

Validation of the Acuros XB dose calculation algorithm versus Monte Carlo for clinical treatment plans

Lone Hoffmann^{a)}

Department of Oncology, Aarhus University Hospital, Aarhus 8000, Denmark

Markus Alber

Department of Oncology, Aarhus University Hospital, Aarhus 8000, Denmark

Section for Medical Physics, Department of Radiooncology, University Clinic Heidelberg, Heidelberg 69120, Germany
Scientific RT GmbH, Munich 81373, Germany

Matthias Söhn

Scientific RT GmbH, Munich 81373, Germany

Ulrik Vindelev Elstrøm

Department of Oncology, Aarhus University Hospital, Aarhus 8000, Denmark

(Received 22 February 2018; revised 14 May 2018; accepted for publication 31 May 2018;
published 6 July 2018)

Purpose: The two distinct dose computation paradigms of Boltzmann equation solvers and Monte Carlo simulation both promise in principle maximum accuracy. In practice, clinically acceptable calculation times demand approximations and numerical short-cuts on one hand, and modeling the beam characteristics of a real linear accelerator to the required accuracy on the other. A thorough benchmark of both algorithm types therefore needs to start with beam modeling, and needs to include a number of clinically challenging treatment plans.

Methods: The Acuros XB (v 13.7, Varian Medical Systems) and SciMoCa (v 1.0, Scientific RT) algorithms were commissioned for the same Varian Clinac accelerator for beam qualities 6 and 15 MV. Beam models were established with water phantom measurements and MLC calibration protocols. In total, 25 patients of five case classes (lung/three-dimensional (3D) conformal, lung/IMRT, head and neck/VMAT, cervix/IMRT, and rectum/VMAT) were randomly selected from the clinical database and computed with both algorithms. Statistics of 3D gamma analysis for various dose/distance-to-agreement (DTA) criteria and differences in selected DVH parameters were analyzed.

Results: The percentage of points fulfilling a gamma evaluation was scored as the gamma agreement index (GAI), denoted as $G(\Delta D, DTA)$. $G(3,3)$, $G(2,2)$, and $G(1,1)$ were evaluated for the full body, PTV, and selected organs at risk (OARs). For all patients, $G(3,3) \geq 99.9\%$ and $G(2,2) > 97\%$ for the body. $G(1,1)$ varied among the patients. However, for all patients, $G(1,1) > 70\%$ and $G(1,1) > 80\%$ for 68% of the patients. For each patient, the mean dose deviation was $\Delta D < 1\%$ for the body, PTV, and all evaluated OARs, respectively. In dense bone and at off-axis distance > 10 cm, the Acuros algorithm yielded slightly higher doses. In the first layer of voxels of the patient surface, the calculated doses deviated between the algorithms. However, at the second voxel, good agreement was observed. The differences in $D(98\%PTV)$ were $< 1.9\%$ between the two algorithms and for 76% of the patients, deviations were below 1%.

Conclusions: Overall, an outstanding agreement was found between the Boltzmann equation solver and Monte Carlo. High-accuracy dose computation algorithms have matured to a level that their differences are below common experimental detection thresholds for clinical treatment plans. Aside from residual differences which could be traced back to implementation details and fundamental cross-section data, both algorithms arrive at identical dose distributions. © 2018 The Authors. Medical Physics published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine. [https://doi.org/10.1002/mp.13053]

Key words: Boltzmann equation solver, dose computation algorithm, Monte Carlo

1. INTRODUCTION

Patient outcome is dependent on dosimetric accuracy. In the treatment planning process, inaccuracies originate, among other things, from CT calibration and dose calculation, whereby especially modeling of the lateral electron dose deposition and of the radiation source and collimation have been challenges.^{1,2}

The algorithm classes of both linear Boltzmann equation solvers (LBES) and Monte Carlo (MC) simulation promise maximum accuracy of radiation transport within the patient.¹

The dose planning algorithm Acuros XB (Varian Medical Systems, Palo Alto, CA) is based on the solution of the linear Boltzmann transport equation.³ Over the years, its performance has been tested in water phantoms,⁴⁻⁶ in



heterogeneous phantoms,^{7–10} and in anthropomorphic breast phantoms,³ lung phantoms,^{11,12} and head and neck phantoms.^{10,13} Dosimetric one-dimensional (1D) and two-dimensional (2D) measurements and MC calculations have been used for the validation. MC-based benchmarking has been performed using CT datasets of 4 lung cancer patients¹⁴ and 10 patients treated at a variety of anatomical sites.¹⁵ In both studies, the gamma agreement index (GAI) with 2% dose difference and 2 mm distance-to-agreement (DTA) were reported, and the GAI ranged from 82% in a patient with lung cancer to 98% in a patient with pancreas cancer. Given these studies, it can be concluded that in previous version, Acuros has been on a par with other advanced algorithms (convolution/superposition, collapsed cone), but it has not exceeded their accuracy to reach the level of a well-implemented MC algorithm.

