Assessment of Lidar Point Cloud Simulation Using Phenomenological Range–Reflectivity Limits for Feature Validation

RELINDIS ROTT[ID] AND SELIM SOLMAZ[ID] (Senior Member, IEEE)

Department of Electrics, Electronics and Software, Virtual Vehicle Research GmbH, 8010 Graz, Austria

CORRESPONDING AUTHOR: R. ROTT (e-mail: relindis.rott@ v2c2.at)

ABSTRACT We present an assessment of simulated lidar point clouds based on different phenomenological range–reflectivity models. In sensor model development, the validation of individual model features is favorable. For lidar sensors, range limits depend on surface reflectivities. Two phenomenological feature models are derived from the lidar range equation, for clear and adverse weather conditions. The underlying parameters are the maximum ranges for best environment conditions, based on sensor datasheets, and a maximum range measurement for attenuation conditions. Furthermore, an assessment of different feature models is needed, similar to unit tests. Therefore, resulting point clouds are compared with respect to the total number of corresponding points and the number of points with no correspondences for pair-wise cloud comparison. Applications are presented using a point cloud lidar model. Results of the point cloud comparison are demonstrated for a single scene or time step and an entire scenario of 40 time steps. When a reference point cloud is provided by the sensor manufacturer, feature validation becomes possible.

INDEX TERMS Adverse weather, atmospheric attenuation, automotive domain, lidar sensor models, model feature validation, synthetic point cloud comparison, virtual testing.

I. MOTIVATION

THE DEVELOPMENT of autonomous vehicles and advanced driver assistance systems requires extensive testing. Exhaustive testing cannot be covered by real-world tests only, but also requires virtual validation methods. For driving functions, different virtual testing requirements and regulations exist, e.g., [2], while for virtual validation of environment perception systems and perception sensor models, the development of reliable validation procedures in simulation is still an open topic.

This topic is addressed in different research projects [3], [4], [5] and publications [6], [7]. General validation metrics for radar, lidar, and camera sensors are discussed by Schaermann et al. [7]. In particular, they distinguish between validation procedures for object list and raw-data type sensor models. Hanke et al. [6] presented a real-time capable lidar sensor measurement model based on raytracing. The output of the model is a point cloud that is used to generate an occupancy grid which is a suitable abstraction layer for further processing the data for the automated driving pipeline. Based on these, data types sensor model validation procedures are presented. An overall error of the simulated and real-world data occupancy grid is computed together with the Barons and Pearson correlation. Deviations of simulated and real-world data are discussed and different projection methods for grid sampling

are compared to reduce the simulation error. In their work, a 4-layer lidar sensor is considered as an example, where the layers describe the vertical resolution and field of view of the sensor.

Meanwhile, there are many automotive lidar sensors on the market which provide a much larger field of view and a high angular resolution. Many sensors offer a 360° horizontal field of view. To address the challenge of high-resolution point clouds and to be able to compare point clouds independently of the sensor field of view and the total number of points, applicable metrics are needed.

A good overview of validation strategies in simulation-based automotive testing is presented in a survey by Donà and Ciuffo [8]. It summarizes the different types of models needed for virtual validation and possible methods to assess their performances. Different methodologies are discussed, which occur along the testing pipeline, such as model verification, operation validation, and comparison with measurement data. Moreover, different metrics for point cloud validation are summarized.

Best practices for general simulation validation can be found in the NATM document [9] which focuses on new methods for automated driving topics, or, exemplary for general modeling and simulation, in a memorandum [10] by the EU Aviation Safety Agency (EASA). In Section II of the NATM document [9], validation procedures are discussed for radar and lidar sensors. Similarly, a distinction is made between explicit and implicit validation methods, which refer to raw-data comparison (e.g., measurement data and sensor output) and postprocessed perception data, such as detections or object lists. For raw-data comparison, such as point clouds, they propose Euclidean distance and Pearson correlation.

In the field of computer graphics, the Hausdorff distance [11] is applied for point set comparison of 3-D objects. Another distance measure used for point clouds in this field is the Wasserstein distance [12].

A prerequisite for realistic environment perception modeling are sensor models for radar, lidar, and camera. They are needed as input for perception algorithm testing in virtual environments. However, there is no standard on how the quality of sensor model outputs can be assessed. This leads to the research question on how to compare sensor model outputs. We discuss this for a lidar point cloud model using different range–reflectivity feature models.

A. NEED FOR FEATURE VALIDATION

For sensor model developers, the perception sensor is often a black box that is only known to the manufacturer, apart from the physical detection principle. A good overview and discussion of radiometric processing of general lidar sensors are provided in [13]. However, for autonomous systems, a full sensor system simulation is mostly not necessary. The simulation scope can be limited to predominant sensor effects, with a high impact on the system under test, which
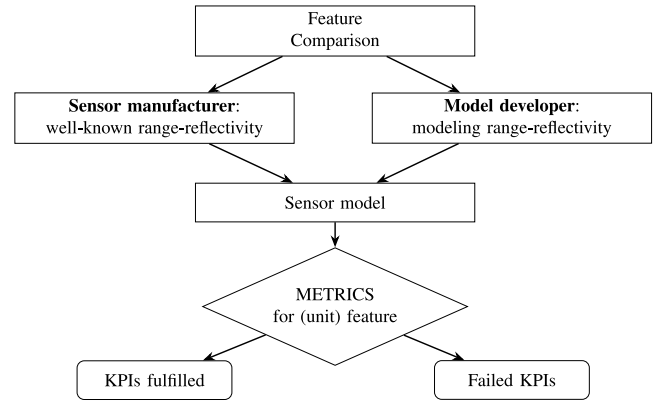
FIGURE 1. Overview of suggested process for feature validation. The sensor model validation process is simplified by validating sensor effects as units. Two models of a feature are compared in virtual environments. Key performance indicators (KPIs) need to be defined in advance.

are typically object perception algorithms. A certain sensor effect model can be interpreted as a model feature.

Any sensor model consists of different model features. Each feature needs to be tested separately to validate the model. In Fig. 1, an overview of a possible assessment strategy is shown, to validate a modeled feature as a unit. While the manufacturer often has detailed knowledge about the sensor properties, these insights often cannot be shared with third parties.

In this study, we use a phenomenological sensor model feature for lidar sensors as an example to demonstrate a possible strategy for feature validation in virtual environments. The sensor effect considered is the dependence of maximum range, or range limit, on the target reflectivity of the surface detected by the lidar sensor.

This phenomenological lidar sensor feature can be applied independently to different types of lidar sensor models. For demonstration, we use an existing physical sensor model, LiMOX [14], [15], based on raytracing of 3-D environments, to generate point clouds. Moreover, we consider the impact of adverse weather conditions on the maximum range.

