The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B2-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission II, 4–11 July 2026, Toronto, Canada

Evaluation of Systematic and Random Errors in Occupancy Grid Maps

Yuguang Liu[1], Marko Radanovic[1], Krista A. Ehinger[2], Kourosh Khoshelham[1]

[1] Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Australia
(yuguangl@student.unimelb.edu.au; m.radanovic@unimelb.edu.au; k.khoshelham@unimelb.edu.au)
[2] School of Computing and Information Systems, The University of Melbourne, Australia
(kris.ehinger@unimelb.edu.au)

Keywords: Map evaluation, Occupancy grid maps, LiDAR, Pose uncertainty, Map quality.

Abstract

Map evaluation for occupancy grid mapping (OGM) is critical in the field of high-definition mapping of the road environment for autonomous vehicles. Existing methods cannot adequately evaluate the systematic and random errors that might be present in OGM. This article introduces two evaluation metrics for OGM under LiDAR position uncertainty: Mean Signed Distance (MSD) and Mean Absolute Deviation (MAD). MSD quantifies systematic displacement of occupied cells, while MAD measures random error exhibited as boundary thickening. Unlike classification-based, probabilistic, and geometric metrics, MSD and MAD directly isolate displacement and thickening effects in OGM. We validate both metrics in a controlled synthetic environment and on a real indoor LiDAR dataset, showing better performance than conventional metrics.

1. Introduction

OGM is a widely used approach to representing environments in the field of Simultaneous Localization and Mapping (SLAM). It discretizes the environment into a grid of cells, and each cell stores a probability value to indicate the occupancy probability (Elfes, 1989). In SLAM, the LiDAR pose is estimated rather than known (Shu et al., 2025), and the pose estimation process inevitably involves errors from the Inertial Measurement Unit, odometry, and Global Navigation Satellite System, as well as from calibration and timing issues and environmental disturbances (Sier et al., 2023). As a result, LiDAR poses are inherently uncertain and never exactly known. The pose uncertainty propagates through the inverse sensor model (ISM) and distorts the spatial distribution of occupied cells, having negative impacts on robotics tasks such as navigation, localization, motion planning, and collision avoidance (Joubert et al., 2015). Pose uncertainty includes both position and orientation uncertainty, and this paper focuses on the positional component.

SLAM algorithms model, propagate, and reduce the pose uncertainty by multi-sensor fusion (Jang et al., 2024), particle filters (Ivanjko and Petrovic, 2004), graph constraints (Hess et al., 2016), and so on. However, pose uncertainty cannot be fully removed, and it still affects maps.

Most evaluation methods in OGM rely on classification-based (Ronecker et al., 2024a), probabilistic (Pearson et al., 2022), and geometric metrics (Dantanarayana et al., 2016), but these metrics are not designed to capture the effects caused by the LiDAR position uncertainty: systematic outward displacement of obstacles and random boundary thickening. Therefore, current evaluation metrics struggle to quantify and differentiate between random and systematic errors. This differentiation is important because random errors are inevitable and must only be quantified, whereas systematic errors must be identified and eliminated from the map. To address this gap, this work introduces two metrics: MSD to quantify systematic obstacle displacement, and MAD to quantify random obstacle thickening.

The main contributions of this paper are the following.

- We propose two OGM-specific evaluation metrics to separately characterize systematic displacement and random boundary thickening in occupancy grid maps.
- We demonstrate, in a controlled synthetic environment and on a real dataset, that the proposed metrics provide improved separation and interpretability of these two error modes compared to conventional metrics.

This paper is structured as follows. Section II demonstrates the systematic and random errors under the LiDAR position uncertainty. Section III introduces existing evaluation metrics. Section IV proposes two new evaluation metrics, and Section V summarizes and analyzes the evaluation results based on a synthetic environment and a real data set. Section VI concludes the paper.

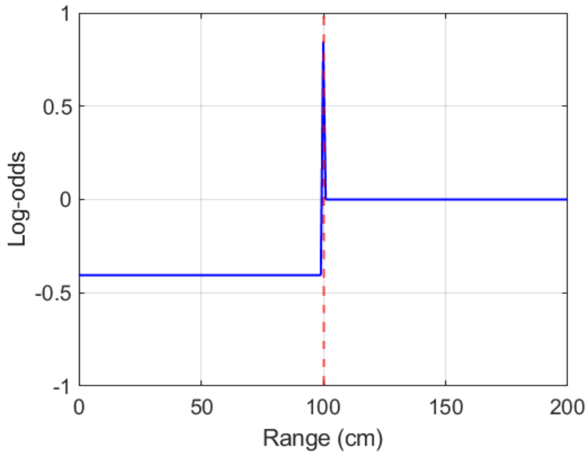
2. Occupancy Grid Map Errors

The OGM errors are classified into two types in this paper: the systematic error appears as outward displacement of occupied surfaces, and the random error shows as boundary thickening around the true surface. Both errors arise because repeated observations are fused by the ISMs used in the occupancy grid algorithm, which assign unbalanced log-odds updates to cells before and after the true distance, leading to a net bias as displacement and increased variance as thickening. In the following sections, we demonstrate these errors in 1D and 2D experiments.

