

LaNoising: A Data-driven Approach for 903nm ToF LiDAR Performance Modeling under Fog

Tao Yang^{1,2}, You Li³, Yassine Ruichek², and Zhi Yan²

Abstract—As a critical sensor for high-level autonomous vehicles, LiDAR’s limitations in adverse weather (e.g. rain, fog, snow, etc.) impede the deployment of self-driving cars in all weather conditions. In this paper, we model the performance of a popular 903nm ToF LiDAR under various fog conditions based on a LiDAR dataset collected in a well-controlled artificial fog chamber. Specifically, a two-stage data-driven method, called LaNoising (la for laser), is proposed for generating LiDAR measurements under fog conditions. In the first stage, the Gaussian Process Regression (GPR) model is established to predict whether a laser can successfully output a true detection range or not, given certain fog visibility values. If not, then in the second stage, the Mixture Density Network (MDN) is used to provide a probability prediction of the noisy measurement range. The performance of the proposed method has been quantitatively and qualitatively evaluated. Experimental results show that our approach can provide a promising description of 903nm ToF LiDAR performance under fog.

I. INTRODUCTION

Having good performance in accurate distance measuring, LiDAR (Light Detection and Ranging) systems become one of the most important sensors used for autonomous vehicles. A typical LiDAR system consists of a laser transmitter, a receiver, and a signal processing unit. The laser transmitter first emits a pulsed laser at a given direction. When encountering an obstacle’s surface, the transmitted laser will be diffused or reflected, depending on specific surface material. Ranges can be measured by processing the echoed signals captured by laser detectors. Currently, the most popular ranging principle is Time-of-Flight (ToF), which uses the time difference between the transmitted and received lasers to measure distance, since the light speed can be regarded as an universal physical constant.

However, the degraded performance of a LiDAR under adverse weather, such as fog, rain, and snow, is a major drawback that limits high-level autonomous vehicles in all weather conditions. Therefore, it has received increasing attention from both academia and industry. The fog has a twofold impacts on LiDAR performance. On the one hand, small droplets in the atmosphere absorb or scatter near-infrared laser light, which increases the attenuation of laser transmission [1]. On the other hand, wet surfaces of objects leads to weaker reflectivity [2]. These two factors result in

The work is supported by the UTBM-Renault joint research project (DENOSAU).

¹Unmanned System Research Institute, Northwestern Polytechnical University, 710072 Xi’an, China yangtaonpu@hotmail.com

²CIAD UMR7533, Univ. Bourgogne Franche-Comté, UTBM, F-90010 Belfort, France firstname.lastname@utbm.fr

³Groupe Renault, France you.li@renault.com

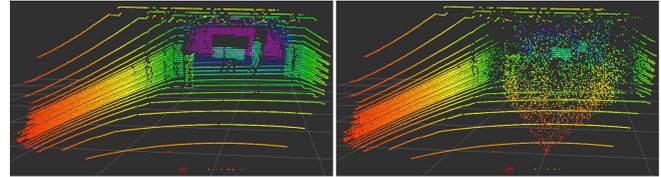


Fig. 1. Point cloud under clear weather conditions (left) and noising simulation for fog environment (right).

lower received laser power, which make subsequent signal processing fail. Nevertheless, due to the previous lack of quantitative meteorological data, it was difficult to effectively analyze LiDAR performance under such conditions.

In this paper, we focus on the performance modeling of a 903nm ToF LiDAR in fog environment. Specifically, a data collection process with a Velodyne VLP-32C LiDAR¹ in a well-controlled artificial fog environment is first described. Then, a two-stage data-driven approach, called LaNoising (la for laser), is presented. In the first stage, given certain visibility describing the level of fog, Gaussian Process Regression (GPR) model is used to predict a possibility distribution of the required minimum visibility for a laser to output a true detection range. If the laser cannot output the true range, in the second stage, Mixture Density Network (MDN) provides a probability prediction of the noisy measurement range. In our method, input can be obtained directly from LiDAR sensory data, and the data-driven statistical models are easy to implement and capable of providing a promising description of LiDAR performance for various objects under fog. For an intuitive understanding, Fig. 1 shows the appearance of the point clouds under clear weather conditions (generated by the Velodyne VLP-32C LiDAR) and simulation for fog environment (using our method), respectively.

The contributions of this paper are twofold. First, we propose a two-stage data-driven method to model the performance of 903nm ToF LiDAR in fog environment. Second, we quantitatively and qualitatively evaluate the proposed method and show promising results towards future autonomous driving under fog conditions. To the best of our knowledge, there is no prior data-driven approach to model the ranging process of a ToF LiDAR under fog, as we present in this paper.