For both, LBES and MC algorithms, the description of the radiation source and the collimating system is a great challenge especially for VMAT treatments, where field sizes can be very irregular, down to single opened leaf pairs, and are fully dynamic both in leaf angle and in beam angle, which requires discretization of these degrees of freedom for analytical algorithms, but not for Monte Carlo. The beam models of both algorithms need to cope with the precise computation of output for very small fields in highly modulated plans, which require precise handling of leaf shapes and positions as well as primary source properties.

In the present study, Acuros was benchmarked against SciMoCa (Scientific RT GmbH, Munich, Germany), a MC product for secondary dose calculation. For both algorithms, 6 and 15 MV beam models of a Varian Clinac linear accelerator in use in our clinic were created. Twenty-five treatment plans were used for the benchmark process. The plans fell into five categories: lung stereotactic/hypofractionated (HF), lung normofractionated (NF), head and neck (H&N), cervix, and rectum cancer, and the delivery techniques used were three-dimensional (3D) conformal, intensity modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT), respectively.

2. MATERIALS AND METHODS

2.A. Patient data and treatment planning

Five patients were selected from each of the following categories: lungHF, lungNF, H&N, cervix, and rectum cancer (see Supplementary Materials Table S1). These groups represent some of the major patient groups treated by radiotherapy. Furthermore, these sites present challenges as regions with large density changes, tiny tumors, air, lung and bony tissues, and tumors close to the skin, combined with stereotactic field sizes, large volumes, and low- and high-photon energies in modulated fields. The patients were randomly selected from the clinical database. For all patients, a free-breathing CT scan with 3-mm slice thickness was acquired. For patients with lung cancer, the mid-ventilation phase of a four-dimensional (4D) CT scan was used for delineation and dose calculation.

The patients were treated using a Varian Clinac (Varian Medical Systems, Palo Alto, CA) equipped with a Millennium 120 multileaf collimator (MLC). The width of the leaves at the isocenter was 5 mm for the 40 most central leaves and 10 mm for the remaining leaves. The beam qualities were 6 and 15 MV.

All treatment plans were created in the Eclipse treatment planning system (Varian Medical Systems, Palo Alto, CA) version 13.7 and grid size of 2.5 mm was used for the calculation. Dose was reported as dose-to-medium. For each patient category, standard techniques including IMRT and VMAT were used for treatment planning (See Supplementary Materials Table S1).

2.B. Modeling of Varian accelerator in Eclipse

The LBES algorithm Acuros has been described in details by Vasiliev et al. and in the reference materials from the vendor.^{3,16} For each voxel, the mass density is derived from the Hounsfield units using ICRP data for the cross-sections.¹⁶ Mass density table version 13.5 was used and it includes air and five organic material composition types: adipose tissue, bone, cartilage, lung, and skeletal muscle.

The clinical implementation of Acuros was based on 14-depth dose curves and profiles measured at five depths. Output factors were measured for 196 square and rectangular fields in the range $1 \times 1 \text{ cm}^2$ to $40 \times 40 \text{ cm}^2$.¹¹ The configuration of the algorithm allowed for a slight change in the width of the penumbra by optimizing the effective spot size. A value of 1 mm in x- and y-direction was found to be optimal for both energies.

The dosimetric leaf gap (DLG) and transmission (T) through the MLCs were measured in solid water by use of the method proposed by Van Esch et al.¹⁷ Measurements were performed at several selected points of a chair-like fluence at five different depths, and the best overall values for DLG and T were selected. Similar results for DLG were obtained by the method proposed by LoSasso et al.¹⁸ Here, DLG was found by extrapolating measurement of dose vs MLC defined field size to zero dose. The maximum deviation between the two methods was 0.2 mm.

2.C. Modeling of a Varian accelerator in SciMoCa

The SciMoCa algorithm combines concepts of the VMC family of MC codes^{19–21} (henceforth denoted *VMC*) with some element of EGSnrc.²² The most significant deviations from general purpose codes like EGSnrc are: (a) material property assignments are derived from CT images, and material compositions are grouped into “lung-like,” “soft-tissue-like,” and “bone-like,” each with variable density, as well as water, titanium, and steel¹⁹; (b) transport parameters are fixed to photon-cutoff energy = 60 keV, electron-cutoff energy = 240 keV, electron energy step = 0.12, and maximum energy = 25 MeV; (c) patient transport is limited to rectilinear grids with minimum dimension of 0.5 mm; (d) variance reduction schemes: kerma approximation (photon energy < 1 MeV),

history repetition, and Russian Roulette are adopted with modifications from Ref. 20; (e) condensed history simulation follows the concepts laid out in Ref. 21 with some modifications for speed-up. The *VMC* algorithms owe a great deal of their performance from simplifying the material property look-up, which reduces memory throughput as the main bottleneck in modern hardware architecture. This causes errors for the dose calculation inside non-tissue equivalent materials, but not around them. For example, air is described as lung-like tissue with the density of air, that is, ascribed the wrong material composition but the correct density.

