\circ, -25^\circ)$, $(-48^\circ, 25^\circ)$, $(70^\circ, 35^\circ)$, $(80^\circ, -25^\circ)$ were chosen, of which three pass partially through cavities in front or next to the PTV. The resolution of the dose computation grid was $(1.75\ \mathrm{mm})^3$, the pencil-beam resolution was $(3\ \mathrm{mm})^2$ at isocentre distance. Again, the pencil beams were arranged on a hexagonal grid. The optimization was subject to biological maximum dose

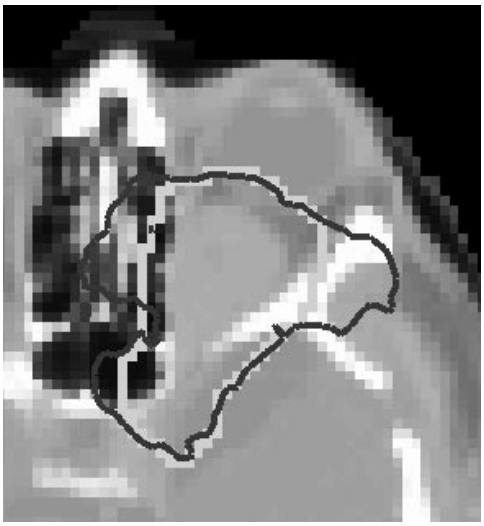

Figure 8. 95% isodose (dark grey) for the same slice as in figure 7 after IM/MC optimization. The algorithm compensated for the scatter losses with the result that the PTV (bright grey) is almost completely within the 95% isodose volume.

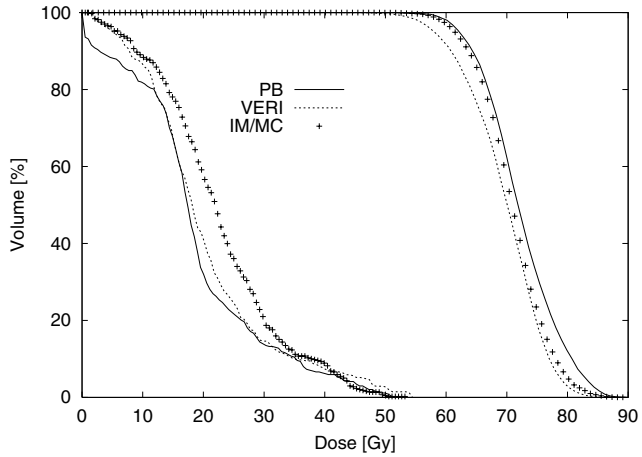

Figure 9. The DVH for the stereotactic case shows a loss of around 5 Gy to approximately 30 per cent of the PTV according to the MC verification (broken curve) of the PB optimized plan (full curve). The other DVH given is for the chiasma, whose dose is slightly higher close to the PTV according to the MC verification than that computed with the PB algorithm, which is a result of underestimated scatter. The IM/MC optimization (crosses) restores the loss to the PTV, but also redistributes the dose to the chiasma. The normalization of the verification dose distribution was independent of the particular case and accurate to within 3% so that the shift of the PTV histogram cannot be attributed to a normalization artefact alone.

constraints including both eyes, the contralateral optic nerve, the chiasma and the brain. A maximum dose constraint on the target volume did not become active, i.e. the dose was too low to have been limited by this constraint.

The PB optimization yielded an homogeneous total volume dose to the PTV of 70.9 Gy, which was diminished to a mere 63.8 Gy according to the EGS4 verification computation

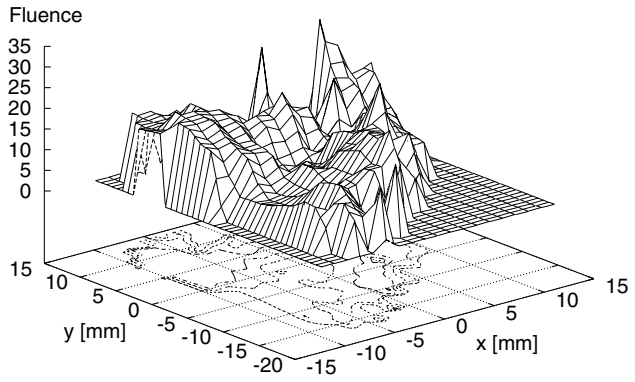

Figure 10. The fluence profile for the field from (−48, 25) degrees for the optical nerve tumour after IM/MC optimization. The field is defined on a plane at 500 mm source distance. The fluence is given in arbitrary units.