II. RELATED WORK

In literature, there are a few researches investigating the impacts of adverse weather on LiDARs. Peynot *et al.* [3] ob-

¹<https://velodynelidar.com/vlp-32c.html>

served that dust particles can be detected by laser sensors and might hide obstacles behind the dust cloud. In [1], Rasshofer *et al.* modeled the LiDAR received laser power with a microphysical model. However, the developed model required modeling of the size, distribution, and optical characteristics of the droplets of fog or rain, which was very challenging, and it was only verified under a few rough visibility of severe weather due to facility limitations. Filgueira *et al.* [4] carried out real-world measurements with a Velodyne VLP-16 LiDAR for a couple of months to evaluate the influences of different rain levels on various reflective materials. The authors found that the returned intensity and the number of detected points decreased with the increasing of rain in all objects. Hasirlioglu *et al.* [5] investigated the effect of exhaust gases at low temperatures on sensor quality and found that exhaust gases were visible for laser scanners and led to degraded performance. In [6], Ashraf *et al.* used a Velodyne HDL-64E LiDAR at $903nm$ wavelength to analyze fog attenuation. An empirical fog attenuation model based on visibility range was applied to the point cloud generated by the LiDAR. However, their attenuation model was not verified by real experiments.

Data is essential to advance related research. To this end, the CEREMA climate chamber was built in France to provide stable conditions at a certain rain or fog level [7], [8]. Leveraging the data collected in this chamber, Bijelic *et al.* [9] tested four state-of-the-art LiDAR models in different controlled conditions and presented disturbance patterns for these LiDAR systems under fog. Kutila *et al.* [10] compared the performance of different LiDARs with two wavelengths of $905nm$ and $1550nm$ under various fog levels. However, the LiDAR performance modeling under fog wasn't given in these work, which remains a gap between their research and what would be required for an implementation that works with actual LiDAR data.

III. DATA COLLECTION

The data collection was carried out with a Velodyne VLP-32C LiDAR in CEREMA's artificial fog chamber, where various visibility of artificial fog can be generated and quantified by a transmissometer for meteorological visibility measurement. In our case, a dense fog with low visibility was firstly generated. Then, by stopping the generation of new water droplets, the dense fog becomes lighter and lighter through dissipation. An example of the change of the fog is shown by the blue curve in Fig. 3. Moreover, as shown in Fig. 2, three boards with calibrated reflectivities [11], a dummy human model, a real car, and two traffic signs were used as experimental objects (for detection purposes). The three calibrated boards with three different colors including black, white, and gray, have the known reflectivities of 0.1, 0.9, and 0.5, respectively, which are used to collect training data for diffuse reflectors. The human model is intentionally dressed with clothes and pants of different materials and colors to reflect different laser reflection characteristics. We placed these objects in front of the LiDAR at various distances (see Fig. 2, maximum $30m$ due to space limitations of CEREMA

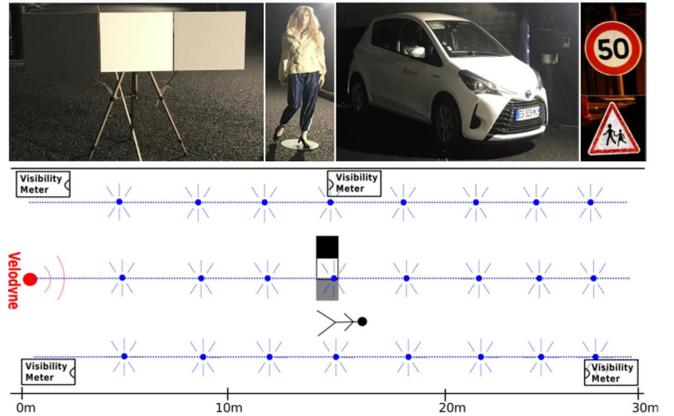


Fig. 2. Different objects and a typical scene setting in CEREMA chamber.

chamber) and recorded LiDAR measurements under different fog conditions. The simulated weather conditions during the data collection and the placements of each object are detailed in Table I.

