doi.org/10.15392/2319-0612.2025.2815
2025, 13(1) | 01-17 | e2815
Submitted: 2024-11-28
Accepted: 2025-02-21

Comparative Study of the Performance of Two Treatment Planning Systems Using Tests from the MPPG 5b Guidelines, TECDOC 1583 and Venselaar's Confidence Limits

Vega D'Espaux, C[a,b*]; Lorenzo, F[b,c]; LaPazMastandrea.F[b]; Larragueta, N[b,d]

[a]Universidad Nacional De Costa.Rica,86-3000, Heredia, Costa Rica
[b]CentroHospitalario Pereira Rossell, 11200, Montevideo, Uruguay
[c]Hospital de Clinicas,11600, Montevideo, Uruguay
[d]Universidad Nacional De Entre Rios, RP11, Oro Verde, Entre Ríos, Argentina

*Correspondence: carlosvega127@hotmail.com

Abstract: This work presents a comprehensive comparative study of the performance of two Treatment Planning Systems (TPS) in radiotherapy: the Eclipse™ TPS by Varian Medical Systems, which utilizes the Analytical Anisotropic Algorithm (AAA), and the MIRS TPS by Nuclemed, which uses the Convolution and Superposition (CS) Algorithm. The evaluation of both systems was conducted following different methods, including point measurements, dose profiles, and measurements on phantoms with heterogeneities. The methodologies recommended by Venselaar, the MPPG 5b protocol by the American Association of Physicists in Medicine (AAPM), and the guidelines of the IAEA TECDOC 1583 for heterogeneity assessment were adopted for this purpose.The results indicated that the Eclipse™ system showed better overall performance in water measurements, with lower variability and greater compliance with gamma criteria (average above 95% in 3%, 2 mm profile tests). For the heterogeneity tests, Eclipse showed weakness in lung measurements, while MIRS encountered difficulty in bone calculations for our beam model and under the evaluation conditions of TECDOC 1583.

Keywords: Radiotherapy, Treatment Planning System Validation, Eclipse, MIRS.

doi.org/10.15392/2319-0612.2025.2815
2025, 13(1) | 01-17 | e2815
Submitted: 2024-11-28
Accepted: 2025-02-21

Estudio comparativo del desempeño de dos sistemas de planificación de tratamientos siguiendo las guias MPPG5b, TECDOC1583 y Limites de Confianza de Venselaar

Resumen: Este trabajo presenta un estudio comparativo exhaustivo del desempeño de dos Sistemas de Planificación de Tratamiento (TPS) en radioterapia: el TPS Eclipse™ de Varian Medical Systems, que utiliza el Algoritmo Analítico Anisotrópico (AAA), y el TPS MIRS de Nuclemed, que utiliza el Algoritmo de Convolución y Superposición (CS). La evaluación de ambos sistemas se llevó a cabo siguiendo diferentes métodos en los que se consideraron medidas puntuales, perfiles de dosis y medidas en fantoma con hetereogeneidades. Con este objetivo, se adoptaron las metodologías recomendadas por Venselaar, el MPPG 5b de la American AssociationofPhysicists in Medicine (AAPM) y las directrices del TECDOC 1583 del OIEA para la evaluación de heterogeneidades.Los resultados indicaron que el sistema Eclipse™ con AAA mostró un mejor desempeño general en medidas en agua, con menor variabilidad y mayor concordancia con los criterios gamma.Para el caso de test de heterogeneidades Eclipse mostro debilidad en las medidas en pulmón y MIRS encontró dificultad para el cálculo en hueso para nuestro modelo de haz y según las condiciones de evaluación del TECDOC 1583.

Palabras Clave: Radioterapia, Validación de Sistemas de Planificación de Tratamiento, Eclipse, MIRS.

1. INTRODUCTION

This work was carried out in the context of commissioning a Varian Unique linear electron accelerator, installed in the radiotherapy service at the Pereira Rossell Hospital in Montevideo, Uruguay. This model has a single 6MV beam and a Varian Millenium 120 collimation system. The commissioning of linear electron accelerators requires rigorous validation of treatment planning systems (TPS). The importance of this validation lies in the fact that any inaccuracies in dose calculations can significantly affect the quality and safety of radiotherapy treatment [1].

This study focuses on evaluating two TPS used at the Pereira Rossell Hospital Center in Montevideo: Varian Medical Systems Eclipse™ and MIRS by Nuclemed. Both TPS use model-based algorithms to calculate radiation dose distributions but differ in the characteristics of these models. Eclipse™ uses the Analytical Anisotropic Algorithm (AAA), while MIRS employs the Convolution and Superposition (CS) algorithm with Collapsed Cones.

The analysis aims to assess the performance of both TPS and their respective algorithms. Validation tests were selected to cover a range of scenarios, from simple cases such as point dose calculations in a homogeneous medium to dose profiles and extending to clinically relevant conditions like dose calculations in the CIRS Thorax Phantom. Specifically, tests 5.5 and 5.8 from the American Association of Physicists in Medicine (AAPM) MPPG 5b guideline[2], as well as the tests outlined in Appendices A and E of the IAEA TECDOC 1583 [3], were employed.

