Towards the elimination of Monte Carlo statistical fluctuation from dose volume histograms for radiotherapy treatment planning

Josep Sempau†‡ and Alex F Bielajew†

† Department of Nuclear Engineering and Radiological Sciences, The University of Michigan, Ann Arbor, Michigan, USA
‡ Institut de Tècniques Energètiques, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Abstract. The Monte Carlo calculation of dose for radiotherapy treatment planning purposes introduces unavoidable statistical noise into the prediction of dose in a given volume element (voxel). When the doses in these voxels are summed to produce dose volume histograms (DVHs), this noise translates into a broadening of differential DVHs and correspondingly flatter DVHs. A brute force approach would entail calculating dose for long periods of time—enough to ensure that the DVHs had converged. In this paper we introduce an approach for deconvolving the statistical noise from DVHs, thereby obtaining estimates for converged DVHs obtained about 100 times faster than the brute force approach described above. There are two important implications of this work: (a) decisions based upon DVHs may be made much more economically using the new approach and (b) inverse treatment planning or optimization methods may employ Monte Carlo dose calculations at *all* stages of the iterative procedure since the prohibitive cost of Monte Carlo calculations at the intermediate calculation steps can be practically eliminated.

(Some figures in this article appear in colour in the electronic version; see www.iop.org)

1. Introduction

With the advent of affordable high-speed computers and the development of faster Monte Carlo based algorithms (Neuenschwander and Born 1992, Neuenschwander *et al* 1995, Keall and Hoban 1996a, b, Kawrakow *et al* 1996, Sempau *et al* 2000), the Monte Carlo (MC) method is becoming an important tool for medical physicists performing photon and electron treatment planning for radiotherapy cancer treatment (Bielajew 1994, 1995, Mohan 1997).

The information provided by a treatment planning system can be represented graphically in a variety of different forms, for example isodose lines and dose volume histograms (DVHs). Given a target volume $\mathcal{V}$ within a patient's body and certain irradiation conditions, the DVH, say $p(x)$, is defined such that $p(x) \, dx$ equals the fraction of $\mathcal{V}$ with a dose in the interval $[x, x+dx]$. Thus, the DVH defines the probability density function (PDF) of finding a dose x in $\mathcal{V}$. The integral DVH

$$F(x) = \int_0^x dx' \, p(x') \tag{1}$$

is also an important representation, since $F(x_2) - F(x_1)$ represents the fraction of $\mathcal{V}$ that receives a dose between x_1 and x_2.

DVHs have been used as a test to compare the accuracy and performance of MC codes (Hartmann-Siantar *et al* 1997). To this end, the patient's geometry, including the target volume, is divided into voxels as provided by CT or MRI scans. Voxels are normally shaped as rectangular boxes with a square base with characteristic lengths of the order of millimetres. As the particle transport simulation is carried out, the dose in each voxel is scored in an accumulator. When a sufficient number of particle histories has been processed, the calculation stops and each voxel is assigned to a dose bin, to which it contributes with a count. The histogram of counts in the different bins corresponds to the discretization of the underlying continuous PDF, $p(x)$. We note that the discretization is performed both in space (voxels) and in dose (bins).

The remainder of the paper is organized as follows. In section 2 we give a general discussion of the statistical noise in the calculated dose in a voxel. In section 2.1 we provide a general proof that the variance of the dose is proportional to the dose and argue that the distribution of dose should be distributed according to a Gaussian. We demonstrate in section 2.2 that calculated DVHs and true (or converged) DVHs, those produced after an infinite number of particle histories have been simulated, are related by a Fredholm equation of the first kind, an integral equation that yields the calculated (finite number of particle histories) DVH after integration over the true DVH and a 'response function', the Gaussian derived in section 2.1. In section 2.3 we test our Gaussian response model and demonstrate its usefulness and in section 2.4 we find practical lower limits of the number of histories for which the model can be employed as well as a determination of the number of histories required to ensure convergence of the calculated DVHs. In section 3 we develop a deconvolution method that allows us to estimate the converged DVH from an approximate DVH obtained with order-of-magnitude fewer particle histories. In section 4 we show how the new method performs on a variety of test cases before the conclusions.

2. Monte Carlo generated statistical noise

2.1. Dose variance in a voxel

Let us suppose, as is usually the case, that MC simulations are carried out with the transport of electrons and positrons turned on, either because they are the primary particles or because they are produced when photons interact with matter. In any case, this implies that all energy depositions are actually made by charged particles. For the sake of brevity, we will refer only to electrons hereafter, since positron transport involves similar methods except for the possibility of annihilation, which does not imply any direct energy deposition.

Various of the currently available MC codes rely on 'mixed' schemes to simulate electron energy losses (Nelson *et al* 1985, Baró *et al* 1995, Sempau *et al* 2000). The basic idea is to separate inelastic interactions into two categories, depending on whether the energy lost is above (hard) or below (soft) some threshold, E_{cut}. Hard events are simulated in a detailed way, that is, they are processed one at a time. Soft events, on the contrary, are simulated in a condensed way, which means that many soft interactions are accumulated in a single fictitious event. Delta rays are generated only by hard events and therefore energy losses larger than E_{cut} involve the generation of secondary particles but no deposition on the spot. This motivates the definition of the restricted stopping power, S_r, or mean energy lost per unit step length for energy losses below E_{cut}. For the majority of applications, including radiotherapy, it is plausible to set E_{cut} equal to the absorption energy, E_{abs}, below which electrons are considered to be absorbed and their kinetic energy deposited locally.

MC calculations provide an exact solution to the Boltzmann transport equation except for the fluctuations inherent in the random nature of the method. These fluctuations may be made arbitrarily small by simulating a sufficiently large number of particle histories. Let x be the dose deposited per primary particle in a given voxel when the number of simulated particle histories is infinite, that is, when there are no statistical fluctuations, and let ΔV and ρ be the voxel volume and average density respectively. After completion of a finite number of histories, N, the average dose $\overline{x}$ defined as the quotient between the MC simulated energy deposition per history, $\Delta E/N$, and the voxel mass, $\rho \Delta V$, is an unbiased estimator of x.

A voxel will be said to have been 'hit' when an electron, either primary or secondary, traverses it. Hereafter, n and ϵ stand for the random variables representing the number of hits produced in a given voxel and the energy deposited by a particle that hits this voxel, respectively. As ΔE and consequently $\overline{x}$ are random variables, they have an associated mean and variance. To investigate these quantities, consider $\overline{x}$ expressed in terms of n and ϵ, namely

$$\overline{x} \equiv \frac{1}{N} \frac{\Delta E}{\rho \Delta V} = \frac{1}{N} \frac{\sum_{i=1}^{n} \epsilon_i}{\rho \Delta V} \tag{2}$$

where ϵ_i is the energy deposited by the ith hit.

For voxel sizes of the order of some millimetres, the probability of observing a multiple hit on a single voxel is small compared with the probability of the voxel receiving a single hit, $\mathcal{P}$. Hence, n follows approximately a binomial distribution with N trials. Moreover, for broad beams, such as those found in radiotherapy applications, $\mathcal{P}$ is small for the majority of voxels and $N\mathcal{P}$ is rather large. These considerations lead to the conclusion that the binomial distribution of n can be approximated by a Gaussian with variance, $\sigma^2(n)$, given by

$$\sigma^2(n) = \langle n \rangle. \tag{3}$$

If n is distributed according to a binomial distribution, it is shown in the appendix that

$$\left\langle \sum_{i=1}^{n} \epsilon_i \right\rangle = \langle n \rangle \langle \epsilon \rangle. \tag{4}$$

That is to say, the average value of the energy deposited in a voxel is given by the product of the average number of hits multiplied by the average energy deposition per hit on the voxel.

This allows equation (2) to be written as

$$\langle \overline{x} \rangle = \frac{\langle n \rangle}{N} \frac{\langle \epsilon \rangle}{\rho \Delta V} \equiv x \tag{5}$$

where we have redefined x to make the notation a little less cumbersome. In the appendix it is also proven that

$$\sigma^2 \left(\sum_{i=1}^{n} \epsilon_i \right) = \langle n \rangle \sigma^2(\epsilon) + \langle \epsilon \rangle^2 \sigma^2(n) \tag{6}$$

indicating that the variance of the Monte Carlo estimated energy deposition in the voxel originates from the two distributed quantities, n and ϵ.

Substituting from equation (3) results in

$$\sigma^2 \left(\sum_{i=1}^{n} \epsilon_i \right) = \langle n \rangle (\sigma^2(\epsilon) + \langle \epsilon \rangle^2) = \langle n \rangle \langle \epsilon^2 \rangle \tag{7}$$

which leads to an expression for the estimated variance of the dose in the voxel

$$\sigma^2(\overline{x}) = \frac{\langle n \rangle \langle \epsilon^2 \rangle}{N^2 \rho^2 \Delta V^2} = \frac{x}{N \Delta V} \frac{\langle \epsilon^2 \rangle}{\rho \langle \epsilon \rangle}. \tag{8}$$
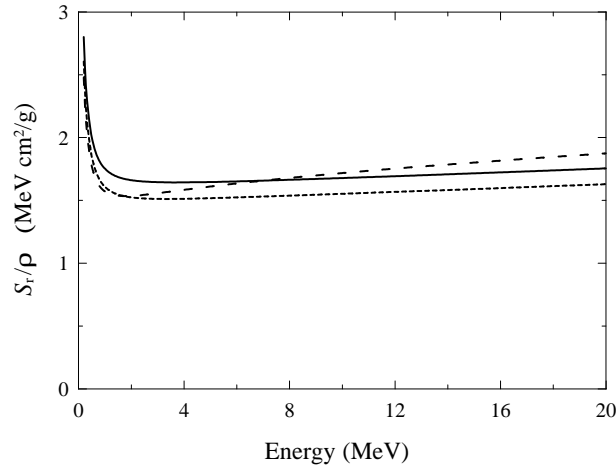
Figure 1. Restricted mass stopping power for water (full curve), bone (short dashed curve) and air (long dashed curve). E_{cut} was set to 200 keV. Data have been obtained from the preparation program associated with the DPM code (Sempau *et al* 2000).

We note that only energy losses due to soft collisions are involved in the distribution of ϵ. The energy straggling due to these soft collisions, which accounts for a dispersion in ϵ when electrons with identical initial conditions travel a given path length, can be neglected for lengths equal to 1 mm or larger. This 'continuous slowing down approximation' (CSDA) has been adopted successfully (Rogers and Bielajew 1990) for many medical physics applications by the EGS4 Monte Carlo code (Nelson *et al* 1985), for *any* path length. Mass restricted stopping powers (S_{r}/ρ) for materials relevant to radiotherapy applications have a very small variation with energy, except at low energies, and are fairly independent of the material as well. This is seen in figure 1. It is plausible to approximate S_{r}/ρ by a constant value for all voxels in the estimate of the variance of the dose to a voxel.