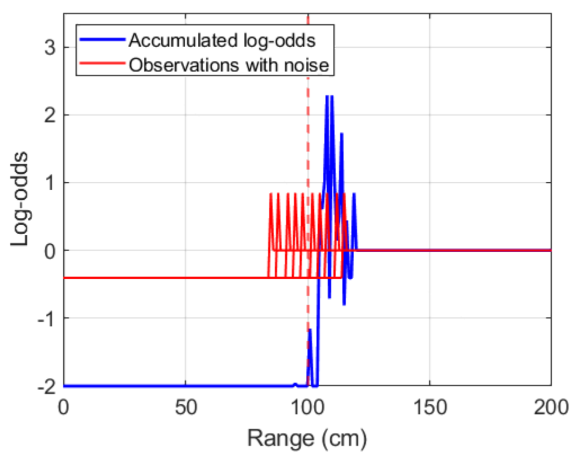
2.1 OGM Errors in 1D

The Octomap ISM (Hornung et al., 2013) is widely used in OGM. It assigns the occupancy probability of grid cells passed through by the laser ray, and the model is defined by (Hornung et al., 2013).

$$L(n \mid z_\ell) = \begin{cases} -0.4, & 0 \le z < z^*, \\ 0.85, & z = z^*, \\ 0, & z^* < z \le z_{max}, \end{cases} \tag{1}$$



(a) Octomap ISM in log-odds.



(b) Noisy observations in red and the accumulated log-odds in blue (σ_{sensor} = 5 cm).

Figure 1. Peak probability shifts and obstacle width increases of the Octomap ISM with LiDAR position uncertainty (cell size = 1 cm, z^* = 100 cm, log-odds clamp $[-2, 3.5]$).

$$L(n \mid z_{1:t}) = \max\left(\min\left(L(n \mid z_{1:t-1}) + L(n \mid z_t),\ l_{\max}\right),\ l_{\min}\right) \tag{2}$$

where n is the grid cell index, z is the distance of the queried cell along the ray, z^* is the returned distance, z_{max} is the maximum perception range of the LiDAR sensor, $L(n \mid z_t)$ is the ISM update given in Eq. (1), and $l_{\min}$ and $l_{\max}$ are the lower and upper clamping bounds of the log-odds values. The log-odds values are -0.4 before the returned distance, 0.85 at the returned distance, and 0 after the returned distance. These correspond to occupancy probabilities of 0.4, 0.7, and 0.5 for free, occupied, and unknown cells, respectively. In addition, a log-odds clamp is provided to define an upper and lower bound on the occupancy estimate: the minimum is -2, and the maximum is 3.5, corresponding to the probabilities of 0.12 and 0.97, respectively.

Consider a 1D environment with a laser beam measuring an obstacle at a true distance of 100 cm. Fig. 1(a) shows how the Octomap ISM assigns the log-odds corresponding to occupancy probabilities to cells before, at, and after the obstacle for a single measurement. The 1D maps in Fig. 1 come with a cell size of 1 cm. The red dashed line indicates an obstacle at a distance of 100 cm from the sensor in both maps. Fig. 1(b) shows how the log-odds are accumulated after 1000 LiDAR sensor observations with position uncertainty. The position uncertainty is simulated by Gaussian noise with a standard deviation of 5 cm (σ_{sensor}). Each red curve represents one noisy observation. The accumulated log-odds in blue are limited by the clamp from -2 to 3.5. As shown in Fig. 1(b), the obstacle indicated by the peak of the accumulated log-odds is wider than the actual obstacle and is significantly moved away from its ground-truth location.

2.2 OGM Errors in 2D Synthetic Environment

Further to OGM errors in 2D, Fig. 2 demonstrates the obstacle shifts and obstacle thickening in binary occupancy grid maps of a controlled synthetic environment with a cell size of 10 cm. The threshold was set at 0.65 (i.e., cells with an occupancy probability greater than 0.65 were classified as occupied). The synthetic dataset is a controlled L-shaped 2D environment designed to illustrate systematic displacement and random boundary thickening under repeated observations. The virtual sensor is placed inside the L-shaped environment, and repeated scans are fused under the same ISM as in Eq. (1). In the reference map shown in Fig. 2(a), the obstacle boundary is represented by an occupied band with a width of 5 dm, and the longer sides of the L-shape are 250 dm.

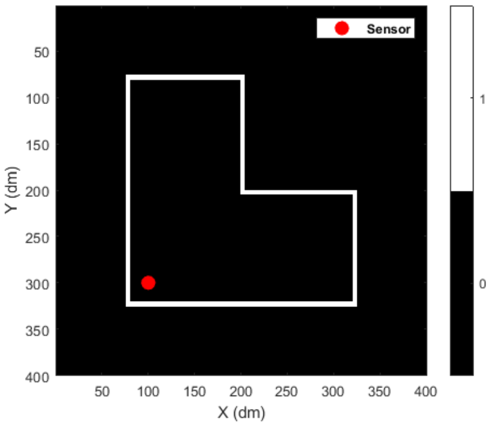
Three synthetic binary occupancy grid maps are generated that represent random error only (R), systematic error only (S), and combined random and systematic errors (RS), as shown in Fig. 2(b), Fig. 2(c), and Fig. 2(d), respectively. The synthetic maps in Fig. 2 are not intended to reproduce a realistic SLAM scene, but serve as controlled canonical cases in which random thickening and systematic displacement can be isolated visually and quantitatively.

Fig. 2(b) shows the random error displayed as the thickening of the surfaces, and Fig. 2(c) shows the systematic error displayed as outward displacement of the surfaces. In addition, Fig. 2(d) is the map that suffers from both errors.