The significant contribution is twofold. One aspect is the derivation of phenomenological range–reflectivity models based on the lidar range equation, for clear, and adverse weather conditions with respect to maximum range measurements. These models are applied in a lidar point cloud simulation. The second aspect is the simple metric proposed for the comparison of the different resulting point clouds. Comparative results are shown for both a scene, and a 40-time-step scenario to demonstrate how the validation of a single simulated feature, the range–reflectivity limit in this case, can be prepared.

By comparing simulation models based on single features rather than validating an entire sensor model, the implementation and validation process can be facilitated and sensor effects decoupled. We limit our study to the comparison of simulation results.

This article is structured as follows. First, we introduce the phenomenological model at best weather conditions in

TABLE 1. Nomenclature.

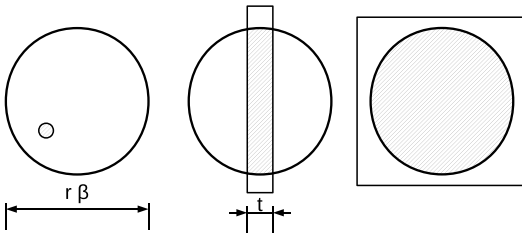| Symbol	Quantity
r	range
d	receiver aperture (diameter)
β	beam divergence (angle)
ζ	cross section of beam and target
ρ	reflectivity
P_r	received signal power
P_t	transmitted signal power
η_{atm}	efficiency of atmospheric transmission
η_{sys}	efficiency of optical system
A	target area
σ	attenuation coefficient
Ω	scattering solid angle

FIGURE 2. Laser beam intersection area: point target, wire target of thickness t, and extended target blocking the entire laser beam. The beam diameter is $r \cdot \beta$.

Section II. Next, we discuss adverse weather influences and their attenuation effects, and how these can be included in the model in Section III. In the subsequent Section IV, common metrics for the assessment of point clouds from literature are summarized. An application of the model feature in a physical lidar sensor simulation is presented in Section V, using two further approaches for comparison. Finally, assessment strategies are shown.

II. FEATURE: PHENOMENOLOGICAL RANGE–REFLECTIVITY DEPENDENCE

For automotive lidar sensors, range–reflectivity values are typically provided by the sensor manufacturer, often for two Lambertian targets of different reflectivities. The functional relation for all reflectivities is basically unknown to the user. For the development of sensor models, this relation is fundamental for simulating realistic sensor outputs.

From the lidar equation [16] a basic approach can be derived. For the following relations, we use a nomenclature defined in Table 1. The received power is a function of the transmitted power, sensor system parameters, atmospheric conditions of the volume the ray passes through, and the ray–surface interaction

$$P_r = P_t \frac{d^2}{4\pi\, r^4 \beta^2}\ \eta_{\mathrm{sys}}\eta_{\mathrm{atm}}\ \zeta. \tag{1}$$

The cross section of the interaction with general materials is $\zeta = (4\pi/\Omega)\,\rho\,A$. For a Lambertian target, which equals an ideally diffuse reflective surface, the cross section is $\zeta_{\mathrm{Lamb}} = 4\rho\,A$. Basically, three cases of target–beam intersections are

distinguished for the interaction of the laser beam with a solid target [17]: a point target, a wire target, or an extended target (see Fig. 2). For the wire target, the target area $A_w = t\, r\beta$, and for the extended target, which blocks the entire laser beam, $A_{\mathrm{ext}} = \pi (r\beta/2)^2$. Thus, the exponential factor of r in the lidar equation varies between $n \in \{2, 3, 4\}$, for extended, wire or point-like targets, respectively. Restructuring the lidar equation to focus on reflectivity ρ and range r yields

$$c_n = \eta_{\mathrm{atm}}\ \frac{\rho}{r^n} \tag{2}$$

where all parameters of the sensor system and target interaction are collected in c_n. The atmospheric attenuation is $\eta_{\mathrm{atm}} = \exp(-2\sigma r)$, an exponential function of r, yielding

$$c_n = \frac{\rho}{r^n}\ e^{-2\sigma r}. \tag{3}$$

For short distances, the atmospheric attenuation factor σ can be neglected and therefore, in clear conditions $\sigma \approx 0$.

However, the further away a target is from the sensor, the larger the area of the beam gets due to the beam divergence. A simplified form of the lidar equation for the case of zero attenuation is

$$c_n = \frac{\rho}{r^n}. \tag{4}$$

To model the maximum detectable range $r_{\max}$ for a given reflectivity of a target surface, the minimal detectable signal strength $P_{r,\min}$ of the receiver is assumed, yielding

$$r_{\max} = \left(\frac{\rho}{c_n}\right)^{\frac{1}{n}}. \tag{5}$$

However, sensor internal processing can influence the detection of objects near and far from the sensor. Without further knowledge of the sensor-specific parameters as well as the exact interaction of the laser beam with the target hit and their cross section, we build a phenomenological model using datasheet parameters and (5) as the basis for a fit function with no significant attenuation effects, thus under optimal visibility conditions.

In Fig. 3, different fit functions are compared for two representative datasheet parameters, provided by the sensor manufacturer.

The fit of (4) yields $n = 3$, see the red curve in Fig. 3, for the exemplary parameters. Therefore, the lidar equation curve fit for the given datasheet values is equal to the fit of $n = 3$. Similarly, fit functions are shown for $n = 2$, which corresponds to the lidar equation for an extended target and $n = 4$, for the point targets, smaller than the lidar beam cross section. These lidar equation fit functions only differ insignificantly from one another for the ranges of interest, except for the linear fit, which is nonphysical and added only for demonstration. However, for retro-reflectors with reflectivities above 100%, the fitted curves show larger deviations.

Further sensor effects, such as dynamic occlusion, due to strong reflectors in close vicinity to weakly reflecting targets, or internal signal processing, have a significant influence on
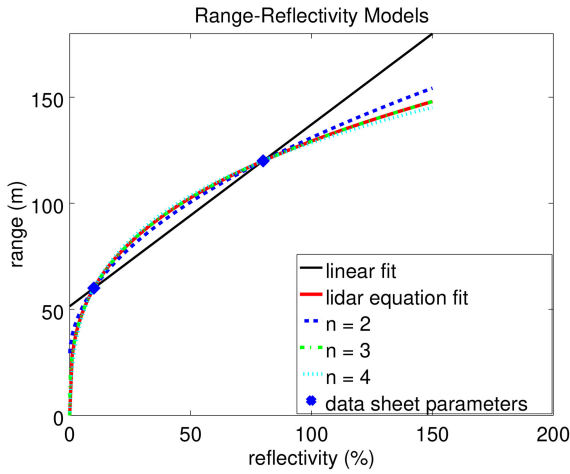
FIGURE 3. Maximum range curve fitting for clear conditions, where the atmospheric attenuation can be neglected, based on two typical datasheet value pairs of reflectivity and maximum range [10%, 60 m] and [80%,120 m].

the maximum detection range, which cannot be covered in this approach.

