

From Propagation to Prediction: Point-level Uncertainty Evaluation of MLS Point Clouds under Limited Ground Truth

Ziyang Xu¹, Olaf Wysocki³, Christoph Holst^{1,2}

¹ Chair of Engineering Geodesy, TUM School of Engineering and Design, Technical University of Munich, Munich, Germany — {ziyang.xu, christoph.holst}@tum.de

² TUM Leonhard Obermeyer Center, Technical University of Munich, Munich, Germany

³ CV4DT, University of Cambridge, Cambridgeshire CB3 0FA, Cambridge, UK — okw24@cam.ac.uk

Keywords: Point Cloud Quality Control, Error Quantification, Machine Learning, Geometric Features, Uncertainty Regression.

Abstract

Evaluating uncertainty is critical for reliable use of Mobile Laser Scanning (MLS) point clouds in many high-precision applications such as Scan-to-BIM, deformation analysis, and 3D modeling. However, obtaining the ground truth (GT) for evaluation is often costly and infeasible in many real-world applications. To reduce this long-standing reliance on GT in uncertainty evaluation research, this study presents a learning-based framework for MLS point clouds that integrates optimal neighborhood estimation with geometric feature extraction. Experiments on a real-world dataset show that the proposed framework is feasible and the XGBoost model delivers fully comparable accuracy to Random Forest while achieving substantially higher efficiency (about 3 times faster), providing initial evidence that geometric features can be used to predict point-level uncertainty quantified by the C2C distance. In summary, this study shows that MLS point clouds' uncertainty is learnable, offering a novel learning-based viewpoint towards uncertainty evaluation research.

1. Introduction

Mobile Laser Scanning (MLS) systems are widely used to acquire 3D point clouds across different fields such as Architecture, Engineering, and Construction (AEC), engineering geodesy, and Geographic Information Systems (GIS) (Xue et al., 2020). Downstream tasks based on these point clouds, including registration, Scan-to-BIM, deformation analysis, and 3D modeling, require not only accurate geometry but also reliable estimates of uncertainty (Geißendörfer and Holst, 2025, Xu et al., 2026). Moreover, errors in point clouds will inevitably accumulate and propagate through downstream workflows, making uncertainty evaluation indispensable (Xu et al., 2025a).

Traditional point clouds uncertainty evaluation methods can be grouped into forward and backward modeling (Shi et al., 2021). Forward modeling relies on variance and covariance propagation and detailed theoretical analysis of involved errors. It models the various error sources to infer the uncertainty of the final point clouds. In MLS systems, four main categories of error sources contribute to uncertainty: instrumental errors, atmospheric errors, object- and geometry-related errors, and trajectory estimation errors (Habib et al., 2009, Schenk, 2001). However, MLS point clouds' uncertainty is also influenced by many other factors, including the observation environment, internal SLAM algorithms, and scanning speed. As a result, accurate and comprehensive modeling of all error sources is rarely feasible (Holst and Kuhlmann, 2016, Shi et al., 2021).

In contrast, backward modeling, which is the more common solution, avoids explicit error modeling and propagation. Instead, it uses accurate reference data to evaluate uncertainty empirically. Its main advantage is that uncertainty is quantified from actual measurements, which supports evaluation procedures that generalize across different systems. The core of backward modeling is to establish reliable correspondences between

the scanned and the reference data to characterize their discrepancies.

Although backward modeling is more applicable than forward modeling, a crucial concern remains. Most backward solutions assume access to high-precision reference data as the ground truth (GT), for example, Terrestrial Laser Scanning (TLS) or high-fidelity as-built models, and then use the GT to evaluate MLS point clouds (Xu et al., 2025b). This assumption is theoretically sound, but in practice, users often face challenges :

- (1) Evaluation depends heavily on the GT and will be problematic if the GT is unavailable, which is frequently the case in real-world applications.
- (2) Obtaining the GT often costs much more than common data collection in terms of money and time, which limits adoption.
- (3) Inherent practical paradox, namely that if the GT is already available, the value of performing MLS data collection for evaluation may be questionable.

To overcome these challenges, learning-based solutions have begun to gain attention (Hartmann et al., 2023, Hartmann and Alkhatib, 2023). The core idea is to recast error quantification as a supervised learning problem, in which a model learns the relationship between each point's error and its local geometric or attribute features. Unlike statistical error propagation or fully GT-based evaluations, this data-driven paradigm requires reference data as GT only during training; once trained, the model no longer depends on it. This design improves flexibility and offers substantial potential for practical deployment. In this context, this study proposes a framework of point-level uncertainty evaluation based on machine learning under limited GT. The main contributions are as follows:

- A learning-based framework is designed to predict point-level uncertainty from its geometric features for MLS point clouds.

- A controlled comparison of XGBoost and Random Forest implementation is conducted, showing comparable accuracy, with XGBoost achieving lower runtime.
- A feature importance analysis reveals how important features correspond to interpretable geometric patterns that drive error growth.

2. Related Works

2.1 Forward Modeling

Forward modeling is based on the variance-covariance propagation law. A detailed theoretical analysis of all the errors of laser scanning systems is presented, in which all kinds of error sources are modeled, and the error characteristics of the final point clouds are obtained. Several forward modeling-based evaluations are realized, considering the error information from the different sensors (Barber et al., 2007, Goel and Lohani, 2015, Xu, 2016). Some studies aim to establish error budgets for controlling point clouds quality (Habib et al., 2009). The ranging precision, based on the intensity information of the point clouds, has also been explored (Stenz et al., 2020, Wujanz et al., 2017). As previously mentioned, forward modeling is theoretically feasible but faces several limitations. Consequently, an increasing number of studies are shifting toward backward modeling.