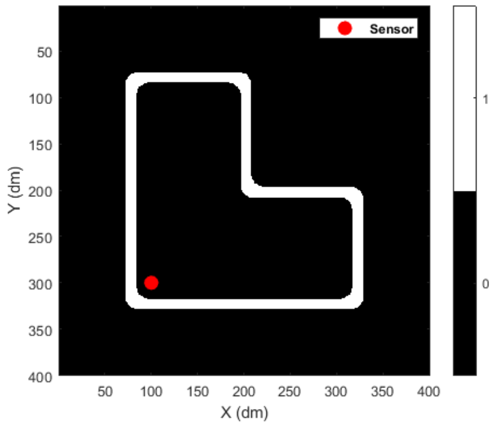
2.3 OGM Errors in 2D Real Dataset

The occupancy map errors in the 2D real dataset are shown in Fig. 3 with a cell size of 10 cm. The real dataset was obtained from the University of Pennsylvania Robotics Specialization repository and contains LiDAR scans and odometry data collected in an indoor corridor environment (Geedh, 2022). The dataset comprises 3701 LiDAR scans, each spanning an angular range from -2.36 to 2.36 rad and containing 1081 rays. Fig. 3(a) is the probabilistic reference map with ground-truth poses using the Octomap ISM, while Fig. 3(b) is the binary reference map converted from Fig. 3(a) using a threshold of 0.65. The obstacles in white cells on both maps are clear and precise.

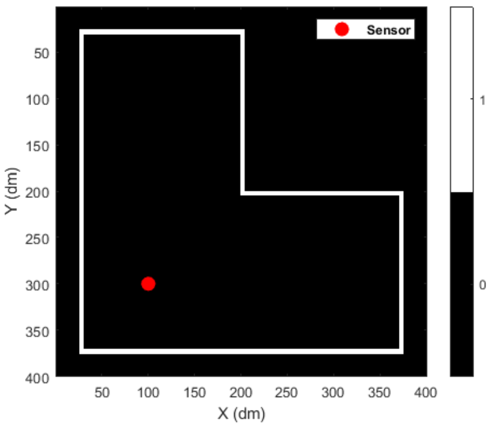
In comparison, Figs. 3(c) and 3(e) are the probabilistic predicted maps generated using the Octomap ISM with different levels of position uncertainty. The position uncertainty is simulated by position noise following a zero-mean Gaussian distribution with a standard deviation of 10 and 20 cm, respectively. Figs.



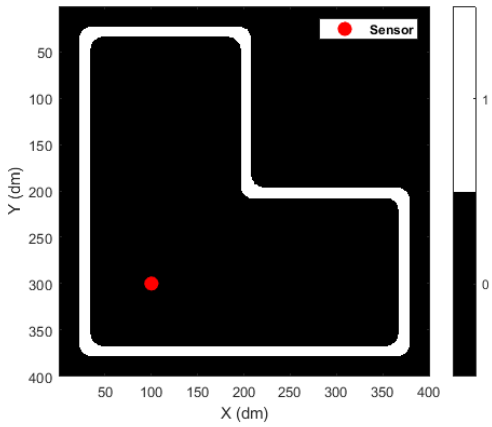
(a) Reference map (width = 5 dm, longer sides’ length = 250 dm).



(b) Map with the random error (Map_R): the occupied surfaces are thicker but still in the correct location (width = 11 dm, longer sides’ length = 250 dm).



(c) Map with the systematic error (Map_S): the occupied surfaces are thin but displaced (width = 5 dm, longer sides’ length = 350 dm).



(d) Map with both systematic and random errors (Map_RS): thicker and displaced occupied surface (width = 11 dm, longer sides’ length = 350 dm).

Figure 2. Systematic and random errors in binary occupancy grid maps of a controlled synthetic environment (cell size = 10 cm).

3(d) and 3(f) are the binary predicted maps converted from their corresponding probabilistic maps with a threshold of 0.65.

As the position uncertainty increases, the probabilistic and binary predicted maps show an increasingly significant peak shift (note the shifted occupied spaces) and thicker obstacles. In summary, the systematic error is presented on the map through outward displacement of the occupied areas, while the random error is reflected as the thickness increase of occupied regions.

3. Review of Existing Metrics

3.1 Classification-based Metrics

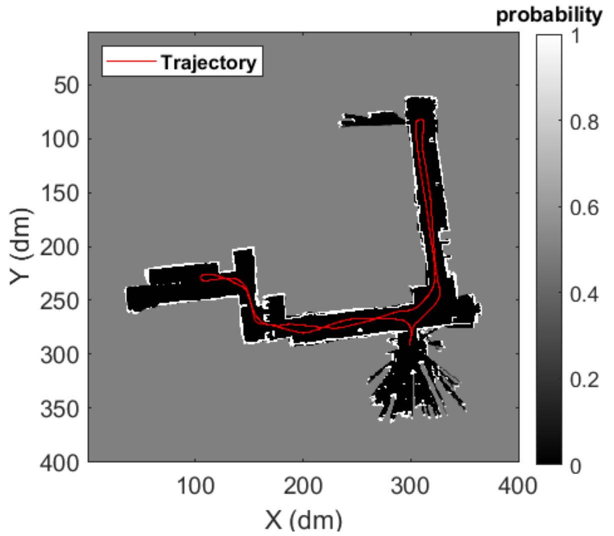
In a probabilistic occupancy grid map, each cell is labeled with a probability value from 0 to 1, showing the probability of being occupied (usually 0.5 denotes unknown). In a binary occupancy grid map, each grid cell is classified as either occupied (1) or free (0). To obtain a binary map from a probabilistic one, a threshold is applied: cells with probability values greater than or equal to the threshold will be occupied, while others are free. The chosen threshold sets the trade-off between false obstacles and missed obstacles. The classification-based metrics always operate on the binary grid maps and evaluate how well the predicted occupancy labels match the ground truth.