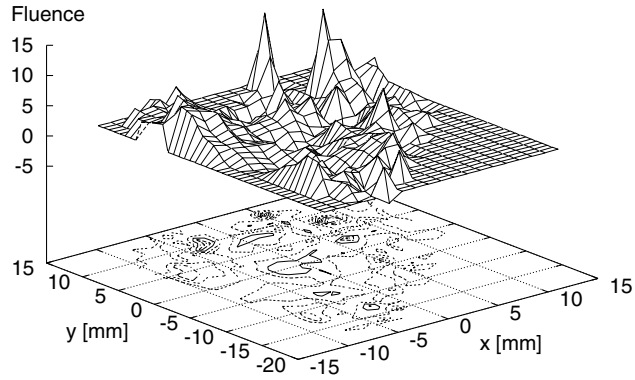

Figure 11. The difference between the fluence profile obtained by PB optimization and the IM/MC optimized fluence profile for the field of figure 10. The fluence peaks in the first quadrant correspond to the frontal sinus, whereas the elevated fluences in the second quadrant correspond to the sphenoidal sinus, both being adjacent to the PTV.

(figure 7). At the same time, the constraint to the chiasma dose was violated. The subsequent IM/MC optimization improved the result to 68.1 Gy (figure 8). Again, active constraints prevented a total restoration of the PB dose.

Figures 7 and 8 show the 95% isodose which was normalized to 70 Gy in a slice where the tumour has intruded into the nasal cavity. The dose rebuild-up of the frontal fields causes an underdosage which spreads about 1 cm from the surface. Figure 9 shows the DVHs of the PTV for the PB dose, the MC verified dose, and the IM/MC optimized dose. Also shown is the set of DVHs for the chiasma, where the effects of the redistribution of fluence from high-dose to low-dose volumes can be seen.

The two fluence profiles shown in figures 10 and 12 and differences from the PB profiles shown in figures 11 and 13 illustrate that to meet the demand for higher primary fluence in some areas, the fluence had to be redistributed. The high-fluence peaks correspond to different sinuses and the nose.

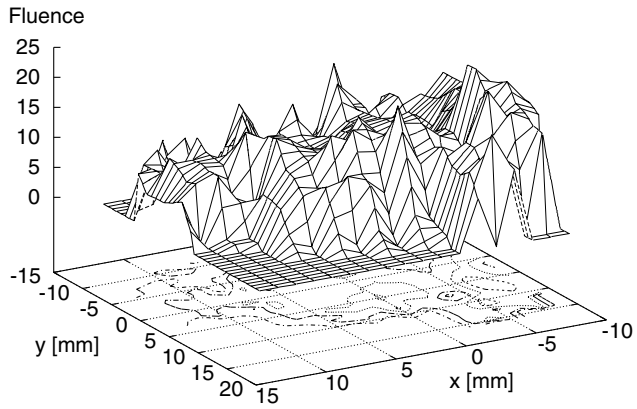

Figure 12. The fluence profile for the field from (70, 35) degrees for the optic nerve tumour after IM/MC optimization.

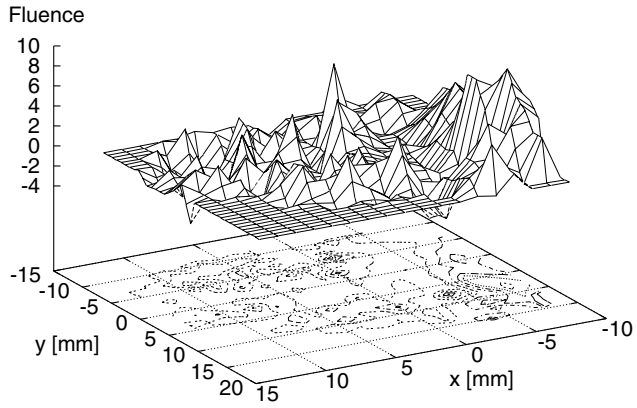

Figure 13. The difference between the fluence profile obtained by PB optimization and the IM/MC optimized fluence profile for the field of figure 12. The increased fluence in the second quadrant compensates partially for the dose rebuild-up area next to the nose as shown in figure 7.

5. Discussion

We demonstrate how an IM algorithm can be combined with an MC algorithm to optimize fluence profiles with the full accuracy of the MC dose computation but with a very small computational overhead. Although the optimization employs a hybrid method to compute the fluence updates during the optimization, the fluence distributions which form the solution of the optimization problem will not be different from those obtained from a purely MC based optimization if certain accuracy conditions are met. A proof of this statement is given in the appendix. It is important to notice that for MC generated fluence updates the same accuracy conditions apply. In order that a full MC optimization acquire a theoretical advantage over the PB/MC hybrid, the relative statistical error must be of the order of the relative error of the PB dose computation at each point of the dose grid. As a consequence more particles may have to be simulated to generate the fluence updates than necessary for the verification of the dose distribution.