III. IMPACT ON LIDAR SENSORS DUE TO ADVERSE WEATHER CONDITIONS

Weather effects need to be distinguished with respect to their effects on the lidar sensor. Precipitation is one class of weather effects covering rain, snow, and hail. The size of the drops, snowflakes, or hailstones can vary over a wide range and thus influence the sensor detection principle significantly but also unsystematically [18]. Fog is another phenomenon with more homogeneous properties or droplets. Background infrared (IR) light caused by the sunset, sunrise, or other ambient IR light sources has a very different impact on the sensor, compared to precipitation. The background IR light increases the sensor noise, and the sensor becomes less sensitive to sources of low reflectivity. Apart from isotropic or homogeneous weather conditions, there are dynamic or turbulent weather phenomena, like splash turbulence or effects due to the speed of the vehicle driving in the weather condition. Each phenomenon needs to be studied in detail for a physical model.

A good overview of the physical description of adverse weather conditions for automotive perception sensors can be found in [19]. Lidar sensors allow to differentiate between weather conditions, as demonstrated in [20] for snow, fog, and rain. Different weather conditions and their influence on lidar detection have been studied in recent publications [21], [22], [23], [24]. A study on lidar detection in fog based on measurements can be found in [25]. Environment influences due to weather conditions, such as rain, fog, precipitation, or sunlight on perception sensors, have been studied and analyzed recently in a thesis by Linnhoff [26]. In particular, he studies lidar sensors with extensive data collection and presents stochastic simulation models thereof.

A full-scale analysis of adverse weather conditions for a particular lidar sensor (Cube 1 by Blickfeld) is demonstrated in [27] by applying the Mie scattering theory. Haider et al.

investigated the signal-to-noise ratio under attenuation conditions, as well as detection rates, false positives, and distance errors caused by rain and fog. By modeling a particular lidar sensor of interest, they achieve a sensor model of high accuracy covering various effects along the detection process, including signal processing specifics and the impact of scattering or signal absorption caused by droplets. Some of the results of this system-related model can be used for general lidar sensor modeling. However, not all details are known or can be applied to modeling general lidar sensors.

Data sets for different weather conditions created from a combination of perception sensors, including radar, lidar, and camera sensors with weather classification are publicly available [28], [29], [30], [31], [32], [33]. Unfortunately, there are no systematic data sets for lidar sensors with respect to material reflectivity measurements. Additionally, the classification of weather conditions is often not detailed enough for lidar sensors, as discussed in [18].

Moreover, weather can cause additional surface layers on top of different materials, e.g., ice, snow, frost, moisture, or wetness, which can change the surface reflectivity and need to be modeled through the material parameters, systematically or directly from measurements.

A simplified approach to account for adverse conditions in a sensor model is to consider signal attenuation. The limited range caused by a certain weather condition can be included in the range limit curves directly. For this, a curve fit is generated based on potential weather measurements under different conditions. A generic attenuation model for lidar is presented in the following. It is a phenomenological model based on heuristic parameters, as opposed to rigorous physical derivations [27], such as Mie scattering theory. An advantage of the model is that it is applicable without detailed knowledge of the physical parameters of the weather condition, e.g., droplet size distributions and intensities, or the sensor specifics.

A. ATTENUATION MODEL
In the next step, we want to model attenuation effects caused by adverse environmental conditions. To achieve this, we assume the attenuation coefficient $\sigma \neq 0$, and therefore, the exponential function in (3) is not equal to one anymore. Solving (3) for the range r results in a transcendental function of the form

$$c_n = \frac{\rho}{r^n} \, e^{-2\sigma r} \tag{6}$$

$$r = \left(\frac{\rho}{c_n} \, e^{-2\sigma r} \right)^{\frac{1}{n}}. \tag{7}$$

This transcendental equation is of the form

$$z = a \, e^{bz} + c \tag{8}$$

where a, b, and c are coefficients. A general solution of this equation is provided by the principal branch W_0 of Lambert's W-function

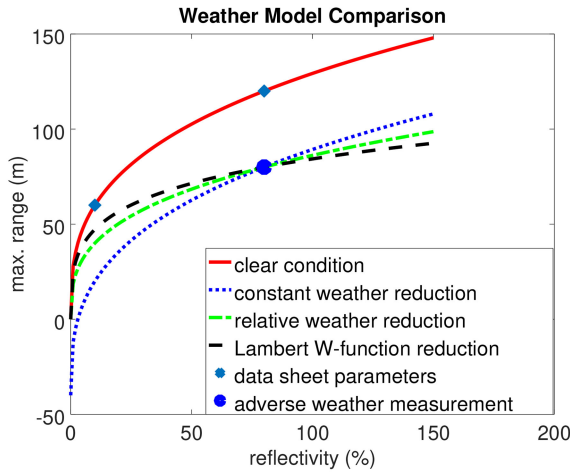$$z = c - \frac{1}{b} W_0 \left[-a \, b \, e^{cb} \right]. \tag{9}$$

FIGURE 4. Comparison of range reduction models using a single measurement. An arbitrary measurement point under a certain adverse weather condition is assumed at [80%, 80 m]. The attenuation model (dashed black), based on the Lambert-W function, is compared to the clear condition curve fit (red). The green and blue curves are relative and constant reduction models, respectively, which are used purely as comparative approaches in Section V-A.

Applied to (7), the range is given by

$$r = \frac{n}{2\sigma} \; W_0 \left[\frac{2\sigma}{n} \left(\frac{\rho}{c_n} \right)^{\frac{1}{n}} \right] \qquad (10)$$

where $c = 0$. Again, r_{max} is determined by the minimal signal strength $P_{r,min}$ detectable by the lidar system, included in $c_n(P_{r,min})$.

A comparable approach to model atmospheric attenuation can be found in the white paper [34] by Allegro Microsystems, where a detailed radiometric description is derived. In their modeling approach, detailed knowledge is assumed about the sensor systems, including the emitted and return signal energies. In automotive lidar sensors, detailed sensor specifications are either unknown to the user, or change during operation in different weather conditions. As discussed in [25], some automotive lidar sensors counteract adverse weather by adapting emitted intensity dynamically.

Currently, there are no parameters provided by the sensor manufacturer on maximum ranges under different weather conditions for lidar sensors on the market. However, missing knowledge can be extracted from measurement data, when surface reflectivities are known. The difference $w(\rho_{ref})$ can be defined as the reduction of the maximum range at the adverse condition compared to best conditions for a known reflectivity ρ_{ref}. Even for only a single data point taken under a defined, adverse weather condition, a curve fit can be generated (see Fig. 4).