2.2 Backward Modeling

Backward modeling relies on accurate reference data as the GT to quantify errors empirically. To achieve this, the fundamental procedure involves accurately and efficiently quantifying predefined uncertainty metrics with respect to the GT. Standard practices compute cloud-to-cloud (C2C), cloud-to-mesh (C2M), and multiscale model-to-model cloud comparison (M3C2) distances to measure discrepancies between scans and the GT (Lague et al., 2013, Heinz et al., 2015, Stenz et al., 2017, Zahs et al., 2022, Xu et al., 2025a). These distance-based evaluations offer simplicity and low cost, but they lack 6 degrees of freedom (DoF) transformation parameters, so they cannot comprehensively represent uncertainty from a 3D perspective. To handle this gap, point-based and line-based methods have been introduced (Poreba and Goulette, 2013, Hauser et al., 2016). Plane-based strategies are also widely used; they exploit planar primitives to establish robust correspondences and to evaluate uncertainty (Shi et al., 2021, Shahraji et al., 2020). The key is using geometric features of point clouds to establish correspondences, followed by quantifying uncertainty estimates for 6 DoF. In the context of MLS, this is often used to evaluate trajectory estimation errors (Xu et al., 2025b).

Owing to its flexibility and ease of use, backward modeling remains the predominant solution. Its reliance on the GT ensures its reliability and accuracy, but also limits scalability to large real-world deployments. Learning-based modeling has therefore emerged as a promising direction.

2.3 Learning-based Modeling

Learning-based uncertainty evaluation is an emerging paradigm in the context of geodesy. It stems from attempts and explorations to reduce or eliminate dependence on the GT using learning techniques. Initial idea used ensemble methods to estimate ranging precision in TLS from features such as intensity and incidence angle (Hartmann et al., 2023, Hartmann and Alkhatib,

2023). Subsequent work extended this idea to deep learning to predict and calibrate TLS ranging accuracy (Hartmann et al., 2024). These investigations mainly focused on TLS distance measurement under laboratory conditions and did not address MLS point clouds. More recently, ensemble learning has been explored for MLS uncertainty by framing the task as a binary classification of point quality in a real-world scenario (Xu et al., 2025c).

To the best of our knowledge, most existing research on evaluating the uncertainty of point clouds primarily centers on forward and backward modeling. Only a limited number of publications utilize machine learning for this purpose on TLS, indicating that this area is still in its early stages of development. This study presents the first learning-based uncertainty modeling applied to a real-world MLS dataset.

3. Methodology

This section presents the proposed framework as shown in Figure 1. The framework starts from data preparation in Section 3.1, then covers a description of feature engineering in Section 3.2. After that, more information about model training and testing is presented in Section 3.3 and Section 3.4, respectively.

3.1 Data Preparation

As seen in the framework, two inputs are MLS point clouds and the corresponding reference point clouds, preferably obtained via TLS. In this study, point-level uncertainty is quantified using the C2C distance. C2C distance is preferred over M3C2 distance for two reasons. First, C2C distance is more responsive to noise at the individual point level and therefore provides a more direct representation of local error characteristics. Second, M3C2 distance requires a user-defined radius, and its results are sensitive to that choice, whereas C2C distance does not involve such a scale parameter.

Since this framework aims to learn a mapping from point-level geometric features to the corresponding uncertainty metric (quantified by C2C distance). To reduce label noise caused by density differences between MLS and TLS data and by residual minor alignment errors (less than 10 mm), only MLS points with a C2C distance less than 80 mm are retained for model training.

3.2 Feature Engineering

As aforementioned, this study aims is to examine whether point-level C2C distance is related to local geometric features. In other words, it investigates whether a point's error behavior can be predicted and interpreted from its geometric feature domain, which makes point-level feature extraction a critical step. Moreover, only with a suitable neighborhood, local geometric features can be estimated reliably at the point level. Selecting this neighborhood remains a fundamental challenge, since different definitions can introduce additional variability into the extracted features.

It has been proven that neighborhoods defined by the k -nearest neighbors (k -NN) offer clear advantages, and the geometric features derived from them can improve point cloud classification performance (Weinmann et al., 2015). Additionally, these geometric features have also been used for other 3D learning tasks to achieve better performance, such as segmentation (Tan et al.,

