INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **50** (2005) 909–922 doi:10.1088/0031-9155/50/5/014

A comparison of Monte Carlo dose calculation denoising techniques

I El Naqa[1], I Kawrakow[2], M Fippel[3], J V Siebers[4], P E Lindsay[1], M V Wickerhauser[1], M Vicic[1], K Zakarian[1], N Kauffmann[5] and J O Deasy[1,6]

[1]Washington University, St Louis, MO, USA
[2]National Research Council of Canada, Ottawa, Ontario, Canada
[3]Univ Tuebingen, Tuebingen, Germany
[4]Virginia Commonwealth University, Richmond, VA, USA
[5]Ecole Polytechnique, Palaiseau, France

E-mail: deasy@wustl.edu

Received 14 May 2004, in final form 14 September 2004
Published 17 February 2005
Online at stacks.iop.org/PMB/50/909

Abstract

Recent studies have demonstrated that Monte Carlo (MC) denoising techniques can reduce MC radiotherapy dose computation time significantly by preferentially eliminating statistical fluctuations ('noise') through smoothing. In this study, we compare new and previously published approaches to MC denoising, including 3D wavelet threshold denoising with sub-band adaptive thresholding, content adaptive mean–median-hybrid (CAMH) filtering, locally adaptive Savitzky–Golay curve-fitting (LASG), anisotropic diffusion (AD) and an iterative reduction of noise (IRON) method formulated as an optimization problem. Several challenging phantom and computed-tomography-based MC dose distributions with varying levels of noise formed the test set. Denoising effectiveness was measured in three ways: by improvements in the mean-square-error (MSE) with respect to a reference (low noise) dose distribution; by the maximum difference from the reference distribution and by the 'Van Dyk' pass/fail criteria of either adequate agreement with the reference image in low-gradient regions (within 2% in our case) or, in high-gradient regions, a distance-to-agreement-within-2% of less than 2 mm. Results varied significantly based on the dose test case: greater reductions in MSE were observed for the relatively smoother phantom-based dose distribution (up to a factor of 16 for the LASG algorithm); smaller reductions were seen for an intensity modulated radiation therapy (IMRT) head and neck case (typically, factors of 2–4). Although several algorithms reduced statistical noise for all test geometries, the LASG method had the best MSE reduction for three of the four test geometries, and performed the best for the Van Dyk criteria. However, the wavelet thresholding

[6] Address for correspondence: WUSM Radiation Oncology, Division of Physics, 4921 Parkview Place, Box 8224, St Louis, MO 63110, USA.

method performed better for the head and neck IMRT geometry and also decreased the maximum error more effectively than LASG. In almost all cases, the evaluated methods provided acceleration of MC results towards statistically more accurate results.

1. Introduction

Accurate radiotherapy dose calculations are increasingly more important because treatment margins in conformal therapy or intensity modulated radiation therapy (IMRT) are often reduced or steep gradients are used near important normal tissue structures and target volumes. Most simplified dose calculation algorithms have reduced accuracy in the presence of air cavities, bones or complex lung density variations (Ahnesjo and Aspradakis 1999). Monte Carlo (MC) methods, in their combination of underlying simplicity and fidelity to basic physics, appear to be a reliable technique for advanced radiotherapy treatment planning (Ma *et al* 2000, 2002, Li *et al* 2001, Jones *et al* 2003). Calculation times, however, are typically longer compared to more approximate methods. Recently, a technique for reducing MC dose computation time many-fold was proposed (Deasy 2000), based on the idea that MC results could be modelled as the sum of the 'true' dose distribution (i.e., the result for an infinite number of source particles) and a noise source (due to statistics of particle counting). Analogous to image restoration problems in the field of image processing, an improved estimate of the true image can then be produced by 'smoothing' or 'denoising' the raw MC result, produced with fewer source particles than needed without denoising, resulting in an accelerated approach to the true dose distribution[7]. The goal of MC denoising algorithms is to be as aggressive as possible in locally smoothing the raw MC result while attempting to avoid the introduction of systematic error ('bias'), especially near sharp features such as beam edges.

Although there is a wide agreement that some form of post-processing denoising is beneficial (Deasy 2000, Deasy *et al* 2002, Kawrakow 2002, Fippel and Nusslin 2003, Miao *et al* 2003), the methods which have been reported to accomplish this are extremely varied. More recent reports have focused on methods which adapt to the local dose distribution characteristics. There is extreme variety even among this category, which is unsurprising. There are many known methods from the image restoration field which apply to this problem (Jahne 1997, Lim 1990). Some of the proposed MC denoising techniques were inspired by image processing methods (e.g., convolution filtering, wavelets, anisotropic diffusion (AD), adaptive median filtering), whereas other methods do not have obvious counterparts in the image processing literature (e.g., Kawrakow's locally adaptive method, Fippel and Nusslin's formulation of denoising as a global optimization problem). All the methods reviewed here go well beyond standard image-processing/image-restoration techniques. In this paper, we compare the performance of several previously proposed denoising methods when applied to a variety of MC datasets. The goal here is not to examine all aspects of MC denoising, but to make a fair comparison of proposed denoising methods in terms of statistical accuracy and computational speed.

[7] The term 'denoising' is preferable to 'smoothing' as there are often nonsmooth features in the dose distribution. Here the term 'true' dose distribution merely means the digital dose distribution produced in the limit of an infinite number of digital source particles.