The IM/MC optimization is essentially independent of the MC dose computation engine. In this instance, the MC code EGS4 was chosen because of its widely acknowledged accuracy and was linked to a recently developed IM algorithm using constrained biological optimization and a finite size PB algorithm. The computation times did not appear to be prohibitively long; they were in the range of 10 h on a basic DEC Alpha workstation with EGS4. However, since computation times scale with the speed of the MC algorithm, total computation times of about an hour can be reached with an highly efficient algorithm such as XVMC (Fippel 1999, Fippel *et al* 1999) or the use of parallel computers, to a better statistical accuracy than in the presented EGS4 computations.

Naturally, the impact of the MC dose calculation is greatest if the target volume is adjacent to low-density surfaces or is partly composed of low-density subvolumes itself. In the two examples presented, the optimized fluence profiles showed evidence for the compensation of secondary electron disequilibrium by primary fluence. In the case of the lung tumour, the algorithm aims to sharpen the apparent broader penumbra of the fields by lifting the fluence at the field edges. In the case of the optic nerve tumour, the optimization compensates for the genuine loss of dose in the PTV in the rebuild-up areas of the beams and those areas where the target volume and nasal cavities are intricately intertwined.

The objection is often made that MC algorithms need more complex accelerator head models to bring their accuracy to bear than PB algorithms. In the case of IMRT, this turns into a virtue: the complex influence of the beam modifying equipment can be more easily modelled with an MC algorithm than with a PB algorithm. Forthcoming work will be concerned with the inclusion of the intensity-modifier equipment into the IM/MC optimization.

Acknowledgments

We are grateful to Dr M Fippel for his continuous support in all matters and to Dr I Kawrakow for supplying the graphical tools for dose display. This work was in part supported by the DFG. MA and MB acknowledge the support of the Deutsche Krebshilfe e.V.

Appendix

It has to be shown that the sequence of fluences $\Phi^k, \Phi^{k+1}, \ldots$ being generated by the optimization algorithm utilizing T_{PB} for derivatives has a decreasing sequence of objective function values, i.e.

$$f(T_{\mathrm{MC}}\Phi^{k+1}) < f(T_{\mathrm{MC}}\Phi^k) \tag{7}$$

if the relative error of the PB dose calculation is bounded everywhere. In the course of the proof we need to resort to intermediate dose distributions

$$\tilde{D}^{k+1} = D^k + T_{\mathrm{PB}}(\Phi^{k+1} - \Phi^k) \tag{8}$$

$$= D^{k+1} - (T_{\mathrm{MC}} - T_{\mathrm{PB}})(\Phi^{k+1} - \Phi^k) \tag{9}$$

which are the MC dose plus the corrections made in the current iteration as computed by the PB algorithm. It does not matter if these dose distributions are ever calculated during optimization. Since we assume convergence of the algorithm at every iteration, $f(\tilde{D}^{k+1}) < f(D^k)$. If $f(D^{k+1}) \leqslant f(\tilde{D}^{k+1})$ equation (7) is then proved. If the converse is true, we prove equation (7) starting with

$$f(D^{k+1}) - f(\tilde{D}^{k+1}) < f(D^k) - f(\tilde{D}^{k+1}). \tag{10}$$

With equation (8) we obtain for the left-hand side

$$\begin{aligned}\text{l.h.s.} &= f(D^{k+1}) - f(D^{k+1} - (T_{\mathrm{MC}} - T_{\mathrm{PB}})(\Phi^{k+1} - \Phi^k)) \\ &< \partial_D f|_{D^{k+1}}(T_{\mathrm{MC}} - T_{\mathrm{PB}})(\Phi^{k+1} - \Phi^k) \qquad (11)\end{aligned}$$

by first-order expansion, bearing in mind that f is convex. Likewise, we obtain together with equation (7)

$$\text{r.h.s.} > f(D^k) - f(D^{k+1}) \qquad (12)$$

$$> -\partial_D f|_{D^{k+1}} T_{\mathrm{MC}}(\Phi^{k+1} - \Phi^k). \qquad (13)$$

Equations (11) and (12) combine to yield as a necessary condition for equation (7)

$$\partial_D f(D)|_{D^{k+1}}(T_{\mathrm{MC}} - T_{\mathrm{PB}})(\Phi^{k+1} - \Phi^k) \leqslant -\partial_D f(D)|_{D^{k+1}} T_{\mathrm{MC}}(\Phi^{k+1} - \Phi^k) \qquad (14)$$

$$2\partial_D f(D)|_{D^{k+1}}(T_{\mathrm{MC}} - T_{\mathrm{PB}})(\Phi^{k+1} - \Phi^k) \leqslant -\partial_D f(D)|_{D^{k+1}} T_{\mathrm{PB}}(\Phi^{k+1} - \Phi^k). \qquad (15)$$

This condition depends on the fluence update $\Phi^{k+1} - \Phi^k$, which is itself dependent on the algorithm and can basically be an arbitrary vector as long as

$$-\partial_D f(D)|_{D^{k+1}} T_{\mathrm{PB}}(\Phi^{k+1} - \Phi^k) > 0 \qquad (16)$$

is fulfilled. Hence, algorithm-independent convergence can be guaranteed if the condition

$$\tfrac{1}{2}|(\partial_D f(D)|_{D^k} T_{\mathrm{PB}})_i| \geqslant |(\partial_D f(D)|_{D^k}(T_{\mathrm{MC}} - T_{\mathrm{PB}}))_i| \qquad (17)$$

is met.

