Relationship between IMRT Complexity, Monte Carlo Dose Verification, and Phantom-Based QA In SBRT Plans

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Benedikt Rudek, PhD[1,2]
[1]Jamaica Plain VA Medical Center, [2]Boston University School of Medicine

U.S. Department of Veterans Affairs
Veterans Health Administration
Boston Healthcare System

Boston University School of Medicine

ABSTRACT

Purpose: To evaluate correlations among publicly accessible IMRT plan complexity metrics and their relationships with plan uncertainty, independent dose calculation, phantom-based QA, and log-file analysis.

Methods: Ten complexity metrics based on dose, monitor units (MU), and MLC parameters (e.g., open area, perimeter, leaf travel) were computed for 5-fraction lung SBRT plans using Radformation's ClearCheck and two open-source ESAPI scripts. All plans used 6 MV energy and were calculated with AcurosXB v16.1. Delivery was performed on a Varian TrueBeam, and measured with an SNC SRS-MapCHECK. Absolute gamma passing rates were evaluated with 3% dose / 1 mm distance / 10% threshold criteria. Independent 3D dose calculations were performed in RadMonteCarlo for initial and log-reconstructed plans using 2%/1 mm/10% criteria for the full matrix and PTV. Plan uncertainty was qualitatively assessed using ±3 mm x/y/z isocenter shifts in the Eclipse Plan Uncertainty tool.

Results: Eight metrics incorporating MLC open area were strongly correlated with each other (Pearson r = 0.56–1.0) but not with total MU or total leaf travel. The MU-duty cycle showed the strongest correlation (r = 0.72) with Monte Carlo dose differences. Log-based dose variations were minimal. Complexity metrics showed weak correlation with phantom-based QA results (max r ≈ 0.3). Plans with higher complexity exhibited more pronounced dose streaks and hotspots in nearby critical structures under simulated isocenter shifts.

Conclusions: For clinically approved SBRT IMRT plans, complexity scores showed limited correlation with consistently high gamma pass rates (>98% at 3%/1 mm). Simple indicators—such as MU-duty cycle—may provide sufficient warning of potential over-modulation and support improved plan robustness.

What do complexity metrics mean and how do they correlate?

Complexity metrics are used in intensity modulated radiation treatment planning to assess how modulated, deliverable and accurate a plan is. Radformation's ClearCheck and two ESAPI scripts provided on GitHub by Varian Medical Affairs Applied Solutions (MAAS)[1] and the Universitätsspital Zürich (USZ)[2] were used to extract metrics for 92 lungs SBRT plans (221 fields). For a subset of this data, the gamma passing rate was calculated with an independent 3D Monte Carlo dose algorithm of the original plan and based on log files, and for measurements with the SNC SRS-MapCheck. The complexity metrics can be sorted into the three categories:

1. **Fluence Metrics:**

These metrics analyze the modulation and intensity variations of the overall fluence map or beam profile rather than individual machine components.

- **Complexity Score (ClearCheck)**: Evaluates intensity modulation gradients or pixel variations across the total composite fluence map per beam.
- **Average Area (MAAS)**: Quantifies the average opening size relative to the overall planned radiation field or fluence profile.

2. **Deliverability Metrics:**

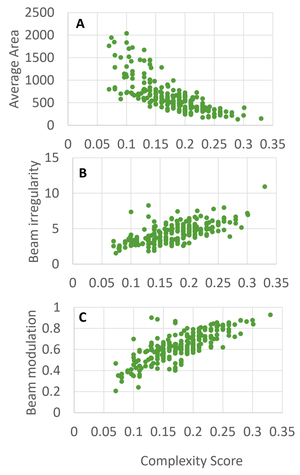
These metrics assess specific dynamic machine parameters, mechanical movements, and multileaf collimator (MLC) aperture configurations.

- **MU Duty Cycle (ClearCheck)**: Calculates the total MU divided by the prescribed dose..
- **ALPO - Average Leaf Pair Opening (MAAS)**: Directly tracks the physical distance between opposing MLC leaf pairs during delivery.
- **BI - Beamlet Interactivity / Variability (MAAS)**: Evaluates mechanical MLC sequencing variations between adjacent control points.
- **BA - Boundary Area (MAAS)**: Calculates the ratio of MLC perimeter to aperture area, indicating complex leaf sequencing.
- **BM - Beam Modulation / Leaf Travel (MAAS)**: Measures the total physical travel distance and speed of MLC leaves.
- **MMO - Mean Multileaf Collimator Opening (Zurich)**: Tracks the physical geometric openings of the MLC leaves across the plan.
- **OTL - Overall Travel of Leaves (Zurich)**: Sums the total distance travelled by the MLC leaves in mm per arc normalized on control points.

3. **Accuracy Metrics**

- **RadMonteCarlo Gamma Pass Rates**: Evaluates independent 3D dose calculation accuracy against the TPS algorithm.
- **Log File Analysis**: Compares actual machine delivery parameters against intended plan parameters.
- **SNC SRS MapCHECK (PSQA) Gamma Pass Rates**: Measures actual delivered dose distribution accuracy against the calculated plan.