The MPPG 5b guideline was followed for the commissioning of the linear accelerator. However, since the MPPG 5b provides limited specificity regarding the suggested tests for simple geometries, the tests from Appendix E of the TECDOC 1583 were adopted for this purpose. On the other hand, to evaluate heterogeneities, the MPPG 5b recommends measurements using slab phantoms. Instead, we opted for the tests from Appendix A of the TECDOC 1583, as these tests, conducted with the CIRS Thorax Phantom, are more representative of clinical conditions.

2. MATERIALS AND METHODS

Venselaar's confidence limit is a tool designed to summarize large volumes of data generated in TPS validation, reducing complexity to a metric that reflects average accuracy and dose calculation variability. Instead of analyzing each data point individually, measurements are grouped into categories based on beam geometry and measurement point location (on-axis, off-axis, and out-of-field measurements), with specific tolerances for each group.

Venselaar proposes tolerances based on field complexity, distinguishing between open fields and those with wedges, both on-axis and off-axis. Tolerances are defined in terms of mean deviation (systematic errors) and standard deviation (variability due to random errors), and the confidence limit is calculated as follows:

$$\Delta = |\text{average deviation}| + 1.5 \times SD \qquad \text{Ec 1}$$

This approach balances the influence of systematic and random errors, providing an overall view of TPS performance.

For this study, we used the validation measurements suggested by Appendix E of TECDOC 1583, grouping them according to Venselaar's measurement sets. The tolerances proposed by Venselaar and those of TECDOC 1583 are detailed in Table 1, showing values according to field type (open or wedged, in our case) and measurement point position.

For this initial evaluation, a PTW dosimetry system was used, consisting of a Unidos E electrometer and a Farmer chambermodel TN 30013, using the Sun Nuclear 3D Scanner for the spatial positioning of the point measurements.

Table 1: Tolerances from Appendix E of TECDOC 1583 (first and second columns) and Venselaar's confidence limits (third and fourth columns)

Group	Tolerance TECDOC1583 Open	Tolerance TECDOC1583 Wedge	ToleranceVenselaar Open	ToleranceVenselaarWedge
Onthebeamaxis	2%	3%	2%	3%
Off thebeamaxis	3%	3%	3%	3%
Out of the field edges	3%	4%	3%	4%

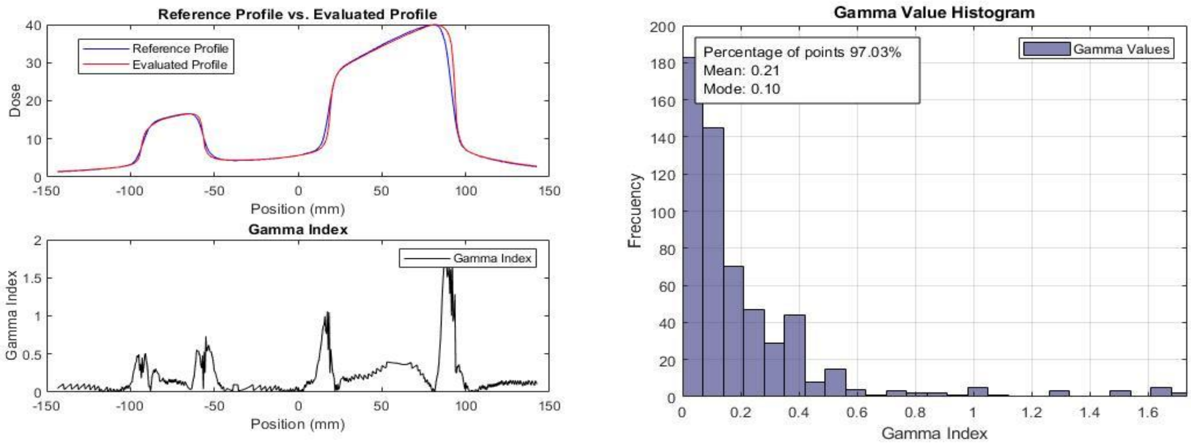
In the comparative analysis using MPPG 5b beam scanning tests, tests 5.5 (Complex Conformation) and 5.8 (Oblique Incidence) were evaluated for both open fields and those with a 60° wedge. A global Gamma analysis with a 3%, 2 mm agreement criterion, normalized at Dmax, was performed between measured and calculated profiles, with no threshold

While the percentage of points meeting this criterion is an important indicator for dosimetric evaluation, it alone does not provide a complete picture of system performance. Therefore, other parameters were included in the analysis, such as the mean and mode of the Gamma Index. The mean provides an overall perspective of system performance, while the mode identifies the most common index values.

Measurements for this purpose were taken using a cylindrical ionization chamber model SNC125c from Sun Nuclear. This detector offers a volume of 0.125 cc with a smaller diameter than ionization chambers of the same volume from other brands, allowing measured profiles with less volumetric averaging.

A Matlab-based application was developed, inspired by the work of Duong, T. T. et al[5], which allows extracting the dose profile to be evaluated from the RTDose files exported by the TPS. This application also enables Gamma comparison between measured and calculated profiles, displaying in one figure the overlay of both profiles with the Gamma comparison, and in another figure the respective histogram of evaluated points, showing the percentage of points meeting the criterion, their mean, and mode (see Figure 1).