2. Methods

2.1. Denoising methods

2.1.1. Locally adaptive Savitzky–Golay filtering. In Kawrakow (2002), a three-dimensional generalization of a Savitzky–Golay filter[8] was proposed with an adaptive window size (independent x, y and z widths), selected based on a chi-square comparison between the local filtered voxel value and the raw MC value to limit the effect of systematic bias. Essentially, the size of the local linear-quadratic curve-fitting region is made as large as possible while still not introducing bias detectable according to the local chi-square test. Moreover, it was demonstrated that this technique could be applied in multi-batch mode to provide further improvements: the individual batch processed locally adaptive Savitzky–Golay (LASG) results (typically from one of four divided batches) are combined, voxel-by-voxel, according to estimated uncertainties of the batch results.

For these tests, the maximum window size for the 1D filtering phase was seven voxels; for the 3D phase it was five. The maximum accepted chi-square value was set to 1 (except in one case when the output of DOSXYZ was known to underestimate uncertainties and the chi-square cutoff was set to 1.7). The implementation was that described in Kawrakow (2002).

2.1.2. Content adaptive median hybrid filters (CAMH). In a new approach, linear filters were combined with the median operation to produce hybrid median filters. This technique is more fully described in a conference proceeding (El Naqa *et al* 2003), but is briefly described here. The filter is a weighted sum of a local linear filter and the median operation. This can be done in a fixed way (linear median-hybrid filtering) or in a way which depends on the local image (dose) characteristics (content-adaptive median hybrid (CAMH) filtering). The basic idea is that in regions with strong second derivative features, the median operator is used (although based on a set of mean values computed along sampled directions); in smoother regions the mean operator is preferred. Generally, median filters outperform the moving average and other linear filters in removal of impulsive noise (outliers) and preservation of edges, but they fail to provide the same degree of smoothness in homogeneous regions, thus motivating an adaptive combination with mean value (linear) filters. Here, the motivation is purely the preservation of sharp features in high second derivative regions. However, CAMH does not make use of uncertainty information.

For these tests, spatial derivatives were computed after smoothing with Gaussian kernels of widths (σ) 0.1 and 1 and window sizes of 3 and 5 voxels, for the derivative and the smoothing kernels, respectively. The window widths were 3 or 5 voxels.

2.1.3. Wavelet threshold denoising (WTD). Deasy *et al* (2002) studied wavelet denoising using the computationally efficient 9,7-biorthogonal basis. 3D dose distributions were transformed into wavelet space on a slice-by-slice basis; then wavelet coefficients below a selected threshold were set to zero ('threshold denoising'), and smoothed dose images were reconstructed from the remaining coefficients. The wavelet code used here, however, is a fully 3D realization of the same fast biorthogonal 9,7-transform. 'Spin cycling,' the averaging of wavelet results for all possible one-voxel (nearest-neighbour) shifts, is applied. The threshold values were derived by minimizing Stein's unbiased risk estimate (SURE) in each sub-band (wavelet decomposition support size) (Donoho and Johnstone 1995). Essentially, SURE attempts to tune the wavelet denoising threshold automatically to achieve estimated

[8] Savitzky–Golay filters are equivalent to fitting a windowed region to a polynomial in x, y, and z. See Deasy (2000) or Kawrakow (2002) for further description.

minimum mean-square-error (MSE) for each wavelet sub-band. The statistics of the noise were determined using the median absolute deviation (Fodor and Kamath 2001).

2.1.4. Anisotropic diffusion (AD). An adaptive denoising method based on anisotropic diffusion was proposed by the University of Wisconsin-Madison group (Miao *et al* 2003). The 'true' dose distribution is modelled analogous to heat transport, but with a diffusivity which varies as a function of position and transport direction. Thus, diffusion can be lowered in directions across edges (reducing smoothing) and increased in directions parallel to an edge (thus preserving the edge). In the implementation of Miao *et al*, the diffusivity was also a function of the MC noise level.

For these tests, and based on trial runs, algorithm parameters were set as follows (see Miao *et al*, for definitions): the threshold K was set to 1.75 times the estimated uncertainty[9] of the noise; the integration constant Δt was set to 0.15; the number of iterations was chosen between 4 and 7 to achieve the best trade-off between blurring/biasing and noise suppression. The implementation (in heavily-vectorized Matlab code) was by the Washington University group.

2.1.5. Noise reduction as an optimization problem: the iterative reduction of noise method (IRON). Fippel and Nusslin (2003) proposed a denoising method in which denoising is formulated as a minimization problem with a smoothing term related to the local curvature (greater smoothing where there is less curvature) and a restoration penalty term designed to control the introduction of bias. The Tuebingen implementation was used for all tests reported here. The weighting factor parameter was of the order of 1% and only values larger than 5% of the maximum dose were denoised.

2.2. Denoising metrics

It has been shown that the statistical variance can be written as

$$\sigma^2(d) = d\frac{C}{N}, \tag{1}$$

where d is the mean dose in that voxel, N is the number of source histories and C is a proportionality constant which depends on the voxel material (Sempau and Bielajew 2000). Applying a pre-denoising square-root transformation reduces the local dependence of the noise level on the dose level (Deasy 2000). However, experience showed that the transform was quantitatively unimportant and is therefore not applied here.

We define a mean-square-error improvement ratio (IR) as

$$\mathrm{IR} = \frac{\mathrm{MSE}(d_{\mathrm{ref}} - d^i)}{\mathrm{MSE}(d_{\mathrm{ref}} - \hat{d}^i)}, \tag{2}$$

where 'MSE' is the mean-square error with respect to the benchmark (high density) MC result (d_{ref}), d^i is a noisy (low density) sample realization and $\hat{d}^i$ is the result of denoising d^i. This improvement factor was computed only for voxels with dose values equal to or greater than 50% of the maximum dose in the benchmark image. This was done to numerically emphasize the effect of denoising on the higher dose regions, where the accuracy of the dose distribution features is more likely to impact treatment plan review. The selection of 50% as the threshold is unimportant; we expect similar results for thresholds which are between 10% and 90%.