Taking these approximations into account

$$\epsilon = \rho \left(\frac{S_{\text{r}}}{\rho} \right) \ell \tag{9}$$

where ℓ stands for the path length traversed through a voxel. Consequently, equation (8) can be rewritten as

$$\sigma^2(\overline{x}) = \frac{x}{N \Delta V} \left(\frac{S_{\text{r}}}{\rho} \right) \frac{\langle \ell^2 \rangle}{\langle \ell \rangle} = x \frac{(S_{\text{r}}/\rho)\eta}{N \Delta V^{2/3}} \tag{10}$$

where the dimensionless parameter

$$\eta \equiv \frac{1}{\Delta V^{1/3}} \frac{\langle \ell^2 \rangle}{\langle \ell \rangle} \tag{11}$$

can, in principle, change from voxel to voxel depending on the voxel shape and on the degree of isotropy of the radiation field at the voxel location.

For all practical radiotherapy situations, however, η has a very restricted range of variation and can be considered as approximately constant inside the target volume. To justify this assertion it will be considered that, due to the voxel smallness, electrons travel along almost straight lines and that particle flux is homogeneous inside a given voxel. Thus, $\langle \ell^2 \rangle$, $\langle \ell \rangle$ and then η can be readily obtained by means of an MC numerical integration for any radiation

field. Using this approach, we find for monodirectional beams and cubic voxels, that η spans the interval [0.86, 1.00] as the beam direction changes. As any electron radiation field can be thought of as a continuous superposition of monodirectional beams, η is bounded by the same interval for any field.

In practice, collimated radiation coming out of an accelerator head will enter the patient's geometry nearly parallel to a voxel face, say $z = 0$ where the z-axis is directed along the patient's body. This includes the cases in which the direction is normal to $x = 0$ or $y = 0$ (which gives $\eta = 1$), or it goes along the diagonal $(1, 1, 0)$ ($\eta = 0.94$). As electrons penetrate deeper into the patient, the electron angular distribution becomes progressively more isotropic as the kinetic energy decreases and the electrons scatter. For a completely isotropic field, $\eta = 0.90$. It is interesting to note that, in the isotropic case, the mean value of ℓ equals $4V/S$ for *any* region of volume V limited by a convex surface† of area S. When the region is a cube with unit side, $\langle \ell \rangle = 4V/S = 2/3$.

Hence, η is likely to have values close to 1 in the region where the radiation enters the patient's body and progressively decrease until nearly reaching the isotropic value 0.90 at the end of the electron range. Taking into account that target volumes cover only a fraction of the region where radiation is transported, it is clear that this range of variation will usually be further limited. Moreover, the decrease in η may somewhat compensate for the increase in S_r/ρ at low energies, keeping the product even less sensitive to variations across voxels. All these considerations justify the plausibility of the assumption that η may be treated as a constant.

Taking this assumption into account, equation (10) can be recast as

$$\sigma^2(\overline{x}) = x \frac{C}{N} \tag{12}$$

where the parameter

$$C \equiv \frac{(S_r/\rho)\eta}{\Delta V^{2/3}} \tag{13}$$

is a constant for a given target. This equation demonstrates that the estimate of the variance in dose in a voxel is approximately proportional to the dose itself, inversely proportional to the number of histories and inversely proportional to the voxel surface. From the central limit theorem (Lindeberg 1922, Feller 1967) it can be concluded that, when the number of simulated histories is large enough, $\overline{x}$ has a Gaussian distribution. For the sake of notational simplicity, the quantity $\sigma(\overline{x})$ will be denoted by $\sigma(x)$ hereafter.

It is important to notice that when electron transport is turned off so that photon interactions produce energy depositions on the spot, equation (9) loses its validity and the expression for $\sigma(x)$ is no longer applicable. This is due to the fact that, in this case, $\sigma(\epsilon)$ comprises the fluctuations of ℓ *and* the fluctuations of the energy deposited by photons along a given path. The second contribution can be made negligible for photons as well by using interaction forcing or some equivalent method and the validity of the former analysis can be restored. A study of the influence of variance reduction techniques on the present model will not be pursued herein. It suffices to say that equation (12) seems unlikely to hold in general when non-analogue methods are applied unless all particles entering any voxel carry similar statistical weights, which will then appear as a multiplicative factor in equation (12) modifying the value of C.

---

† Every element of surface area must be able to be joined to any other surface element by an uninterrupted straight line.
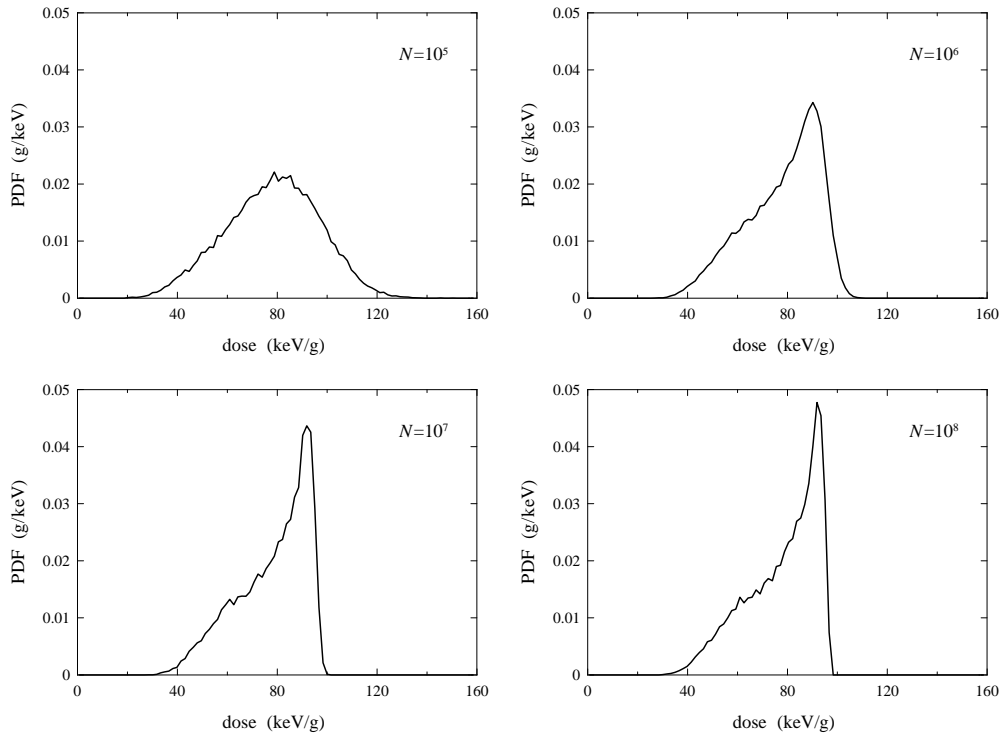
Figure 2. MC simulated DVHs for 10 MeV electrons impinging normally on a water phantom formed by $128 \times 128 \times 128$ 1 mm^3 cubic voxels. The electron field on the water surface was collimated to a 5×5 cm^2 square. The target volume was defined as a field-centred parallelepiped of $4 \times 4 \times 2$ cm^3, with the shortest distance in the direction of the incident beam and located between 2 and 4 cm under the water surface. The dose bin width was roughly 1% of the full scale. Different cases correspond to different numbers of histories N.

2.2. DVH convolution

The statistical 'noise' produced by the fluctuation of $\overline{x}$ about its mean can have a profound effect on the DVH. An ideal calculation of the dose in each voxel would reproduce its true value, say y, and the corresponding true DVH, which will be represented by p. The MC simulation represents such an ideal only when the number of histories, N, tends to infinity. For finite N, the calculated dose x fluctuates about its mean y with the variance σ^2 given by equation (12). Consequently, the MC calculated DVH, $\tilde{p}$, may differ from p.

Let us define a response function $g(x; y)$ such that $g(x; y) \, dx$ gives the probability that a voxel with a true dose y yields a measure in the interval $[x, x + dx]$ for a given value of N. The MC calculated fraction of voxels with a dose in this same interval is $\tilde{p}(x) \, dx$. This fraction includes contributions from voxels that have a true dose y not in the interval under consideration, but that have been associated with it by virtue of the statistical fluctuations. With this interpretation of g, $\tilde{p}$ and p are related by

$$\tilde{p}(x) = \int_{\mathcal{D}} dy \, p(y) g(x; y) \tag{14}$$

with $\mathcal{D}$ representing the domain where $p(y) \neq 0$. Considering g and $\tilde{p}$ as known quantities, the former is a Fredholm equation of the first kind (Arfken and Weber 1995) in the unknown p.
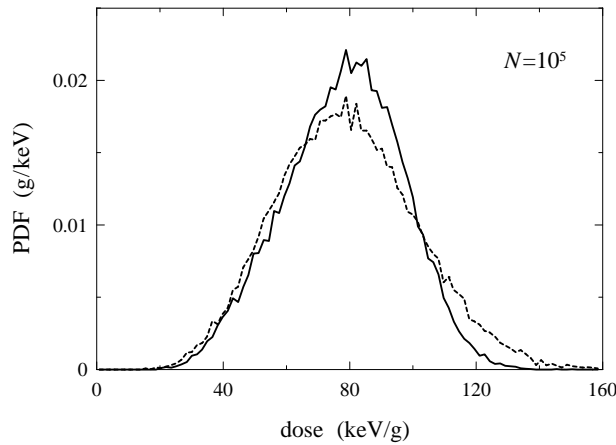
Figure 3. DVHs obtained with DPM (full curve) and Penelope (dashed curve) for 10^5 histories. Source, target and phantom are the same as in figure 2.