A confusion matrix is used to evaluate the performance of a classification model. It can provide a detailed evaluation of a predicted occupancy grid map by comparing its predicted states (occupied or free) to the ground truth states of the environment for each cell (Erkent et al., 2019). Four possible outcomes for comparison are listed:

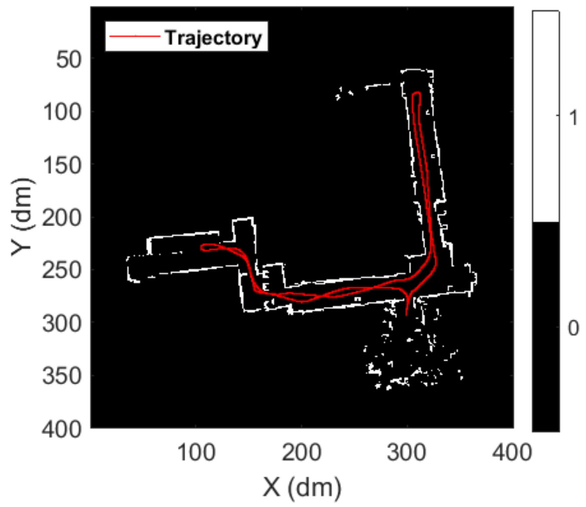
- True Positives (TP): Cells that are predicted as occupied and are actually occupied in the ground truth.
- True Negatives (TN): Cells that are predicted as free and are actually free in the ground truth.
- False Positives (FP): Cells that are predicted as occupied but are actually free in the ground truth.
- False Negatives (FN): Cells that are predicted as free but are actually occupied in the ground truth.

3.1.1 Precision

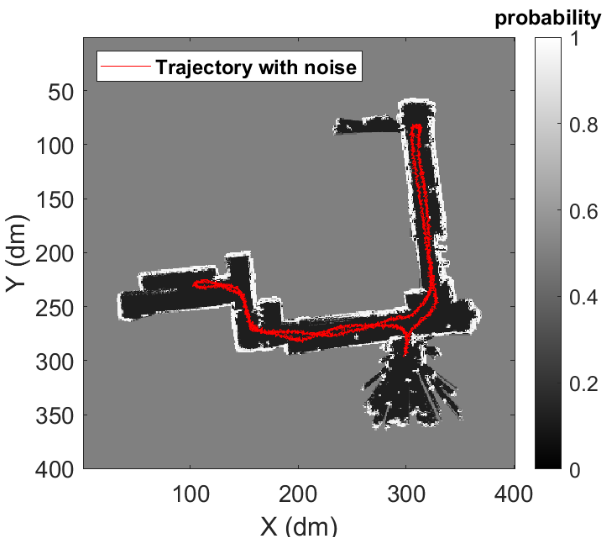
$$\text{Precision} = \frac{TP}{TP + FP}. \tag{3}$$



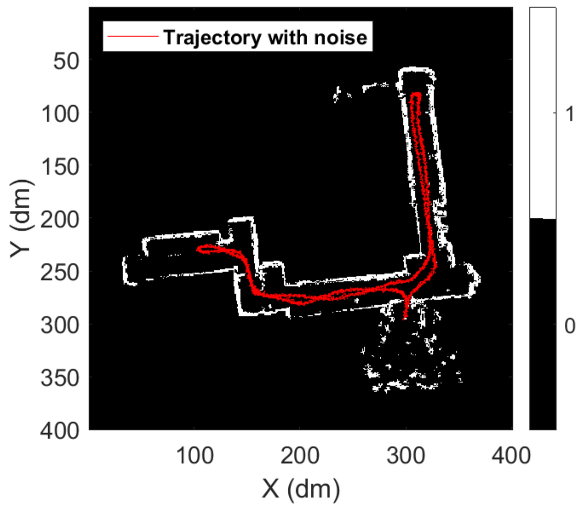
(a) Probabilistic reference map.



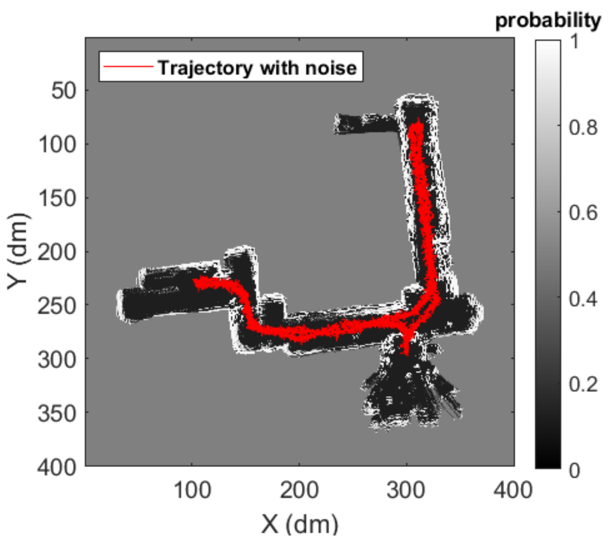
(b) Binary reference map (threshold = 0.65).



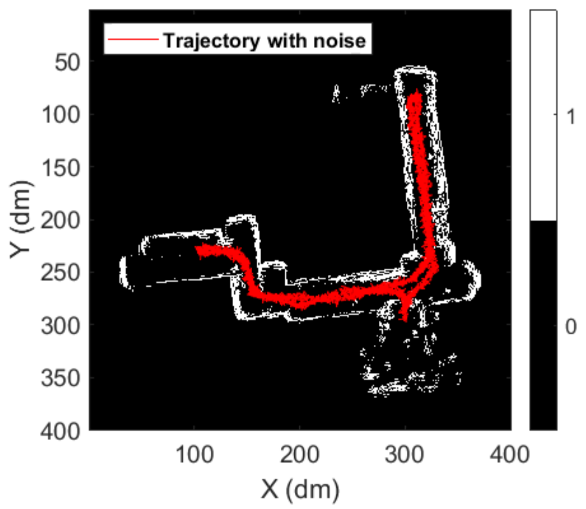
(c) Probabilistic predicted map generated using the Octomap ISM with position uncertainty (σ_{sensor} = 10 cm).