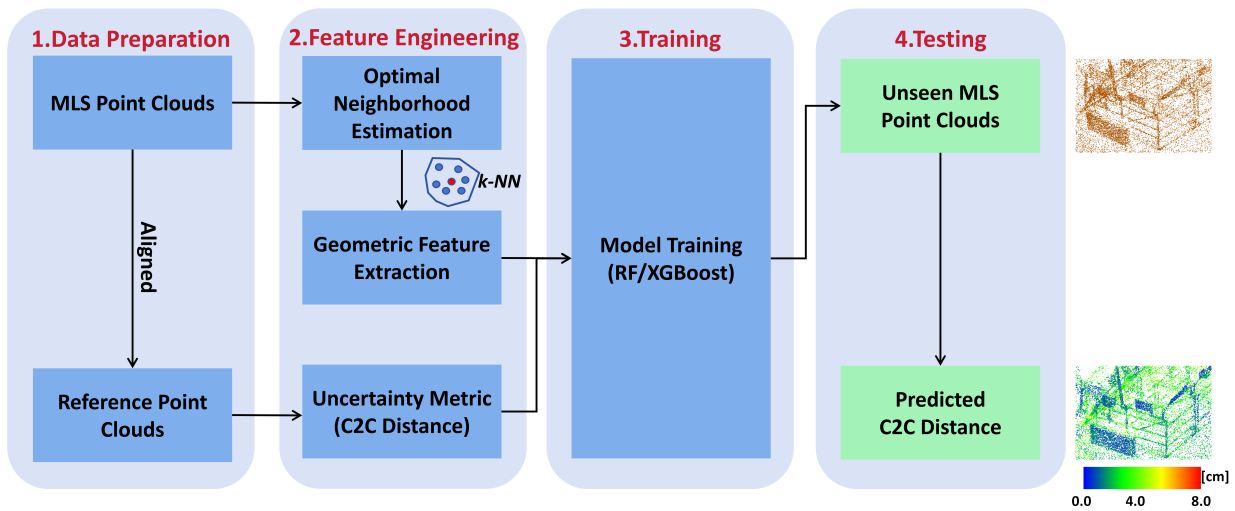


Figure 1. Proposed learning-based point-level uncertainty evaluation framework.

2024, Robert et al., 2023, Wysocki et al., 2023). Motivated by this evidence, this study adopts the same optimal neighborhood estimation strategy. For each point in the input MLS point clouds, the optimal number of neighbors $OptN$ is first estimated. Based on $OptN$, the corresponding 3D and 2D geometric features are then computed. In total, 26 features will be used for training to predict uncertainty. Table 1 lists basic definitions of these geometric features. Detailed calculations are outside the scope of this study. Please refer to these publications (Weinmann et al., 2014, Weinmann et al., 2015).

3.3 Training

After geometric feature extraction, each point P_i is represented by a 27-dimensional feature vector $\mathbf{x}_i \in \mathbb{R}^{27}$ that concatenates its 26 geometric features and the corresponding optimal neighborhood size $OptN_i$. Stacking all samples yields the design matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T \in \mathbb{R}^{n \times 27}$ and the target vector $\mathbf{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$, where y_i denotes the C2C distance of point P_i . The task is formulated as supervised regression,

seeking a function $f: \mathbb{R}^{27} \rightarrow \mathbb{R}$ such that $\hat{\mathbf{y}} = f(\mathbf{X})$ approximates \mathbf{y} . Two commonly used ensemble models, Random Forest (RF) (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016), are employed to obtain the predictions $\hat{\mathbf{y}}$.

RF and XGBoost are selected for two reasons. From a technical standpoint, both are tree-ensemble regressors that capture nonlinear relationships and feature interactions, handle mixed-scale inputs with minimal preprocessing, and are robust to outliers. RF reduces variance through bootstrap aggregation and random feature subspaces, whereas XGBoost improves accuracy and efficiency through gradient boosting with regularization and optional GPU acceleration. Both also support model attribution, using permutation importance for RF and SHAP values for XGBoost. From an empirical standpoint, prior studies report promising results with RF and XGBoost on TLS measurement (Hartmann et al., 2023, Hartmann and Alkhatib, 2023).

To clarify the training protocol, Tables 2 and 3 list the key hyperparameters used in this study. The XGBoost setup employs histogram-based boosting with GPU execution and early stopping. The Random Forest configuration adopts memory-safe settings on the CPU. Additionally, a 5-fold cross-validation strategy was employed. For each iteration, the model is trained on four folds and validated on the held-out fold. The averaged

results across the five folds were reported to provide a comprehensive and spatially unbiased evaluation of model performance.

3.4 Testing

The regression performance is tested with the aforementioned 5-fold cross-validation strategy using five complementary metrics: root mean squared error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (1)$$

emphasizing larger errors due to the quadratic penalty; mean absolute error (MAE),

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2)$$

reflecting the average absolute deviation and is less sensitive to outliers than RMSE; median absolute error (MedAE),

$$\text{MedAE} = \text{median}_i |y_i - \hat{y}_i|, \quad (3)$$

providing a robust estimate of the typical absolute error; coefficient of determination (R^2),

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad (4)$$

measuring the proportion of variance in y explained by the predictions; and tolerance-aligned accuracy $P@m$ for $m \in \{10, 20, 30, 40, 50\}$ mm,

$$P@m = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{|y_i - \hat{y}_i| \leq m\}, \quad (5)$$

reporting the percentage of points whose absolute error does not exceed m mm, which aligns directly with common tolerance thresholds. All metrics are computed on the predicted values \hat{y}_i and references y_i .