The treatment head simulation employs five virtual sources (primary, primary collimator scatter, other head and flattening filter scatter, backscatter into the monitor chamber, and electron contamination) and is an evolution from previous models.^{23–25} The source properties are determined from BEAMnrc simulations of the generic accelerator head type and modified by an interpolation method described in Ref. 25 to fit a specific, real-world treatment machine. The MLC transport model comprises generic leaf shapes, which are configurable for tongue and groove, leaf tip curvature, and interleaf leakage to match measured cross-profiles. Photon transport through the leaves is explicit with suppression of scatter that would not reach the patient.

The 6 and 15 MV beam modalities of the Varian Clinac were commissioned on the basis of 5 depth dose curves ($40 \times 40 \text{ mm}^2$ - $400 \times 400 \text{ mm}^2$), output factors, and cross-profiles of the maximum field size. The leaf positions and primary source diameter ($2\sigma(6 \text{ MV}) = 1.0 \text{ mm}$, $2\sigma(15 \text{ MV}) = 1.2 \text{ mm}$) were calibrated for simulation from standard DLG measurements,²⁶ which is an essential step especially for highly modulated treatments. A grid size of 2.5 mm was used for the calculation.

The treatment plans were exported to SciMoCa and the dose was recalculated using the DICOM image, structure set, and plan information. Aside from CT-based material assignment, mass density can also be assigned on a per-structure basis. The statistical uncertainty was chosen such that the volume receiving at least 70% of the maximum dose had a statistical uncertainty smaller than 0.5%.

2.D. Plan comparison

The 3D dose distributions from Acuros and SciMoCa were compared in ProSoma (MedCom GmbH, Darmstadt, Germany) using a global gamma evaluation^{27,28} with suppression of doses below 5% of the maximum dose. The percentage of points fulfilling the gamma evaluation was scored as the GAI. Different threshold criteria were used: dose difference $\Delta D = 3\%$, 2% , and 1% , and DTA = 3, 2, and 1 mm. In total, three values of GAI denoted as $G(\Delta D, \text{DTA})$ were obtained: $G(3,3)$, $G(2,2)$, and $G(1,1)$. For all 25 plans, GAI was scored inside the full body, in the PTVs, and for selected organs at risk (OARs). For each of the five patient classes, lungHF, lungNF, H&N, cervix, and rectum, the median GAI for all five patients was calculated. Furthermore, the mean dose difference (ΔD) and the standard deviation (SD) were

calculated for each patient, and the median value of ΔD for all five patients was reported for each patient class. Acuros was used as a reference in GAI and ΔD calculations.

$G(3,3)$ to $G(1,1)$ were plotted in radar plots for the full body and the PTVs (CTVs for H&N) for each of the five patient classes.

DVH parameters for PTV and selected OARs were calculated. For each of the five patient classes and for each dose parameter, the median dose and range for all five patients were calculated by Acuros and SciMoCa. The dose difference (reference: Acuros) was calculated for each plan and dose parameter, and for each class of patients; the median dose difference and minimum and maximum dose difference were also reported. The following DVH parameters were evaluated: dose covering 98% of the PTV; $D(98\% \text{PTV})$, high dose to spinal cord, esophagus and mandible evaluated as dose to 2% of the OARs; $D(2\% \text{OAR})$ and mean dose to lung, bladder, and bowel; and $D(\text{mean OAR})$.

3. RESULTS

For all patients, GAIs, mean ΔD , and SD were evaluated. In Table I, the median values are shown for the full body and for the PTV or PTV1 (CTV1 for H&N) evaluation. The standard deviation of ΔD is reported for the patient with the median value of ΔD .

In the patient cohort, $G(2,2)$ is fulfilled for a minimum 97.0% of the evaluated points for the full body and $G(1,1) \geq 71.7\%$. The mean dose deviates less than 1% and SD range from 0.6 to 1.5%.

In Table II, median values for GAI and ΔD are shown for relevant OARs of each patient class. For all 25 evaluated patients and OARs, $G(2,2) \geq 97.3\%$ and $G(1,1) \geq 74.3\%$, and $\Delta D < 1\%$. For the bones, Acuros yielded a slightly higher dose than SciMoCa.

TABLE I. Median gamma agreement index, dose difference, and standard deviation for the body and the PTV (CTV for H&N).