[9] The term 'uncertainty' here means the estimated repeat-run standard deviation of dose values in the high dose region.

If denoising is fast relative to the MC compute time (usually the case), the IR serves as an estimate of how much longer the MC simulation would need to be run to achieve the same MSE compared to starting with a noisy realization and denoising. However, if denoising computation time is significant compared to the MC simulation time, it should be included explicitly in the calculation of an efficiency improvement ratio (Kawrakow 2002).

Other criteria besides improvements in the MSE can also be applied (Kawrakow 2002). In addition to MSE improvement, we also examined the smoothness of isodose lines, the behaviour of the maximum dose variation from the reference and the fraction of points failing the Van Dyk criteria (Van Dyk *et al* 1993). In areas of low gradient, a voxel passes the test if it is within a given percentage of the reference dose ($\Delta\%$); in areas of high gradient (taken here as a maximum slope greater than 30°), a voxel may also pass the test if *any* point (not just a voxel) within a distance y agrees with the reference dose to within $\Delta\%$. In practice, we interpolated points to a 1 mm grid in areas of high gradient.

2.3. Dose distribution test cases

2.3.1. Heterogeneous phantom with electron beam. A highly challenging denoising dataset used here was originally described by Kawrakow (2002). The phantom comprises cubic voxels of width 2 mm. The mostly water phantom contains an aluminium block and an air cavity at a depth of 1 cm. The source was 20 MeV monoenergetic electrons. In the resulting dose distribution, there is a very sharp penumbra at shallow depths, gradually widening towards the practical range of the electrons. Distributions with uncertainties of 5.1%, 3.2% and 2.2% were denoised with a benchmark distribution of 0.2% uncertainty.

2.3.2. Intensity modulated radiation therapy head and neck dose distribution. The head and neck dose distribution was generated using VMC++ (Kawrakow 2001) with five equi-spaced coplanar beams, using a 6 MV photon spectrum, and beamlet widths of 0.5 cm $\times$ 0.5 cm at isocentre. The target volume and CT data were accessed using our research treatment planning system, CERR ('computational environment for radiotherapy research', Deasy *et al* (2003)). VMC++ has been integrated with CERR, resulting in a convenient system for producing MC-generated beamlets (Lee *et al* 2003). Optimization was performed using in-house algorithms. The calculated total MC dose distributions had uncertainties of 3.3% and 6.6%. A benchmark was generated with 40 million source photons with a photon splitting factor of 40 (effectively 1.6 billion photons) resulting in a 0.2% uncertainty.

2.3.3. Computed-tomography-based lung dose distributions. Two lung CT-based test cases were used: the 'lungA' distribution was computed at Washington University as part of an ongoing project to recompute archived lung treatment plans with MC methods (uncertainties 5.7%, 1.8%, 0.9%, and 0.2% for the reference image). The 'lungB' distribution, provided by Virginia Commonwealth University, was computed for MC uncertainties of 7.2% and 3.7%, with a reference image uncertainty of 0.4%. Only a single batch was used for the lungB tests.

3. Results

3.1. Isodose smoothing

Sample isodose contour results at the mid-plane of the electron-beam phantom are shown in figure 1. It is seen that many of the denoising algorithms produce surprisingly smooth isodose