#	Feature	Formula / Definition	Range	Unit	Notes
1	linearity	$(l_1 - l_2)/l_1$	[0, 1]	–	line-like strength
2	planarity	$(l_2 - l_3)/l_1$	[0, 1]	–	plane-like strength
3	scattering	l_3/l_1	[0, 1]	–	λ_3/λ_1
4	omnivariance	$(l_1 l_2 l_3)^{1/3}$	[0, 1/3]	–	geometric mean of normalized EVs
5	anisotropy	$(l_1 - l_3)/l_1$	[0, 1]	–	directionality contrast
6	eigenentropy	$-\sum_{i=1}^3 l_i \ln l_i$	[0, ln 3]	–	larger if more isotropic
7	sum_EVs	$\lambda_1 + \lambda_2 + \lambda_3$	≥ 0	m ²	total 3D variance
8	change_of_curvature	l_3	[0, 1]	–	noise-sensitive
9	Z_vals	z	–	m	absolute height
10	radius_kNN	r_{3D} to the k -th neighbor	≥ 0	m	neighborhood scale (3D)
11	density	$\frac{k+1}{\frac{4}{3}\pi r_{3D}^3}$	≥ 0	points/m ³	volumetric density
12	verticality	$1 - v_z$ (eigenvector of λ_1)	$\approx [0, 2]$	–	uses major axis, not $1 - n_z $
13	delta_Z_kNN	$z_{\max} - z_{\min}$ in k -NN	≥ 0	m	vertical relief in k -NN
14	std_Z_kNN	unbiased σ_z in k -NN	≥ 0	m	vertical roughness
15	radius_kNN_2D	r_{2D} : max XY dist among first $k+1$	≥ 0	m	projected k -NN radius
16	density_2D	$\frac{k+1}{\pi r_{2D}^2}$	≥ 0	points/m ²	planar density
17	sum_EVs_2D	$\mu_1 + \mu_2$	≥ 0	m ²	total variance in XY
18	EV_ratio	$m_2/m_1 = \mu_2/\mu_1$	[0, 1]	–	XY anisotropy; smaller \rightarrow linear
19	frequency_acc_map	points per XY grid cell	≥ 0	–	global 2D grid
20	delta_z	$z_{\max} - z_{\min}$ per XY cell	≥ 0	m	global grid relief
21	std_z	sample σ_z per XY cell	≥ 0	m	global grid roughness
22	EV3D_1	$l_1 = \lambda_1/(\sum \lambda)$	[0, 1]	–	largest normalized 3D EV
23	EV3D_2	$l_2 = \lambda_2/(\sum \lambda)$	[0, 1]	–	middle normalized 3D EV
24	EV3D_3	$l_3 = \lambda_3/(\sum \lambda)$	[0, 1]	–	smallest normalized 3D EV
25	EV2D_1	$m_1 = \mu_1/(\mu_1 + \mu_2)$	[0, 1]	–	largest normalized 2D EV
26	EV2D_2	$m_2 = \mu_2/(\mu_1 + \mu_2)$	[0, 1]	–	smallest normalized 2D EV

Notes: (1) k is selected per point by entropy minimization: $H(k) = -\sum_i p_i \ln p_i$ with $p_i = \lambda_i/(\lambda_1 + \lambda_2 + \lambda_3)$; pick $k = \arg \min H(k)$. The k -NN includes the query point, thus $n = k + 1$. (2) Unbiased covariance is used when $n > 1$ (scaled by $n/(n - 1)$). 3D eigenvalues (EV) are sorted as $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ and normalized $l_i = \lambda_i/(\lambda_1 + \lambda_2 + \lambda_3)$. For 2D (XY), $\mu_1 \geq \mu_2 > 0$ and $m_i = \mu_i/(\mu_1 + \mu_2)$. (3) Verticality uses the major eigenvector (λ_1); a more “surface-normal” variant would use $1 - |n_z|$ with n the eigenvector of λ_3 . (4) r_{3D} is the distance to the k -th neighbor in 3D; r_{2D} is the maximum XY distance among the first $k+1$ neighbors (ordered by 3D distance). (5) Grid features (19–21) are computed on a global XY raster with a given cell size and are independent of k .

Table 1. Basic definitions and explanations of 26 point-level geometric features.

Hyperparameter	Value	Notes
n_estimators	100	Number of trees
max_depth	20	Depth cap
max_samples	0.5	Subsample per tree
class_weight	Balanced	Handles class imbalance
n_jobs	1	Single-thread
random_state	Random seed	Reproducibility

Table 2. Key Random Forest hyperparameters.

Hyperparameter	Value	Notes
max_depth	8	Tree depth
eta	0.05	Learning rate
subsample	0.8	Row sampling per round
colsample_bytree	0.8	Feature sampling per tree
num_boost_round	1000	Maximum boosting rounds
early_stopping_rounds	50	Patience on validation loss
seed	Random seed	Reproducibility

Table 3. Key XGBoost hyperparameters.

4. Results and Analysis

This section introduces the experiment dataset in Section 4.1 and presents quantitative results in Section 4.2, then visualizes the predicted results in Section 4.3, and gives feature importance analysis in Section 4.4.

4.1 Experiment Dataset

All data was collected in an assembly hall at the BMW factory, an indoor industrial environment that combines complex geometries and includes both reflective and absorbent materials. An overview of the scanned area is shown in Figure 2. MLS point clouds were acquired from the Emesent Hovermap ST-X

unit¹. Reference point clouds were acquired with a Trimble X9 terrestrial laser scanner². After this, using a high-precision total station control network and black & white targets, two datasets are accurately aligned within the same coordinate system (Xu et al., 2025b). More information about the experiment scenario can be found in Table 4.

4.2 Quantitative Results

Table 5 reports the core performance metrics averaged over five folds with 95% confidence intervals. RF and XGBoost achieve

¹ <https://emesent.com/hovermap-series/>

² <https://geospatial.trimble.com/en/products/hardware/trimble-x9>

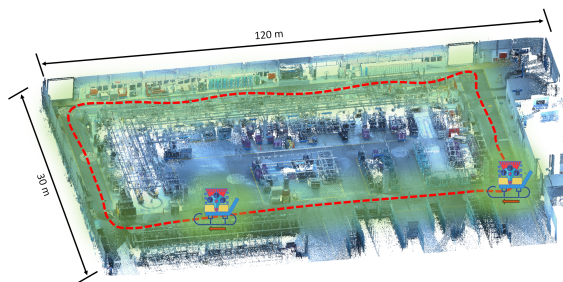


Figure 2. Overview of the scanned area. Red line indicates the scanning path.

Scanning Area	About 3,500 m ²
Number of Points	5,000,000
MLS System	EMESENT HOVERMAP ST-X
Accuracy Specification*	±15mm
Number of black & white targets	12

* manufacturer's specification in typical environment.