Figure 1: On the left, Gamma comparison view between measured crossline profile (Reference Profile) and calculated profile (Evaluated Profile) for test 5.5 with a 60° physical wedge. On the right, the histogram showing the respective percentage values of points that meet the Gamma criterion, the Mean, and the Mode

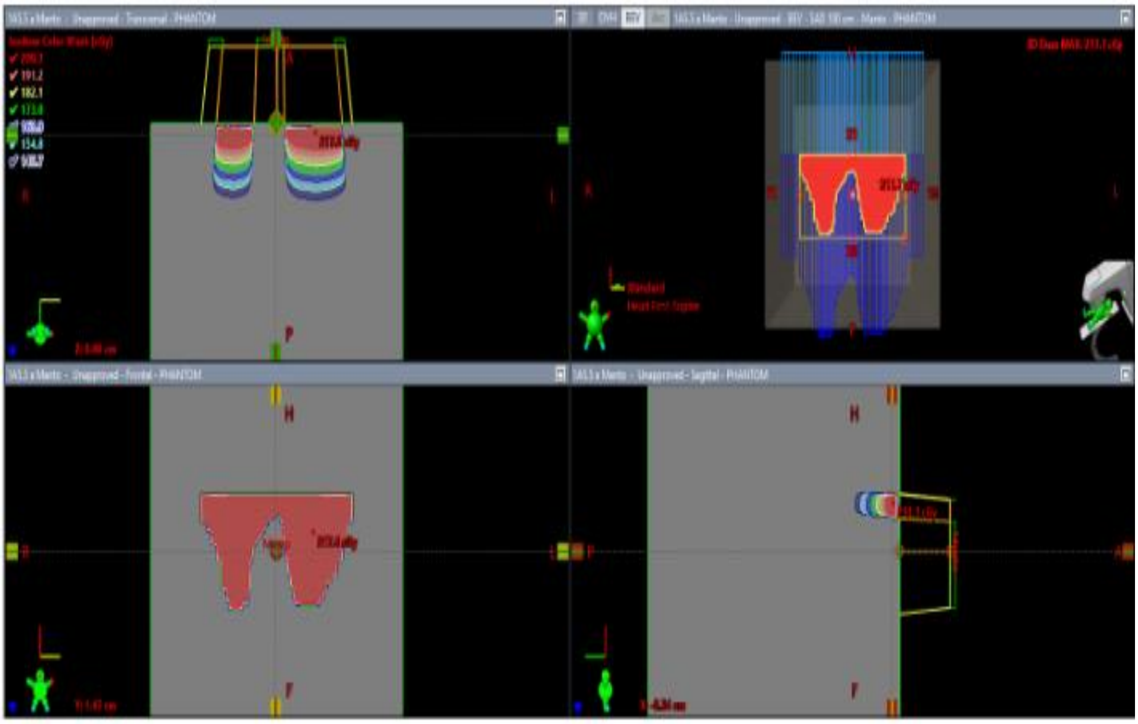


For test 5.5, which evaluates MLC leaf transmission and Output Factor effects, a large field was generated with extensive blocking. MPPG 5b recommends following the Photon Test 3 procedure from IAEA TRS430. Profiles were measured at depths of 5 cm, 10 cm, and 25 cm, with a point dose measurement taken at a point within each profile to convert relative dose to absorbed dose by multiplying it by the point dose. Additionally, a depth dose scan was performed for each case, transformed to absorbed dose using the same point measurements. The MPPG5b guide recommends performing comparisons in absorbed dose rather than relative dose.

The MPPG5b guide for this type of test establishes a different criterion for tolerances, considering whether evaluation points are in a uniform dose zone, outside the field, or in a high-gradient area. In our case, to evaluate both algorithms, we used a fixed gamma criterion of 3%, 2mm for the entire dose profile.

Due to the mechanical characteristics of the 3D Scanner phantom, the profiles were scanned along the field axes, generating Crossline and Inline profiles.

Figure 2: Image of MPPG5b test 5.5 showing different sections of the virtual phantom generated in Eclipse



To assess heterogeneities, the Thorax CIRS Model 002LFC phantom with Farmer chamber inserts was used. The eight tests suggested by TECDOC 1583 were performed, exposing the TPS to various calculation complexities with heterogeneity changes.

The MIRS system works by default with a dose calculation matrix of 128x128x128. The DICOM RT file exported from this TPS only contains the 128x128x128 matrix, not allowing export of a matrix with smaller voxels, which results in lower resolution in dose profiles for this analysis but does not necessarily impact dose calculation. The dose profiles obtained from the export of the DICOM RT files have a resolution of 3 mm. For the planning module, the MIRS TPS allows defining smaller calculation matrices, achieving higher resolution in areas of interest for patient treatment.

While the Matlab application interpolates missing points for exported profiles, this interpolation is linear and, therefore, increases errors when dose gradients are high, and the resolution does not capture dose changes.