Alternatively, $g(x; y)$ can be viewed as the MC calculated $\tilde{p}$ obtained when the underlying p corresponds to a delta function centred at y, that is, when the doses deposited in the different voxels of the target volume have exactly the same expected value y. According to equation (12), this implies that all the voxels have nearly the same deposited dose PDF, a Gaussian centred at y with a variance given by that equation. Therefore, from this point of view it is clear that $g(x; y)$ is also represented by this same Gaussian, namely

$$g(x; y) = \frac{1}{\sigma(y)\sqrt{2\pi}} \exp\left(-\frac{(x-y)^2}{2\sigma^2(y)}\right) = \sqrt{\frac{N}{2\pi Cy}} \exp\left(-\frac{N(x-y)^2}{2yC}\right). \tag{15}$$

As N tends to infinity, σ tends to zero, $g(x; y)$ collapses into a delta function, $\delta(x - y)$, and $\tilde{p}$ converges to p, which will be called the converged DVH hereafter. For values of y much larger than C/N, $g(x; y)$ can be approximated by a function of $(x - y)$ only and the Fredholm equation (14) becomes a familiar convolution integral. Although in practical situations this limit may not be attained, the term convolution will be retained to designate the process that transforms p into $\tilde{p}$.

The effect of this convolution is to make $\tilde{p}$ wider than p. Since the variance decreases with increasing N, this widening is expected to be significant at relatively small values of N. In figure 2 a comparison of DVHs obtained for the same conditions but with different number of histories presented. Even for the same value of N, different mechanisms of simulating soft energy losses can produce different values of C and thus yield apparently different DVHs. For instance, in homogeneous media, electron steps can be taken independently of voxel boundaries and soft energy losses produced along a given step length can be deposited at some point chosen randomly on the electron path. This means that all the energy lost in a step is deposited in only one voxel, irrespective of the number of them that have been traversed. (This is the current scheme adopted in the Penelope code.) Alternatively, the particle can be stopped every time it crosses a voxel boundary, in which case some energy deposition is obtained every time a voxel is hit. (This is the current scheme adopted in the DPM code.) Given a sufficient number of histories, both methods produce the correct mean value, but their variances are different. The Penelope scheme produces a larger variance. Our analysis strictly applies only to the DPM approach. However, when DVH convergence is employed to compare the reliability of MC codes the subtleties must be properly taken into account.

Figure 3 presents an example of this danger. Two different MC codes, DPM and Penelope, are used to calculate a DVH in water. These codes have shown very good agreement in calculations of depth and lateral dose in water phantoms. Surprisingly enough, the two DVHs present significant differences for the same number of histories. The reason is that, in this particular case, Penelope's energy-deposition model is less efficient than DPM's.

It is important to realize that the effect of the convolution is essentially different from the customary MC uncertainty that appears in the form of error bars, e.g. in the calculation of depth dose profiles. In this case, uncertainties produce only fluctuations of the MC calculated data points in the direction of the ordinate axis and therefore they can be denoted with *vertical* error bars. In contrast, the convolution of DVHs is a process in which the uncertainty of the dose in a voxel affects the abscissa value that will be assigned to it. Hence, this uncertainty is *horizontal*.

The convolution of DVHs is a phenomenon somewhat analogous to the response of a scintillation detector, where the spectrum of the deposited energy is convolved with a response function, analogous to the action of the integral in equation (14), to produce the spectrum actually measured. In this case, the fluctuations in the light yield and the response of the photoelectric tube are responsible for the nearly Gaussian shape of the response function, which also has a width roughly proportional to the deposited energy. In contrast to the scintillator case, however, the width of the response function can be changed in the process of calculating a DVH by changing the number of histories simulated or the voxel dimensions.

2.3. Testing the model

In this section we present some results intended to clarify the meaning of the model introduced above and verify its accuracy. All of them have been obtained with a new MC code called DPM (Sempau *et al* 2000), an acronym for Dosimetry Planning Method. DPM is based on the physics contained in EGS4 plus some newer developments intended to improve electron multiple scattering and thus speed up the simulations considerably. It is optimized for radiotherapy applications, and in the tests performed so far it has been found to agree to with about 3% with other well-known standard codes. However, the inaccuracies that DPM or any other MC code may have are irrelevant here, for what is being investigated refers to the basic nature of the MC method.

In order to verify the accuracy of the approximations leading to equation (12), the variance of the deposited dose, x, has been calculated using MC methods for each voxel. As the abscissa, or dose axis, x, of the DVH is divided into a number of equispaced bins, a mean variance for each bin can be obtained by averaging this quantity over all voxels falling within the bin. Fitting a straight line to the data points representing σ^2 versus x yields a value of the constant C and the the validity of the linear fit can be judged.

The variance of each individual voxel, σ^2, can be estimated using

$$\sigma^2 \simeq \frac{1}{N} \left[\frac{\sum_{i=1}^{N} x_i^2}{N} - \left(\frac{\sum_{i=1}^{N} x_i}{N} \right)^2 \right] \tag{16}$$

where x_i represents the dose scored in the voxel during the simulation of the ith history and N is the number of histories simulated. In principle, to carry out the sums in the numerators, three different counters, say `xtmp`, `xscore` and `x2score` must be defined. `xtmp` accumulates the individual energy depositions produced by all the particles of only one shower. Once the shower is completed, its value is added to `xscore`, which contains $\sum x_i$, and its square is added to `x2score`, which represents $\sum x_i^2$. Then `xtmp` is set to zero before a new shower starts. Although this simple procedure represents an insignificant computational cost for the
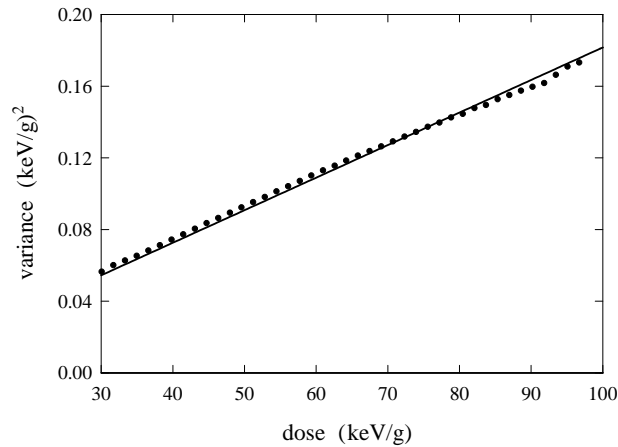
Figure 4. The variance in dose for the same arrangement as in figure 2 and $N = 10^8$. Dots represent 'experimental' MC data and the full line is a linear fit of the type $\sigma^2 = xC/N$. The value of the fitting parameter C and the correlation coefficient are 181.6 MeV g^{-1} and $r = 0.9998$ respectively.

majority of MC calculations, this is not true for radiotherapy applications since the number of counters is three for each voxel, and the total number may amount to many millions. Accessing the computer memory after each history can slow down the execution of the code rendering the calculation infeasible.

To overcome this difficulty, we employed the following method†. We define a counter, say `ihist`, that identifies the current shower history. The novelty of the method consists in using an integer label, say `last`, for each voxel in addition to the three counters described above. `last` identifies the last shower that contributed to the voxel by storing its `ihist` value. Every time there is a deposition ϵ in a given voxel, the current `ihist` is compared with `last`. If they match, the deposition is added to the temporary counter `xtmp`. If they differ, `xtmp` is added to `xscore`, its square is added to `x2score`, the temporary counter `xtmp` is set to ϵ and `last` is updated to the current `ihist`. This counting mechanism ensures that every time a shower contributes to a voxel, the contribution from any former shower has been stored correctly. The simulation time is not severely affected, because the new method involves only an additional `if` every time a deposition is made, plus an additional array to store the integer label `last` for each voxel.

A calculation of variance is presented as a function of the dose x in figure 4. Our model, equation (12), predicts a linear variation, in good agreement with what it is observed.

Since the Gaussian shape of g is related to the validity of equation (12), which states that all voxels with the same dose have the same variance, a test of the model consists in calculating the dispersion of this variance. Suppose that K of the voxels in the target volume happen to be in a certain dose bin. Considering their variances σ_k^2 with $k = 1 \ldots K$ as a sample of some population, the quantity

$$s^2(\{\sigma_{k=1\ldots K}^2\}) = \frac{K}{K-1}\left[\frac{\sum \sigma_k^4}{K} - \left(\frac{\sum \sigma_k^2}{K}\right)^2\right] \tag{17}$$

is a measure of the dispersion in the variance.

† We are indebted to Professor Francesc Salvat of the University of Barcelona for supplying us with the idea for this method.
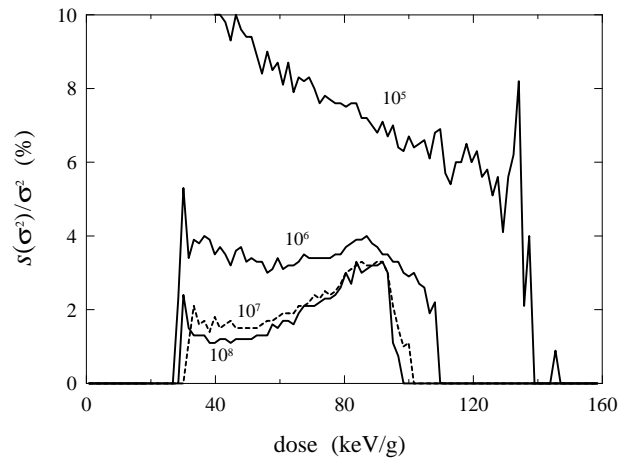
Figure 5. Relative dispersion of the variance for different number of histories, N.
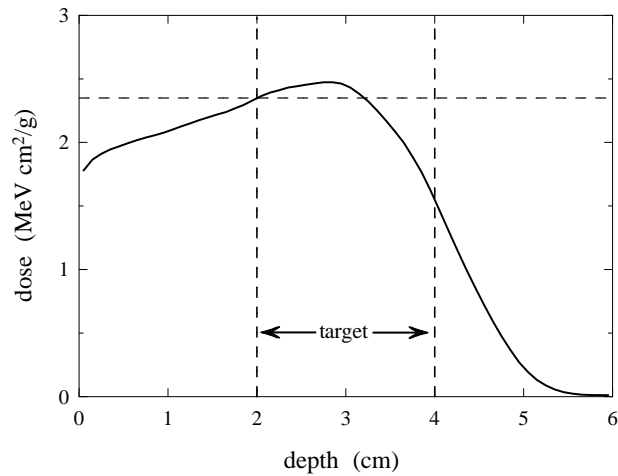

Figure 6. The axial depth-dose profile for the same problem conditions as in figure 2. The two vertical dashed lines indicate the location of the target volume and the horizontal dashed line sets the limit that divides the target area in two regions, one with a one-to-one correspondence between depth and dose (below the horizontal dashed line) and the other with a two-to-one correspondence (above the horizontal dashed line).

Figure 5 presents the dispersion of the variance as a function of the dose. The small relative dispersion seen validates, at least for this particular arrangement, the premise upon which equation (15) was derived.