The Complexity Score was used as benchmark for per-field metrics: Average area, ALPO, BA and MMO decrease with CS (Fig. 1A), while BI (Fig. 1B) and BM (Fig. 1C) increase with CS. There is no correlation between CS and OTL.



	CS	aArea	ALPO	BI	BA	BM	MMO	OTL
CS	1	0.73	0.77	0.66	0.735	0.789	0.778	0.278
avgArea		1	0.87	0.25	0.996	0.53	0.87	0.107
ALPO			1	0.57	0.869	0.645	0.937	0.107
BI				1	0.25	0.662	0.508	0.526
BA					1	0.518	0.856	0.124
BM						1	0.758	0.497
MMO [mm]							1	0.178
OTL [mm]								1

Table 1. Absolute Pearson correlation coefficient for field-based complexity metrics.

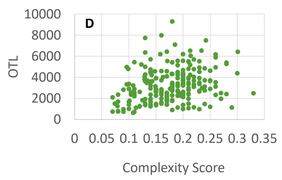


Figure 1. A/B/C/D Avg. Area, BI, BM and OTL plotted versus Complexity Score.

Do complex plans lead to low passing scores in independent dose calculation or measured plan QA?

Most departments calculate the independent dose verification and measure plan QA per plan and not per field. Thus, the field-based metrics have been averaged in Table 2 for comparison with the Monte Carlo calculation and PSQA. The MU duty cycle is easily calculated per plan and used as benchmark. Fig. 2 shows that it linearly correlates with the CS.

The correlation between complexity metrics and PSQA gamma passing scores is poor due to three reasons:

1. **2D vs 3D washout**: The metrics calculate microscopic 2D features, whereas PSQA measures a 3D dose distribution that results from overlapping hundreds of these 2D distributions where gradients wash out and phantom scatter dominates.
2. **Measurement Conditions:** The detector spacing under-samples small differences in MLC positions. An offset in dose or array calibration, or daily output variations are larger than dosimetric differences due to plan complexity.
3. **Gamma Analysis:** The Distance-to-Agreement criteria allows pixels in high-gradient regions to pass easily, masking the impact of a highly complex MLC sequence, which is why the 3% / 2mm criteria in Table 1 shows the lowest correlation.

However, there have been publications (such as Han et al.[3]) using large data sets and machine learning, where multiple metrics were used to predict on a Receiver Operating Characteristic (ROC) curve whether plans were likely to fail PSQA.

The 3D 2% / 1mm MC-dose verification correlates poorly with 2D complexity metrics but reasonably well with the MU duty cycle (Fig. 3). The passing rate drops for MU duty cycles larger than about 3.

Lung SBRT plans optimized with AXB 16.1 and recalculated with AXB v18.1 using the new enhanced leaf modelling show higher PTV mean doses. The difference in mean dose (Fig. 4.) scales with MU duty cycle illustrating the underlying differences in small field modelling.

	CS	MU total	MU-DC	aArea	ALPO	BI	BA	BM	MMO	OTL	MC2%	PSQA 2%1mm	PSQA 3%1mm	PSQA 3%2mm	3%1mm relative
CS	1	0.1	0.82984	0.77	0.81	0.7	0.77	0.81	0.82	0.24	0.481	0.059	0.15	0.06	0.24
MU total		1	0.3137	0.12	0.09	0.02	0.14	0.18	0.11	0.3	0.414	0.205	0.26	0.17	0.14
MU duty cycle			1	0.56	0.67	0.58	0.56	0.67	0.69	0.34	0.723	0.121	0.19	0.02	0.21
avgArea				1	0.92	0.34	1	0.58	0.92	0.07	0.208	0.069	0.06	0.13	0.07
ALPO					1	0.6	0.91	0.7	0.98	0.11	0.22	0.025	0.16	0.05	0.16
BI						1	0.35	0.71	0.57	0.47	0.258	0.081	0.11	0.04	0.25
BA							1	0.58	0.91	0.08	0.238	0.056	0.06	0.13	0.07
BM								1	0.77	0.44	0.363	0.154	0.23	0.01	0.23
MMO [mm]									1	0.16	0.279	0.033	0.16	0.05	0.16
OTL [mm]										1	0.171	0.213	0.24	0.20	0.20
MC2%1mm											1	0.323	0.37	0.02	0.18

Table 2. Pearson correlation coefficient for plan-based complexity metrics, MU verification and PSQA.

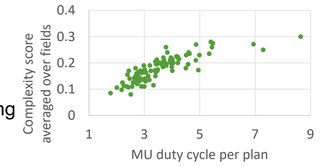


Fig. 2. Complexity Score vs. MU duty cycle.