For beam modeling in each TPS, it is necessary to indicate that this process presents differences for each system. For Eclipse, a series of machine parameters are preconfigured. For MIRS, modeling involves manual entry of these parameters and an iterative beam modeling process to match measured profiles with calculated ones. This makes beam modeling in the MIRS TPS user-dependent. Therefore, in this study, when referring to MIRS performance, we are considering our specific model, which could differ for another user.

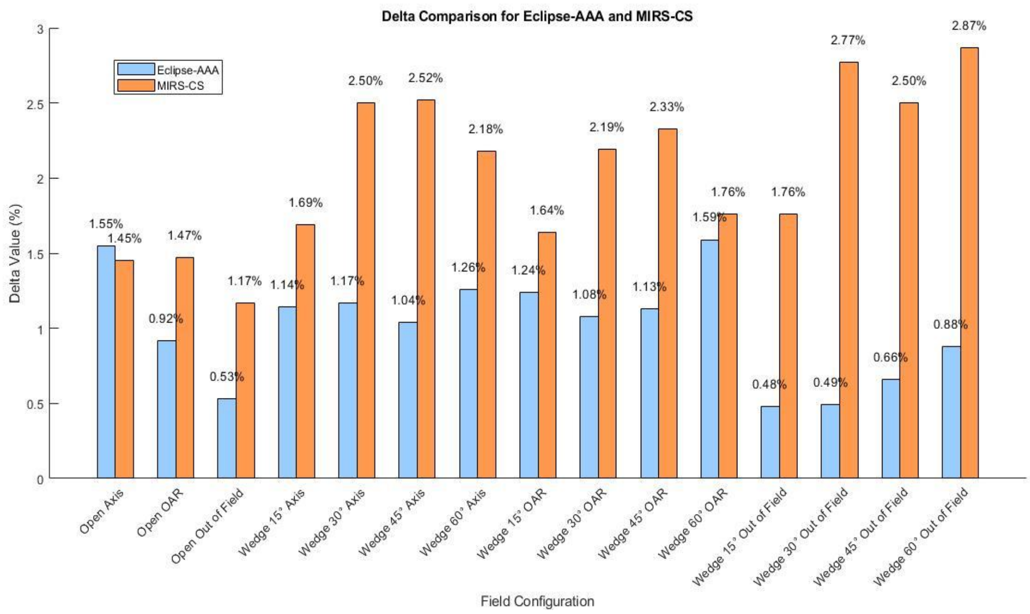
3. RESULTS AND DISCUSSIONS

3.1. VenselaarConfidence Limits

A total of 20 point measurements were taken for each open field and 28 for each field with a wedge. Given that there are 3 open fields and 12 fields with a wedge, the total number of measured points used for this comparison was 396.

Figure 3 shows a bar graph of the confidence limit results for both TPS, grouped by open or wedged fields, as well as by on-axis, off-axis, and out-of-field points. On average, the Eclipse AAA results for this parameter were closer to zero in most groups, indicating a better match between measured and calculated data. This analysis also showed that the deviations in Venselaar's confidence limit were higher for MIRS with CS.

Figure 3: Bar graph for the different groups of points considered for evaluation using confidence intervals



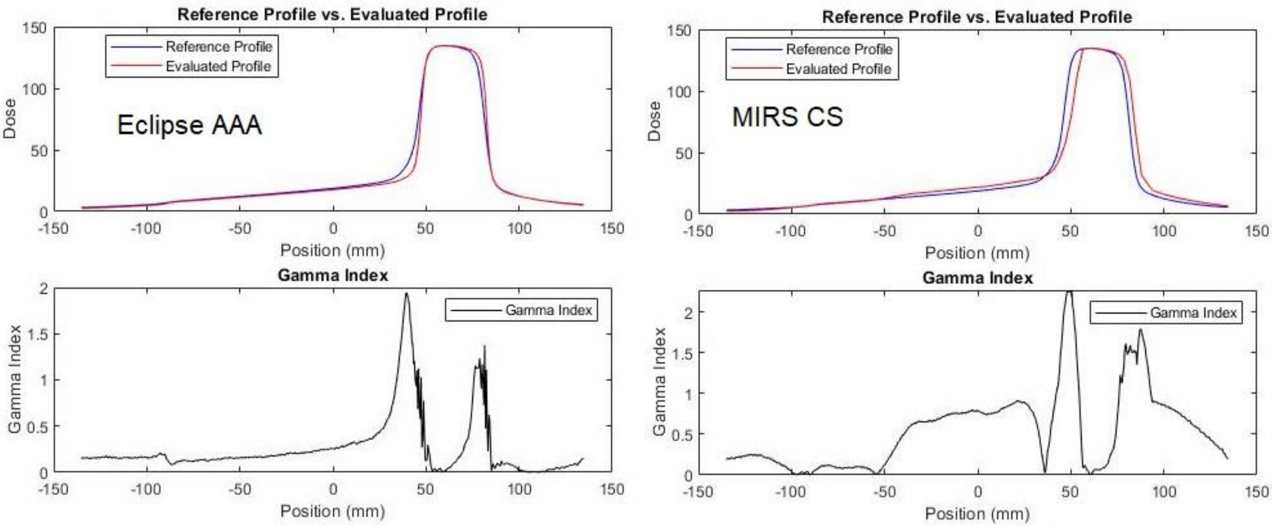
3.2 MPPG5b Tests 5.5 and 5.8

Figure 4 displays the Gamma comparison between profiles calculated by both TPS and the measured profile; these are inline profiles for test 5.5 at a depth of 10 cm. The positions below 50 mm on the graph correspond to the field blocked by the multi-leaf collimator (MLC), and positions greater than 100 mm correspond to the field blocked by the primary collimator. In this case, the least agreement between both TPS and the measurement is in the penumbra region.