(d) Binary predicted map converted from the probabilistic predicted map (σ_{sensor} = 10 cm, threshold = 0.65).



(e) Probabilistic predicted map generated using the Octomap ISM with position uncertainty (σ_{sensor} = 20 cm).



(f) Binary predicted map converted from the probabilistic predicted map (σ_{sensor} = 20 cm, threshold = 0.65).

Figure 3. Systematic and random errors in predicted probabilistic and binary occupancy grid maps with LiDAR position uncertainty (cell size = 10 cm, probability clamp $[0.12, 0.97]$).

Precision defined in Eq. (3) measures the proportion of cells that are truly occupied (TP) out of all cells predicted as occupied (TP + FP), ranging from 0 to 1. High precision means that almost every predicted occupied cell really corresponds to a true wall or object, and there are few false obstacles. This matters in mapping because an FP can block a feasible path and make the planner overly conservative, and high precision can ensure that the predicted obstacles are trustworthy (Ronecker et al., 2024b). However, precision only focuses on the correctness of occupied cells, with no attention to the correct detection of free cells. Also, precision is spatially blind, which means that it counts label agreement cell by cell but does not consider where the occupied cells are located, how far they are displaced, or how obstacle geometry is deformed, so two maps with identical precision can have very different geometric or topological structures.

3.1.2 Recall

$$\mathrm{Recall} = \frac{TP}{TP+FN}. \tag{4}$$

Recall defined in Eq. (4) measures the proportion of truly occupied cells (TP) out of all actual occupied cells (TP + FN), ranging from 0 to 1. High recall means the map rarely misses obstacles, reducing the risk of planning through false free space for navigation (Ronecker et al., 2024a). Compared to precision, recall ignores false obstacles (FP) in the map and is also spatially blind.

3.1.3 F1 Score

$$\mathrm{F1} = \frac{2\,\mathrm{Precision}\cdot\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} = \frac{2TP}{2TP+FP+FN}. \tag{5}$$

F1 score in Eq. (5) is the harmonic mean of precision and recall, measuring the balance between avoiding false obstacles (FP) and avoiding missed obstacles (FN), ranging from 0 to 1. A high F1 score indicates that the map has few false obstacles and few missed obstacles (Jin et al., 2024). On the other hand, F1 score ignores the true free space (TN) in the map and is also spatially blind. Additionally, under class imbalance, like a binary grid map dominated by free cells, F1 score provides a single-class view for the occupied class and thus might hide poor performance on the free class.

3.1.4 Intersection over Union (IoU)

$$\mathrm{IoU} = \frac{TP}{TP+FP+FN}. \tag{6}$$

IoU in Eq. (6) is the intersection over union between the predicted occupied set and the true occupied set: the intersection is TP and the union is TP + FP + FN, ranging from 0 to 1. A high IoU means the predicted obstacle regions align well with the ground truth overall (Ronecker et al., 2024b). IoU reflects both false obstacles (FP) and missed obstacles (FN), but it ignores the true free space (TN) in the map and remains spatially blind. Moreover, IoU is sensitive to map alignment, and even small misalignments between maps can reduce IoU substantially.

3.1.5 AUC-PR

The precision-recall curve is a curve that calculates precision and recall for every threshold. AUC-PR is the area under the precision–recall curve, ranging from 0 to 1. For map evaluation, a high AUC-PR means that across all thresholds, the binary grid map keeps precision high while achieving high recall, which indicates that predicted obstacles are usually real (low FP) and most true obstacles are found (low FN) (Sodhi et al., 2019). It is also threshold-free because it considers performance across all possible thresholds. Regarding the limitations, AUC-PR is also spatially blind and ignores free space correctness (TN).

3.2 Probabilistic Metrics

Probabilistic metrics measure how well the predicted probabilities in the generated map reflect the true occupancy likelihood. They work with continuous occupancy probabilities rather than binary values.

3.2.1 Mean Square Error (MSE)

$$\mathrm{MSE} \;=\; \frac{1}{N}\sum_{i=1}^{N}\bigl(p_i-g_i\bigr)^2. \tag{7}$$

MSE in Eq. (7) measures the mean squared error between the predicted occupancy probabilities $p_i \in [0,1]$ and the known ground-truth values $g_i \in [0,1]$ for each evaluated cell, generating a value from 0 to 1. It quantifies the average squared difference between the predicted and actual occupancy probabilities, with a lower MSE indicating that the two probability sets match closely (Pearson et al., 2022). As limitations, the score is spatially blind, treats all cells equally so class imbalance might dominate the average, and does not distinguish between systematic misalignment and calibration errors.

3.2.2 Map Score

$$\mathrm{MapScore} \;=\; 1 \;-\; \frac{1}{N}\sum_{i=1}^{N}\bigl|\,p_i-g_i\,\bigr|. \tag{8}$$

Map score in Eq. (8) measures the cell-wise similarity between the predicted probabilities $p_i \in [0,1]$ and the ground-truth values $g_i \in [0,1]$, making a value from 0 to 1 (Collins et al., 2007). A high map score means the predicted map's cell probabilities closely match the reference values on average. In addition, it has the same limitations as MSE.