	$G(3,3)$ (%)	$G(2,2)$ (%)	$G(1,1)$ (%)	ΔD (%)	SD (%)
LungHF, body	100	100	99.0	−0.2	0.6
LungNF, body	100	99.8	97.5	0.1	1.0
H&N, body	99.9	99.5	89.2	−0.3	1.5
Cervix, body	100	97.9	78.2	0.7	0.8
Rectum, body	100	99.8	95.8	−0.1	0.9
LungHF, PTV	100	99.9	92.5	−0.7	1.0
LungNF, PTV	100	99.8	89.7	−0.4	0.6
H&N, CTV1	100	99.7	94.5	−0.1	0.6
H&N, CTV2	100	99.7	95.2	−0.1	0.7
H&N, CTV3	100	99.9	91.3	0.2	0.7
Cervix, PTV1	100	99.9	95.4	−0.1	0.7
Cervix, PTV2	100	100	96.2	−0.1	0.6
Cervix, PTV3	100	99.9	94.2	−0.4	0.7
Rectum, PTV	100	99.7	85.3	−0.6	0.7

Radar plots are shown in Fig. 1 for the full body and PTV or PTV1 (CTV1 for H&N) for each of the five patient classes. The plots illustrate G(3,3), G(2,2), and G(1,1) for each of the five patients. Similar values of G(3,3) and G(2,2) are seen for all the patients. The lowest values of G(2,2) for the full-body evaluation was seen in a patient with cervix cancer. For this group of patients, the range of G(2,2) was [97.0;99.2%]. The highest values of G(2,2) was seen for lungHF patients where the range of G(2,2) was [99.9;100%]. In 88% of the patients, $G(2,2) \geq 99.0\%$ for the full-body evaluation. Values for G(1,1) varies among the patients. However, G(1,1) is always greater than 70%, and in 68% of the patients, $G(1,1) \geq 90\%$.

TABLE II. Median gamma agreement index, dose difference, and standard deviation for selected organs at risk.

	G(3,3)	G(2,2)	G(1,1)	ΔD	SD
LungHF, bones	100	100	99.5	0.1	0.8
LungNF, bones	100	100	99.0	0.2	0.7
H&N, bones	100	100	88.8	0.8	0.6
Cervix, bones	100	100	93.5	0.4	0.6
Rectum, bones	100	100	97.2	0.4	0.6
LungHF, esophagus	100	100	100	0.0	0.3
LungNF, esophagus	100	100	99.2	0.2	0.6
H&N, esophagus	100	100	92.7	0.9	0.8
LungHF, lung	100	100	98.6	-0.3	0.6
LungNF, lung	100	99.9	97.4	0.1	0.6
Cervix, bowel	100	99.5	76.8	0.8	0.6
Rectum, bowel	100	99.9	97.4	0.0	0.6
Cervix, bladder	100	100	96.9	0.1	0.6
Rectum, bladder	100	99.9	92.7	0.1	0.8

As seen in Fig. 1, only minor differences exist in GAI among the patients. A color plot of the gamma evaluation is shown in Fig 2 for the full-body evaluation for each patient class. For each class, the patient with median G(1,1) was selected for illustration.

Patients with cervix cancer are treated with large field sizes due to the involvement of paraaortic lymph nodes. A dose difference plot of the patient with median G(1,1) is shown in Fig. 3. Dose deviations are primarily seen in the most cranial and the most caudal part of the irradiated region. Here, Acuros calculated approximately 1% higher dose than SciMoCa. A similar trend was seen for all CERVIX patients.

In all patients, dose difference was observed at the patient outline. Up to 3% higher dose was calculated by SciMoCa. The phenomenon was confined to the first voxel layer.

A maximum deviation of 1.9% was seen for D(98%PTV) and for 76% of the patients, the deviation was below 1% (see Supplementary Materials Table S2). For the lungHF patients, the dose calculated by SciMoCa was in median 1.6% (0.8 Gy) higher than for Acuros. For the spinal cord and mandible, the median dose difference was within 1%, and Acuros yielded a slightly higher dose than SciMoCa. Mean doses to lung, bladder, and bowel were in median within 0.6%, except for the mean bowel dose for the cervix patients where higher dose was calculated by Acuros for four patients. The bowel is partly located in the most caudal part, where Acuros was seen to deliver a higher dose than SciMoCa (see Fig 3).

4. DISCUSSION

The accuracy of the Acuros dose calculation algorithm has been benchmarked against the SciMoCa algorithm. While