However, a better agreement is observed for Eclipse, which could be due to Eclipse's modeling of the MLC, taking into account both its geometric and physical characteristics. On the other hand, the lower dose resolution in MIRS affects the percentage of points passing the Gamma function.

Figures 5 and 6 show the average Gamma evaluations of each profile for tests 5.5 and 5.8, respectively.

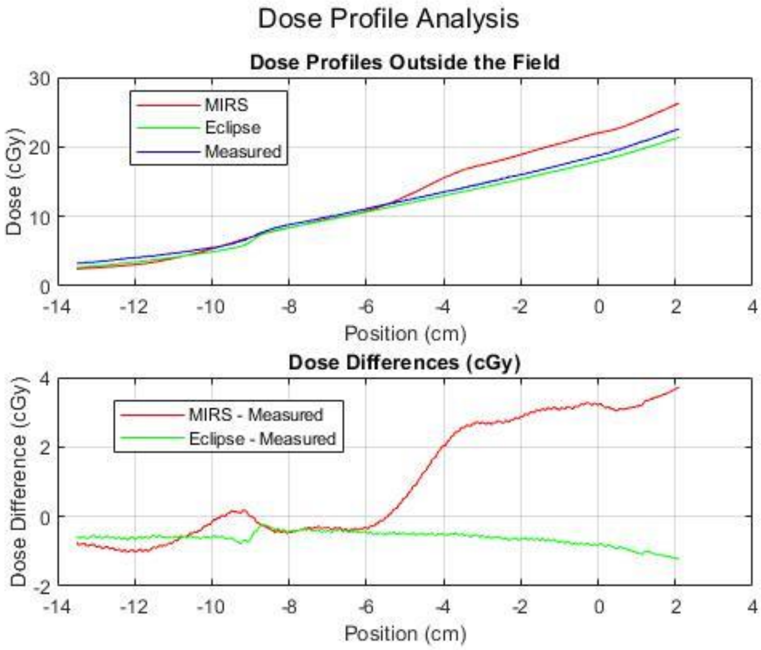
Figure 4: Inline dose profiles were evaluated with a global Gamma criterion of 3%, 2 mm, normalized at Dmax, at a depth of 10 cm for MPPG5b test 5.5. Left: Eclipse AAA; Right: MIRS CS



The analysis of test 5.5 reveals that MIRS has greater difficulty achieving agreement with measurements in inline profiles than in crossline profiles. This may be due to the fact that the source spot size is circular, and during beam modeling in the TPS, and more effort was focused on matching crossline profiles than inline profiles. Another possible reason could be that the MLC penumbra modeling in the TPS is more accurate in the transverse direction than in the longitudinal direction.

For test 5.8, a weakness was observed in both treatment planning systems when reproducing measurements for the inline profile at a depth of 25 cm. This profile is completely outside the irradiation field and poses the additional challenge of oblique beam incidence. Although these are out-of-field dose points, the ability of TPS to accurately reproduce the dose in these conditions is important, especially for patients with pacemakers. It is essential to carefully monitor the dose received by these devices, as inadequate exposure could induce electrical currents and cause frequency mismatches in their operation [6].

Figure 5: Dose Profiles and Differences Outside the Field for MLC-Shaped Field: MPPG5b Test 5.5 at 10 cm Depth., Inline Profile



Taking the dose outside the field from Test 5.5 of the MPPG5b guideline as an example, Figure 5 illustrates the dose behavior in this region, which is shaped by the multileaf collimators (MLC). MIRS shows an area of overestimation, with a maximum difference of 3.725 cGy observed near the beam's penumbra, while Eclipse exhibits a maximum difference of 1.209 cGy. The maximum percentage deviations in this region are 25.8% for MIRS and 19.0% for Eclipse.

Tables 2 and 3 present the mean and standard deviations percentualsof all evaluated points for both tests, showing that, in general, Eclipse's means are closer to optimal results than those of MIRS. Eclipse also displays less variability between results, which is an additional factor to consider when evaluating a TPS.

Table 2: Means and standard deviations of all evaluated points for MPPG5b test 5.5 on both TPS

Algorithm	Mean percentage	Standard Deviation of Percentage	Mean of Mean	Standard deviation of Mean
MIRS CS	95.13	4.47	0.4	0.15
EclipseAAA	96.04	3.49	0.24	0.12

Table 3: Means and standard deviations of all evaluated points for MPPG5b test 5.8 on both TPS.