TABLE I
DATA COLLECTION SETTINGS

Conditions	Objects	Placements
Fog (10-500m visibility), clear weather conditions	Calibrated boards	7.5-27.5m (every 2.5m)
	Human model	7.5-27.5m (every 2.5m)
	Car	10-25m (every 5m)
	Traffic signs	10-25m (every 5m)
	Background (walls)	Unchanged

For LiDAR data, each emitted laser is indexed with a ring number and an azimuth angle. The ring number and azimuth of the laser hitting on the objects were manually extracted and saved as Regions of Interest (RoIs) for further analysis [13]. For each placement (corresponding to a recording round) in Table I, the collected data can be represented as:

$$r_{i,j}(t), I_{i,j}(t), V(t), t | (i, j) \in \text{RoIs}, t \in [t_0, t_1] \quad (1)$$

where i indicates the ring number; j represents the azimuth angle; $r_{i,j}(t), I_{i,j}(t)$ and $V(t)$ mean the range, intensity, and visibility measurements at time t , respectively, while t_0 and t_1 are the start and end times of the recording round.

Fig. 3 shows how the measured ranges change with fog visibility for a laser hitting on the white calibrated board placed at $15m$. In which the red dots indicate the measured ranges $r(t)$ at different time t , the blue line represents the visibility $V(t)$, and the orange circle represents the disappear visibility V_{dis} , which is the minimum visibility that allows the ranging process to return the actual distance measurements.

IV. PROPOSED METHOD

In this section, we first give the theoretical model of ToF LiDAR, then propose a data-driven framework that describes the ToF LiDAR ranging process in fog environment.

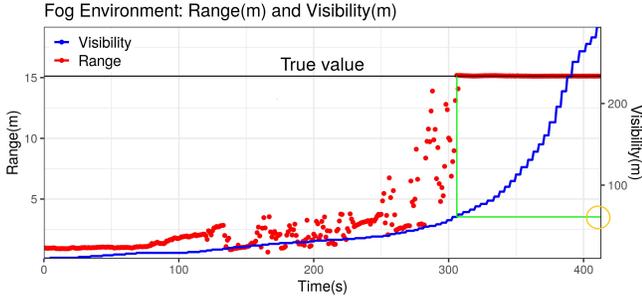


Fig. 3. Range measurements change with increasing visibility for a laser in fog environment.

A. Theoretical Model of ToF LiDAR

The physical scheme of a LiDAR is represented by the LiDAR power equation [1]:

$$P_R = C \frac{\beta}{R^2} \exp\left(-2 \int_{r=0}^R \alpha(r) dr\right) \quad (2)$$

where P_R represents the power of a received laser return at distance R ; C is a constant related to the light speed, laser transmit power, laser detector's optical aperture area, and overall system efficiency; β indicates the reflection efficiency of the object surface; α indicates the extinction coefficient of the LiDAR signal.

The signal processing system in LiDAR is required to detect the true return signal with a low signal-to-noise ratio. Typically, a thresholding algorithm is applied [12]. Under the homogeneity assumption of fog, the extinction coefficient α [1] and object surface reflectivity β [2] are both influenced by fog. According to the received laser power function, the ranging process of ToF LiDAR can be modeled as:

$$r(t) \sim f(\alpha(t), \beta(t) | R, \beta) \quad (3)$$

where R represents the object distance; β indicates the reflectivity of the object surface under clear weather conditions; $r(t)$, $\alpha(t)$, $\beta(t)$, and $V(t)$ are the detection range of a laser, extinction coefficient, object surface reflectivity, and meteorological visibility measurement under fog at time t , respectively. β is related to the laser intensity I under clear weather conditions. Both $\alpha(t)$ and $\beta(t)$ can be characterized by $V(t)$ [13]. Thus, Eq. 3 can be further expressed by:

$$r(t) \sim g(V(t) | R, I) \quad (4)$$

However, since the signal processing unit is usually a black box embedded inside the sensor (for commercial purposes), the analytical form of $g(\cdot)$ can hardly be obtained. Therefore, it is necessary to adopt a data-driven approach to break through this problem, as $g(\cdot)$ is critical to model noise under adverse weather conditions.

B. Data-driven Approach

Our approach is summarized in Algorithm 1. Specifically, we consider a set of N laser points (i.e. a point cloud) under clear weather conditions, and each point $n \in N$ is processed

independently by our nosing module. The inputs of this module is a triplet $\langle R, I, V_{vir} \rangle$ formed by the detection range R , intensity I , and virtual visibility V_{vir} , extracted from point n , while V_{vir} is the visibility we want to simulate and set manually. The output is the predicted detection range r . Since the diffuse reflectors and retro-reflectors don't share the same reflection features, they are distinguished by the intensity I (i.e. $I < 100$ for diffuse reflectors, and $I \geq 100$ for retro-reflectors according to the Velodyne VLP-32C LiDAR User Manual) and applied with different models. The GPR models predict the distribution of the disappear visibility in order to decide whether the corresponding ranging process can return the true distance measurements. The MDN models output the detection range distribution, if the true object distance cannot be displayed through the first stage.