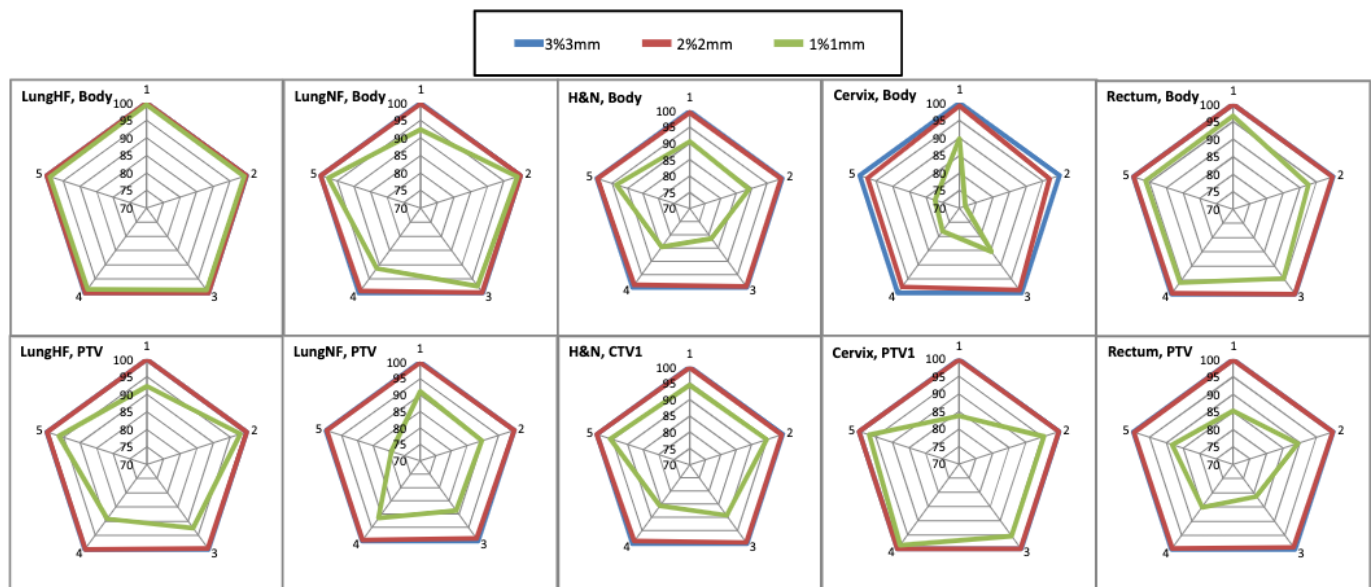


FIG. 1. Radar plot of G(3,3), G(2,2), and G(1,1), the full body (top row), and for PTV or PTV1 (CTV1 for H&N) (bottom row). All plots show data for five patients (numbered 1 to 5) in the categories lungHF, lungNF, H&N, cervix, and rectum. G(3,3) and G(2,2) coincides for all patients except for the body evaluation for the cervix patient. [Color figure can be viewed at wileyonlinelibrary.com]

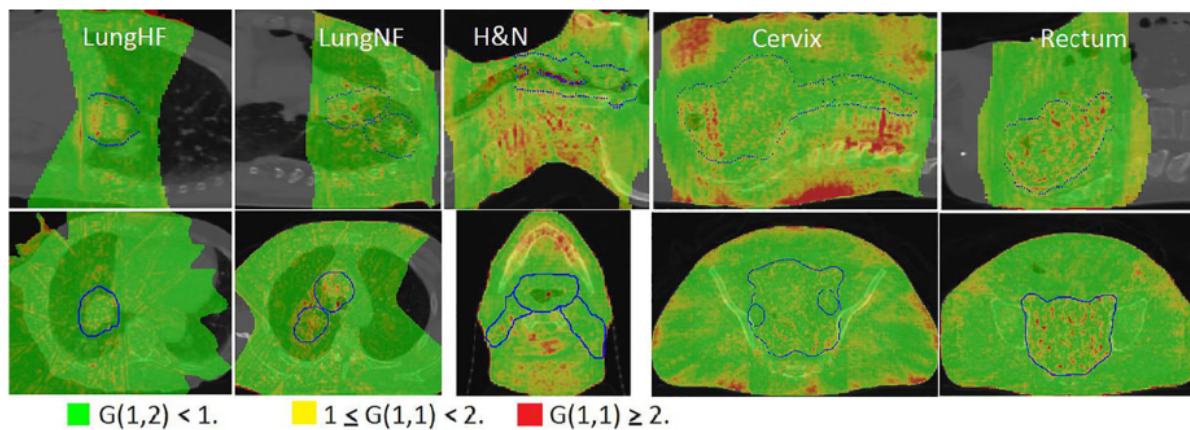


FIG. 2. Color plot of $G(1,1)$ for lungHF, lungNF, H&N, cervix, and rectum. Top row: sagittal view. Bottom row: transversal view. Green: $G(1,1) < 1$. Orange: $1 \leq G(1,1) < 2$. Red: $G(1,1) \geq 2$. PTV is shown in blue. [Color figure can be viewed at wileyonlinelibrary.com]