3.2.3 Negative Log-Likelihood (NLL)

$$\mathrm{NLL} \;=\; -\,\frac{1}{N}\sum_{i=1}^{N}\Bigl[\,g_i\,\log(p_i) \;+\; (1-g_i)\,\log\bigl(1-p_i\bigr)\Bigr]. \tag{9}$$

NLL in Eq. (9) computes the Bernoulli cross-entropy between the predicted occupancies $p_i \in [0,1]$ and the ground-truth labels $g_i \in [0,1]$ (Ramos and Ott, 2016), giving a value in $[0,+\infty)$. A low NLL means the occupied cells are usually assigned probabilities close to 1 and free cells close to 0 (Snyder et al., 2024). For limitations, NLL is also spatially blind, sensitive to probability noise, and unbounded above, which requires numerical stabilization when p_i is near 0 or 1.

3.3 Geometric Metrics

Classification-based metrics such as IoU and F1 score only tell whether each grid cell is labeled as occupied or free, and they ignore where those occupied cells lie in space and how they are arranged. In mapping, however, spatial alignment matters: two maps can have the same number of occupied cells but with walls shifted a few cells apart or distort their shapes. To capture these effects, geometric metrics are applied to measure how far two obstacle maps differ in physical space.

Geometric metrics treat the occupied cells in the predicted map and the ground-truth map as point sets:

$$
\begin{aligned}
A &= \{(u_i, v_i)\ :\ \text{points occupied in the predicted map},\\
B &= \{(u_j, v_j)\ :\ \text{points occupied in the ground-truth map},\\
a &\in A\ :\ \text{a single occupied point in the predicted map},\\
b &\in B\ :\ \text{a single occupied point in the ground-truth map}.
\end{aligned}
$$

3.3.1 Chamfer Distance (CD)

$$
d_{\mathrm{CD}}(A,B) = \frac{1}{|A|}\sum_{a\in A}\min_{b\in B}\|a-b\| \;+\; \frac{1}{|B|}\sum_{b\in B}\min_{a\in A}\|b-a\|. \tag{10}
$$

Chamfer distance measures the average nearest-neighbor distance between the predicted and ground-truth occupied cells (Dantanarayana et al., 2016). For each occupied cell in one map, it finds the distance to the nearest occupied cell in the other map and then averages both directions. Chamfer distance reflects overall geometric similarity between the two maps. As limitations, it hides outliers and ignore structural failures such as broken or missing walls.

3.3.2 Hausdorff Distance (HD)

$$
\begin{aligned}
h(A,B) &= \max_{a\in A}\min_{b\in B}\|a-b\|,\\
d_{\mathrm{HD}}(A,B) &= \max\{\,h(A,B),\, h(B,A)\,\}.
\end{aligned} \tag{11}
$$

Hausdorff distance measures the maximum spatial deviation between the predicted and ground-truth occupied cells (Hu et al., 2025). For each occupied cell in one map, it finds the distance to the nearest point in the other map, then takes the maximum over both directions, so it captures the largest geometric mismatch between the two maps. A low Hausdorff distance means every obstacle is close to its ground-truth position. A low Chamfer distance means that, on average, the predicted occupied cells lie very close to the corresponding ground-truth occupied cells across the map. As limitations, Hausdorff distance is sensitive to outliers or single noisy cells, and it also ignores structural failures on the map.

3.3.3 Earth Mover's Distance (EMD)

$$
\begin{aligned}
EMD(P,Q) \;=\; \min_{\gamma\geq 0}\; & \sum_{i=1}^{m}\sum_{j=1}^{n}\gamma_{ij}\,\|u_i - v_j\|,\\
\text{subject to } & \sum_{j=1}^{n}\gamma_{ij} = p_i \quad (i=1,\dots,m),\\
& \sum_{i=1}^{m}\gamma_{ij} = q_j \quad (j=1,\dots,n),\\
& \sum_{i=1}^{m}p_i = \sum_{j=1}^{n}q_j = 1,
\end{aligned} \tag{12}
$$

where u_i and v_j are the cell centers of the predicted map and the ground-truth map, respectively; $p_i \in [0,1]$ and $q_j \in [0,1]$ are the non-negative occupancy probabilities used as masses normalized to sum to one; and γ_{ij} is the transport plan that moves mass from u_i to v_j at cost $\|u_i - v_j\|$, typically Euclidean distance.

Earth mover's distance measures the minimum work required to transform the predicted occupancy distribution into the ground truth (Hu et al., 2025). A low EMD means that the predicted probability mass lies close to the true mass spatially, and only a small amount of mass needs to be moved over short distances, thus indicating a small average misalignment of obstacles. As limitations, EMD requires non-negative masses and equal total mass; it is computationally heavier than Chamfer/Hausdorff distance, sensitive to global misregistration, and does not capture structural failures.

In summary, existing metrics mainly report overall agreement, probability consistency, or geometric discrepancy. However, they do not explicitly decompose map errors into directional bias (systematic displacement) and spread around the mean boundary location (random thickening). In classical statistics, systematic effects may be reflected by signed residual means, but such quantities are not commonly formulated for occupancy grid boundary evaluation. This gap motivates the proposed signed-distance formulation.

4. Proposed Evaluation Metrics

Evaluation metrics introduced in the previous section provide a single score and do not differentiate between random and systematic errors. This is a problem: random errors are partly unavoidable and are often handled by averaging or smoothing, while systematic errors should be corrected or eliminated because map applications such as localization and collision avoidance are more sensitive to systematic displacement. Two evaluation metrics are proposed to independently quantify the systematic error and random error caused by the position uncertainty in OGM.