Since there is no measurement database available, the model is limited to a subset of data, which can potentially be generated through a simple measurement procedure. This can be achieved by measuring the maximum range of a target of known reflectivity ρ_{ref} at the adverse weather condition. The maximum range of the target at clear weather conditions can either be measured additionally, taken from a fitting curve, or a known datasheet parameter.

IV. METRICS

For the validation of the phenomenological models, usually real-world point clouds are used. Point clouds detected by a particular sensor, however, typically contain sensor-specific attributes, which differ from ideal point clouds. The goal of virtual feature validation is, to validate a feature independently of building a full sensor system simulation of a specific sensor. Ideally, the sensor manufacturer could provide a synthetic point cloud built upon a validated feature generated from an identical scene or scenario, which can then be used for a direct comparison of lidar point clouds.

To assess the feature that the range–reflectivity models introduce to the point cloud generated, a metric to compare point clouds is needed. From the literature, the following metrics have been applied to point cloud comparison problems: Pearson correlation [6], [7], Hausdorff metric [11], or Wasserstein metric [12]. However, direct comparison can lead to high computational demands, depending on the size and density of the point clouds, which need to be compared.

In feature validation, two cases are of interest:
1) arbitrary point cloud comparison;
2) point cloud comparison for identical ray patterns.

Our application is the comparison of point clouds generated by our lidar point cloud sensor model and thus belongs to the second type.

Synthetic point clouds can be generated, using well-defined ray directions, in identical 3-D environment scenes based on identical material surface properties. Apart from raytracing models, there are lidar models based on rasterization, which operate in the graphics pipeline of an environment simulator. These models often have access to additional information which allows for the sorting of rays and hits. An example of a depth-buffer-based model is presented in [35], where a pixel association is available. This allows for the comparison of different point clouds within the same simulation environment, according to their internal sorting and alignment information.

It simplifies the point cloud comparison extremely, if the point association of the point clouds is available, e.g., via ray directions or indices. If this is not the case, a nearest neighbor search is needed to find associated points. This can be done using hierarchical space decomposition, such as k-d trees or octrees, but is still very costly.

V. APPLICATION OF RANGE–REFLECTIVITY FEATURE

The presented phenomenological models for clear and adverse weather conditions represent a lidar sensor feature, which can be used for different types of lidar sensor models, physical models, such as raytracing or reflection-based sensor models, but also object-based models. In the following, we apply the feature to LiMOX, a high-fidelity lidar model based on raytracing [15].

LiMOX [14], [15] is a point cloud lidar sensor model, using the Nvidia OptiX engine for parallel raytracing. A point cloud is generated from the interaction of a ray with a reflective surface. The ray–surface interaction is performed
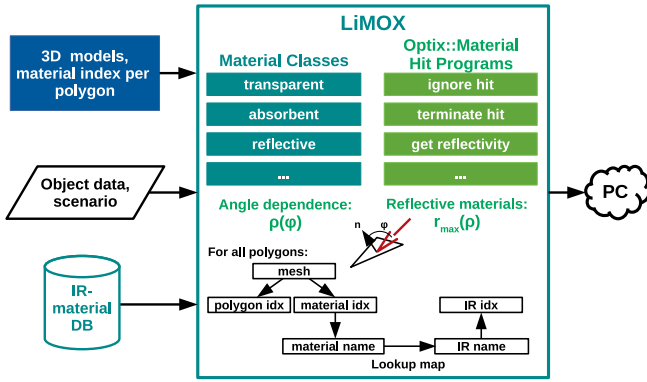
FIGURE 5. Overview of lidar point cloud (PC) simulation with LiMOX. 3-D models, object data, and IR material databases are loaded. For the ray–surface interaction, material classes correspond to different hit programs. For each polygon, an infrared material from the IR database is assigned using a lookup mapping table.

FIGURE 6. Program flow of the stand-alone raytracing implementation in OptiX. The abbreviation rt stands for raytracing and refers to OptiX-API functions.

on the polygon level of the 3-D geometry surface meshes in the environment. Each polygon holds a visual material as an index. A lookup table can be used to assign corresponding infrared (IR) materials to the visual materials of the mesh.

LiMOX uses material classes to distinguish between transparent, absorbent, retroreflective, and general reflective materials. Depending on the material class a material belongs to, a different hit program is called. In OptiX, the hit program is used to define the ray–surface interaction.

An IR material database is used to assign each material a reflectivity in the infrared spectrum. Depending on the material database used, the bidirectional reflectance distribution function model is assumed. For database entries without angular dependence (e.g., NASA ECOSTRESS), the surface is modeled by an ideal diffuse material via the Lambertian law. This is not necessary when a material of the angle-dependent spectral database [36] is used. This database offers angular reflectance measurements with a 10° angular resolution for [0°, 90°) incidence angle. Depending on the incidence angle of the ray, and the material of the surface polygon, a reflectivity can be retrieved for each lidar point. An overview of the basic components and processes in LiMOX is shown in Fig. 5.

Fig. 6 shows the general flow of the raytracing program. First, a hierarchical space decomposition is performed to limit the search space of possible intersections in the ray direction. Next, possible intersections of polygons with the ray are checked. If an intersection is reported, the material assigned to the polygon is retrieved and the corresponding hit program of the material class it belongs to is called. For reflective materials, a reflectivity is retrieved from the material database and the maximum range limit is compared to the detected range. If the range exceeds the maximum range for the reflectivity, the intersection is not added to the point cloud. Special treatment can be added for retroreflective materials depending on the application and considered sensor, e.g., range or reflectivity adaptations.

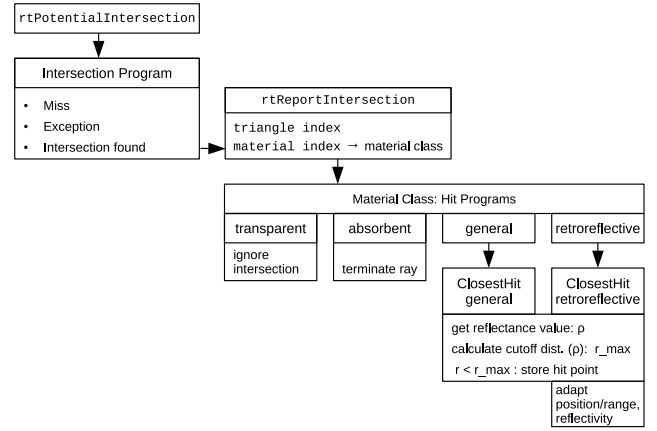In this application, all surfaces are modeled with the same Lambertian target of a chosen reflectivity. A Lambertian target corresponds to an ideal diffuse surface. For this type of material, the Lambert cosine law can be applied to determine the reflectivity under different incidence angles

$$\rho(\phi) = \rho(0) \cdot \cos(\phi). \tag{11}$$

The reason for using the same material reflectivity for all surfaces is that the effects of angular and material dependence can be separated, which supports the validation and helps to study the effects separately. For realistic 3-D environments, different surface materials are assigned and an analysis of surface types and appropriate materials together with a mapping strategy is necessary.