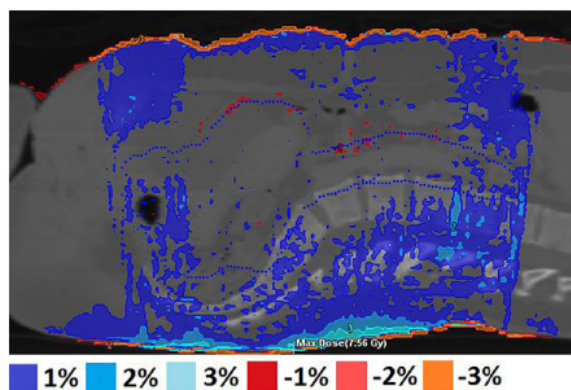


FIG. 3. Dose difference plot (Acuros minus SciMoCa) for the cervix patient with the median value of $G(1,1)$ for the full body. Dark blue: 1%, blue: 2%, light blue: 3%, dark red: -1%, red: -2% and orange: -3%. [Color figure can be viewed at wileyonlinelibrary.com]

both algorithms obviously embody fundamentally different concepts, the differences in implementation of the radiation sources and the treatment of dynamic delivery go even deeper. It is therefore fair to say that both algorithms are fundamentally different, and yet their agreement is remarkably high. Residual differences in patient plans are so small that it would be difficult to determine the more accurate algorithm experimentally, as dosimetry errors for complex plans tend to be in the same order of magnitude. Note that at a MC statistical uncertainty of 0.5% and a voxel size of 2.5 mm, already 5% of all dose points exceed a $G(1,1)$ criterion. Given inevitable discrete numerical errors, a $G(1,1)$ pass rate of 90–95% is probably the conceptual maximum.

Gamma index computation has become the customary tool for quantifying the difference between two dose distributions, but may not be without bias (see Ref. [29,30] for a comprehensive discussion). In the current context, it is important to make a few points. First, both dose computations used the same dose grid resolution of 2.5 mm in all directions, and compute the same quantity, that is the average dose over a voxel (instead of a point dose), so that both distributions show the same averaging effects around steep (field edge)

gradients. Second, as pointed out in Ref. [29], a DTA criterion that is smaller than the voxel size requires interpolation or specially devised algorithms for gamma calculation that can handle this situation. The software used here, ProSoma, implements an algorithm that eliminates interpolation errors altogether.²⁸ Third, the presence of noise in the “compare” dose distribution can bias the gamma passing rate toward lower values.³⁰ While all gamma pass rates reported here set the analytical dose distribution as reference and the Monte Carlo distribution as “compare,” the bias cannot be large as the pass rates are largely the same if the roles are reversed. This hints at the fact that analytical dose computation results are also not as smooth as physics would have them, owing to the discrete nature of the calculations. Finally, a threshold limit of 5% was selected for the cutoff value for low doses. In the AAPM TG119 report on IMRT commissioning of IMRT,³¹ a threshold value of 10% was recommended. No difference in $G(3,3)$ for the body evaluation was found between the two threshold limits and for $G(2,2)$, a maximum difference of $\pm 0.3\%$ was found. For $G(1,1)$, a higher threshold value in median resulted in 0.3% higher $G(1,1)$ pass rate, with a maximum increase of 3.0% in one patient.

Some systematic differences remain and can be discussed. First, Acuros employs ICRP material parameters, while SciMoCa employs ICRU material parameters. Some differences that can be observed especially in dense, cortical bone may originate from the fundamental cross-section and stopping power data. In this context, it is important to note that both algorithms were configured to report dose-to-medium — apparently, this configuration is much more general than dose-to-water, for which both algorithms would need to employ different conversion methods. Some apparent differences in dose inside air cavities can be attributed to the fact that MC uncertainty scales with the inverse of the square-root of the relative mass density, in other words, statistical noise in airways is roughly 30 times greater than in the surrounding tissue.

Second, some differences could be seen at the patient outline. These highlight the fact that density grids are usually downsampled from the original CT grids, which calls for some handling of the patient surface. SciMoCa’s grid

resampling observes strict conservation of mass, which together with the restriction to voxelized geometries leads to a “halo” of voxels with a variable, low density around the patient outline. Apparently, this is handled differently in Acuros. However, both algorithms agree already on the second layer of voxels, which speaks for a consistently good handling of contamination electrons.

For patients with cervix cancer, Acuros yielded slightly higher dose in the most cranial and the most caudal part of the field. The patients were treated with 6–9 coplanar IMRT fields with a collimator rotation of 5 degrees. The deviations emerged at approximately 10 cm from the isocenter. At this point, the width of MLCs changes from 5 mm to 10 mm and a slight difference in the calculation of the transmission may result in these differences. Furthermore, the off-axis spectrum softening is handled differently between the two systems.