4.1 Mean Signed Distance (MSD)

Let us define the occupied set on the shifted–thickened map by $\Omega = \{\, i \mid P_{\text{shift}}(i) > \tau \,\}$, and N is the number of occupied cells. For each cell $i \in \Omega$, its signed nearest–neighbour distance to the reference occupancy set Ω_{ref} is

$$
d_i \;=\; s_i \min_{j\in\Omega_{\text{ref}}}\big\|\mathbf{x}_i - \mathbf{x}_j\big\|_2, \tag{13}
$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ denote the spatial coordinates of the occupied cell i in the evaluated map and the occupied cell j in the reference map, respectively. The sign s_i encodes the motion direction of the obstacle boundary:

$$
s_i = \begin{cases} -1, & \text{the obstacle moves toward the sensor,}\\ +1, & \text{the obstacle moves away from the sensor.} \end{cases} \tag{14}
$$

The MSD is then

$$
\mathrm{MSD} \;=\; \frac{1}{N}\sum_{i\in\Omega} d_i. \tag{15}
$$

Large negative values indicate inward displacement, whereas large positive values indicate outward displacement; the magnitude quantifies the systematic shift. More importantly, if the map only contains random errors (thickened surfaces), the negative and positive distances cancel out, resulting in an MSD of 0. Therefore, MSD only quantifies the systematic error.

4.2 Mean Absolute Deviation (MAD)

To measure thickness independent of the shift, we remove the mean from each signed distance and average the absolute residuals. With $r_i = d_i - \mathrm{MSD}$,

$$\mathrm{MAD} \;=\; \frac{1}{N}\sum_{i=1}^{N}|r_i| \;=\; \frac{1}{N}\sum_{i=1}^{N}|d_i - \mathrm{MSD}|. \qquad (16)$$

A larger MAD reflects larger random errors and therefore thicker occupied regions; values near zero correspond to narrow, sharply defined obstacle surfaces.

Metrics	Map_R	Map_S	Map_RS
Precision ↑	0.451	0.181	0.084
Recall ↑	0.999	0.255	0.261
F1 Score ↑	0.621	0.212	0.127
IoU ↑	0.450	0.119	0.068
MSE ↓	0.038	0.058	0.098
Map Score ↑	0.923	0.942	0.847
NLL ↓	0.116	1.606	0.811
CD [dm] ↓	0.584	36.302	35.243
HD [dm] ↓	7.000	70.711	70.036
EMD [dm] ↓	1.969	62.493	62.828
MSD [dm] ↓	0.076	39.165	39.034
MAD [dm] ↓	16.357	1.201	16.363

Note. Arrows indicate preferred direction: ↑ higher is better; ↓ lower is better. For MSD we evaluate the magnitude (|MSD|), smaller is better.

Table 1. Evaluation results of 2D synthetic environment.

Metrics	Predicted Map 1 σ_{sensor} = 10 cm	Predicted Map 2 σ_{sensor} = 20 cm
Precision ↑	0.349	0.184
Recall ↑	0.602	0.393
F1 Score ↑	0.442	0.250
IoU ↑	0.284	0.143
AUC-PR ↑	0.225	0.076
MSE ↓	0.010	0.017
Map Score ↑	0.981	0.967
NLL ↓	0.683	0.704
CD [cm] ↓	7.964	17.487
HD [cm] ↓	110.445	114.018
EMD [cm] ↓	140.682	191.243
MSD [cm] ↓	7.401	19.176
MAD [cm] ↓	10.165	19.737

Note. Arrows indicate preferred direction: ↑ higher is better; ↓ lower is better. For MSD we evaluate the magnitude (|MSD|), smaller is better.

Table 2. Evaluation results of 2D real dataset.

5. Results

To evaluate the effectiveness of the proposed metrics, we calculate these for the occupancy grid maps shown in Section II. Table 1 presents the evaluation results for the synthetic maps shown in Fig. 2(b), Fig. 2(c), and Fig. 2(d) compared to the reference map Fig. 2(a). For Map_S and Map_RS, the classification-based metrics decrease sharply indicating that they are sensitive to geometric displacement caused by the systematic error, but fail to indicate the random errors in Map_R. The results of probabilistic metrics are quite similar for maps with random errors, systematic errors, and both. The geometric metrics reflect the systematic error well for Map_S and Map_RS, but they do not reflect the random error in Map_RS compared to Map_S. In contrast, MSD and MAD separately and consistently capture the dominant systematic displacement and random thickening patterns in the three maps.

Table 2 presents the evaluation results of the predicted maps in Fig. 3 compared to the reference maps Fig. 3(a) or Fig. 3(b). As the position uncertainty increases, both errors increase, presented as the growing outward displacement and thickening obstacle surfaces in the predicted maps. The classification-based metrics decrease, and the geometric metrics climb, indicating deteriorating map quality, but they cannot quantify obstacle displacement and boundary thickening separately. In contrast, MSD and MAD dynamically capture their corresponding errors as the position uncertainty varies. In addition, MAD is not intended to equal the wall thickness directly. It is the mean absolute deviation of signed boundary distances after removing the mean shift. Therefore, MAD should be interpreted as an indicator of boundary thickening rather than the physical wall thickness itself.

The numerical values in Table 1 and Table 2 are influenced by the selected ISM and clamping parameters. Their effects have been analyzed in more detail in our related work (Liu et al., 2026). In this paper, the Octomap ISM is used as a representative and widely adopted baseline. However, the proposed evaluation framework is not restricted to this specific ISM, because MSD and MAD are computed from the resulting map boundaries rather than from the analytic form of Eq. (1) itself.