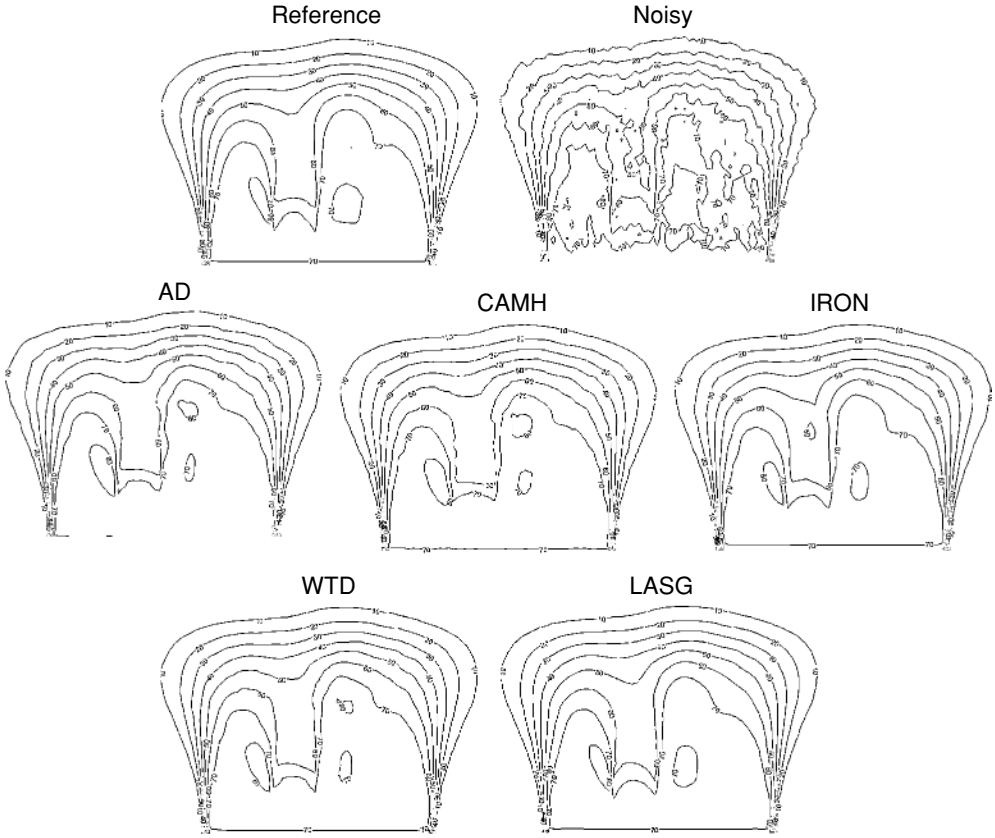

Figure 1. Denoising and isodose-line smoothing. Examples of isodose contours of a plane from the 3D heterogeneous phantom with 20 MeV incident source electrons, including a 'reference image' (0.2% uncertainty), a 'noisy run' and denoising results. Denoising using various methods results in a vast improvement in isodose line intelligibility and accuracy. The LASG result is the closest to the reference image.

line contours given the noisy (3.2% uncertainty) input. However, small hot spots near the entrance and the heterogeneities are difficult to denoise, and there are significant variations between the methods. In particular, the LASG method performs significantly better than other methods in this case.

3.2. Greyscale smoothing

As an example of the effect of denoising on greyscale dose display, figure 2 shows the raw and wavelet denoised results for the lungB dose distribution (3.7% uncertainty). As expected, the beam boundaries are much sharper and the dose distribution interior regions are much smoother for the denoised results. Similar improvements would be expected for colourwash dose displays.

3.3. The introduction of bias

For a test of the introduction of bias by the different denoising techniques, figure 3 shows, for the IMRT head and neck case, images of the difference between denoised dose distributions

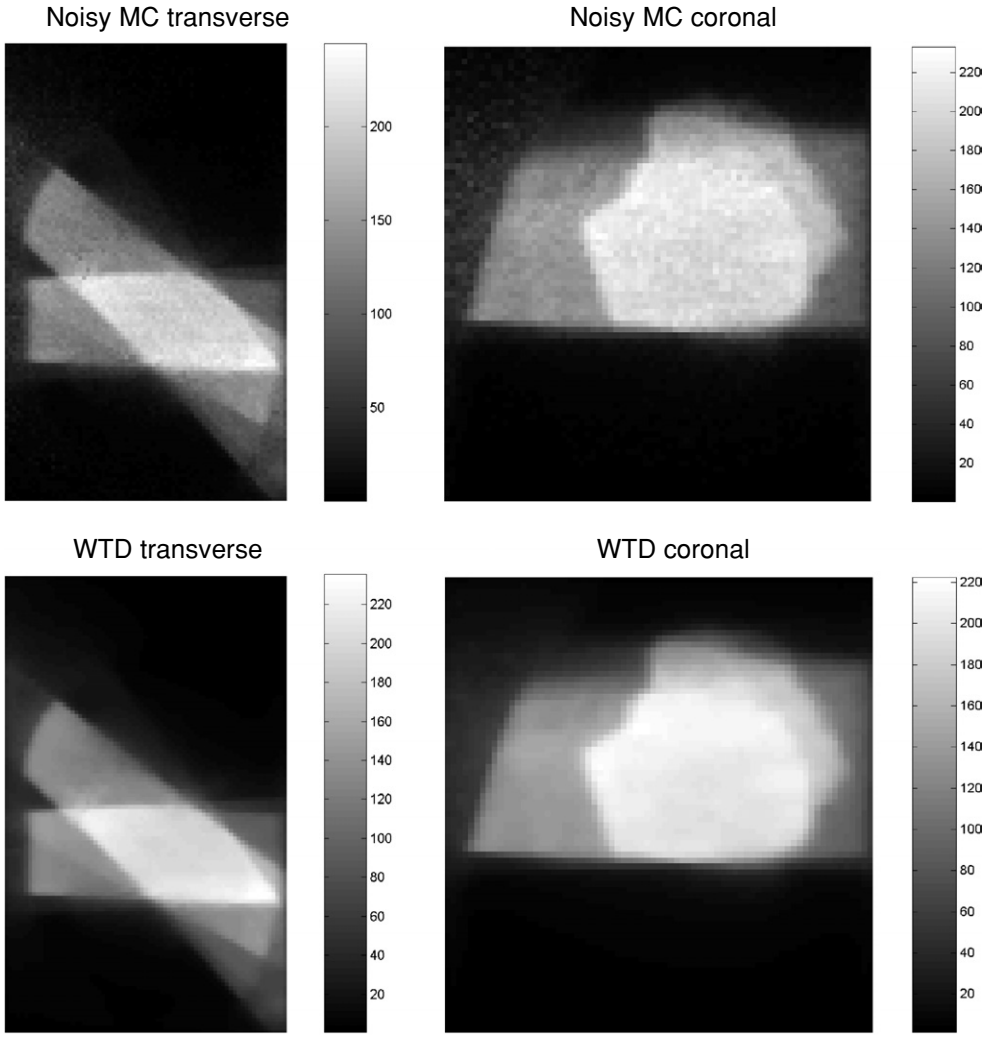

Figure 2. Denoising and greyscale dose display of the LungB case. Top row: original noisy treatment distribution (3.7% uncertainty). *Left* is a transverse data cut and *right* is a coronal slice. Bottom row: an example of a denoised greyscale display (WTD).

and the reference MC image. Many beam edges show bias, which is expected due to their high-frequency characteristics.

3.4. MSE improvement ratios

Figure 4 summarizes MSE improvement ratios for the four different test cases. The different techniques change ranking depending on the dose distribution. However, the LASG method always performs well, having the best performance in three out of four geometries and a close second in the other (the optimized IMRT test).