Note that as N increases the relative dispersion stops decreasing indicating that the DVH has converged and hence, the fluctuations become too small to move voxels to neighbouring dose bins. The bump observed at high doses in the cases 10^7 and 10^8 for a dose around 85 keV g^{-1} can also be understood as a consequence of the model. Figure 6 shows the depth dose curve for the problem under consideration. For doses below a certain limit, the correspondence between dose and depth is such that to each dose there corresponds only one depth. Above this limit, however, every dose corresponds to two possible depths, one deeper than the point where the maximum occurs and the other at a shallower depth. These two regions
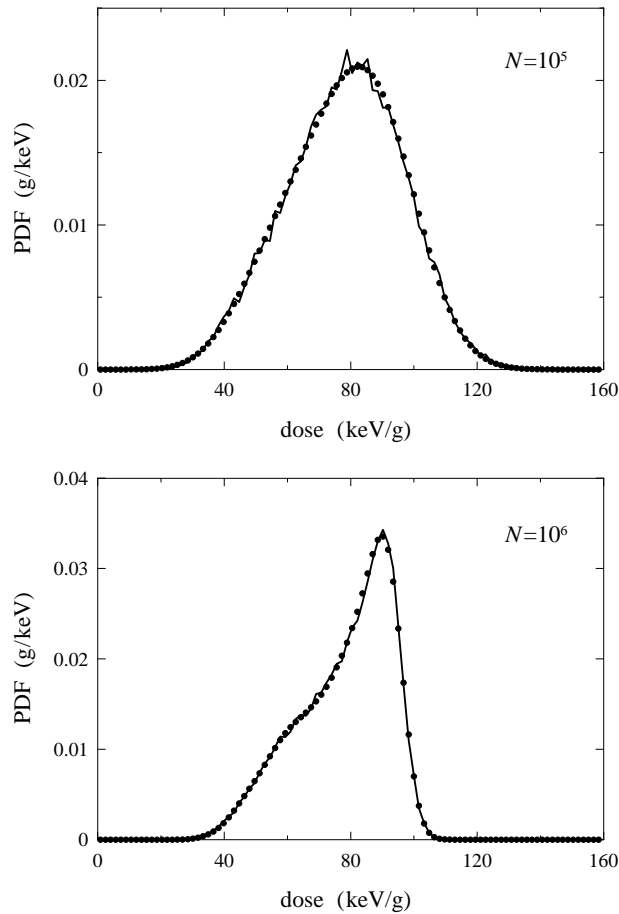
Figure 7. Comparison between a raw MC DVH (full curve) and the convolved DVH (dots) obtained using equation (14) for the problem considered in figure 2 and two different values of N.

have different radiation fields, the deeper one being more isotropic and with lower average kinetic energies. Hence, the variance of the set of these two types of voxels will show larger dispersions than the variance corresponding to doses where only one kind of dose-to-depth correspondence participates. However, even in the dual case, the relative dispersion remains quite low—about 3%. This same duality could also be the cause of the slight departure of the MC calculated variance from the linear fit, due to the fact that at depths shallower than the position of the depth-dose maximum, the lesser isotropy implies less variance, as reflected in the high dose region of figure 4.

A numerical calculation of the convolution integral in equation (14) taking a known converged DVH (p) as input data can be easily done. The comparison of the convolved DVH obtained in this way with a raw MC-calculated DVH provides a more direct test of the model. Figure 7 shows the result of this test. The agreement is excellent.

2.4. Convergence criteria

We may use the previous analysis to define robust DVH convergence criteria. A naive way to decide when the DVH has converged is to simulate an increasingly larger number of histories,

$N_1 < N_2 < \ldots$ up to a point where the variation of p in every bin becomes smaller than a predetermined value δ. For an analytic function $f(z)$ with $z \equiv 1/N$, this procedure can be justified by considering its truncated Taylor expansion around $z_\infty = 0$

$$f(z_i) \simeq f(0) + f'(0)z_i \qquad i = 1, 2, \ldots. \tag{18}$$

Consider now $z_{i+1} = z_i - \Delta z_i$. If

$$|f(z_{i+1}) - f(z_i)| = |f'(0)|\Delta z_i < \delta \tag{19}$$

then

$$|f(z_{i+1}) - f(0)| = |f'(0)|z_{i+1} < \delta \frac{z_{i+1}}{\Delta z_i} = \delta \frac{N_i}{N_{i+1} - N_i} \tag{20}$$

which gives an upper limit for the error $|f(z_{i+1}) - f(0)|$. For instance, if the number of histories is doubled, $N_{i+1} = 2N_i$ with N_i large, and a variation in the quantity of interest f of less than 1% is obtained, then the error associated with $f(N_{i+1})$ can also be expected to be less than 1%.

A different approach consists in assessing the convergence by analysing the dose variance associated with the voxels. Convergence is reached when a large proportion of voxels have a σ much smaller than the bin width, Δx, even for the largest dose participating in the DVH, x_{max}. This method is somewhat superior to the naive scheme introduced above, because it does not rely on the small magnitude of the (unknown) higher-order derivatives of the DVH as N tends to infinity and also because the number of histories needed to reach convergence can be readily predicted even before starting up the simulation. Indeed, since the variation of the variance is known to behave as $1/N$ for large N, from equation (12) we have

$$\sigma_{max}^2 = x_{max} \frac{C}{N} \ll \Delta x^2 \Rightarrow N \gg x_{max} \frac{C}{\Delta x^2} \equiv N_{lim}. \tag{21}$$

In order to ensure convergence, a number of histories N_{conv} larger than N_{lim} should be simulated, say $N_{conv} = 10 \times N_{lim}$. For the example shown in figure 2, this recipe gives $N_{conv} = 10^8$. It is advisable to validate this prediction with a direct evaluation of the variance σ_{max}^2 while the simulation is in progress.

A convenient measure of the degree of convergence is given by the dimensionless parameter

$$\alpha \equiv \frac{N \Delta x}{C}. \tag{22}$$

Substituting N with N_{conv} results in

$$\alpha_{conv} \equiv N_{conv} \frac{\Delta x}{C} = 10 \frac{C x_{max}}{\Delta x^2} \frac{\Delta x}{C} = 10 \frac{x_{max}}{\Delta x} = 10I \tag{23}$$

where I stands for the number of bins in which the DVH abscissa has been divided. For the example of figure 2, $\alpha_{conv} = 10^3$.

A lower limit to the applicability of the analysis presented in this work can be established in terms of α. When the number of histories N is small, the distribution of the dose in a given voxel is not Gaussian, and our model fails to reproduce the observed results. This failure is first noted in the low-dose region of the DVH, where our Gaussian model would predict standard deviations larger than the dose itself, giving rise to unphysical negative doses. A practical limit can be then found by imposing

$$\sigma(\Delta x) < \Delta x \tag{24}$$

which means that the standard deviation in the first bin has to be smaller than its width. This condition yields

$$\Delta x \frac{C}{N} < \Delta x^2 \qquad \Rightarrow \qquad 1 < \frac{N \Delta x}{C} = \alpha \tag{25}$$

which shows that $\alpha = 1$ is the minimum value of this parameter that needs to be considered. In the example shown in figure 2 this limit implies $N > 10^5$.

3. DVH deconvolution

3.1. Discretized Fredholm equation

Let us consider a discretized version of the Fredholm equation (14). Suppose that the DVH abscissa is divided into I dose bins. Let $P = \{p_i\}_{i=1\ldots I}$ be the discretized DVH array defined by

$$p_i \equiv \frac{1}{\Delta x} \int_{x_i \pm \Delta x/2} \mathrm{d}x \, p(x) \tag{26}$$

where the subscript $x_i \pm \Delta x/2$ is shorthand for the interval $[x_i - \Delta x/2, x_i + \Delta x/2]$. Analogously, the convolved DVH can be discretized to give the array $\tilde{P}$, whose components are defined by

$$\tilde{p}_i = \frac{1}{\Delta x} \int_{x_i \pm \Delta x/2} \mathrm{d}x \, \tilde{p}(x) = \frac{1}{\Delta x} \int_{\mathcal{D}} \mathrm{d}y \, p(y) \int_{x_i \pm \Delta x/2} \mathrm{d}x \, g(x; y). \tag{27}$$

Substituting p by its discretized representation, P, which by definition is constant over each interval $x_i \pm \Delta x/2$, yields

$$\tilde{p}_i \simeq \sum_j g_{ij} p_j \qquad \text{or} \qquad \tilde{P} = G \cdot P. \tag{28}$$

The $I \times I$ matrix G has components

$$g_{ij} = \frac{1}{\Delta x} \int_{x_i \pm \Delta x/2} \mathrm{d}x \int_{y_j \pm \Delta x/2} \mathrm{d}y \, g(x; y) \tag{29}$$

where

$$x_i \equiv \Delta x (i - 0.5) \qquad \text{and} \qquad y_j \equiv \Delta x (j - 0.5) \tag{30}$$

are the centres of the corresponding bins.

Applying the change of variables $x = x' \Delta x$, $y = y' \Delta y$ transforms equation (29) into the more compact form

$$g_{ij}(\alpha) = \sqrt{\alpha} \int_{[i-1,i]} \mathrm{d}x' \int_{[j-1,j]} \mathrm{d}y' \exp\left(-\alpha \frac{(x' - y')^2}{2y'} \right) \tag{31}$$

where α stands for the convergence parameter defined before. For $\alpha \to \infty$ the term g_{ij} tends to δ_{ij}, the Kronecker delta symbol, and $P' = P$. Notice that, apart from the number of bins I, α is the only quantity needed to define G and, consequently, the degree of convergence reached.