Previous versions of Acuros were validated against MC and measurements with somewhat lower agreement in a number of publications. In water phantoms, a maximum dose deviation of 2% to measurements^{4,11} or MC calculations^{7,9} was found, even for small fields.^{5,6,27} Similarly in heterogeneous slab phantoms, dose deviations below 2% were seen for profiles and depth dose curves except in the vicinity of air cavities where dose deviations increased up to 4.5%.^{3,7,12} Anthropomorphic breast, lung, and H&N phantoms have been used for 3D dose evaluation. In the breast phantom, a very good agreement to MC calculation was found for a treatment plan consisting of two tangential fields and $G(2,2) = 99.9\%$. Film measurements of a variety of treatment plans delivered to lung phantoms showed $G(3,3)$ pass rates in the range 96% to 100% with Acuros overestimating dose to the lung by up to 5%.^{11,12} Finally, for the head and neck phantom measurements, considerably lower pass rates were observed. Film measurements of IMRT and VMAT plans delivered to an anthropomorphic phantom showed $G(5,3)$ pass rates of 89–95% with the lowest pass rate for VMAT plans.¹⁰ Slightly better results were found by Kan *et al.* with $G(3,3) = 91\%$ and $G(2,2) = 78\%$.¹³ A comparison to MC for 10 patients treated at different anatomical regions and with different treatment techniques found GAI below 90% for the PTV in four of the patients. The low GAI values were found for brain, breast, lung, and vaginal cancer patients.¹⁵ Most of these studies were based on former versions of Acuros (v10 or v11). For each new version, minor improvements are made to the algorithm.¹⁵ The findings of the former studies differ considerably, showing large discrepancies between Acuros and MC calculations for a number of cases. We can only conclude that continuous improvements leading up to v13.7 were responsible for the consistently higher agreement seen here.

Both algorithms were commissioned on the basis of water phantom measurements and MLC calibration experiments alone. A good agreement for these water measurements may not necessarily be a marker of quality, as it could also result from over-fitting. The fact that the algorithms agree to the same extent for both static and dynamic, regular and very irregular, small and large fields suggests that the distinctly different beam

model commissioning processes of both algorithms arrive at accurate and generalizable radiation source models.

In this study, high-quality measurements of the parameters used for the beam model were used. Different measuring devices were used for depth dose curves, profiles, and output factors depending on the field size.¹¹ This is important for creating the source model and for obtaining the extremely good agreement between the algorithms.

In the current study, measurements of DLG and T were used for the configuration of both algorithms. This is of utmost importance for both IMRT and VMAT calculations, building on very minute field openings. A very good agreement was found for all five different treatment sites and techniques, and $G(2,2) > 97\%$ for all plans and in 88% of the patients $G(2,2) \geq 99\%$. In most points, agreement within 1% and 1 mm was found and $G(1,1) \geq 90\%$ for 68% of the patients. The discrepancies between the two algorithms may stem from subtle differences in the modeling of MLC properties such as leaf position calibration, interleaf leakage, where production tolerances begin to play a role, or leaf transmission, where (undisclosed) material composition and cross-sections are associated with uncertainties. Furthermore, different definitions of the tissue compositions were used. These factors could only be resolved with precision measurements, which are in turn challenging and associated with errors. Given these sources of uncertainties for dose computation for a real-world treatment machine, it would be difficult to achieve a much higher agreement between dose algorithms than seen in this study. This needs to be seen as an indication of the high quality of either algorithm.

5. CONCLUSION

Extremely good agreement was seen between the Boltzmann equation solver and MC. High-accuracy dose computation algorithms have matured to a level that their differences are below common experimental detection thresholds for clinical treatment plans. Aside from residual differences which could be traced back to implementation details and fundamental cross-section data, both algorithms arrive at identical dose distributions.

ACKNOWLEDGMENT

MedCom GmbH, Darmstadt, Germany is acknowledged for providing a free license of the ProSoma software.

CONFLICTS OF INTEREST

MA and MS jointly founded and own Scientific RT GmbH.

^{a)}Author to whom correspondence should be addressed. Electronic mail: lone.hoffmann@aarhus.rm.dk; Telephone: (+45) 78464452; Fax: (+45) 78464522.