3.5. The Van Dyk test

Figure 5 summarizes the fraction of voxels which pass the 2% (low gradient) or 2 mm distance-to-agreement (high gradient) criteria. (Note that 2% is a very low threshold for comparison

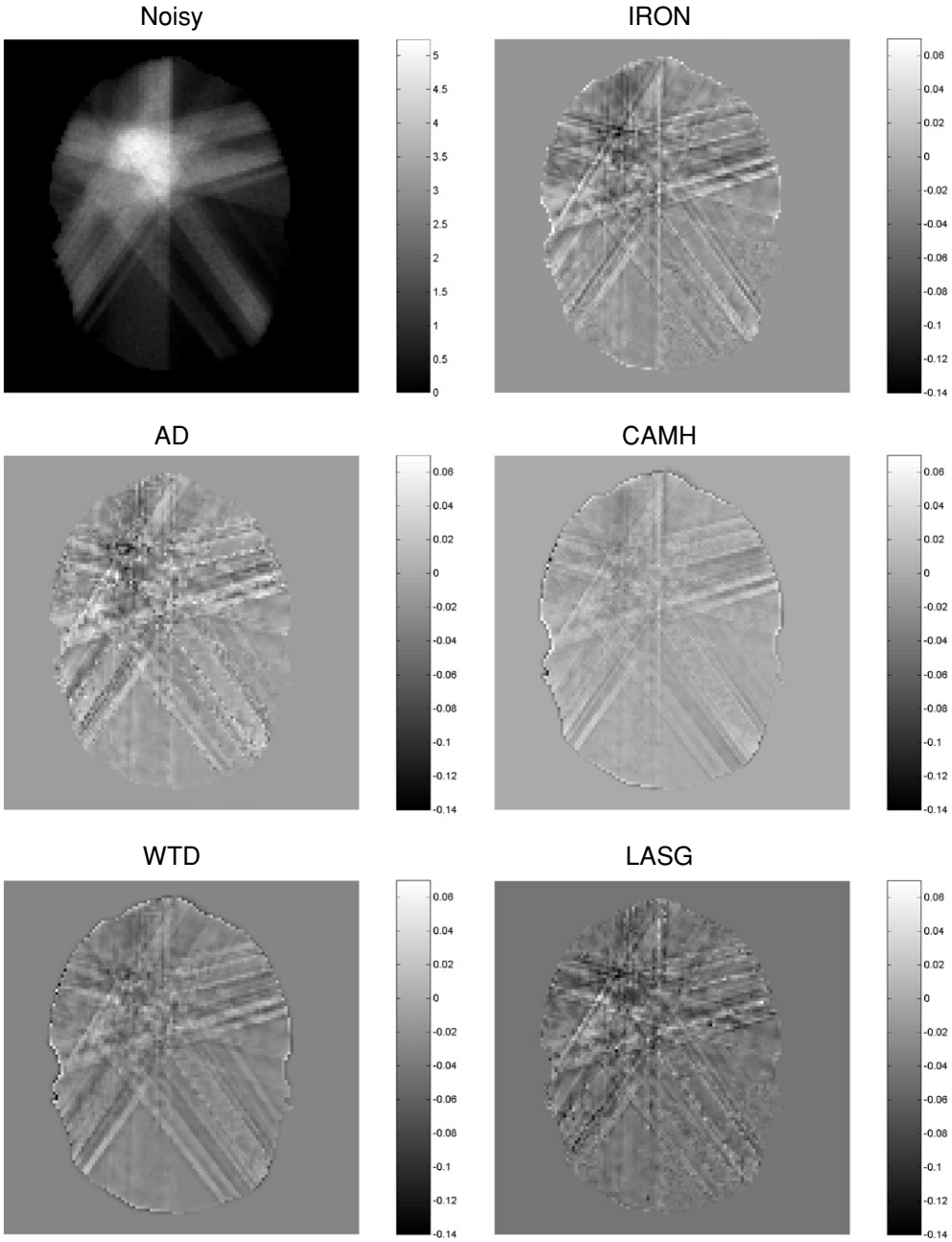

Figure 3. Smoothing versus bias. Difference images of a slice of the denoised IMRT dose distribution (3% uncertainty) with respect to the reference MC result: (denoised voxel − reference voxel)/reference voxel. The 'Noisy' dose greyscale image is shown for reference. Streaks indicate that bias is present, typically near beamlet edges.

purposes and is not meant to indicate acceptable accuracy.) Similar to the MSE results, the only algorithm to distinguish itself was the LASG, which in all tests performed better or nearly as well than the other algorithms.