Algoritmo	Mean percentage	Standard Deviation of Percentage	Mean of Mean	Standard deviation of Mean
MIRS CS	81.84	25.2	0.67	0.74
EclipseAAA	91.4	19.38	0.39	0.36

Figure 6: Average Gamma Index using the 3%, 2mm criterion for MPPG5b test 5.5. Evaluations of open fields and 60° wedged fields for both TPS are shown. Evaluations are presented for crossline (CL) profiles, inline (IL) profiles, and depth dose (PDD).

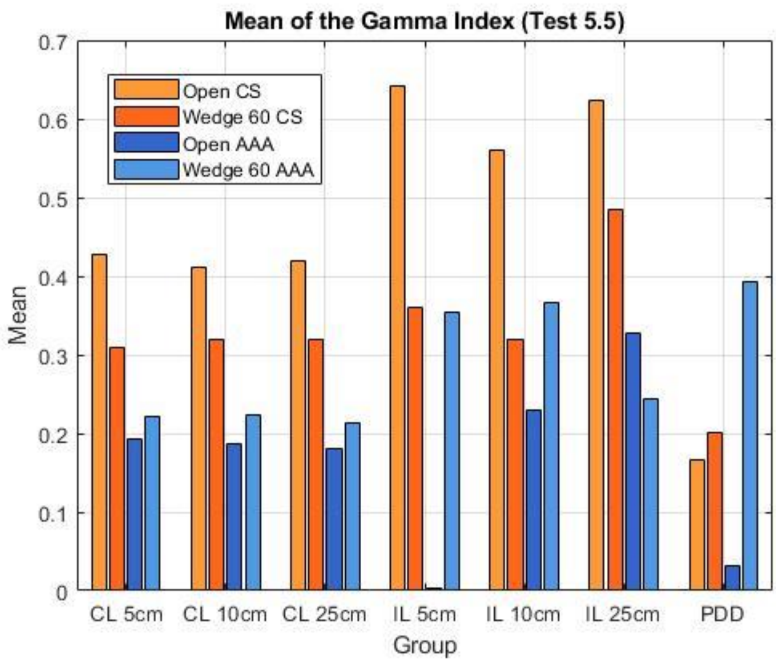
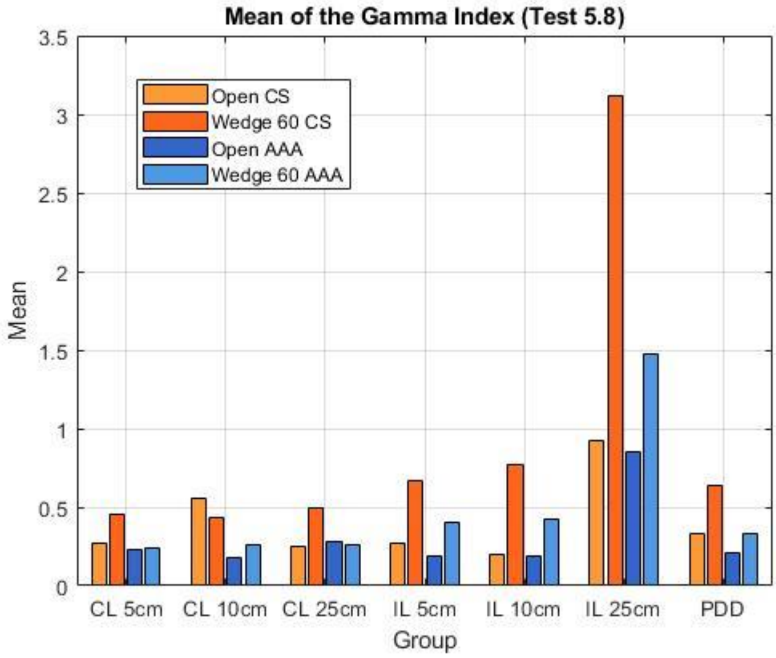


Figure 7: Average Gamma Index using the 3%, 2mm criterion for MPPG5b test 5.8. Evaluations of open fields and 60° wedged fields for both TPS are shown. Evaluations are presented for crossline (CL) profiles, inline (IL) profiles, and depth dose (PDD)



3.3 HeterogeneityTests

For this test, it is expected that CS would outperform AAA, as the characteristics of algorithm make it superior when calculating in heterogeneous conditions, as established by AAPM Task Group 329[7].

As shown in Table 4, discrepancies considered out of tolerance for CS appear in cases 2 and 4 of TECDOC 1583. In case 2, the beam passes through the lung, encounters a lack of tissue, and has a wedge as a beam modifier. As observed in Figure 7, the measurement point is located in an area with dose gradients in two directions. Therefore, the result could improve if a smaller chamber were used.

In case 4, CS shows a discrepancy in bone tissue dose in a posterior field. This result relates to the normalization point used in this test, which is at a lower dose, thus increasing the result. If only the percentage differences between the calculated and measured doses are considered, the percentage difference is 3.3%.

For AAA, discrepancies are shown in the lung for case 4, both at the beam's entry and exit. This result can be attributed to the fact that, being a pencil-beam type algorithm, it does not resolve lung dose accurately due to the lack of radiation transport corrections from multiple directions.