It is worth noticing that the discretized $\tilde{p}$, equation (27), is correctly normalized to 1 provided that the largest dose present in the calculated DVHs, $x_{\max}$, is large enough to ensure that the integral of $g(x; y)$ over the interval $[0, x_{\max}]$ is also normalized for all values of y in the domain $\mathcal{D}$ where $p(y) \neq 0$. Indeed, in this case

$$\Delta x \sum_i \tilde{p}_i = \int_0^{x_{\max}} \mathrm{d}x \, \tilde{p} = \int_{\mathcal{D}} \mathrm{d}y \, p(y) \int_0^{x_{\max}} \mathrm{d}x \, g(x; y) = \int_{\mathcal{D}} \mathrm{d}y \, p(y) = 1. \tag{32}$$

Assuming that $\mathcal{D}$ is fully contained in $[0, x_{\max}]$, the former condition is fulfilled for values of α not too small, for then the corresponding Gaussian does not significantly extend its tail beyond $\mathcal{D}$.
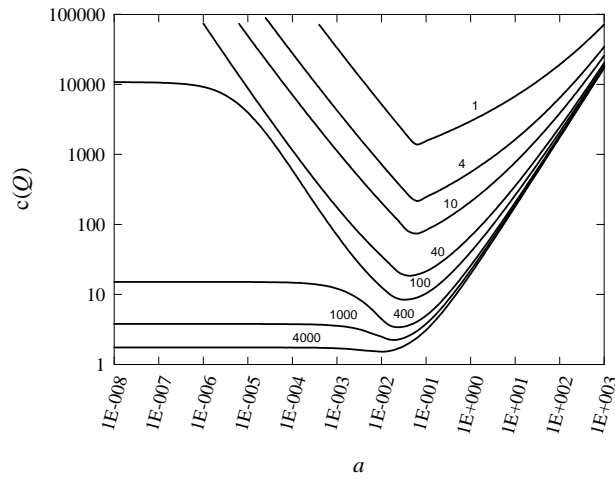
Figure 8. Condition number c(Q) as a function of *a* for selected α values.
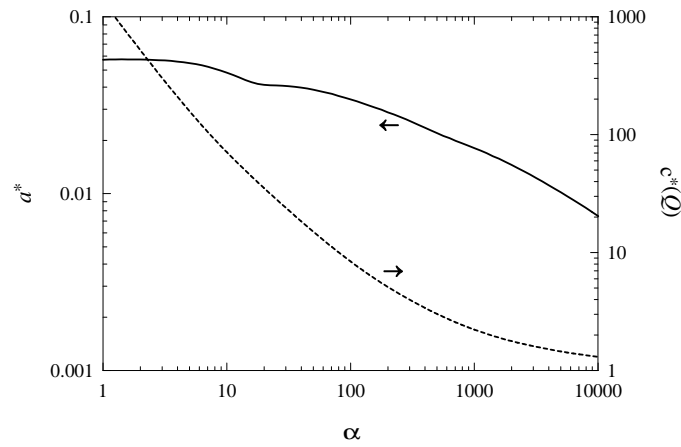

Figure 9. Value of a^* as a function of α (full curve) and the corresponding minimum condition number $c^*(Q)$ (dashed curve) when the DVH contains 100 dose bins.

3.2. Inversion of the Fredholm equation

In principle, the converged DVH can be obtained from a MC simulation with *any* finite value of N by inverting the Fredholm equation (28), which in turn involves the inversion of the matrix G. Unfortunately, the solution P obtained in this way is extremely sensitive to small variations in $\tilde{P}$ except when convergence has been almost reached. Hence, the numerical process is unstable for values of α smaller than α_{conv}, given by equation (23).

In order to characterize this ill-posed problem it is convenient to introduce the 'condition number' of a linear operator A, $c(A)$, defined as

$$c(A) = \|A\| \, \|A^{-1}\| \tag{33}$$

where $\|A\|$ stands for the spectral norm of A (Lewis 1991), which measures how much A stretches or shrinks the length of a unitary vector $\hat{v}$. It can be seen that if $c(A) \gg 1$ then large variations in $\hat{v}$ can produce very small variations in its image, and therefore the inversion of A is an ill-posed problem. Actually, when A is singular $c(A) \equiv +\infty$.
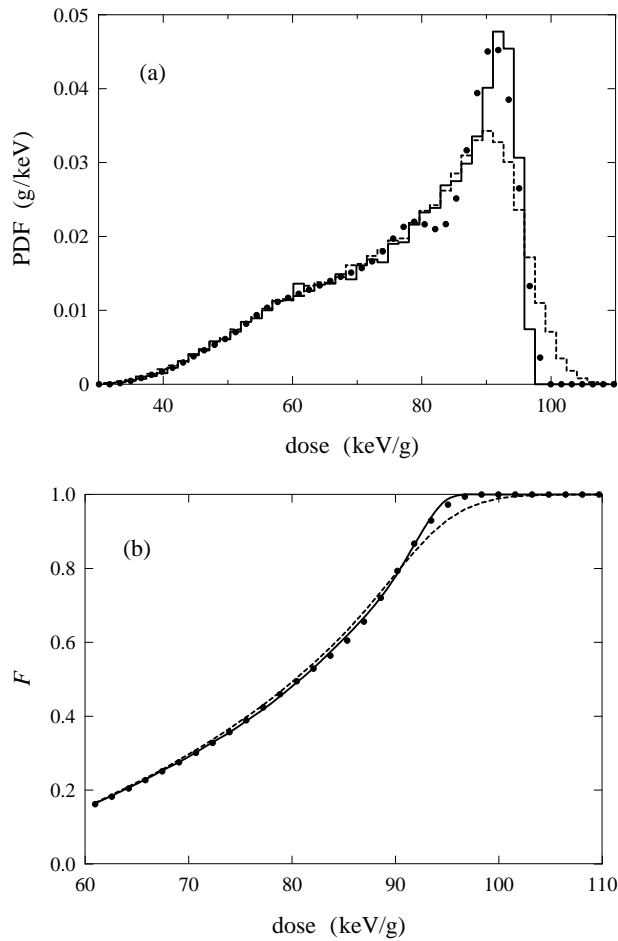
Figure 10. Comparison between the converged (full curve), deconvolved (dots) and raw $N = 10^6$ MC calculation (dashed curve) results for the water phantom problem described in figure 2. DVHs correspond to figure labelled (*a*) and integral DVHs (F) to (*b*). The MC measured value of C used in the deconvolution process was 181.3 MeV g^{-1}.

The analysis of ill-conditioned matrices can be carried out with the aid of the Singular Value Decomposition theorem, or SVD (Press *et al* 1992). SVD guarantees that any square matrix G has a decomposition of the form $G = TWS$, where T and S are square and unitary and W is square and diagonal. The elements in the diagonal of W are called singular values, or SVs. SVs can always be chosen to be positive or zero. If there is a null SV, G is singular. It can be shown that, since T and S preserve the Euclidean norm, the condition number $c(G)$ is the ratio between the largest and the smallest SVs.

For small and moderate values of α, the condition number of the matrix G with elements defined by equation (29) is very large. Indeed, the convolution process, except when the variance is small, or equivalently, when α is large, smears out the spikes present in the original data and therefore, if consecutive components of $\hat{v}$ oscillate around zero, cancellations produce an image with norm close to 0, which gives rise to a large value of $\|G^{-1}\|$. As the norm of G is normally close to 1, $c(G)$ is also very large (see equation (33)), that is, G is ill-conditioned.
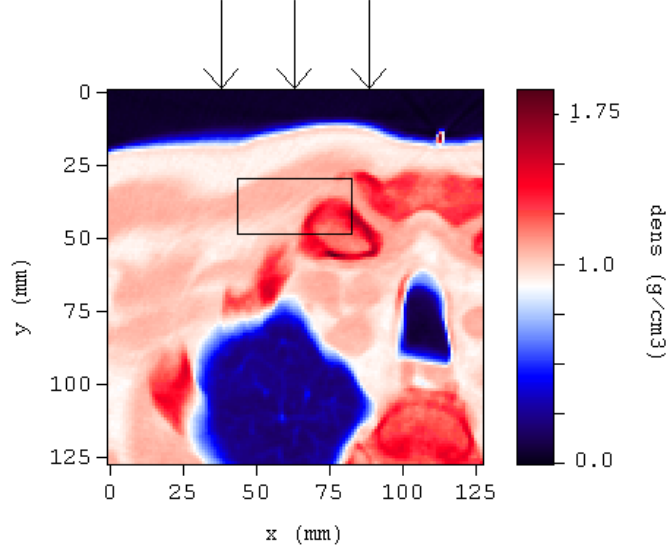
Figure 11. CT slice with a rectangular cut of the target volume to show its location. Densities in the target range from 0.9 up to 1.6 g cm^{-3}. The incoming 5×5 cm^2 normal electron field has an energy of 10 MeV and is bounded to the region indicated by the arrows.

The instability of the solution found by inverting equation (28) shows up in the form of very large oscillations in the deconvolved DVH, which completely mask the true solution. The fact that this unwanted noise has an oscillatory behaviour suggests a different approach, namely, an approximation p to the converged DVH can be found as the outcome of a minimization problem. The quantity to minimize is

$$\phi \equiv \int \mathrm{d}x \left[(\tilde{p} - Gp)^2 + a\Delta x^4 \left(\frac{\mathrm{d}^2 p}{\mathrm{d}x^2} \right)^2 \right] \tag{34}$$

where a is a dimensionless parameter. For $a = 0$ the solution is $p = G^{-1}\tilde{p}$ and the original ill-conditioned inversion problem is recovered. For values of $a > 0$, the participation of the second derivative of p in ϕ penalizes highly oscillating solutions, the penalty being somewhat modulated by the value chosen for a.

The discretization of equation (34) yields

$$\phi = (\tilde{P} - GP)^t (\tilde{P} - GP) + a\Delta x^4 P^t D^t D P = \tfrac{1}{2} P^t Q P + R^t P + \tilde{P}^t \tilde{P} \tag{35}$$

where D is the discrete second derivative operator

$$D = \frac{1}{\Delta x^2} \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \ldots \\ 1 & -2 & 1 & 0 & 0 & \ldots \\ 0 & 1 & -2 & 1 & 0 & 0 & \ldots \\ & & & \ldots \\ & \ldots & 0 & 0 & 1 & -2 & 1 & 0 \\ & & \ldots & 0 & 0 & 1 & -2 & 1 \\ & & \ldots & 0 & 0 & 1 & -2 & 1 \end{pmatrix} \tag{36}$$

which is singular, and

$$R = -2G^t \tilde{P} \qquad \text{and} \qquad Q = 2(G^t G + a\Delta x^4 D^t D). \tag{37}$$

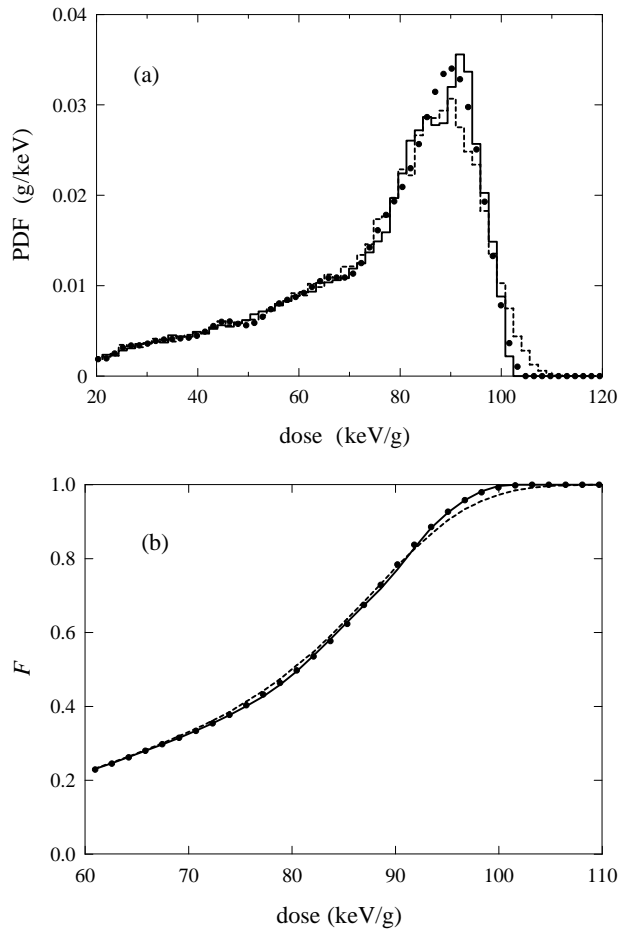Note that Q is symmetrical.