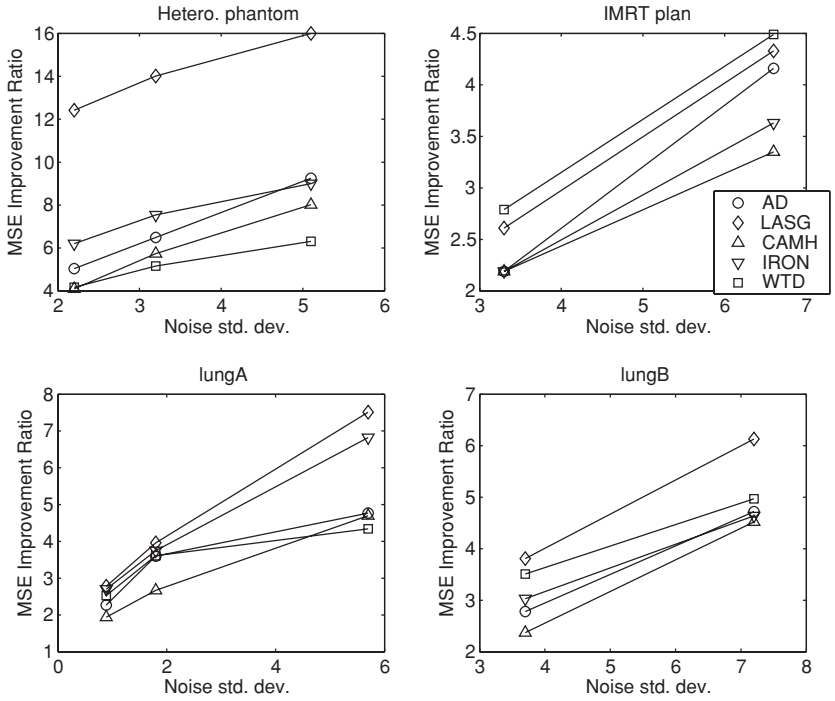

Figure 4. MSE improvement ratios for the four test cases. LASG performs the best in three cases and nearly the best in the fourth case (WTD performs the best for the IMRT case). MSE ratios decrease with decreasing uncertainties apparently due to the residual effect of bias in the denoising MSE term.

3.6. Reduction in maximum error

Figure 6 shows the maximum error, defined as the maximum difference from the reference image divided by the maximum reference image dose. For this metric, the LASG underperforms compared to the other methods (except CAMH which performs poorly). This is apparently due to the fact that the chi-square test will show a lack of fit in the local region when a statistical 'outlier' is smoothed; hence smoothing of outliers is inhibited. CAMH has a poorer performance, and actually makes the maximum error worse as seen in the lung cases, apparently due to its failure near 'corners,' which can result in increased bias.

3.7. Improvements due to batching

Batched denoising refers to denoising of fractional independent MC runs (batches) individually, then combining the batch results into a single dose distribution. The process of applying denoising to batches can be beneficial if the algorithm locally smoothes more aggressively when local noise in that batch happens to be statistically reduced. This effect is further enhanced if the algorithm combines batch-denoising results weighted relative to their filter window sizes/estimated uncertainties, rather than straight arithmetic averaging. Among current denoising techniques, only the Kawrakow method carries along estimated batch uncertainties. Batching resulted in an improvement factor of 2 in the case of the 3-D heterogeneous electron-beam phantom but had a negligible effect in the lungA case. For the other methods, we applied batching by arithmetic mean values, which produced significant

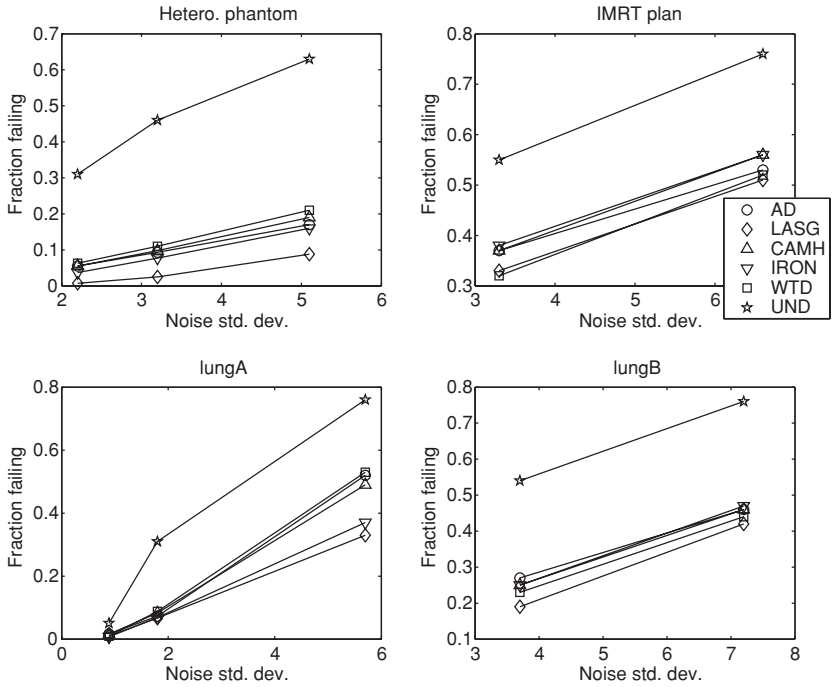

Figure 5. Fraction of voxels failing a 2%/2 mm criterion for the four test cases. Relative algorithm performance is similar to the MSE improvement ratio tests. 'UND' refers to the fraction failing in the undenoised (raw) results.

improvements only in the case of the content adaptive median hybrid filter, which improved by a factor of 1.1 in the case of 3D electron-beam phantom, and a factor of 1.2 in the lung case.

3.8. Computation time

The denoising algorithms vary with respect to computational speed. We have attempted to relate programs running on different hardware platforms (Intel and AMD CPUs), operating systems (Windows and Linux) and differing clock speeds, to estimate approximate speed of denoising for a hypothetical dose distribution of size $100 \times 100 \times 100$ (see table 1). Note that LASG runs approximately three times faster without batching.