In the following, we use two simple range-reduction models, which are not physically motivated, but purely for demonstration and comparison purposes. They can be interpreted as a first and second approximation of the measured range reduction $w(\rho_{\mathrm{ref}})$ for a known material reflectivity, an absolute and a relative reduction. Alternative range–reflectivity limit curves can be generated and resulting point clouds can be compared to the previously derived models (see Fig. 4).

For the absolute reduction model, the range reduction is constant over all reflectivities

$$r_{\max}(\rho, w) = r_{\mathrm{clear}}(\rho) - w \tag{12}$$

where $r_{\mathrm{clear}}(\rho)$ is the range limit for best conditions, and $r_{\max}(\rho, w)$ is the maximum range for the considered weather condition. Negative $r_{\max}$ values are set to zero.

The other option is to calculate the relative reduction of the maximum range at the reference reflectivity, at which the measurement was taken, and apply it across the entire reflectivity range

$$r_{\max}(\rho, w) = r_{\mathrm{clear}}(\rho) \cdot \left(1 - \frac{w}{r_{\mathrm{clear}}(\rho_{\mathrm{ref}})}\right). \tag{13}$$

With the different range–reflectivity models, it is possible to generate point clouds. As a typical Lidar sensor, the Ouster-OS1 sensor with 128 layers over 45° vertical field of view is chosen, with a ray pattern of 512 points per
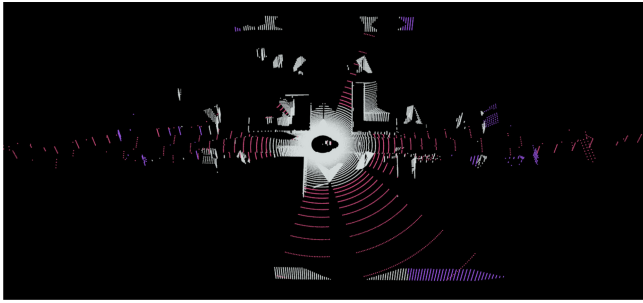
FIGURE 7. Point clouds with no reflectivity–range limit (pink, violet, and white), compared to the range limit fit at clear conditions (violet and white) and attenuation model, based on the Lambert W function (white only), for all surfaces modeled by 50% Lambertian targets.
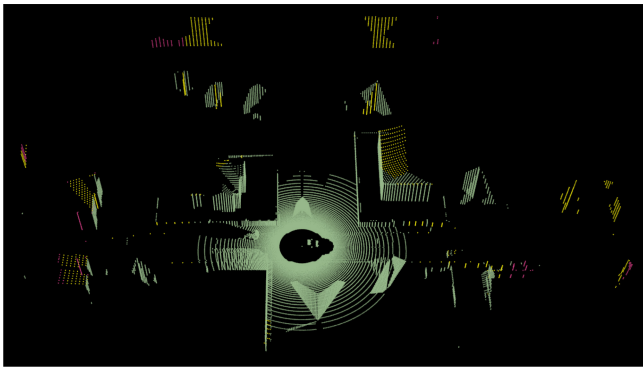

FIGURE 8. Comparison of attenuation modeled by the Lambert W function (red, yellow, and green), relative reduction (yellow and green), and constant reduction (green only), for all surfaces modeled by a Lambertian target of 50% reflectivity.

line in a 360° field of view, corresponding to a vertical resolution of 0.35° and a horizontal resolution of 0.7°. The static scene used is from an urban scenario consisting of a coarse static environment meshes with houses, trees, fences, and two vehicles on opposite sides of the road. All presented range–reflectivity limits are used for point cloud generation. In Fig. 7, the clear weather fit is compared to the point cloud with no range limit and the attenuation model. Adverse weather conditions are shown in Fig. 8, for the attenuation model, and the relative and constant range-reduction approaches. The next section discusses possible and useful measures to compare the resulting point clouds quantitatively.

A. ASSESSMENT

The phenomenological features are applied in LiMOX. When generating the point cloud, the distance of each point from the sensor origin is checked. When the distance exceeds the limit given by the maximum range of the feature model, the point is not added to the point cloud. In OptiX, it is straightforward to implement different attenuation functions, determine the distances of each hit point, and apply this comparison to each intersection point. In this manner, different point clouds are generated for each attenuation model. Noise or false positive detections are not considered in this example.

TABLE 2. Parameters for point cloud comparison.

Parameter	Definition
N_c	number of corresponding/associated points
N_{nc}	number of points with no corresponding point, or beyond a distance limit L
$D = \sum_i^{N_c} d_i$	sum of Euclidean distances d_i of two corresponding points above a noise threshold h and below an association limit L

The respective point cloud sets $\mathbf{P}$ are subsets of one another, for all surface materials below 80% reflectivity

$$\mathbf{P}_{\text{const}} \subseteq \mathbf{P}_{\text{rel}} \subseteq \mathbf{P}_{\text{att}} \subseteq \mathbf{P}_{\text{clear}} \subseteq \mathbf{P}_{\text{no\_limit}}. \qquad (14)$$

For general point cloud comparison, a certain noise tolerance needs to be considered and the resulting deviation of corresponding points needs to be neglected accordingly within the tolerance limits.

B. PROPOSED METRICS

For the comparison of point clouds with similar relations as in (14), metrics are needed. Many widely used metrics only apply to point cloud sets of equal size, such as Pearson correlation, minimal Euclidean distance, or root mean-square error [8]. A comparison of equal-sized sets often applies to the comparison of simulated with measured data, if only those points are considered for the comparison that can be associated with one another.

When comparing sets of different sizes, the Hausdorff distance δ can be used, which is the largest distance of all distances d of one point in each set to the closes point in the other set, hence

$$\delta(A, B) := \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(A, b)\}. \qquad (15)$$

However, the calculation of δ demands a high computational effort in particular for large sets.

Another interesting metric from transport theory is the Wasserstein metric, which is typically used for comparing two distributions. It reflects the cost of transferring one distribution into the other. To calculate this measure, the distributions need to be normalized, however, and the transfer plan is not unique. For discrete distributions, the Euclidean norm ($p = 2$) or the general p-norm can be used to calculate the Wasserstein distance, which results again in the restriction to sets of equal size.

Due to these restrictions, a metric is needed, which can be easily applied and does provide a comparison measure for sets of different sizes, and for the particular case of comparing a set with its subset.

As a general approach, we define parameters for a point cloud comparison in Table 2. A corresponding point is a point within a chosen distance limit to a point of the compared point cloud. For the ideal, synthetic point clouds generated in LiMOX, this limit is zero, since there is no noise model included and the detection points result from

ray–surface intersections. In general, this limit is typically sensor-specific, depending on the accuracy or resolution.