Figure 12. Comparison between the converged (full curve), deconvolved (dots) and raw MC calculation with $N = 10^6$ (dashed curve) for the CT geometry shown in figure 11. DVHs correspond to figure labelled (*a*) and integral DVHs (*F*) to (*b*). The MC measured *C* was 176.3 MeV g^{-1}.

Setting the derivative of equation (35) to zero yields the exact solution

$$P = -Q^{-1}R. \tag{38}$$

The combination of *D* and *G* produces a quadratic form *Q* which, for suitable values of *a*, is no longer ill-conditioned and the inversion process becomes stable. A drawback of this approach is that the resulting *P* is not guaranteed to be a valid PDF function, that is, the conditions

$$p_i \geqslant 0 \qquad \text{and} \qquad \Delta x \sum p_i = 1 \tag{39}$$

may not be fulfilled. A direct application of equation (38) shows that negative values and unnormalized DVHs do appear, although the violation is, generally speaking, small. On the other hand, it would be desirable to include in the calculation the constraints represented by equation (39), for this information could probably improve the final solution by forbidding small oscillations of *P* around zero.

The minimization of the quadratic ϕ with the constraints in equation (39) renders the problem to a form that is usually referred to as quadratic programming (Press *et al* 1992, NAG: Numerical Algorithms Group, Inc.), or QP. Solving QP problems is a routine task that
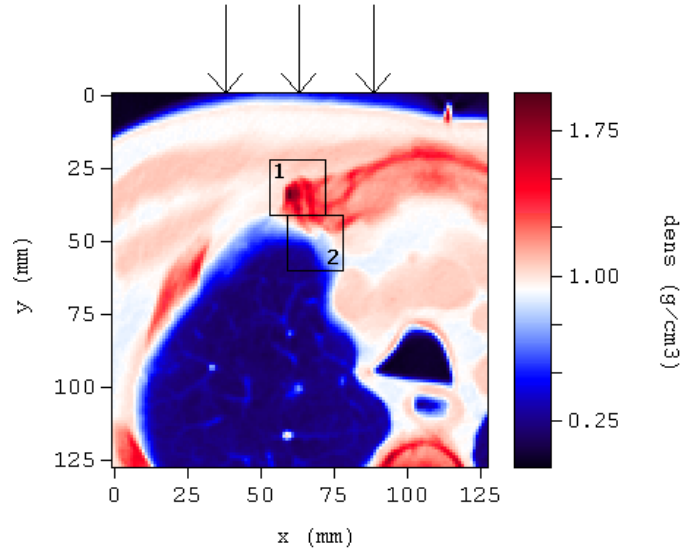
Figure 13. CT slice with two target volumes superimposed. Densities in target 1 span the interval [0.85,1.94] g cm^{-3} and in target 2 [0.20, 1.40] g cm^{-3}. Electrons impinged normally on a 5×5 cm^2 square limited by the arrows. Their initial energy was 18 MeV.

can be performed with the aid of iterative algorithms contained in standard numerical libraries (NAG 1996).

3.3. Optimizing the deconvolution

To completely define Q, the value of a must be chosen. Intuitively, a good a should produce a low condition number for Q, indicating the stability of the solution. To test this hypothesis, the variation of c(Q), which is a function of α and a, has been studied and is represented in figure 8. It is found that for each value of α there is a value of a, named a^*, for which the condition number is at its minimum, c*(Q). The function $a^*(\alpha)$ is represented in figure 9.

   Generally speaking, the deconvolution process turns out to work well with $a = a^*$, provided that c*(Q) does not exceed a few hundred. According to figure 9, this implies that the deconvolution process can be initiated when a value of α of some tens (for DVHs with 100 bins) has been reached. Notice that the relation between α and a^* does not depend on the target volume or on the irradiation conditions. A table relating α with a^* can therefore be calculated beforehand and stored for later use.

4. Results

In the examples that follow, the DVH and its associated integral DVH, F, are studied. To this end, a discretized approximation to F given by

$$F(x) = \int_0^x \mathrm{d}x' \, p(x') \simeq \Delta x \sum_{i=1}^{(x/\Delta x)+1} p_i \tag{40}$$

is used. The maximum error made when comparing a certain approximated F with the converged value will be taken as a figure of merit of the agreement between the two curves.
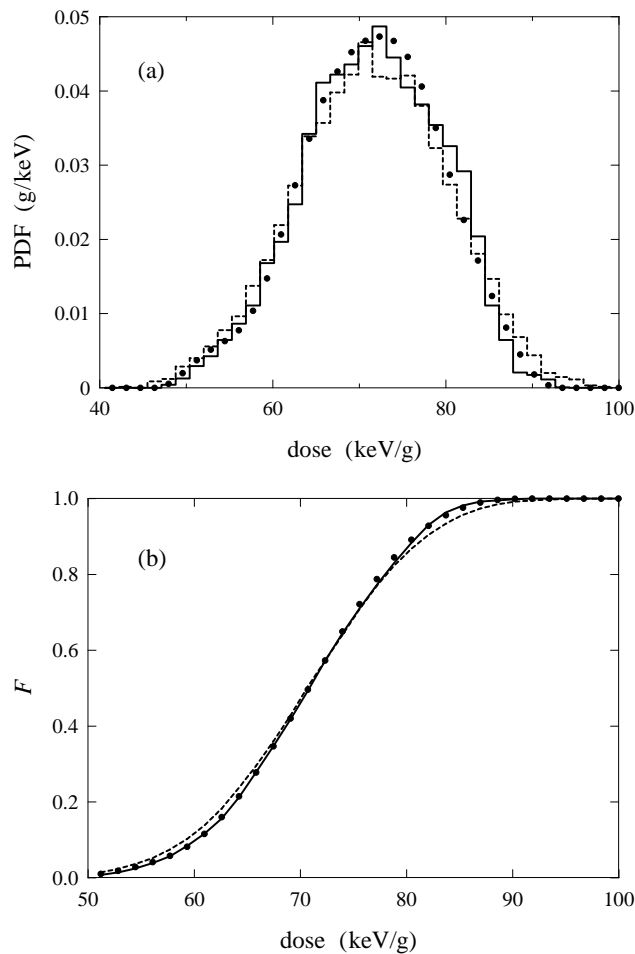
Figure 14. Converged (full curve), deconvolved (dots) and raw $N = 10^6$ MC (dashed curve) results for target 2 in the CT geometry shown in figure 13. DVHs correspond to the figure labelled (*a*) and integral DVHs (*F*) to (*b*).

In figure 10 the deconvolution of the $N = 10^6$ MC calculation for the water phantom case described in figure 2 is presented. The maximum error in *F* was reduced from 0.056 for the raw MC simulation with $N = 10^6$, to 0.015 for the deconvolved result.

In the following examples the deconvolution analysis is applied to real CT data. These data correspond to a 12.8^3 cm^3 cube, which was extracted from the CT scans of a patient, divided into 1 mm^3 voxels. In each case, a cut of the full 3D geometry is presented to show the location of the target volume and the incident beam. For the sake of simplicity, only three materials were defined—air, water and bone. Thus, lung and soft tissue are considered as water with variable density. Hounsfield numbers determined which material and density were assigned to each voxel.

The first example corresponds to figure 11. The target volume covers a region encompassing soft tissue and bone, thus presenting an additional difficulty of mixing materials with significantly different average atomic numbers. The deconvolution process yielded the results shown in figure 12. The maximum error in *F* was reduced from 0.026 for the $N = 10^6$
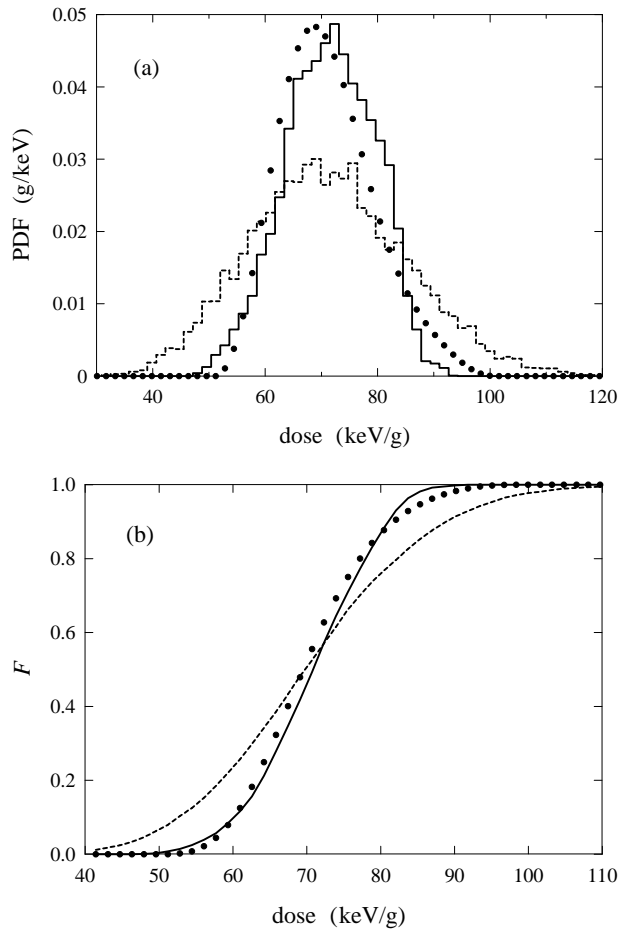
Figure 15. Converged (full curve), deconvolved (dots) and raw $N = 10^5$ MC (dashed curve) results for target 2 in the CT geometry shown in figure 13. DVHs correspond to the figure labelled (*a*) and integral DVHs (*F*) to (*b*).