Table 4: Results of TECDOC 1583 heterogeneity tests by IAEA

Case	Point	CS MIRS	AAA Eclipse	Tolerance	Field
Case 2	1	-3.29%	1.73%	3%	LeftLat
Case 4	10	-4.63%	-0.97%	3%	Posterior
Case 4	6	1.32%	4.21%	3%	LeftLat
Case 4	6	1.36%	-3.07%	3%	RightLat

Figure 8: Case 2 planning in MIRS TPS calculated with CS on CIRS Thorax Phantom.

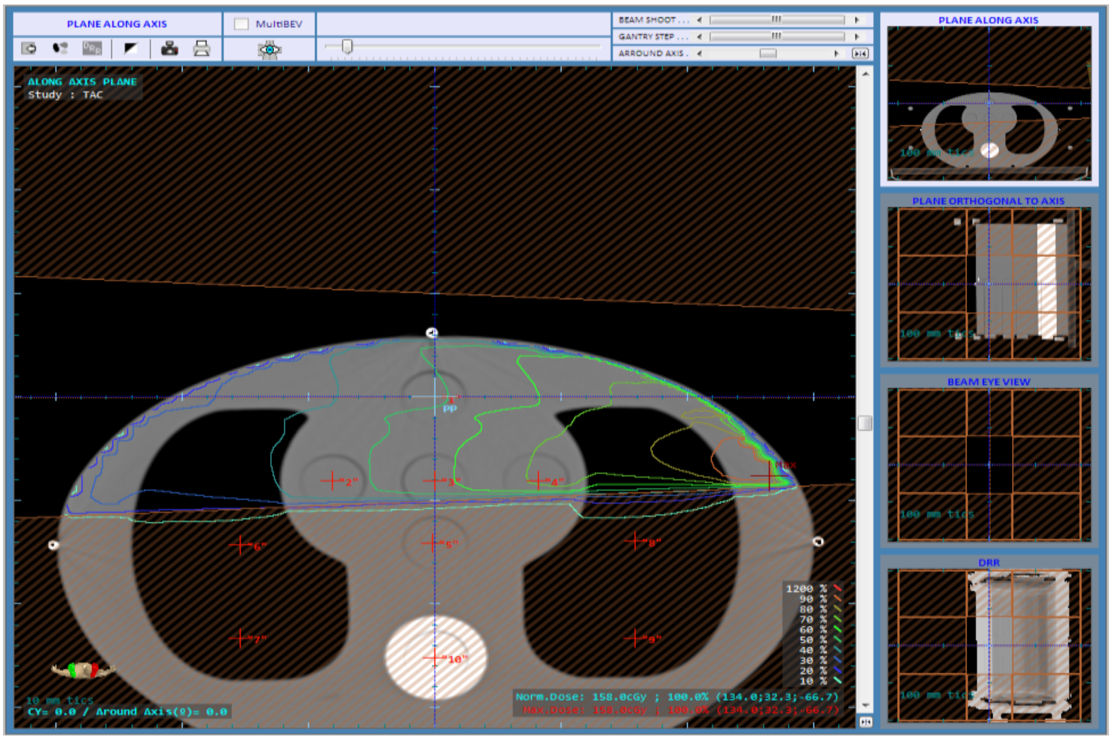
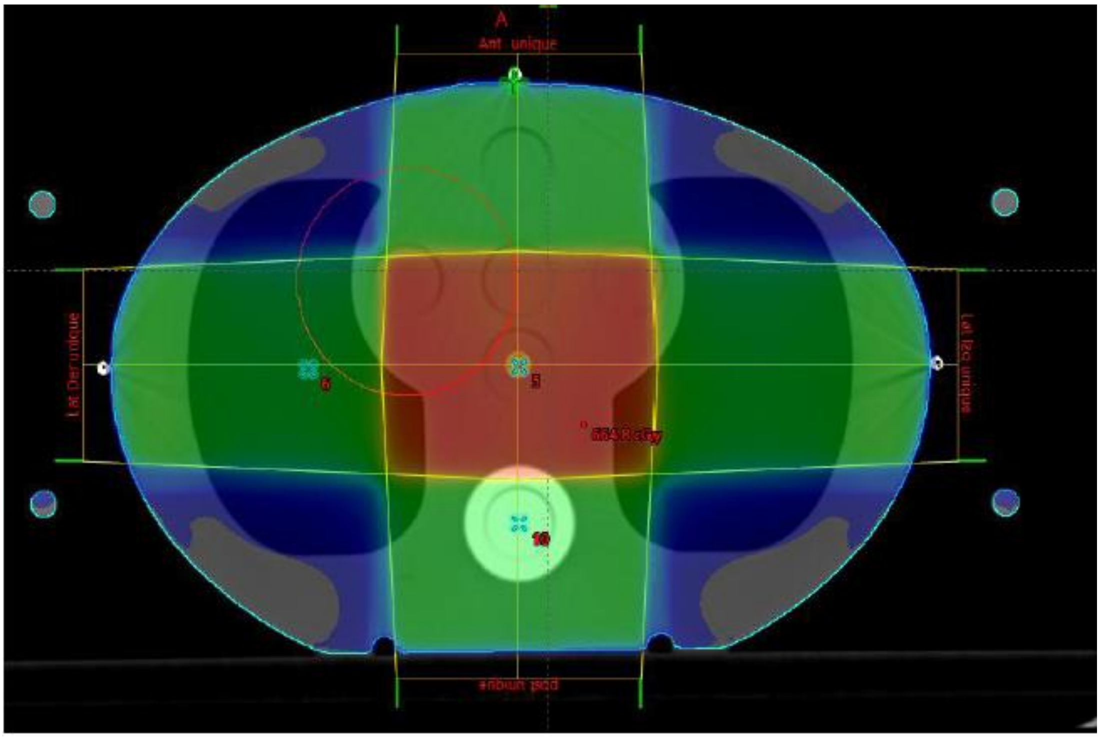