Metrics applicable to our point cloud examples include the number of corresponding points N_c of two point clouds and the number of points with no correspondence N_{nc} in the other point cloud. A distance of all corresponding points is not applicable in this case, since ideal point clouds are generated by raytracing where intersection points are stored and stochastic effects, such as false positives or noise, are not included here. The reason for this is that these effects are highly data-driven and therefore product-specific.

An interesting parameter is the ratio f_c of the total number of points with a corresponding point N_c in the other cloud to the number of points with no corresponding or equivalent point N_{nc}

$$f_c = \frac{N_{nc}}{N_c}. \qquad (16)$$

This measure is not normalized; however, it provides a quantitative assessment of the clouds' similarities and differences. Equation (17) lists the different value ranges with the respective relation of the compared point counts. The higher the similarity of the compared clouds, the lower the value. For values above one, there are more noncorresponding points, while below one, there are more corresponding points. A particular advantage is, that there is no reference cloud, thus the comparison is symmetrical or commutative

$$f_c \begin{cases} = \infty, & \text{for } N_c = 0 \\ > 1, & \text{for } N_{nc} > N_c \\ = 1, & \text{for } N_c = N_{nc} \\ < 1, & \text{for } N_{nc} < N_c \\ = 0, & \text{for } N_{nc} = 0. \end{cases} \qquad (17)$$

The absolute numbers depend highly on the chosen scene and the alignment of the lidar sensor and its ray pattern, since not all rays hit a surface or return a reflection.

C. APPLICABILITY OF THE METRIC

The proposed metric f_c is applicable to sets with known point association, e.g., an ordered set of rays and hit points in our case. Within the OptiX API, a set of rays is generated according to the ray pattern of the sensor, where each ray corresponds to a specific direction in space. For each ray, a return signal is recorded, either a hit point or a miss. This structure allows to associate each intersection point stored to the ray direction. Therefore, the association of points from different point clouds with identical ray pattern and ray origin is implicit and can be performed according to their indices.

Moreover, the metric could also be applied to arbitrary sets of point clouds, by defining regions of overlap, where both point clouds yield similar results, and regions with no correspondences in the compared point set. In this case, a nearest-neighbor search can be skipped and only points need to be counted in each region and summed up to N_c and N_{nc}.

TABLE 3. Ratio f_c of the pair-wise comparison of point clouds for clear condition, attenuation model, relative and constant reduction, and no range limit, evaluated for a single scene or time step.

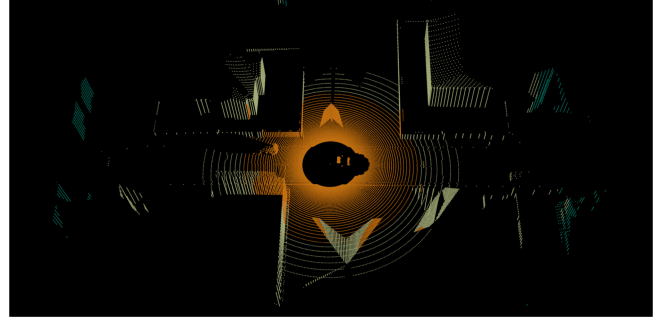| 50% targets	clear	attenuation	relative	constant
no limit	$\frac{2\,039}{51\,555}$	$\frac{3\,031}{50\,563}$	$\frac{3\,359}{50\,235}$	$\frac{4\,911}{48\,683}$
clear	-	$\frac{992}{50\,563}$	$\frac{1\,320}{50\,235}$	$\frac{2\,872}{48\,683}$
attenuation	see above	-	$\frac{328}{50\,253}$	$\frac{1\,880}{48\,683}$

FIGURE 9. Comparison of attenuation modeled by the Lambert W function (green, yellow, and orange), percentage function (yellow and orange), and constant reduction (orange only), for all targets of 10% reflectivity.

VI. RESULTS

A. EVALUATION OF SCENE

In Fig. 7, the point cloud associated to the comparison of the clear model (violet and white) and the unrestricted point cloud (all colors), with no range limit, are shown. Additionally, the comparison of points for clear and attenuation (white) models can be seen.

The related point clouds of the comparison in Table 3 for the different models for attenuation, relative reduction, and constant reduction can be seen in Fig. 8, for which all surfaces are modeled by a 50% Lambertian target.

An evaluation of the ratio f_c for the chosen scene is shown in Table 3, comparing the different point clouds for all surfaces modeled by a 50% Lambertian target.

Only at larger distances, the point clouds differ from one another. This is where the spatial resolution becomes much lower and the point cloud less dense. Therefore, even if the maximum range differs significantly, the number of points at high distances from the sensor is relatively low. This fact can be seen in Table 3, for example, when comparing clear conditions to the constant reduction, for which the difference in maximum range is the largest.

To show the impact of different target materials on the attenuation models and their dependence on the reflectivity, another visual comparison of the models is shown in Fig. 9, where all surfaces are modeled by 10% Lambertian targets.

B. EVALUATION OF ENTIRE SCENARIO

To compare two point clouds for a single static scene helps to get an idea about the different underlying range–reflectivity models. The comparison ratio f_c provides the user with a

TABLE 4. Mean values (± standard deviation) over a 40-time-step scenario for the ratio f_c of the pair-wise comparison of point clouds for clear condition, attenuation model, relative and constant reduction, and no range limit.

50% targets	clear	attenuation	relative	constant
no lim.	$\frac{2\,176\ (\pm 256)}{51\,184\,(\pm 443)}$	$\frac{2\,877\ (\pm 274)}{50\,482\,(\pm 408)}$	$\frac{3\,144\ (\pm 251)}{50\,215\,(\pm 450)}$	$\frac{4\,399\ (\pm 347)}{48\,961\,(\pm 532)}$
clear	-	$\frac{702\ (\pm 186)}{50\,482\,(\pm 408)}$	$\frac{969\ (\pm 221)}{50\,215\,(\pm 450)}$	$\frac{2\,223\ (\pm 338)}{48\,961\,(\pm 532)}$
atten.	see above	-	$\frac{267\ (\pm 104)}{50\,215\,(\pm 450)}$	$\frac{1\,521\ (\pm 259)}{48\,961\,(\pm 532)}$

TABLE 5. Animations generated, showing the scenario point clouds, based on pairwise comparison of selected range–reflectivity models.

50% targets	clear (red)	attenuation (yellow)	constant (green)
no_limit (white)	nolimit_clear.gif	-	-
clear (red)	-	clear_att.gif	-
attenuation (yellow)	see above	-	att_const.gif

quantitative measure for the point difference. However, since a generated point cloud is highly dependent on the 3-D environment it is based upon, a more general approach is favorable.