MC to 0.013 for the deconvolved DVH. The enhanced flux of backscattered electrons in the tissue located upstream the bone structure contained in the target gives rise to an increase in the dose in that region. As a result, the DVH shows a shoulder next to the main peak, which the deconvolved DVH cannot detect from the input $N = 10^6$ simulation data. Despite this fact, the integral DVH is considerably improved by the deconvolution algorithm. It is worth noticing that the MC measured value of C is now slightly lower than in the pure water phantom case presented before. This is probably due to the smaller mass restricted stopping power of the bone (see figure 1 and equation (13)).

The next example is presented in figure 13. The target volume 2 now covers bone, tissue and lung, thus including large variations in both atomic number and density. The result of the application of the deconvolution is shown in figure 14. The maximum error in F was reduced from 0.030 to 0.015. Notice that since soft tissue is downstream of the bone, the DVH does not clearly show a two-peaked structure as before. It is interesting to repeat the deconvolution of the latter case but now taking as input data the $N = 10^5$ raw MC data. The result is shown in figure 15. The maximum error in F was reduced from 0.142 for the MC to 0.062.
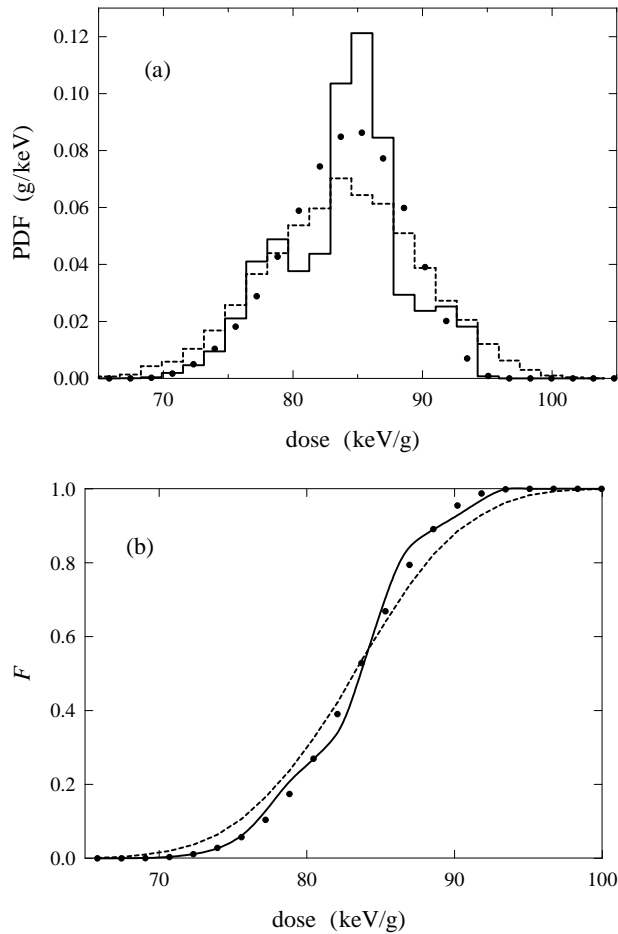
Figure 16. Converged (full curve), deconvolved (dots) and raw $N = 10^6$ MC (dashed curve) results for target 1 in the CT geometry shown in figure 13. DVHs correspond to the figure labelled (*a*) and integral DVHs (F) to (*b*).

As the deconvolution process penalizes high values of the second derivative, it is clear that close sharp peaks will not be reproduced accurately. Integral DVHs, however, vary in a much smoother way and, furthermore, cannot exhibit peaks since F is a monotonically increasing function by definition, thus being much less sensitive to this problem. The following cases exemplify this observation.

Let us consider target volume 1 in figure 13. As illustrated in figure 16, the deconvolved DVH fails to 'detect' the presence of the two small peaks accompanying the central one. Despite this fact, the associated integral DVH still closely matches the converged F, with a reduction in the maximum error from 0.103 for the raw MC to 0.051.

It is worth emphasizing that this loss of accuracy is due to the numerical difficulties involved in carrying out the deconvolution algorithm (minimization of equation (34)) and not to the failure of the convolution model, which is still in very good agreement with the observations. Figure 17 proves this assertion. It represents the application of equation (14) in the forward direction, that is, taking the converged DVH as the input data and comparing the resulting $\tilde{p}$ for $N = 10^6$ with the MC result for the same number of histories.
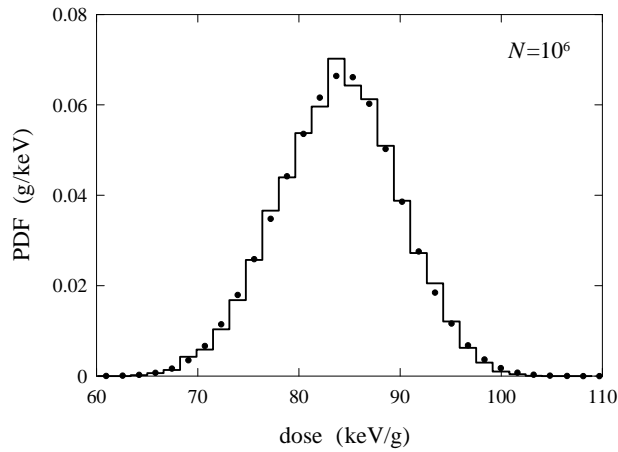
Figure 17. Comparison between the MC measured DVH for $N = 10^6$ (full curve) and the convolved DVH (dots) for target 1 in the CT shown in figure 13.
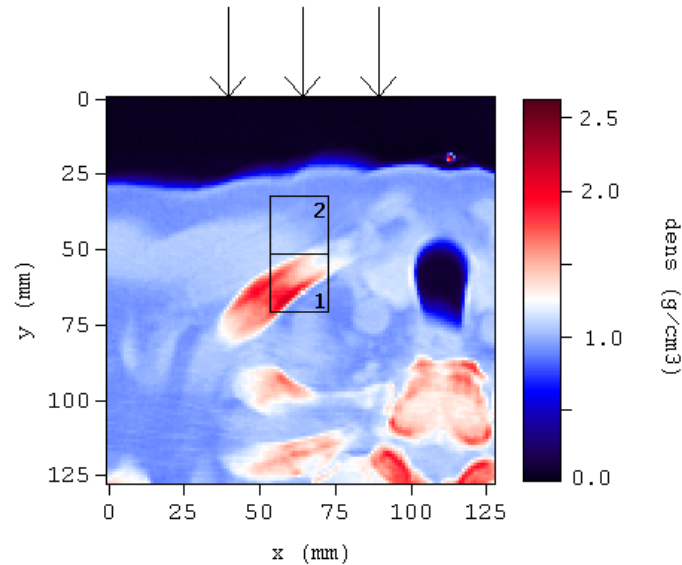

Figure 18. CT slice with two target volumes. Densities in target 1 cover the range from 0.89 to 2.12 g cm^{-3} and in target 2 from 0.93 to 1.39 g cm^{-3}. Electrons enter the geometry normally through the 5×5 cm^2 square limited by the arrows. Their initial energy is 15 ± 3 MeV, Gaussian distributed.

The CT data presented in figure 18 show another example of this loss of accuracy. Figure 19 presents the comparison of DVHs for target volume 1. The deconvolution algorithm is now capable of identifying the presence of two peaks, but reproduces their heights only approximately. Despite this fact, the improvement in the integral DVH is still noticeable, the maximum error being reduced from 0.027 for the raw MC to 0.019.

The same problem is also present to a lesser extend in target 2. Figure 20 presents the corresponding results. In this case, the maximum error in F was reduced from 0.073 for the raw MC, to 0.026 for the deconvolved DVH.
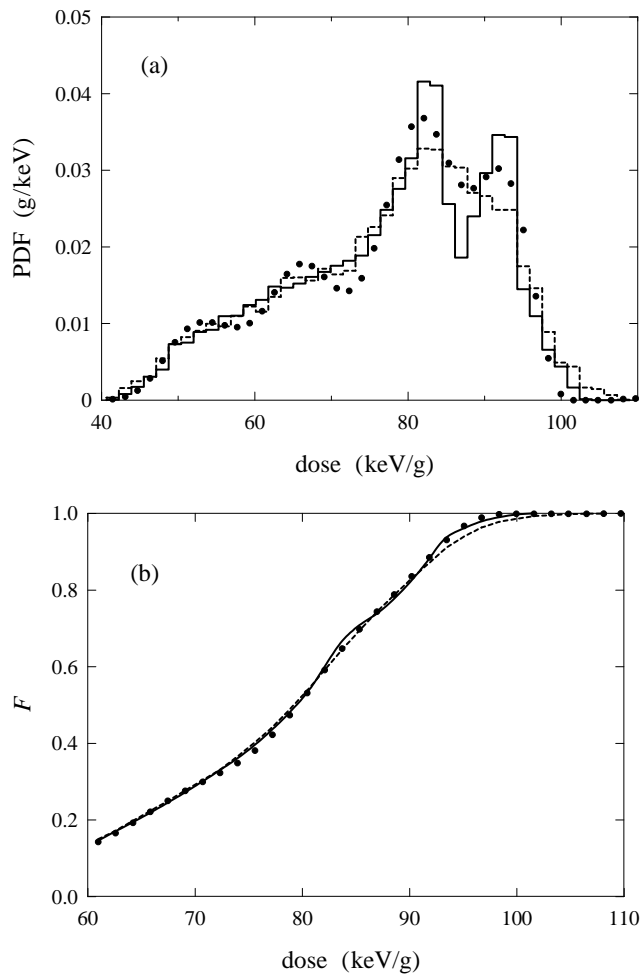
Figure 19. Converged (full curve), deconvolved (dots) and raw $N = 10^6$ MC calculation (dashed curve) results for target volume 1 defined in figure 18. DVHs correspond to the figure labelled (*a*) and integral DVHs (*F*) to (*b*).

5. Conclusions

The process of selecting of a radiotherapy treatment involves the examination of many configurations before one is finally chosen. Although MC calculations, the most accurate method known so far to obtain dose distributions, are starting to become fast enough for employment in routine work, the excessive time that high accuracy calculations demand will still limit the study of possible treatments to a few when an extensive search for an optimal solution is undertaken.