Table 4. Basic information on experiment scenario.

virtually identical accuracy: both obtain an RMSE of 10.9 mm; the MAE differs by 0.1 mm; the MedAE is 4.0 mm for both; and the R^2 values are close with overlapping intervals. The tolerance-aligned results in Table 6 are similarly close across all thresholds: RF is slightly higher at P@10-20 mm, whereas XGBoost is slightly higher at P@40-50 mm, and all differences lie within the reported confidence intervals.

The principal distinction is computational efficiency: the average runtime per fold is 210.5 s for RF and 74.6 s for XGBoost. The efficiency advantage of XGBoost arises from both algorithmic design and GPU support. First, XGBoost employs histogram-based split finding with quantile binning, which reduces the number of candidate thresholds per feature and lowers computation and memory traffic. Second, XGBoost significantly elevates performance by harnessing the power of parallel processing, efficiently accumulating gradients and Hessians across multiple GPU threads.

Metric	RF	XGBoost
RMSE (mm) ↓	10.9 ± 0.6	10.9 ± 0.6
MAE (mm) ↓	6.5 ± 0.4	6.6 ± 0.3
MedAE (mm) ↓	4.0 ± 0.2	4.0 ± 0.2
R^2 ↑	0.36 ± 0.02	0.37 ± 0.03
Runtime / fold (s) ↓	210.5	74.6

Table 5. Regression performance (mean ± 95% CI).

Overall, the two models provide fully comparable prediction accuracy, whereas XGBoost offers a clear efficiency advantage, making it a more feasible solution for practical application. Therefore, subsequent visualization results and feature importance analyses will focus primarily on XGBoost.

4.3 Visualized Results

Figure 3 compares predicted C2C distances from XGBoost with the GT. Most of the predictions closely follow the GT, particularly on smooth and continuous areas. To examine local behavior, five objects with distinct geometric structures are selected from Figure 3 and shown in detail in Figure 4. Across these objects, it maintains fine-scale fidelity and achieves accurate estimates even in geometrically intricate areas.

	RF	XGBoost
P@10 mm	0.832 ± 0.015	0.823 ± 0.013
P@20 mm	0.945 ± 0.007	0.941 ± 0.007
P@30 mm	0.973 ± 0.004	0.973 ± 0.004
P@40 mm	0.983 ± 0.003	0.985 ± 0.003
P@50 mm	0.990 ± 0.002	0.992 ± 0.002

Table 6. Proportion within absolute error thresholds (mean ± 95% CI).

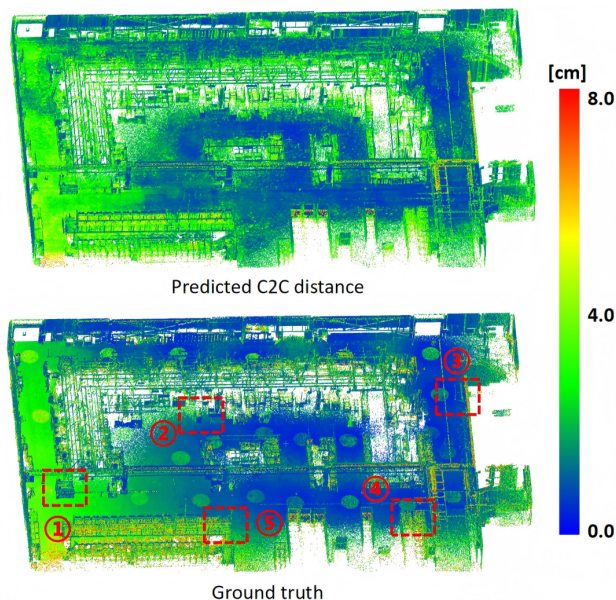


Figure 3. Predicted results from XGBoost model. The outlined red boxes indicate selected representative objects in the experimental scene; further details are shown in Figure 4.

To further quantify and visualize the prediction performance of XGBoost, Figure 5 reports the absolute error $|y - \hat{y}|$ and the residual $r = \hat{y} - y$ of the predicted C2C distances with respect to the GT. The results corroborate the overall accuracy of the predictions, with minor errors across most areas. Increased variability is observed in areas with highly complex or sparse geometry, such as slender rod-like objects and thin linear features, where predictions appear less stable. This behavior is consistent with the known inherent sensitivity of the C2C distance calculation in such structures, which complicates both measurement and prediction.

4.4 Feature Importance Analysis

SHAP Feature Importance Analysis Figure 6 summarizes global contributions by mean SHAP value, indicating that height-variation and sampling-density features dominate (Z_vals , $delta_z$, std_z , $density_2D$, $density$) XGBoost prediction. These are followed by eigenvalue-derived shape measures and orientation features, while neighborhood-scale features such as $radius_kNN$ show moderate but consistent influence. Figure 7 indicates that the top four features have more red points concentrated on the left side of the horizontal axis, which means high feature values tend to decrease the predicted C2C distance. This pattern identifies height and sampling density as the two dominant factors. The underlying reasons are two-fold. First, higher density usually reflects shorter range, a more favorable incidence angle, or slower platform motion, which stabilizes neighborhood statistics and reduces nearest-neighbor

