Efficient Validation of Photon Output Factors Using a Continuous Two-Dimensional Surface Model

B. Wolthuis[1,2], M. Schmidt[3], T. Mazur[3], E. Sajo[2], R. Guthier[2], P. Black[4], L. Farrell[1], A. Wood[3], S. Pani[3], Y. Hao[3], T. Sudlow[3], N. Knutson[3]
[1]Northwell Health, [2]University of Massachusetts Lowell, [3]Washington University in St. Louis School of Medicine, [4]Wake Forest University School of Medicine

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

STUDY APPROACH

Acquire High-Density Measurements
Establish the underlying OF behavior

Evaluate Reduced Sampling Geometries
Vary both measurement number and distribution

Fit Continuous Surface Models
Reconstruct values between measured field sizes

Validate Against Independent Measurements
Quantify model and measurement agreement

Identify an Efficient Measurement Strategy
Minimize burden while maintaining accuracy

BACKGROUND

Accurate output factors (OFs) are essential for commissioning and validating treatment planning system (TPS) beam models. Conventional commissioning relies on discrete OF tables acquired through numerous water-tank measurements across clinically relevant field sizes. Although recent automation strategies have reduced the time and effort required for data acquisition, determining the optimal number and distribution of measurements needed to accurately characterize the OF surface remains an open question.

PURPOSE

To evaluate whether a continuous log-log surface can accurately model photon output factors using strategically reduced measurement subsets while maintaining commissioning-grade accuracy.

Can commissioning-grade accuracy be maintained using substantially fewer, strategically selected measurements?

METHODS & MATERIALS

Experimental Setup

Linear accelerator: 6 MV FFF TrueBeam (Varian)
Water tank: Blue Phantom 2 (IBA)
Detector: 0.13 cc ionization chamber (IBA)
Electrometer: PC Electrometer (SNC)
Acquisition: Automated XML delivery with continuous data logging
Analysis: Python

Data Acquisition

Two independent 729-point high-density datasets were acquired across jaw-defined field sizes from 3 x 3 cm² to 40 x 40 cm². Field-size combinations were selected such that OF changes between neighboring sampled fields did not exceed 1% in the X, Y, or diagonal directions, resulting in greater measurement density in regions of rapid OF variation.

Model Development

Polynomial, neural network, and log-log surface models were evaluated for continuous prediction of OF as a function of X- and Y-jaw field size. Model performance was assessed using maximum percent difference, MAPE, RMSE, and R^2.

Edge + X reduced the measurement count by 73% relative to the published reference set while maintaining comparable accuracy.

SAMPLING STRATEGIES

Reduced measurement subsets were evaluated to determine how measurement number and distribution of measured field-size combinations across the X-Y jaw space affected reconstruction of the high-density OF surface. Measurement density was varied within each strategy.