The deconvolution procedure presented in this work remedies this problem. Approximate DVHs and their corresponding integrals can be obtained in a time orders of magnitude smaller than that of an equivalent MC simulation and hence a larger number of cases can be investigated. This information can be employed to decide which treatments deserve a more extensive and accurate MC analysis.
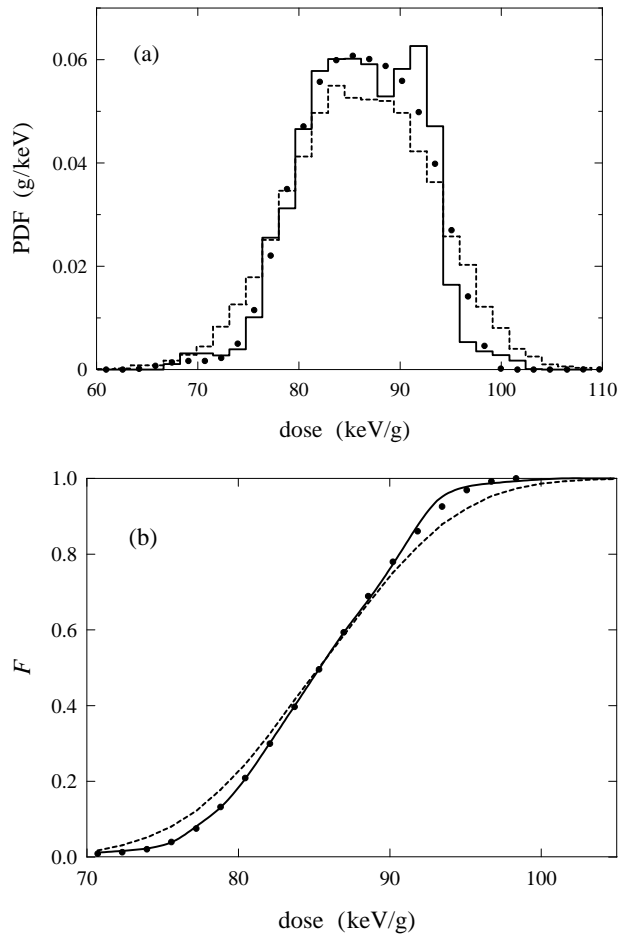
Figure 20. Converged (full curve), deconvolved (dots) and raw $N = 10^6$ MC calculation (dashed curve) results for target volume 2 defined in figure 18. DVHs are shown in (*a*) and integral DVHs in (*b*).

The proposed method can be summarized in the following steps:

(a) The parameter a^* as a function of α (figures 8 and 9) is calculated and stored in tabular form for a mesh dense enough to allow numerical interpolation. This process can take some time—of the order of 1 min on a 300 MHz UltraSparcII Sun workstation employing 100 dose bins. However, for a given number of DVH bins it has to be done only once and for all.

(b) The number N of histories to simulate, or equivalently α (equation (22)), are set such that the minimum condition number $c^*(Q)$, Q being the matrix defined by equation (37), does not exceed some 300 units. For histograms with 100 bins, a convenient value is $\alpha = 10$ (see figure 9), so that

$$N = 10 \times \frac{C}{\Delta x}. \tag{41}$$

Notice that the constant C (see equation (13)), roughly equal to some 180 MeV g^{-1}, may vary slightly from case to case. It is then advisable to obtain an 'experimental' value of

C for the problem in hand by fitting the variance $\sigma^2(x)$ to a straight line, as expressed in equation (12).

(c) The deconvolution is carried out by minimizing the quadratic in equation (35) subject to the constraints in equation (39). The outcome, P, is a medium-accuracy approximation to the converged DVH. Both P and its associated integral DVH, can be used to assess the relevance of the treatment under consideration, allowing a fast selection of the best candidates.

In the examples presented above, the converged α_{conv} was estimated to be 10^3 if DVHs were to be obtained for 100 bins. On the other hand, the deconvolution is completed in a negligible amount of time when compared with a full MC simulation. Hence, if the deconvolved result is accepted as a good approximation to the converged DVH, the proposed procedure has reduced the calculation time by a factor of the order of 100.

Alternatively, the former procedure can be used repeatedly for an increasingly larger number of histories until the *deconvolved* DVH or, preferably, its integral F, has converged in the sense expressed by equation (20). This approach ensures a high-accuracy result at the expense of reducing the improvement in computational efficiency.

Acknowledgments

Appendix: the variance of a binomial sum

Let n be a binomial random variate (RV) corresponding to N trials with a probability $\mathcal{P}$ of having a 'hit' in each trial. Its mean and variance can be written as

$$\langle n \rangle = N\mathcal{P} \qquad \sigma^2(n) = N\mathcal{P}(1 - \mathcal{P}) = \langle n \rangle(1 - \mathcal{P}) \tag{42}$$

respectively. Let q be another RV with mean and variance

$$\langle q \rangle = \int \mathrm{d}q \, p(q)q \qquad \text{and} \qquad \sigma^2(q) = \langle q^2 \rangle - \langle q \rangle^2 = \int \mathrm{d}q \, p(q)q^2 - \langle q \rangle^2 \tag{43}$$

respectively where $p(q)$ represents the probability density function (PDF) of q.

Now let Q be the RV defined by

$$Q = \sum_{i=1}^{n} q_i \tag{44}$$

where q_i is the value obtained from the ith sampling of q. This is equivalent to defining Q as the outcome of the following process:

(a) An RV is sampled that yields 1 with probability $\mathcal{P}$ and 0 with probability $1 - \mathcal{P}$.

(b) If it yields 1, q is sampled and the result added to a counter, Q. If it is 0, no action is taken.

(c) If the process has been iterated fewer than N times, start again.

(d) Deliver Q.

Q can also be expressed as

$$Q = \sum_{i=1}^{N} q_i' \tag{45}$$

where q' is a new RV with PDF p' given by

$$p'(q') = \mathcal{P} p(q') + (1 - \mathcal{P})\delta(q') \tag{46}$$

with δ being the Dirac delta PDF. Note that q' is nil with probability $1 - \mathcal{P}$ and is equivalent to q otherwise. The form of equation (45) is preferable because it contains a fixed summation limit, which simplifies the analysis of its mean and variance.

From equation (46) it follows that

$$\langle q' \rangle = \int dq' \, p'(q')q' = \mathcal{P}\langle q \rangle \tag{47}$$

and

$$\langle q'^2 \rangle = \int dq' \, p'(q')q'^2 = \mathcal{P}\langle q^2 \rangle \tag{48}$$

where use has been made of equation (43).

Hence, according to equation (45) the mean value of Q can be expressed as

$$\langle Q \rangle = N\langle q' \rangle = N\mathcal{P}\langle q \rangle = \langle n \rangle\langle q \rangle \tag{49}$$

and its variance as

$$\begin{aligned} \sigma^2(Q) = N\sigma^2(q') = N\mathcal{P}[\langle q^2 \rangle - \mathcal{P}\langle q \rangle^2] &= \langle n \rangle[\sigma^2(q) + (1 - \mathcal{P})\langle q \rangle^2] \\ &= \langle n \rangle\sigma^2(q) + \sigma^2(n)\langle q \rangle^2 \end{aligned} \tag{50}$$

where equation (42) has been employed.

Equations (49) and (50) can be interpreted as follows. The expected value of the sum (Q) of a 'binomial' number (n) of samples of a RV (q) is the same as if n were set equal to its mean value. However, the variance of Q has two terms. The first one corresponds to the variance that we would get if q were sampled a *fixed* number of times equal to $\langle n \rangle$. The second term is then the contribution to the variance due to the fact that n is not fixed but fluctuates around its expected value. This last term equals the variance of n times the expected value of q squared.

References

Arfken G B and Weber H J 1995 *Mathematical Methods for Physicists* 4th edn (San Diego: Academic)

Baró J, Sempau J, Fernández-Varea J M and Salvat F 1995 PENELOPE: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter *Nucl. Instrum. Methods* B **100** 31–46

Bielajew A F 1994 Monte Carlo modeling in external electron-beam radiotherapy—why leave it to chance? *Proc. 11th Conf. on the Use of Computers in Radiotherapy* (Madison, WI: Medical Physics Publishing) pp 2–5

——1995 EGS4 timing benchmark results: why Monte Carlo is a viable option for radiotherapy treatment planning *Proc. Int. Conf. on Mathematics and Computations, Reactor Physics, and Environmental Analyses* (La Grange Park, IL: American Nuclear Society Press) pp 831–7

Feller W 1967 *An Introduction to Probability Theory and its Applications* vol I, 3rd edn (New York: Wiley)

Hartmann-Siantar C L *et al* 1997 Lawrence Livermore National Laboratory's PEREGRINE project *Proc. 12th Conf. on the Use of Computers in Radiotherapy* (Madison, WI: Medical Physics Publishing) pp 19–22

Kawrakow I, Fippel M and Friedrich K 1996 3D electron dose calculation using a voxel based Monte Carlo algorithm *Med. Phys.* **23** 445–57

Keall P J and Hoban P W 1996a Super-Monte Carlo: a 3-D electron beam dose calculation algorithm *Med. Phys.* **23** 2023–34

——1996b Superposition dose calculation incorporating Monte Carlo generated electron track kernels *Med. Phys.* **23** 479–85

Lewis D W 1991 *Matrix Theory* (Singapore: World Scientific)

Lindeberg J W 1922 Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitrechnung *Math. Z.* **15** 211–25

Mohan R 1997 Why Monte Carlo? *Proc. 12th Conf. on the Use of Computers in Radiotherapy* (Madison, WI: Medical Physics Publishing) pp 16–18

NAG: Numerical Algorithms Group 1996 *Software: The NAG Fortran Library* version 17

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

Neuenschwander H and Born E J 1992 A Macro Monte Carlo method for electron beam dose calculations *Phys. Med. Biol.* **37** 107–25

Neuenschwander H, Mackie T R and Reckwerdt P J 1995 MMC—a high-performance Monte Carlo code for electron beam treatment planning *Phys. Med. Biol.* **40** 543–74

Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in Fortran* 2nd edn (New York: Cambridge University Press)

Rogers D W O and Bielajew A F 1990 Monte Carlo techniques of electron and photon transport for radiation dosimetry *The Dosimetry of Ionizing Radiation* vol 3, ed K Kase, B Bjärngard and F Attix (New York: Academic) pp 427–539

Sempau J, Wilderman S J and Bielajew A F 2000 DPM, a fast, accurate Monte Carlo code optimized for photon and electron radiotherapy treatment planning dose calculations *Phys. Med. Biol.* to be submitted