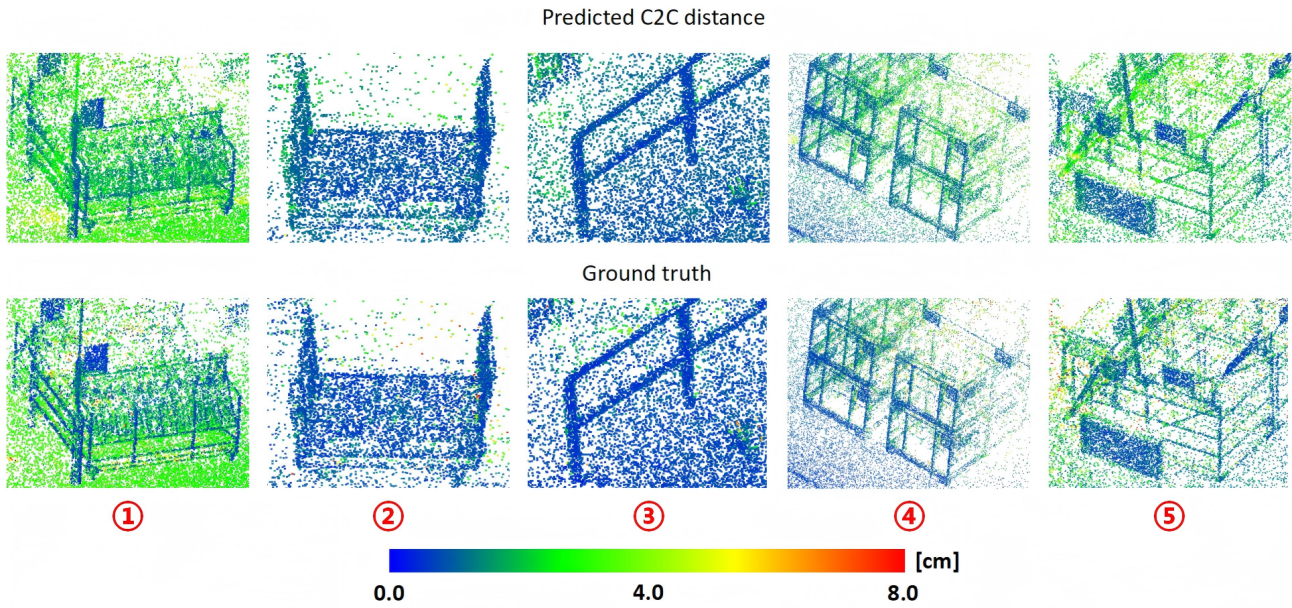


Figure 4. Detailed comparison of XGBoost prediction on selected objects. The numbers correspond to the objects shown in Figure 3.

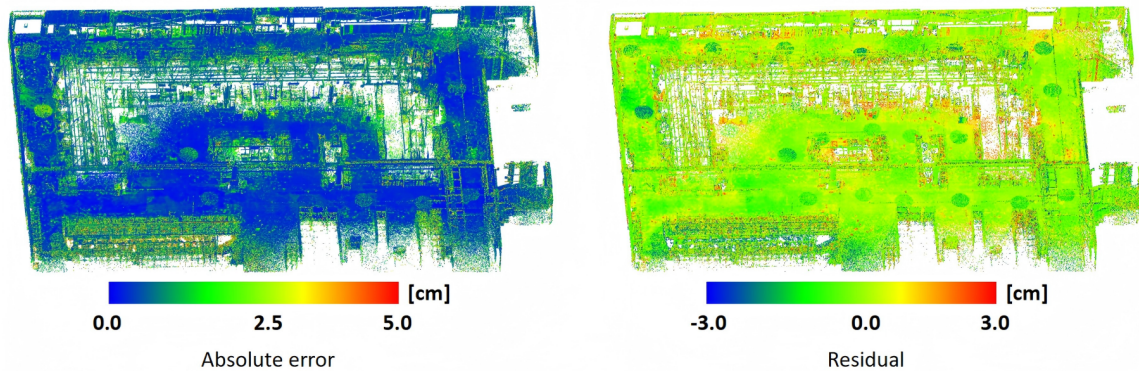


Figure 5. Absolute error (left) and residual (right) of the predicted C2C distance with respect to the GT from XGBoost model.

mismatches; therefore, higher density leads to lower predicted C2C distance. Second, in this dataset, there is no ceiling; the point clouds only consist of ground and surrounding structures, and most errors originate from the vertical direction (Xu et al., 2025b). Points at higher Z_vals therefore mainly belong to walls or elevated components that are more regularly shaped and better observed, which explains why larger Z_vals and, in this scene, larger local height variation $delta_z$ are associated with lower predicted C2C distance.

In summary, the SHAP-based feature analysis of the XGBoost model is only intended to indicate the direction of change in the predicted C2C values; detailed quantification of effect sizes and investigation of scene-specific reasons require additional studies and lie beyond the scope of this study.

Permutation Feature Importance Analysis To further compare RF and XGBoost under the same criterion, a model-agnostic measure is required. Permutation importance under RMSE provides such a baseline: each feature is shuffled on the held-out fold, and the increase in RMSE is recorded. This protocol treats both models identically and reveals differences in how they exploit features.

As seen in Figure 8, XGBoost shows large $\Delta RMSE$ for several groups of features, including eigenvalue- and density-based

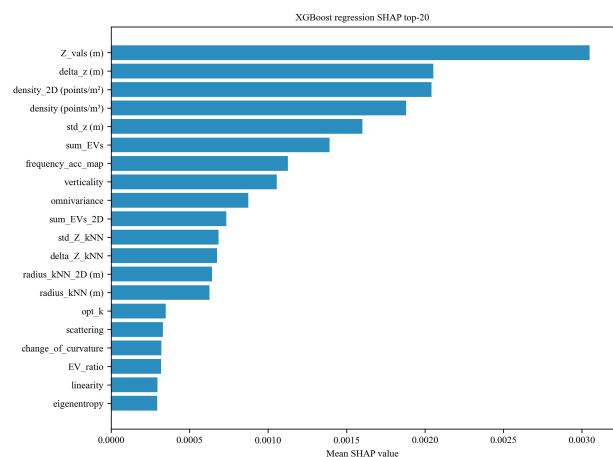


Figure 6. Top 20 most important features of XGBoost model.