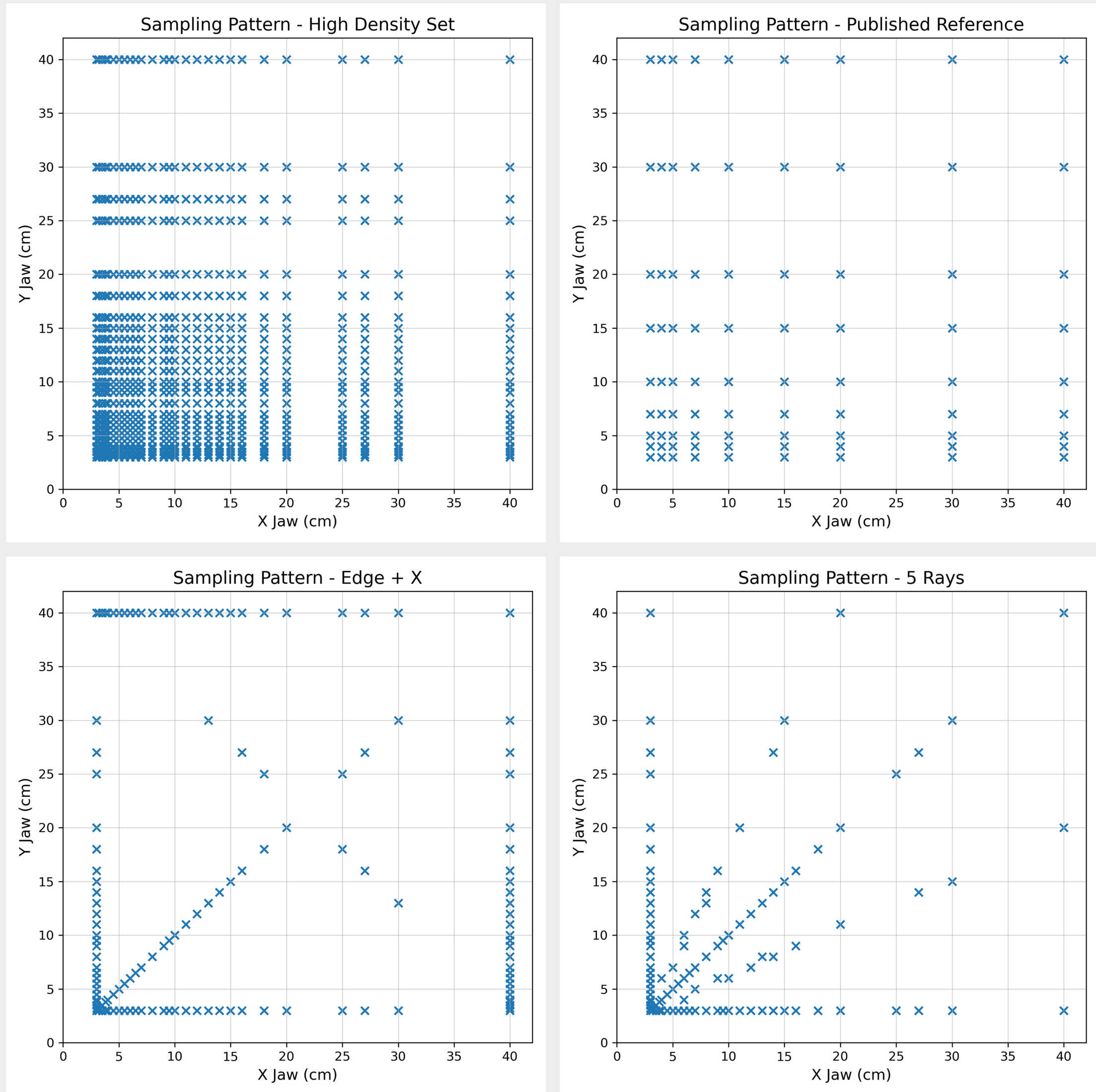


Figure 1. Measurement geometries evaluated in this study, including the high-density dataset, the published reference geometry, and two representative alternative subset strategies (Edge + X and 5 Rays).

RESULTS

Model Selection

The log-log surface model demonstrated the best overall agreement with the high-density reference dataset and was selected for subsequent subset analysis. Polynomial regression showed substantial systematic error, while neural network performance degraded as measurement density was reduced.

Table 1. Statistical comparison of polynomial, log-log, and neural network regression models evaluated using the published reference field-size set (81 measurements each).

	Polynomial Regression	Log-Log Regression	Neural Network Regression
Max % difference	3.67%	0.59%	0.81%
MAPE	1.10%	0.14%	0.23%
RMSE	0.0133	0.0017	0.0028
R^2	0.8978	0.9983	0.9954

Subset Performance

Measurement distribution strongly influenced sampling efficiency. Edge + X reproduced the high-density reference using only 22 measurements, achieving performance comparable to the 81-point published reference set and the 181-point 5 Rays strategy.

Table 2. Performance of representative sampling strategies spanning efficient, conventional, and ray-based geometries.

Strategy	Measurements	Max % Diff	MAPE (%)	RMSE	R^2
Published Reference	81	0.59	0.14	0.0017	0.9983
Edge + X	22	0.50	0.13	0.0016	0.9985
5 Rays	181	0.50	0.13	0.0016	0.9986

External Validation

The 22-point Edge + X subset reproduced an independently acquired high-density dataset with a maximum difference of 0.49%.

MODEL VALIDATION

The fitted log-log surface closely reproduced the high-density OF dataset across the evaluated field-size range, with no dominant systematic trend in the residual distribution.