For a test engineer, who uses different kinds of input point clouds, it is helpful to estimate the differences of two point clouds not only for a single scene but for the entire scenario that it is applied to. Therefore, we present a simple scenario of 40 time steps (or 40 scenes) and evaluate the number of corresponding and noncorresponding points for all time steps separately. Next, the mean values are taken over all time steps. As a result, we can show the nominator and denominator of the ratio f_c (see Table 4).

The exemplary scenario consists of an urban street surrounded by houses, fences, and simple trees, together with two vehicles driving along the street in opposite directions, passing by each other. The geometry meshes for the urban street environment are provided by courtesy of the environment simulator VTD [37], by Hexagon.

Since the scenery and the 3-D environment do not change abruptly during the scenario, the number of corresponding and the number of noncorresponding points stay roughly the same in each scene and the mean values over all time steps (Table 4) are close to the values for a single time step (Table 3). Moreover, the total number of points in each time step can vary, since over different time steps a different number of miss-hits can occur. These are rays of the ray pattern that do not hit a target or produce a reflection.

In conclusion, our measure f_c is not only reflecting the different point clouds generated from the phenomenological models but also the 3-D environment itself. If the 3-D environment offers only a limited extension and there are no objects at larger distances, the point clouds generated from different reflectivity range models might not differ at all. In this case, the ratio will become zero, even though the underlying phenomenological models differ significantly. This fact can be useful for the application engineer, to estimate the impact of different models with respect to a scenario of interest.

To show the entire scenario used in the evaluation, we generated three animations to compare the range–reflectivity models pair-wise (see Table 5). Similarly as before, the larger point cloud includes the smaller one.

VII. CONCLUSION AND DISCUSSION

In sensor model development, validation strategies are still mostly missing. The motivation behind this work is to support the virtual validation of model features individually.

Range–reflectivity limits for lidar sensor models are derived from the lidar range equation, for best or clear conditions as well as for adverse weather conditions using an attenuation model based on a single measurement and datasheet parameters. These are applied as a feature in a physical sensor model, LiMOX [15], based on raytracing. To reduce the impact of different surface materials, all surfaces are modeled by a single Lambertian target, where the reflectivity depends only on the incidence angle. To compare different attenuation models, two cases were presented, one for which all surfaces are modeled by 50% Lambertian targets and the other using 10% Lambertian targets for all surfaces of the 3-D meshes.

For the presented special case of comparing synthetic point clouds in identical 3-D environments, which are fully controllable, a nearest neighbor search is not needed, since the shooting pattern, the detection angles, as well as the sensor alignment are well defined and known. Since the ideal point clouds generated by the range–reflectivity models have a subset relation to one another, we compare the total number of points.

For arbitrary point cloud comparison, the sum of cumulative distances of associated points can be a further parameter of interest. Given that the sensor manufacturer provides a point cloud generated in the same virtual environment, feature validation becomes feasible.

As a comparison measure, the ratio f_c is introduced based on the total number of noncorresponding over corresponding points of two point clouds. The value depends on the point clouds and simultaneously helps to estimate the impact of different scenes. Thus, we provide results for a single scene, which corresponds to a single time step. Additionally, an entire scenario is evaluated by building a mean value over all time steps of the scenario. The evaluation of the entire scenario yields a more comprehensive impression of the feature comparison.

While the comparison metric f_c only reflects the differences of point clouds in a 3-D environment, key performance indicators need to be chosen separately. The appropriate choice highly depends on the concrete application of the point cloud, e.g., testing an object perception algorithm. This choice lies within the responsibility of the application engineer.

VIII. OUTLOOK

By using a single Lambertian surface material for all surfaces in this application, the surface material properties were decoupled from the angular reflectance dependence. For realistic 3-D models and resulting point clouds, an assignment of appropriate surface materials is planned. Prerequisites for this are the analysis of a mapping strategy of material properties to the 3-D meshes and a material database from which infrared reflectance can be derived, e.g., [38] or [39].

As a next step after this virtual validation, measurement data is needed. For this, data sets including material surface properties, in particular, the reflectance is needed, under different weather conditions. Mostly, there are no data sets including reflectance measurements for this purpose. An interesting data set by Linnhoff et al. from TU Darmstadt is publicly available [33]. It targets different automotive perception sensors under various weather conditions.

To estimate the infrared reflectance of different types of materials predominant in automotive applications, a database has been generated by Ritter et al. [39], holding surface reflectance measurements under different incidence angles. This supports the classification of different materials detected in typical data sets and the estimation of the reflectivity of surfaces where the measurement parameter is missing. The data set has been published and can be downloaded open access via Zenodo [40]. It is being continuously extended by new material measurements.

Another direction of future research is to apply the proposed feature validation method and the metric f_c to the comparison of arbitrary point clouds as mentioned in Section V-C. For this, a manual or automatic selection of regions where the point clouds overlap, and regions of no overlap, needs to be defined. Moreover, the sum of all points within the defined regions has to be built. This is of particular interest for comparing large point clouds, where an automatic procedure is necessary.

When comparing simulation and measurement data, another aspect to consider is limiting the comparison to regions that are represented in simulation ground truth accordingly. Any regions of 3-D geometries that are only present in one of the point clouds, either real-world or simulation, need to be excluded from the comparison, as the model can only represent the fidelity of the underlying geometry.

ACKNOWLEDGMENT

REFERENCES

[1] R. Rott, "Phenomenological range-reflectivity limits for Lidar models and feature validation approach," in *Proc. IEEE Int. Autom. Veh. Validation Conf.*, 2023, p. 4011.

[2] *UN Regulation No. 157—Automated Lane Keeping Systems (ALKS)*, Nations Econ. Commission Europe, Geneva, Switzerland, 2021, pp. 75–137.

[3] "VV Methoden." 2023. [Online]. Available: https://www.vvm-projekt.de/

[4] "SET level—Simulation-based development and testing of automated vehicles." 2023. [Online]. Available: https://setlevel.de/

[5] J. Reckenzaun et al., "Transnational testing, operation and certification of automated driving systems: Perspective from testEPS and central system EUREKA projects—Mid-term results," in *Proc. IEEE Int. Autom. Veh. Validation Conf. (IAVVC)*, 2023, pp. 1–8.

[6] T. Hanke et al., "Generation and validation of virtual point cloud data for automated driving systems," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2017, pp. 1–6.

[7] A. Schaermann, A. Rauch, N. Hirsenkorn, T. Hanke, R. Rasshofer, and E. Biebl, "Validation of vehicle environment sensor models," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2017, pp. 405–411.

[8] R. Donà and B. Ciuffo, "Virtual testing of automated driving systems: A survey on validation methods," *IEEE Access*, vol. 10, pp. 24349–24367, 2022.