Figure 9: Case 4 planning in Eclipse TPS calculated with AAA on CIRS Thorax Phantom.Point 6 in lung-equivalent tissue and the normalization point (Point 5) used for this test are highlighted



4. CONCLUSIONS

In water measurements, Eclipse showed better performance than MIRS, with less variability in the results. However, in the heterogeneity tests, designed to simulate complex conditions resembling real patient scenarios, the differences between the systems were not as significant. It was observed that MIRS results could enhance if a smaller chamber were used, reducing the impact of high dose gradients on measurements. Furthermore, MIRS's convolution and superposition algorithm proved robust under these conditions, confirming its capability to handle complex clinical scenarios, making it a reliable tool for 3D planning in patients with significant heterogeneities.

Regarding the protocols used for this comparison, it can be said that MPPG5b provides a range of challenging conditions where the radiation beam is subjected to varying complexities. This allows for a deeper understanding of the strengths and weaknesses of the models and calculation algorithms.

Although the literature highlights convolution and superposition algorithms as superior to pencil-beam-style algorithms, we can also conclude that the versatility of a TPS does not depend solely on its calculation algorithm. The quality of the dose calculation is closely linked to the beam modeling within the TPS. Our conclusions are drawn from the perspective of our own beam models created as users of Eclipse and MIRS.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Centro Hospitalario Pereira Rossell for providing the facilities and support necessary for the development of Medical Physics within the Radiotherapy Department.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] *Van Dyk, J., Rosenwald, J.-C., Fraass, B., Cramb, J., Ionescu-Farca, F., & Sharpe, M.* IAEA Technical Reports Series No. 430: Commissioning And Quality Assurance Of Computerized Planning Systems For Radiation Treatment Of Cancer. **Medical Physics - MED PHYS,** v. 33, p. 561-561, 2006.

[2] *Geurts, M. W., Jacqmin, D. J., Mihailidis, D. N., Ohrt, J. D., Wingreen, N. E., Jones, L. E., & Smilowitz, J.* B. AAPM Medical Physics Practice Guideline 5.b: Commissioning and QA

of treatment planning dose calculations, Megavoltage photon and electron beams. Journalof Applied Clinical **Medical Physics**, v. 23,p. e13641, 2022.

[3] International Atomic Energy Agency. **Commissioning of Radiotherapy Treatment Planning Systems: Testing for Typical External Beam Treatment Techniques**. Vienna: IAEA, 2008. ISBN 978-92-0-100508-3.\

[4] *Venselaar, J., &Welleweerd, H*. Application of a test package in an intercomparison of the photon dose calculation performance of treatment planning systems used in a clinical setting. **RadiotherapyandOncology**, v. 60, n. 2, p. 203-213, 2001.

[5] *Duong, T. T., Omer, H., Le, C. Q., Nguyen, X. H., Truong, V. M., Abdelmoneim, S., Essam, M., Hind, T., Tamam, N., & Bradley, D. A*. An open-source software for calculating 1D gamma index in radiation therapy. **Journalof King Saud University - Science**, v. 35, p. 102937, 2023.

[6] *Miften, M., Mihailidis, D., Kry, S. F., Reft, C. S., Esquivel, C., Farr, J. B., Followill, D. S., Hurkmans, C., Liu, A., Gayou, O., Gossman, M. S., Mahesh, M., Popple, R. A., Prisciandaro, J. I., & Wilkinson, J. D*. Management of radiotherapy patients with implanted cardiac pacemakers and defibrillators: A report of the AAPM TG-203**. Medical Physics**, v. 46, n. 12, p. e757-e788, 2019.

[7] *Kry, S. F., Feygelman, V., Balter, P., Knöös, T., Ma, C.-M. C., Snyder, M., Tonner, B., & Vassiliev,* O. N. AAPM Task Group 329: Reference dose specification for dose calculations: Dose-to-water or dose-to-muscle? MD Anderson Cancer Center, Houston, TX, USA; Moffitt Cancer Center, Tampa, FL, USA; Skåne University Hospital and Lund University, Malmo, Sweden; Fox Chase Cancer Center, Philadelphia, PA, USA; Wayne State University, Detroit, MI, USA; Ackerman Cancer Center, Jacksonville, FL, USA, 2019.

LICENSE