4. Discussion

We have not attempted to identify the pre-denoising uncertainty necessary to produce clinically useable treatment plans, which will require further research. However, based on results shown here, we judge that pre-denoising uncertainties of greater than 5% are probably too large. Rather, uncertainties of 2–3% may produce clinically useful post-denoising results. In that region, the typical improvement ratio in MSE for the CT-based cases was 2.5–3. Maximum error of denoised distributions can still be large for raw MC uncertainties of 3% (max. error up to about 15%, see figure 6). The Achilles' heel of MC may be said to be the time necessary to reduce the maximum error to a low level. Maximum error

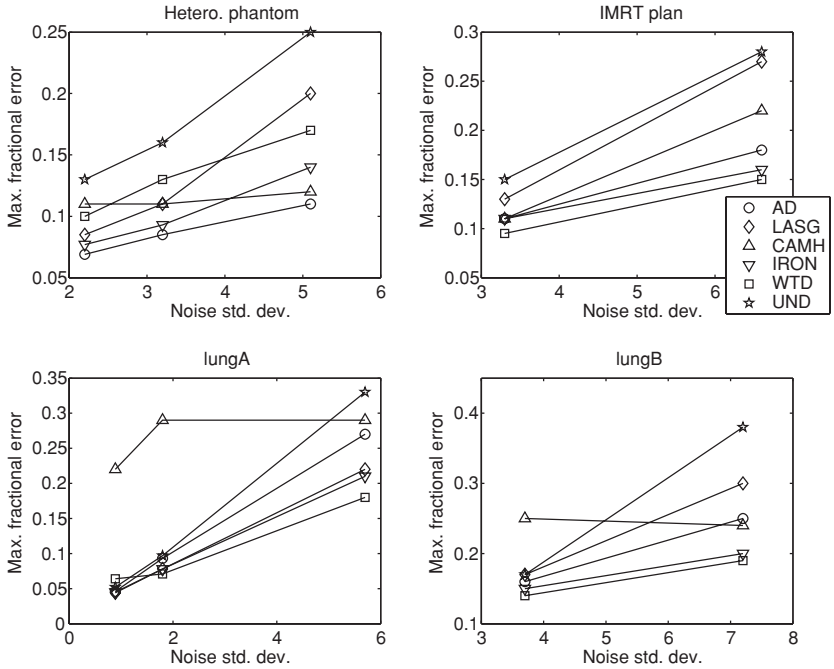

Figure 6. Maximum fractional dose error for the four test cases. Relative algorithm performance is somewhat different to the other tests: WTD is the best performer in three of four cases. 'UND' refers to the maximum dose error in the undenoised (raw) results.

Table 1. Approximate denoising algorithm run-times, estimated for a Pentium IV, 2 GHz computer, in microseconds per voxel, or total time in seconds for a dose distribution of size $100 \times 100 \times 100$. The result quoted for IRON includes a typical speed-up factor due to denoising only voxels with dose greater than 5% of the maximum. The result quoted for LASG is based on using four batches. The slowness of the CAMH algorithm is due partly to implementation in Matlab for prototyping.

Method	Microsecond per voxel, or seconds for a size $100 \times 100 \times 100$ dose distribution
WTD	8
AD	10
IRON	20
LASG (one batch)	22
LASG (four batches)	65
CAMH	900

increases slightly slower than logarithmically with the number of voxels (see Kawrakow (2002)). Fortunately, many treatment planning metrics, such as the mean dose, are relatively insensitive to moderate errors in a small number of voxels (except, of course, the minimum or maximum dose metrics). Establishing the correct level of acceptable uncertainty for clinical MC use would ultimately require consideration of the relevant clinical outcome endpoints and the corresponding probability models.

Relative performance as measured either by improvements in mean-square error or the '2%/2 mm' criteria tended to behave similarly, though differences were seen. In those tests, the LASG method was the clear leader, performing well even when it did not yield the best metric (wavelet threshold denoising (WTD) performed slightly better for the IMRT case).

However, several other algorithms outperformed LASG in reducing maximum error, including WTD, iterative reduction of noise (IRON) method and the AD algorithm. In particular, WTD reduced maximum error the most in three of four test cases. Therefore, it is difficult to declare a 'winner' in the absence of a more definite preference for one metric over another.

Denoising computation times varied considerably (see table 1). The overall efficiency gain from denoising also depends critically on how long the MC calculation takes to reach some threshold of acceptably low noise prior to denoising. This varies widely between different codes and applications. For limited situations, MC run times are currently significantly less than 1 min (for example, using VMC++ to simulate electron beams). In that case, denoising run time could be a significant fraction of the overall run-time and the fastest denoising algorithms (such as the wavelet or anisotropic diffusion methods) automatically gain some advantage. However, for run times longer than a couple of minutes, as currently seen for photon beam calculations for VMC++ and for other MC codes, denoising computation time variations are of less significance (except for the slow prototype CAMH algorithm). In that case, the improvement in MC run times due to the use of denoising is dominated by the effectiveness of the denoising. Improvement ratios by themselves are therefore usually a fair measure of photon MC computation time acceleration with respect to mean-square error. Of course, as computers become faster, relative computation times will improve for both denoising algorithms and MC algorithms in the same proportions. Denoising will therefore retain its overall acceleration factor effect on computation time, but the magnitude of the time advantage will decrease.

Although we have focused on the mean-square error and the '2%/2 mm' criteria as fair comparative measures of algorithm performance, isodose line smoothness is also important. Much of the mean-square error is due to systematic bias (as shown in the normalized error images in figure 3) which would be unchanged from denoising run to denoising run. This implies that, with respect to isodose line noise alone, the acceleration is greater than the quoted mean-square-error improvement ratios (again, see figure 1). This aspect of denoising was also apparent in Deasy *et al* (2002), wherein the image roughness of the denoised result (defined as the square root of the median of the squared Laplacian) was nearly constant and independent of initial noise smoothness for 2D wavelet denoising.