By the same token, a necessary condition for convergence can be derived by starting with the converse of equation (7) as sufficient condition for divergence. We obtain

$$2|\partial_D f(D)|_{D^k} T_{\mathrm{PB}}(\Phi^{k+1} - \Phi^k)| > |\partial_D f(D)|_{D^k}(T_{\mathrm{MC}} - T_{\mathrm{PB}})(\Phi^{k+1} - \Phi^k)|. \qquad (18)$$

Both conditions are met, if the vectors $(T_{\mathrm{PB}})_i$ and $(T_{\mathrm{MC}} - T_{\mathrm{PB}})_i$ are aligned 'reasonably well', i.e. lie in the hyperplane perpendicular to $\partial_D f$. The progress of the optimization is equivalent to a rotation of the vector $\partial_D f|_{D^k}$ such that it is perpendicular to the vectors $(T_{\mathrm{PB}})_i$. If the algorithm 'PB' uses primary fluence only, a large number of non-zero entries in $(T_{\mathrm{MC}})_{ji}$ will be zero in $(T_{\mathrm{PB}})_{ji}$, in other words, $\partial_D f|_{D^k}$ will be perpendicular to a subspace only, leaving convergence totally dependent on spurious large elements of $\partial_D f|_{D^k}$. By extending the vectors $(T_{\mathrm{PB}})_i$ to include scatter, optimization will take into account the whole vector space, thus reducing the potential error to the order of magnitude $|T_{\mathrm{MC}} - T_{\mathrm{PB}}|_\infty$.

References

Alber M, Birkner M, Laub W and Nüsslin F 2000 Constrained biological optimization of IMRT *Phys. Med. Biol.* submitted

Alber M and Nüsslin F 1999 An objective function for radiation treatment optimization based on local biological measures *Phys. Med. Biol.* **44** 479–93

Bogner L, Scherer J and Herbst M 1999 An inverse Monte Carlo optimization algorithm for conformal radiotherapy *Phys. Med.* **15** 111–19

Fippel M 1999 Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm *Med. Phys.* **26** 1466–75

Fippel M, Laub W, Huber B and Nüsslin F 1999 Experimental investigation of a fast Monte Carlo photon beam dose calculation algorithm *Phys. Med. Biol.* **44** 3039–54

Jeraj R and Keall P 1999 Monte Carlo-based inverse treatment planning *Phys. Med. Biol.* **44** 1885–96

Kutcher G T and Burman C 1989 Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method *Int. J. Radiat. Oncol. Biol. Phys.* **16** 1623–30

Lovelock D M J, Chui C S and Mohan R 1995 A Monte Carlo model of photon beam used in radiation therapy *Med. Phys.* **22** 1387–94

Ma C M, Reckwerdt P, Holmes M and Rogers D W O 1996 DOSXYZ users manual *NRCC Report*

Mohan R 1997 Why Monte Carlo *Proc. Int. Conf. on the Use of Computers in Radiation Therapy* (Madison, WI: Medical Physics Publishing) pp 16–18

Nahum A E 1995 Three-D dose calculation (iii); application of Monte Carlo methods *3D Workshop 95* (Heidelberg: DKFZ) pp 89–97

——1997 Conformal therapy needs monte-carlo dose computation *Proc. Challenges in Conformal Radiotherapy (ESTRO)* pp 1–11

Nelson W R, Hirayama H and Rogers D W O 1985 The EGS4 code system *SLAC Report* 265 (Stanford, CA: Stanford Linear Accelerator Center)

Preiser K, Bortfeld T, Hartwig K, Schlegel W and Stein J 1997 A new program for inverse treatment planning *Proc. 12th ICCR* pp 425–8

Wang X-H, Mohan R, Jackson A, Leibel S A, Fuchs Z and Ling C C 1995 Optimization of intensity-modulated 3d conformal treatment plans based on biological indices *Radiother. Oncol.* **37** 140–52

Xing L, Hamilton R J, Spelbring D, Pelizzari C A, Chen G T Y and Boyer A L 1998 Fast iterative algorithms for three-dimensional inverse treatment planning *Med. Phys.* **25** 1845–9