REFERENCES

- Ahnesjö A, Aspradakis MM. Dose calculations for external photon beams in radiotherapy. *Phys Med Biol*. 1999;44:R99–R155.
- Knöös T, Wieslander E, Cozzi L, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. *Phys Med Biol*. 2006;51:5785–5807.
- Vassiliev ON, Wareing TA, McGhee J, Failla G, Salehpour MR, Mourtada F. Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. *Phys Med Biol*. 2010;55:581–598.
- Fogliata A, Nicolini G, Clivio A, Vanetti E, Mancosu P, Cozzi L. Dosimetric validation of the Acuros XB advanced dose calculation algorithm: fundamental characterization in water. *Phys Med Biol*. 2011;56:1879–1904.
- Fogliata A, Nicolini G, Clivio A, Vanetti E, Cozzi L. Accuracy of Acuros XB and AAA dose calculation for small fields with reference to RapidArcVR stereotactic treatments. *Med Phys*. 2011;38:6228.
- Kron T, Clivio A, Vanetti E, et al. Small field segments surrounded by large areas only shielded by a multileaf collimator: comparison of experiments and dose calculation. *Med Phys*. 2012;39:7480.
- Bush K, Gagne IM, Zavgorodni S, Ansbacher W, Beckham W. Dosimetric validation of Acuros[®] XB with Monte Carlo methods for photon dose calculations. *Med Phys*. 2011;38:2208–2221.
- Fogliata A, Nicolini G, Clivio A, Vanetti E, Cozzi L. Dosimetric evaluation of Acuros XB advanced dose calculation algorithm in heterogeneous media. *Radiat Oncol*. 2011;6:82.
- Han T, Mikell JK, Salehpour M, Mourtada F. Dosimetric comparison of Acuros XB deterministic radiation transport method with Monte Carlo and model-based convolution methods in heterogeneous media. *Med Phys*. 2011;38:2651.
- Han T, Mourtada F, Kisling K, Mikell J, Followill D, Howell R. Experimental validation of deterministic Acuros XB algorithm for IMRT and VMAT dose calculations with the Radiological Physics Center's head and neck phantom. *Med Phys*. 2012;39:2193.
- Hoffmann L, Jørgensen MK, Muren LP, Petersen JBB. Clinical validation of the Acuros XB photon dose calculation algorithm, a grid-based Boltzmann equation solver. *Acta Oncol*. 2012;51:376–385.
- Han T, Followill D, Mikell J, et al. Dosimetric impact of Acuros XB deterministic radiation transport algorithm for heterogeneous dose calculation in lung cancer. *Med Phys*. 2013;40:051710.
- Kan MW, Leung LH, So RW, Yu PK. Experimental verification of the Acuros XB and AAA dose calculation adjacent to heterogeneous media for IMRT and RapidArc of nasopharyngeal carcinoma. *Med Phys*. 2013;40:031714.
- Ojala JJ, Kapanen MK, Hyödynmaa SJ, Wigren TK, Pitkänen MA. Performance of dose calculation algorithms from three generations in lung SBRT: comparison with full Monte Carlo-based dose distributions. *J Appl Clin Med Phys*. 2014;15:4662.
- Ojala J, Kapanen M. Quantification of dose differences between two versions of Acuros XB algorithm compared to Monte Carlo simulations — the effect on clinical patient treatment planning. *J Appl Clin Med Phys*. 2015;6:213–225.
- Varian Medical Systems. Eclipse Photon and Electron Algorithms Reference Guide; 2015.
- Van Esch A, Bohsung J, Sovari P, et al. Acceptance tests and quality control (QC) procedures for the clinical implementation of intensity modulated radiotherapy (IMRT) using inverse planning and the sliding window technique: experience from five radiotherapy departments. *Radiother Oncol*. 2002;65:53–70.
- LoSasso T, Chui C, Ling CC. Physical and dosimetric aspects of a multileaf collimation system used in the dynamic mode for implementing intensity modulated radiotherapy. *Med Phys*. 1998;25:1919.
- Kawrakow I, Fippel M, Friedrich K. 3D electron dose calculation using a Voxel based Monte Carlo algorithm (VMC). *Med Phys*. 1996;23:445–457.
- Kawrakow I, Fippel M. Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC. *Phys Med Biol*. 2000;45:2163–2183.
- Kawrakow I. Improved modeling of multiple scattering in the Voxel Monte Carlo model. *Med Phys*. 1997;24:505–517.
- Kawrakow I, Mainegra-Hing E, Rogers DWO, Tessier F, Walters BRB. The EGSnrc Code System: Monte Carlo Simulation of Electron and Photon Transport. Ionizing Radiation Standards, National Research Council Canada, NRCC Report PIRS-701; 2018.
- Sikora M, Dohm O, Alber M. A virtual photon source model of an Elekta linear accelerator with integrated mini MLC for Monte Carlo based IMRT dose calculation. *Phys Med Biol*. 2007;52:4449–4463.
- Sikora M, Alber M. A virtual source model of electron contamination of a therapeutic photon beam. *Phys Med Biol*. 2009;54:7329–7344.
- Sikora M. Virtual Source Modelling of Photon Beams for Monte Carlo Based Radiation Therapy Treatment Planning, PhD thesis. <http://bora.uib.no/handle/1956/5135?show=full>
- Chui C-S, Spiro S, LoSasso T. Testing of dynamic multileaf collimation. *Med Phys*. 1996;23:635–641.
- Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys*. 2003;30:2455–2464.
- Ju T, Simpson T, Deasy JO, Low DA. Geometric interpretation of the gamma dose distribution comparison technique: interpolation-free calculation. *Med Phys*. 2008;35:879–887.
- Miften M, Olch A, Mihailidis D, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: recommendations of AAPM Task Group No. 218. *Med Phys*. 2018;45:e53–e83. <https://doi.org/10.1002/mp.12810>
- Clasie BM, Sharp GC, Seco J, Flanz JB, Kooy HM. Numerical solutions of the γ -index in two and three dimensions. *Phys Med Biol*. 2012;57:6981–6997.
- Ezzell GA, Burmeister JW, Dogan N, et al. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. *Med Phys*. 2009;36:5359–5373.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Table S1. Patient and treatment plan characteristics.

Table S2. DVH parameters for PTV coverage and dose to selected OARs.