Algorithm 1 LaNoising: a two-stage data-driven approach for 903nm ToF LiDAR performance simulation under fog

Input: $\langle R, I, V_{vir} \rangle$

Output: r

- 1: **if** $I < 100$ **then**
- 2: $V_{dis} \sim GPR_{dif}(R, I)$; // c.f. Section IV-B.1
- 3: **if** $V_{dis} < V_{vir}$ **then**
- 4: $r \leftarrow R$;
- 5: **else**
- 6: $r \sim MDN_{dif}(R, I, V_{vir})$; // c.f. Section IV-B.2
- 7: **end if**
- 8: **else**
- 9: $V_{dis} \sim GPR_{ret}(R, I)$; // c.f. Section IV-B.1
- 10: **if** $V_{dis} < V_{vir}$ **then**
- 11: $r \leftarrow R$
- 12: **else**
- 13: $r \sim MDN_{ret}(R, I, V_{vir})$; // c.f. Section IV-B.2
- 14: **end if**
- 15: **end if**

1) *GPR for Disappear Visibility Modeling*: A GPR model is used to predict whether a laser can successfully output the true or false range, given certain visibility. The input vector is defined as $\mathbf{x} = [R, I]$. The output of $GPR(R, I)$ is the predicted probability distribution of disappear visibility V_{dis} . If a sample from $GPR(R, I)$ is smaller than the manually set visibility V_{vir} , we will keep the estimated detection range r as the original one:

$$r = R, \text{ s.t. } V_{dis} < V_{vir} \quad (5)$$

where $V_{dis} \sim GPR(R, I)$

Gaussian process (GP) [14] model is a non-parametric machine learning tool with an infinitive dimensional Gaussian distribution and entirely defined by a mean function $m(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (6)$$

where usually we assume $m(\mathbf{x}) = 0$. For the regression task, given a set of measurements $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where \mathbf{x}_i denotes an input vector and y_i denotes a scalar output, we

assume $y = f(\mathbf{x}) + \varepsilon(\mathbf{x}); \varepsilon(\mathbf{x}) \sim \mathcal{N}(0, \sigma_n^2)$. The goal is to predict the output $f(\mathbf{x}_*)$ at an arbitrary \mathbf{x}_* . The posterior of $f(\mathbf{x}_*)$ can be expressed as:

$$\begin{aligned} f(\mathbf{x}_*) &\sim \mathcal{N}(E\{f(\mathbf{x}_*)\}, \text{var}\{f(\mathbf{x}_*)\}) \\ E\{f(\mathbf{x}_*)\} &= K_*^T [K + \sigma_n^2 I]^{-1} y \\ \text{var}\{f(\mathbf{x}_*)\} &= k(\mathbf{x}_*, \mathbf{x}_*) - K_*^T [K + \sigma_n^2 I]^{-1} K_* \end{aligned} \quad (7)$$

where $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j); K_* = (k(\mathbf{x}_1, \mathbf{x}_*), \dots, k(\mathbf{x}_n, \mathbf{x}_*))^T; y = (y_1, \dots, y_n)^T$.

We found that, using a combined kernel is easier to achieve satisfactory fitting results for our tasks, which is the product of several common kernels as expressed by:

$$k(\mathbf{x}, \mathbf{x}') = k_{Constant}(\mathbf{x}, \mathbf{x}') \times k_{Matern}(\mathbf{x}, \mathbf{x}') \times k_{DotProduct}(\mathbf{x}, \mathbf{x}') \quad (8)$$

where $k_{Constant}$, k_{Matern} , and $k_{DotProduct}$ are the constant kernel, Matern 3/2 kernel, and dot-product kernel respectively:

$$\begin{aligned} k_{Constant}(\mathbf{x}, \mathbf{x}') &= c \\ k_{Matern}(\mathbf{x}, \mathbf{x}') &= a \left(1 + \frac{\sqrt{3}d}{l} \right) \exp\left(-\frac{\sqrt{3}d}{l}\right), d = \|\mathbf{x} - \mathbf{x}'\|_2 \\ k_{DotProduct}(\mathbf{x}, \mathbf{x}') &= \sigma_0^2 + \mathbf{x} \cdot \mathbf{x}' \end{aligned} \quad (9)$$

where c, a, l, σ_0 are the hyperparameters that can be obtained by maximizing a marginal likelihood $\log p(y)$. Finally, two different GPR models, i.e. GPR_{dif} and GPR_{ret} , are developed for diffuse reflectors and retro-reflectors, respectively.