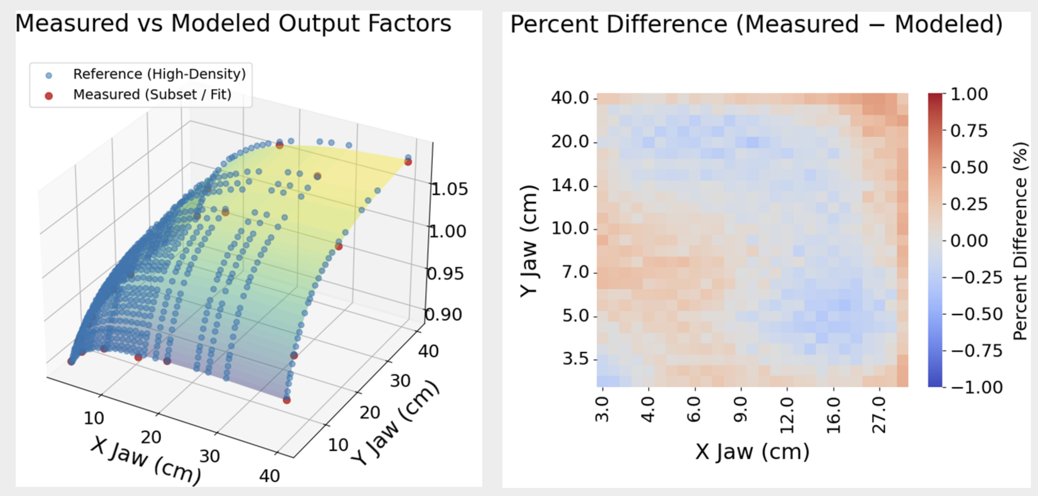


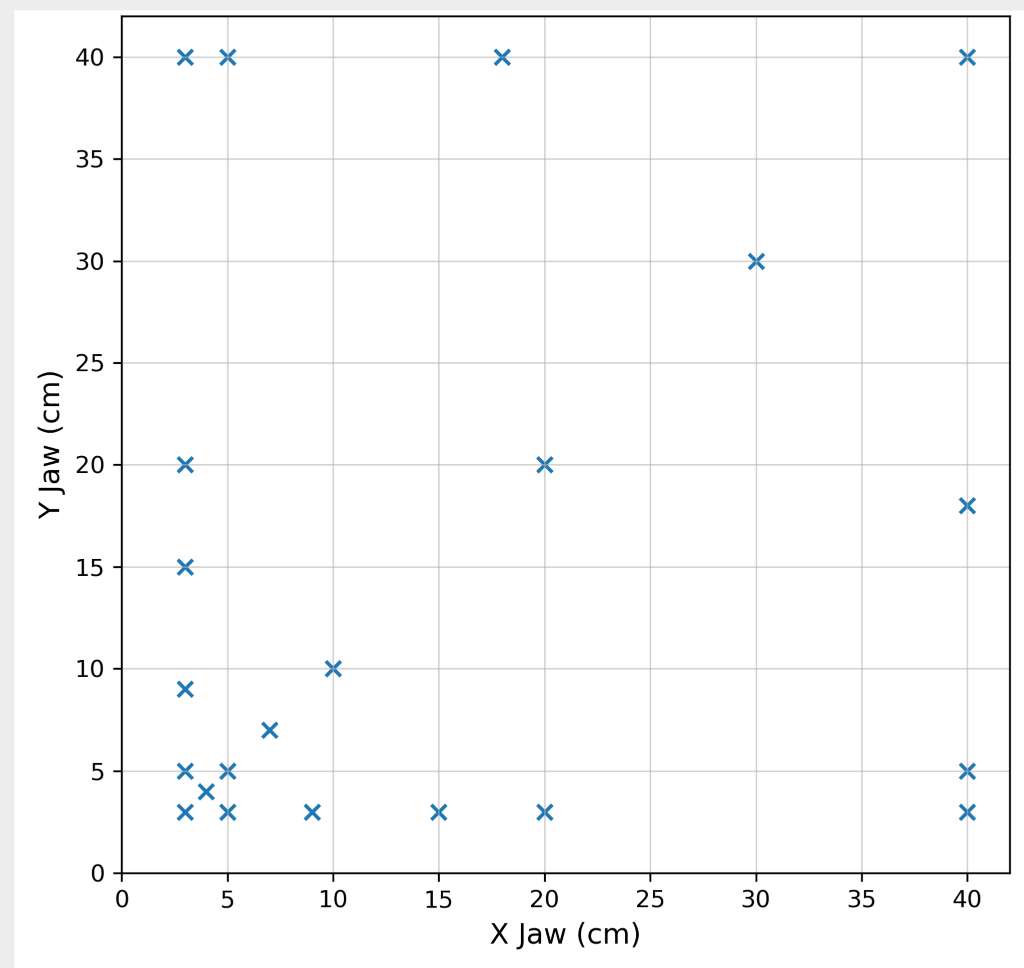
Figure 2. Comparison of measured output factors with the fitted log-log surface (left) and spatial distribution of percent differences between measured and modeled values (right).

DISCUSSION

Measurement distribution was more important than measurement number alone. The Edge + X geometry efficiently sampled regions of rapid output-factor variation and the boundaries of the field-size domain, maintaining accuracy with substantially fewer measurements. Comparable performance with independently acquired data supports the robustness of the approach beyond the dataset used for model development.

RECOMMENDED CONFIGURATION

Edge + X: 22 measurements at 50 MU per field



Maximum model difference: 0.50% | Combined uncertainty: ≈0.35% (1σ)

CONCLUSIONS

A continuous log-log surface accurately modeled photon output factors across the evaluated field-size range and remained robust when derived from substantially reduced measurement subsets. The optimized 22-point Edge + X strategy maintained commissioning-grade accuracy while reducing the measurement count by 73% relative to the published reference set. These findings support the use of strategically selected measurements and continuous surface modeling to improve the efficiency of output factor commissioning and validation.