ones. This indicates broad and effective use of complementary information. RF exhibits smaller $\Delta RMSE$ overall, with noticeable reliance on $density_2D$ and Z_vals and limited sensitivity to other features. The XGBoost permutation ranking is also consistent with the SHAP analysis, which also highlights height, density, and roughness as dominant factors. Differ-

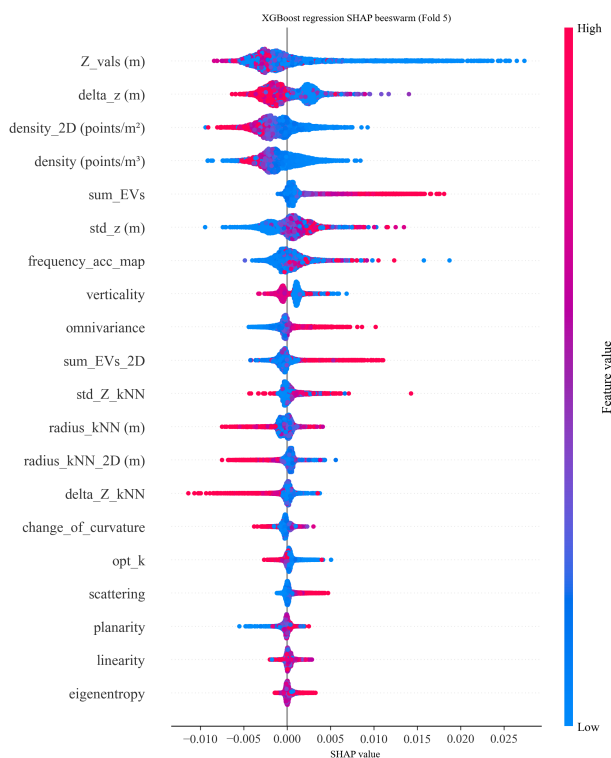


Figure 7. SHAP beeswarm of XGBoost model.

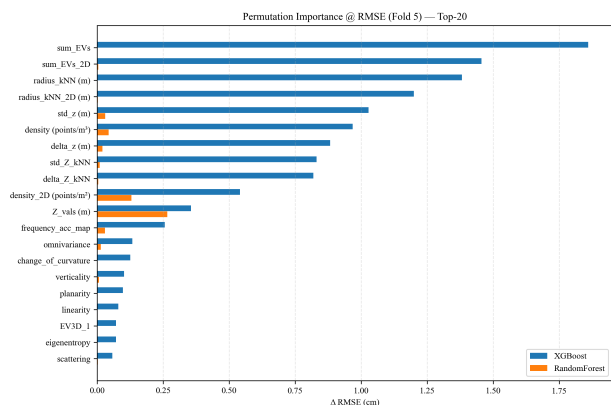


Figure 8. Permutation feature importance analysis on RF and XGBoost.

ences between the two models are linked to their learning mechanisms: gradient boosting with regularization tends to place greater weight on leading features, whereas bagging with depth limits tends to average weaker trees and is comparatively robust to the influence of individual features.

5. Discussion

Advantages Learning-based modeling has opened up a new research frontier and offers three key advantages: First, it effectively reduces reliance on the GT, making it more practically relevant. Second, it preliminarily demonstrates that uncertainty is learnable, broadening the scope of current point cloud uncertainty research. Third, compared to rigid error propagation, it provides a more flexible, task-customizable framework, allowing data users to define error metrics and design prediction models from their perspective.

Challenges Despite the promising results given by the proposed framework, several challenges remain. The generalization ability of the framework has not yet been verified across diverse datasets and platforms, so external validation is required. The necessity of using all 26 geometric features is also unclear (Weinmann et al., 2013); systematic ablation and feature pruning may reveal smaller subsets that achieve comparable or better accuracy with improved efficiency. In addition, the observed R^2 is modest, which calls for further investigation into potential causes such as label noise, feature distribution, and protocol choices. This study focuses specifically on using the C2C distance as a quantitative metric for uncertainty; other distance metrics, such as the C2M distance, are also potential avenues for future research.

6. Conclusion

This study aims to predict point-level uncertainty in MLS point clouds using machine learning. It introduces a learning-based framework that integrates optimal neighborhood estimation with geometric feature extraction and validates the framework on a real-world dataset. The results indicate that point clouds' uncertainty is learnable and XGBoost achieves accuracy comparable to RF while providing higher efficiency. Additionally, feature-importance analyses offer initial evidence of relationships between geometric features and the C2C distance. Overall, the study explores a novel learning-based paradigm for point clouds uncertainty modeling.

For the future, it is imperative to explore the generalizability and transferability of learning-based modeling. Additionally, investigating other advanced learning strategies, such as symbolic regression (Cranmer, 2023), Transformer-based or physics-informed learning, or Graph Neural Network (GNN), may yield valuable insights.

Acknowledgment

This research was funded by TUM Georg Nemetschek Institute of Artificial Intelligence for the Built World, project "NERF2BIM", PI Christoph Holst.

References

- Barber, D., Mills, J., Smith-Voysey, S., 2007. Geometric validation of a ground-based mobile laser scanning system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(1), 128-141.
- Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), 5-32.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 785-794.
- Cranmer, M., 2023. Interpretable machine learning for science with PySR and SymbolicRegression.jl. <https://arxiv.org/abs/2305.01582v3>.
- Geißendörfer, O., Holst, C., 2025. Spatio-temporal mode description in LiDAR point clouds. *Proceedings of the 6th Joint International Symposium on Deformation Monitoring - JISDM 2025*, Karlsruhe Institute of Technology, Karlsruhe.