The denoising algorithms reviewed here use several very different strategies, all based on some form of local adaptation. Wavelets, which are often thought of as 'automatically adapting' to local smoothness, nevertheless have the weakest local adaptation, due to the global thresholds selected (although the different wavelet scales have different thresholds). Clearly, there are many different routes to developing a successful MC denoising algorithm. Further developments in denoising algorithms could therefore be expected in many ways, for example, by improving implementation speed or by combining ideas used in the different algorithms.

The raw MC datasets used for these comparisons are available from the corresponding author. We hope that they will help facilitate evaluations of future denoising research.

5. Conclusions

Several denoising algorithms have been reviewed which improve MC dose distribution smoothness and statistical dosimetric accuracy. We compared different MC denoising approaches on datasets with widely varying characteristics. We conclude that:

1. The effective acceleration achieved by denoising depends significantly on the characteristics of the dose distribution. For example, MSE improvement ratios achieved

for the heterogeneous electron beam phantom (at most 16, using LASG) were much higher than improvement ratios achieved for the head and neck IMRT treatment plans (at most 4.5 using WTD). Apparently, much higher improvement ratios are possible when large volumes of the dose distribution are very smooth, as is the case for the heterogeneous phantom distribution. The results also indicate the denoising will be more effective for conformal therapy dose distributions than topologically more complicated IMRT dose distributions. Nevertheless, even IMRT dose distributions have improvement ratios greater than 2.

2. Relative algorithm performance depends on the dose characteristics and the ranking metric. LASG performed well for all the MSE and Van Dyk criteria tests, but less well in reducing maximum dose error. Overall, several algorithms can be judged to perform well, including LASG, WTD, AD and IRON. In particular, WTD was the strongest at reducing maximum error and also was the best performer on the IMRT case for the other metrics. The wavelet method performed well for all the computed-tomography-based dose distributions. The CAMH algorithm, in contrast, introduced undesirable increases in the maximum dose error and had an undistinguished performance in the other tests.
3. Lastly, these tests again support the conclusion that MC denoising is a useful post-processing step which can be used to improve isodose line smoothness and dosimetric accuracy. MC denoising is therefore a desirable component of MC treatment planning systems.

Acknowledgments

This work was supported in part by NIH grants R01 CA90445 and R01 CA98524. Financial disclosure: Washington University has received a US patent for MC dose distribution denoising.

References

Ahnesjo A and Aspradakis M M 1999 Dose calculations for external photon beams in radiotherapy *Phys. Med. Biol.* **44** R99–155

Deasy J O 2000 Denoising of electron beam MC dose distributions using digital filtering techniques *Phys. Med. Biol.* **45** 1765–79

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

Deasy J O, Wickerhauser V and Picard M 2002 Accelerating Monte-Carlo simulations of radiation therapy dose distributions using wavelet threshold denoising *Med. Phys.* **29** 2366–77

Donoho D L and Johnstone I M 1995 Adapting to unknown smoothness via wavelet shrinkage *J. Am. Stat. Assoc.* **90** 1200–24.

El Naqa I, Deasy J and Vicic M 2003 Locally adaptive denoising of MC dose distributions via hybrid median filtering *Proc. IEEE Medical Imaging Conf. (19–25 October, Portland, OR)*

Fippel M and Nusslin F 2003 Smoothing MC calculated dose distributions by iterative reduction of noise *Phys. Med. Biol.* **48** 1289–304

Fodor I K and Kamath C 2001 On denoising images using wavelet-based statistical techniques *LLNL Technical Report* UCRL-JC-142357

Jahne B 1997 *Digital Image Processing* (New York: Springer)

Jones A O, Das I J and Jones F L 2003 A MC study of IMRT beamlets in inhomogeneous media *Med. Phys.* **30** 296–300

Kawrakow I 2001 *Advanced MC for Radiation Physics, Particle Transport Simulations and Applications* ed A Kling *et al* (Berlin: Springer) pp 229–36

Kawrakow I 2002 On the denoising of MC calculated dose distribution *Phys. Med. Biol.* **47** 3087–103

Lee E, Deasy J, El Naqa I, Kawrakow I and Vicic M 2003 Integrating a MC based optimization module into CERR for designing IMRT treatment plans *Med. Phys.* **30** 1491 (abstract)

Li X A, Ma L J, Naqvi S, Shih R P and Yu C 2001 MC dose verification for intensity-modulated arc therapy *Phys. Med. Biol.* **46** 2269–82

Lim J 1990 *Two-dimensional Signal and Image Processing* (Upper Saddle River, New Jersey: Prentice Hall)

Ma C M, Li J S, Pawlicki T, Jiang S B, Deng J, Lee M C, Koumrian T, Luxton M and Brain S 2002 A MC dose calculation tool for radiotherapy treatment planning *Phys. Med. Biol.* **47** 1671–89

Ma C M, Pawlicki T, Jiang S B, Li J S, Deng J, Mok E, Kapur A, Xing L, Ma L and Boyer A L 2000 MC verification of IMRT dose distributions from a commercial treatment planning optimization system *Phys. Med. Biol.* **45** 2483–95

Miao B, Jeraj R, Bao S and Mackie T 2003 Adaptive anisotropic diffusion filtering of MC dose distributions *Phys. Med. Biol.* **48** 2767–81

Sempau J and Bielajew A F 2000 Towards the elimination of Monte Carlo statistical fluctuation from dose volume histograms for radiotherapy treatment planning *Phys. Med. Biol.* **45** 131–57

Van Dyk J, Barnett R B, Cygler J E and Shragge P C 1993 Commissioning and quality assurance of treatment planning computers *Int. J. Radiat. Oncol. Biol. Phys.* **26** 261–73