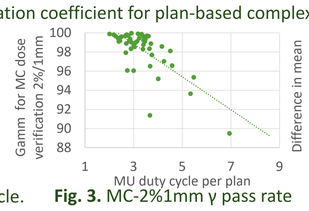


Fig. 3. MC-2%1mm γ pass rate

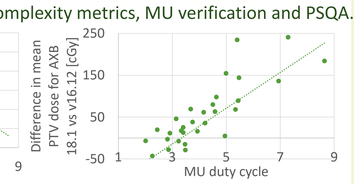


Fig. 4. PTV mean diff. in AXB 18.1

What is the clinical impact of high complexity?

The hypothesis is that high complexity makes treatment plans less robust against setup errors, intra-fractional shifts and anatomic changes. The conformity index (CI), Paddick CI (PCI), gradient index (GI), dose spillage (R50), PTV coverage V100 and D95 were calculated for the approved treatment plan. Then a 3mm shift was introduced in all six cardinal directions using the Eclipse uncertainty evaluation and the dose metrics were again calculated using an ESAPI script and compared to the baseline. The metrics degraded significantly, the degradation was, however, independent of plan complexity, possibly because the relatively large 3mm shift in dose distribution masked small contributions from complex field modulations.

Instead of PTV dose metrics, the total volume of tissue inside a 2cm bounding box around the PTV was calculated that experiences a dose change greater than ±5% of the prescription dose during the shift. This volume increases with average area for geometrical reasons (Fig. 5), with Paddick Conformity index which inherently requires an ultra-steep dose gradient (Fig. 6), and overall MLC leaf travel that indicate the dose grid is built out of many sweeping, overlapping sub-fields that are easily disrupted by shifts (Fig. 7).

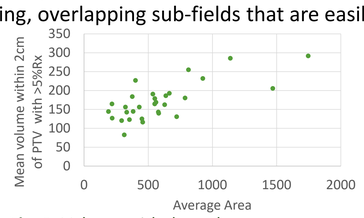


Fig. 5. Volume with dose changes vs. avg. area.

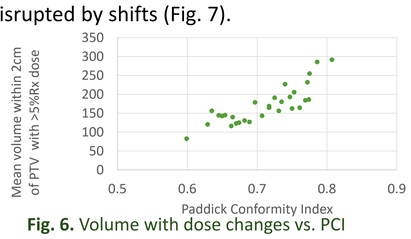


Fig. 6. Volume with dose changes vs. PCI

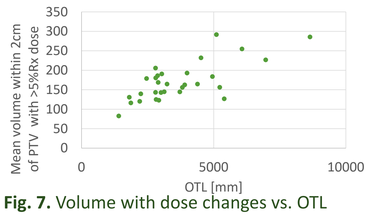


Fig. 7. Volume with dose changes vs. OTL

How can plan complexity be reduced?

- **MU constraints in the optimizer:** Lower total MUs inherently correlate with larger, more stable aperture shapes and reduced delivery time.
- **Optimizer setting (Eclipse):** The Aperture Shape Controller (ASC) works within the Photon Optimizer (PO) algorithm to limit plan complexity by penalizing small, irregular, or jagged beam apertures. A 'Moderate' setting have been shown to reduce unnecessary leaf modulation and cut down total Monitor Units (MUs) while maintaining dosimetric plan quality. The calculation however takes longer.
- **Normal Tissue Objective:** NTO limits complexity indirectly by managing the dose fall-off outside of targets which can replace aggressive constraints on specific OARs. The distance between PTV and start of the fall-off should be 2-3mm.
- **Leakage**: A collimator rotation of 30°-45° (instead of 0°) reduces inter leaf leakage and so does jaw-tracking which also reduces leaf-travel.
- **Minimum field size**: Increasing the minimum field size (e.g. from 0.5 to 2cm) prevents small over-modulated MLC fields, excessive MU, and small-field dose uncertainty.

Conclusion

Correlating complexity metrics with MC calculations, PSQA and clinical dose metrics is complicated, but keeping the simple MU duty cycle <~3 will make plans more robust.

References

1. Medical Affairs Applied Solutions ESAPI plug-in tool to quantify beam modulation complexity as defined in scientific literature: https://github.com/Varian-MedicalAffairsAppliedSolutions/MAAS-PlanComplexity
2. Klara Uher, Stefanie Ehrbar, Stephanie Tanadini-Lang, Riccardo Dal Bello, Reduction of patient specific quality assurance through plan complexity metrics for VMAT plans with an open-source TPS script, Zeitschrift für Medizinische Physik, Volume 34, Issue 4, 2024: https://github.com/medical-physics-usz/ESAPI_VMAT-SBRT_Complexity
3. Han C, Zhang J, Yu B, Zheng H, Wu Y, Lin Z, Ning B, Yi J, Xie C, Jin X. Integrating plan complexity and dosiomics features with deep learning in patient-specific quality assurance for volumetric modulated arc therapy. Radiat Oncol. 2023 Jul 11;18(1):116.

CONTACT

Benedikt Rudek
Jamaica Plain VA Medical Center
150 South Huntington Avenue
Boston, MA 02130-4817
Benedikt.Rudek@va.gov

The author declare no conflicts of interest.

© POSTER TEMPLATE BY GENIGRAPHICS® 1.800.790.4001 WWW.GENIGRAPHICS.COM