- Goel, S., Lohani, B., 2015. Relative contribution and effect of various error sources on the performance of Mobile Mapping System (MMS). *Journal of the Indian Society of Remote Sensing*, 43(3), 639–645.
- Habib, A., Bang, K. I., Kersting, A. P., Lee, D.-C., 2009. Error budget of LiDAR systems and quality control of the derived data. *Photogrammetric Engineering & Remote Sensing*, 75(9), 1093–1108.
- Hartmann, J., Alkhatib, H., 2023. Uncertainty modelling of laser scanning point clouds using machine-learning methods. *Remote Sensing*, 15(9), 2349.
- Hartmann, J., Ernst, D., Neumann, I., Alkhatib, H., 2024. PointNet-based modeling of systematic distance deviations for improved TLS accuracy. *Journal of Applied Geodesy*, 18(4), 613–628.
- Hartmann, J., Heiken, M., Alkhatib, H., Neumann, I., 2023. Automatic quality assessment of terrestrial laser scans. *Journal of Applied Geodesy*, 17(4), 333–353.
- Hauser, D., Glennie, C., Brooks, B., 2016. Calibration and accuracy analysis of a low-cost mapping-grade mobile laser scanning system. *Journal of Surveying Engineering*, 142(4), 04016011.
- Heinz, E., Eling, C., Wieland, M., Klingbeil, L., Kuhlmann, H., 2015. Development, calibration and evaluation of a portable and direct georeferenced laser scanning system for kinematic 3D mapping. *Journal of Applied Geodesy*, 9(4), 227–243.
- Holst, C., Kuhlmann, H., 2016. Challenges and present fields of action at laser scanner based deformation analyses. *Journal of Applied Geodesy*, 10(1), 17–25.
- Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10–26.
- Poreba, M., Goulette, F., 2013. Line segment-based approach for accuracy assessment of MLS point clouds in urban areas. *8th International Symposium on Mobile Mapping Technology*, Tainan, Taiwan.
- Robert, D., Raguét, H., Landrieu, L., 2023. Efficient 3D semantic segmentation with superpoint transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 17195–17204.
- Schenk, T., 2001. Modeling and analyzing systematic errors in airborne laser scanners. *Technical Notes in Photogrammetry*, 19(46).
- Shahraji, M. H., Larouche, C., Cocard, M., 2020. Analysis of systematic errors of mobile LiDAR systems: A simulation approach. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 253–260.
- Shi, B., Bai, Y., Zhang, S., Zhong, R., Yang, F., Song, S., Li, G., 2021. Reference-plane-based approach for accuracy assessment of mobile mapping point clouds. *Measurement*, 171, 108759.
- Stenz, U., Hartmann, J., Paffenholz, J.-A., Neumann, I., 2017. A framework based on reference data with superordinate accuracy for the quality analysis of terrestrial laser scanning-based multi-sensor-systems. *Sensors*, 17(8), 1886.
- Stenz, U., Hartmann, J., Paffenholz, J.-A., Neumann, I., 2020. High-precision 3D object capturing with static and kinematic terrestrial laser scanning in industrial applications—approaches of quality assessment. *Remote Sensing*, 12(2), 290.
- Tan, Y., Wysocki, O., Hoegner, L., Stilla, U., 2024. Classifying point clouds at the facade-level using geometric features and deep learning networks. *Recent Advances in 3D Geoinformation Science*, Springer Nature Switzerland, Cham, 391–404.
- Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286–304.
- Weinmann, M., Jutzi, B., Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 313–318.
- Weinmann, M., Jutzi, B., Mallet, C., 2014. Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 181–188.
- Wujanz, D., Burger, M., Mettenleiter, M., Neitzel, F., 2017. An intensity-based stochastic model for terrestrial laser scanners. *ISPRS Journal of Photogrammetry and Remote Sensing*, 125, 146–155.
- Wysocki, O., Xia, Y., Wysocki, M., Grilli, E., Hoegner, L., Cremers, D., Stilla, U., 2023. Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and bayesian networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 6548–6558.
- Xu, S., 2016. Research on calibration of Mobile Mapping System for land vehicle and its accuracy assessment. PhD thesis, Wuhan University.
- Xu, Z., Fischer, S., Holst, C., 2025a. Uncertainty-aware evaluation and fusion of point clouds for simultaneous scanning of two state-of-the-art indoor MLS systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-G-2025, 1603–1610.
- Xu, Z., Hackl, M., Holst, C., 2025b. PL4U: Automated plane-based uncertainty evaluation and reduction method for indoor Mobile Laser Scanning systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 228, 467–488.
- Xu, Z., Schwab, B., Yang, Y., Kolbe, T. H., Holst, C., 2026. L2M-Reg: Building-level uncertainty-aware registration of outdoor LiDAR point clouds and semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 233, 467–486.
- Xu, Z., Wysocki, O., Holst, C., 2025c. Point-level uncertainty evaluation of mobile laser scanning point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W6-2025, 243–249.
- Xue, F., Lu, W., Chen, Z., Webster, C. J., 2020. From LiDAR point cloud towards digital twin city: Clustering city objects based on Gestalt principles. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167, 418–431.

Zahs, V., Winiwarter, L., Anders, K., Williams, J. G., Rutzinger, M., Höfle, B., 2022. Correspondence-driven plane-based M3C2 for lower uncertainty in 3D topographic change quantification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, 541–559.