[9] "(GRVA) new assessment/test method for automated driving (NATM)—Master document." 2021. [Online]. Available: https://unece.org/transport/documents/2021/04/working-documents/grva-new-assessmenttest-method-automated-driving-natm

[10] "Proposed certification memorandum on modelling & simulation—CS-25 structural certification specifications." 2024. [Online]. Available: https://www.easa.europa.eu/en/document-library/product-certification-consultations/proposed-certification-memorandum-modelling

[11] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A generalized Hausdorff distance based quality metric for point cloud geometry," in *Proc. 12th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2020, pp. 1–6.

[12] K. Kawano, S. Koide, and T. Kutsuna, "Learning Wasserstein isometric embedding for point clouds," in *Proc. Int. Conf. 3D Vis. (3DV)*, 2020, pp. 473–482.

[13] A. G. Kashani, M. J. Olsen, C. E. Parrish, and N. Wilson, "A review of LiDAR radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration," *Sensors*, vol. 15, no. 11, pp. 28099–28128, 2015.

[14] R. Rott, D. J. Ritter, S. Ladstätter, O. Nikolić, and M. E. Hennecke, "LiMOX—A point cloud Lidar model toolbox based on NVIDIA OptiX ray tracing engine," *Sensors*, vol. 24, no. 6, p. 1846, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/6/1846

[15] R. Rott, "Dynamic update of stand-alone Lidar model based on ray tracing using the Nvidia Optix engine," in *Proc. IEEE Int. Conf. Connected Veh. Expo.*, 2022, p. 7006.

[16] A. V. Jelalian, "Laser radar systems," in *Proc. Electron. Aerosp. Syst. Conf.*, 1980, pp. 546–554.

[17] W. Wagner, A. Ullrich, and C. Briese, "Der Laserstrahl und seine Interaktion mit der Erdoberfläche," *Österreichische Zeitschrift für Vermessung Geoinformation*, vol. 91, no. 4, pp. 223–235, 2003.

[18] "DENSE D2.1—Characteristics of adverse weather conditions." Accessed: Feb. 2022. [Online]. Available: https://www.dense247.eu/fileadmin/user_upload/PDF/DENSE_D2.1_Characteristics_of_Adverse_Weather_Conditions.pdf

[19] S. Hasirlioglu, "A novel method for simulation-based testing and validation of automotive surround sensors under adverse weather conditions/submitted by Sinan Hasirlioglu," Ph.D. dissertation, Doktor der technischen Wissenschaften, Universität Linz, Linz, Austria, 2020.

[20] J. R. V. Rivero, T. Gerbich, V. Teiluf, B. Buschardt, and J. Chen, "Weather classification using an automotive LIDAR sensor based on detections on asphalt and atmosphere," *Sensors*, vol. 20, no. 15, p. 4306, 2020.

[21] R. Heinzler, P. Schindler, J. Seekircher, W. Ritter, and W. Stork, "Weather influence and classification with automotive lidar sensors," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2019, pp. 1527–1534.

[22] A. Filgueira, H. González-Jorge, S. Lagüela, L. Díaz-Vilariño, and P. Arias, "Quantifying the influence of rain in LiDAR performance," *Measurement*, vol. 95, pp. 143–148, Jan. 2017.

[23] M. Jokela, M. Kutila, and P. Pyykönen, "Testing and validation of automotive point-cloud sensors in adverse weather conditions," *Appl. Sci.*, vol. 9, no. 11, p. 2341, 2019.

[24] J. Zhao, Y. Li, B. Zhu, W. Deng, and B. Sun, "Method and applications of LiDAR modeling for virtual testing of intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2990–3000, May 2021.

[25] M. Bijelic, T. Gruber, and W. Ritter, "A benchmark for Lidar sensors in fog: Is detection breaking down?" in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2018, pp. 760–767.

[26] C. Linnhoff, "Analysis of environmental influences for simulation of active perception sensors," Ph.D. dissertation, Technische Universität Darmstadt, Darmstadt, Germany, 2023. [Online]. Available: http://tuprints.ulb.tu-darmstadt.de/23116/

[27] A. Haider et al., "A methodology to model the rain and fog effect on the performance of automotive LiDAR sensors," *Sensors*, vol. 23, no. 15, p. 6791, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/15/6891

[28] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," in *Proc. CVPR*, 2020, pp. 11621–11631.

[29] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000km: The Oxford RobotCar dataset," *Int. J. Robot. Res.*, vol. 36, no. 1, pp. 3–15, 2017. [Online]. Available: http://dx.doi.org/10.1177/0278364916679498

[30] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception in bad weather," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 1–7.

[31] "DENSE dataset." Accessed: Mar. 2022. [Online]. Available: https://www.uni-ulm.de/in/iui-drive-u/projekte/dense-datasets/

[32] J.-L. Déziel et al., "PixSet: An opportunity for 3D computer vision to go beyond point clouds with a full-waveform LiDAR dataset," 2021, *arXiv:2102.12010*.

[33] C. Linnhoff, K. Hofrichter, L. Elster, P. Rosenberger, and H. Winner, "Perception sensors in adverse weather—Dataset from TUdatalib." Accessed: Mar. 2023. [Online]. Available: https://www.fzd-datasets.de/weather/

[34] G. M. Williams and A. Huntington, "LIDAR effective range," Allegro Microsyst., Manchester, NH, USA, Rep. MCO-0001135, P0177, 2021. Accessed: Mar. 2023. [Online]. Available: https://www.allegromicro.com/-/media/files/articles/p0177/p0177-lidar-effective-range.pdf?sc_lang=en&hash=8DC4B3DD8737446EFE138208EDF97EC7

[35] M. Kirchengast and D. Watzenig, "A depth-buffer-based Lidar model with surface normal estimation," *IEEE Trans. Intell. Transp. Syst.*, early access, Mar. 12, 2024, doi: 10.1109/TITS.2024.3371531.

[36] D. J. Ritter, "Updating 'angle-dependent spectral reflectance material dataset based on 945 nm time-of-flight camera measurements' with extended data to cover reflectance measurements mainly for vehicle varnish and moss rubber," *Data Brief*, vol. 51, Dec. 2023, Art. no. 109747.

[37] "Virtual test drive." MSC Software. Accessed: Feb. 2022. [Online]. Available: https://www.mscsoftware.com/de/virtual-test-drive

[38] S. K. Meerdink, S. J. Hook, D. A. Roberts, and E. A. Abbott, "The ecostress spectral library version 1.0," *Remote Sens. Environ.*, vol. 230, Sep. 2019, Art. no. 111196.

[39] D. J. Ritter et al., "Angle-dependent spectral reflectance material dataset based on 945 nm time-of-flight camera measurements," *Data Brief*, vol. 48, Jun. 2023, Art. no. 109031. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235234092300149X

[40] D. J. Ritter et al. "Angle-dependent spectral reflectance material dataset based on 945 nm time-of-flight camera measurements." Accessed: Mar. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7467552