6. Conclusion

In this article, we propose two evaluation metrics, MSD and MAD, that quantify systematic and random errors in OGM arising from the LiDAR position uncertainty, respectively. Existing evaluation metrics do not separate obstacle displacement from boundary thickening, whereas MSD and MAD are designed to quantify these two effects separately. Practically, MSD can be used as an indicator of map bias caused by calibration errors, synchronization issues, or trajectory drift, while MAD can be used to evaluate map boundary sharpness and noise sensitivity. The performance of the proposed metrics is verified through a controlled synthetic environment and a real indoor LiDAR dataset. In the future, we will consider map errors caused by orientation uncertainty in OGM, and also extend the evaluation to three-dimensional cases.

References

Collins, T., Collins, J., Ryan, D., 2007. Occupancy grid mapping: An empirical evaluation. *2007 Mediterranean Conference on Control Automation*, 1–6.

Dantanarayana, L., Dissanayake, G., Ranasinge, R., 2016. C-LOG: A Chamfer distance based algorithm for localisation in occupancy grid-maps. *CAAI Transactions on Intelligence Technology*, 1(3), 272-284. https://doi.org/10.1016/j.trit.2016.10.003.

Elfes, A., 1989. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6), 46-57.

Erkent, Ö., Wolf, C., Laugier, C., 2019. End-to-end learning of semantic grid estimation deep neural network with occupancy grids. *Unmanned systems*, 7(03), 171–181.

Geedh, P. O., 2022. pgeedh/RoboticsSpecialization-UPenn-Estimation-and-Learning. original-date: 2022-11-02T01:30:42Z.

Hess, W., Kohler, D., Rapp, H., Andor, D., 2016. Real-time loop closure in 2d lidar slam. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 1271–1278.

Hornung, A., Wurm, K. M., Bennewitz, M., Stachniss, C., Burgard, W., 2013. OctoMap: an efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 34(3), 189–206. http://link.springer.com/10.1007/s10514-012-9321-0.

Hu, X., Wu, J., Jia, M., Yan, H., Jiang, Y., Jiang, B., Zhang, W., He, W., Tan, P., 2025. Mapeval: Towards unified, robust and efficient slam map evaluation framework.

Ivanjko, E., Petrovic, I., 2004. Extended kalman filter based mobile robot pose tracking using occupancy grid maps. *Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference (IEEE Cat. No.04CH37521)*, 1, 311–314 Vol.1.

Jang, H., Kim, T., Ahn, K., Jeon, S., Kang, Y., 2024. Dynamic Occupancy Grid Map with Semantic Information Using Deep Learning-Based BEVFusion Method with Camera and LiDAR Fusion. *Sensors*, 24(9). https://www.mdpi.com/1424-8220/24/9/2828.

Jin, Y., Hoffmann, M., Deligiannis, A., Fuentes-Michel, J.-C., Vossiek, M., 2024. Semantic Segmentation-Based Occupancy Grid Map Learning With Automotive Radar Raw Data. *IEEE Transactions on Intelligent Vehicles*, 9(1), 216-230.

Joubert, D., Brink, W., Herbst, B., 2015. Pose Uncertainty in Occupancy Grids through Monte Carlo Integration. *Journal of Intelligent & Robotic Systems*, 77(1), 5–16. http://link.springer.com/10.1007/s10846-014-0093-y.

Liu, Y., Radanovic, M., Ehinger, K., Khoshelham, K., 2026. Occupancy Grid Mapping with Uncertain LiDAR Data. *IEEE Transactions on Instrumentation and Measurement*, 75. https://doi.org/10.1109/TIM.2026.3667328.

Pearson, E., Doherty, K., Englot, B., 2022. Improving obstacle boundary representations in predictive occupancy mapping. *Robotics and Autonomous Systems*, 153, 104077. https://doi.org/10.1016/j.robot.2022.104077.

Ramos, F., Ott, L., 2016. Hilbert maps: Scalable continuous occupancy mapping with stochastic gradient descent. *The International Journal of Robotics Research*, 35(14), 1717-1730. https://doi.org/10.1177/0278364916684382.

Ronecker, M. P., Diaz, X., Karner, M., Watzenig, D., 2024a. Deep learning-driven state correction: A hybrid architecture for radar-based dynamic occupancy grid mapping.

Ronecker, M. P., Schratter, M., Kuschnig, L., Watzenig, D., 2024b. Dynamic occupancy grids for object detection: A radar-centric approach.

Shu, Z., Bei, S., Dai, J., Li, L., Chen, Z., 2025. MoCap2GT: A High-Precision Ground Truth Estimator for SLAM Benchmarking Based on Motion Capture and IMU Fusion. *arXiv preprint arXiv:2507.12920*. https://arxiv.org/abs/2507.12920.

Sier, H., Qingqing, L., Xianjia, Y., Queralta, J. P., Zou, Z., Westerlund, T., 2023. A Benchmark for Multi-Modal LiDAR SLAM with Ground Truth in GNSS-Denied Environments. *Remote Sensing*, 15(13), 3314. https://www.mdpi.com/2072-4292/15/13/3314.

Snyder, S., Capodieci, A., Gorsich, D., Parsa, M., 2024. Brain Inspired Probabilistic Occupancy Grid Mapping with Vector Symbolic Architectures. *arXiv preprint arXiv:2408.09066*. https://arxiv.org/abs/2408.09066.

Sodhi, P., Ho, B.-J., Kaess, M., 2019. Online and consistent occupancy grid mapping for planning in unknown environments. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7879–7886.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-477-2026