2) *MDN for Ranging Process Modeling*: In the case where the laser cannot correctly output range, MDN will provide a probabilistic prediction of the noisy measurement range. The input vector of MDN is defined as $\mathbf{x} = [R, I, V_{vir}]$. The output of $MDN(R, I, V_{vir})$ is an estimated probability distribution of the detection range r :

$$\begin{aligned} r &\sim MDN(R, I, V_{vir}) \\ \text{s.t. } &V_{dis} \geq V_{vir} \\ \text{where } &V_{dis} \sim GPR(R, I) \end{aligned} \quad (10)$$

which is usually modeled as a set of parameters of mixed Gaussian distributions [15]:

$$\begin{aligned} p(y|\mathbf{x}) &= \sum_{i=1}^m \alpha_i(\mathbf{x}) \phi_i(y|\mathbf{x}) \\ \phi_i(y|\mathbf{x}) &= \frac{1}{(2\pi)^{c/2} \sigma_i(\mathbf{x})^c} \exp\left\{-\frac{\|y - \mu_i(\mathbf{x})\|^2}{2\sigma_i(\mathbf{x})^2}\right\} \end{aligned} \quad (11)$$

where \mathbf{x} is the input vector; m is the number of mixture Gaussian distributions; α_i is the weight of a Gaussian distribution; μ_i and σ_i are the mean and variance, respectively. The negative log-likelihood loss is used in the optimization process, which can be expressed as:

$$\begin{aligned} \mathcal{L}(y|\mathbf{x}) &= -\log(p(y|\mathbf{x})) \\ &= -\log\left(\sum_{i=0}^m \frac{\alpha_i(\mathbf{x})}{(2\pi)^{c/2} \sigma_i(\mathbf{x})^c} \exp\left\{-\frac{\|y - \mu_i(\mathbf{x})\|^2}{2\sigma_i(\mathbf{x})^2}\right\}\right) \end{aligned} \quad (12)$$

We use two fully connected layers followed by the Rectified Linear Unit (ReLU) activation function as the hidden

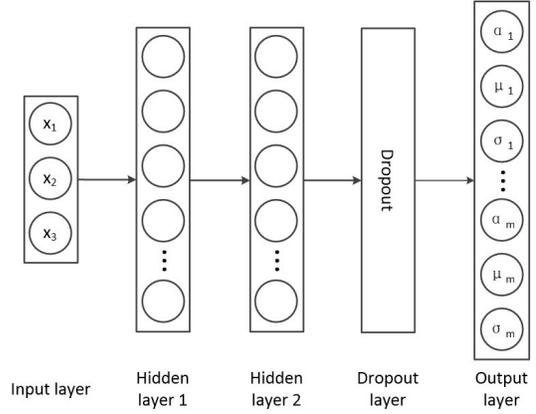


Fig. 4. Architecture of MDN.

layers (see Fig. 4). μ_i is obtained from the fully connected output layer. σ_i and α_i are obtained from the fully connected output layer followed by the Exponential Linear Unit (ELU) and softmax activation functions, respectively. A dropout layer is added between the last hidden layer and the output layer to prevent overfitting during the training process. The hyperparameters and training settings of MDN were obtained through experiments, which are summarized in Table II. Moreover, two different MDN models, i.e. MDN_{dif} and MDN_{ret} , are trained for diffuse reflectors and retro-reflectors, respectively.

TABLE II
HYPERPARAMETERS AND SETTINGS OF MDN

Items	Parameters
Neurons of hidden layer 1	24
Neurons of hidden layer 2	16
Dropout rate	0.25
Number of mixture models	1, 3, 5
Optimizer	Adam
Batch size	3000

V. EXPERIMENTAL RESULTS

In this section, we first illustrate the experimental results of disappear visibility and ranging process estimation, then show simulation results based on Robot Operating System (ROS) for a more intuitive understanding.

A. Disappear Visibility Estimation

Lasers hitting on the calibrated boards $b = \{A, B, C\}$ that were placed at different distances $d = \{7.5m, \dots, 27.5m\}$ (c.f. Table I), are used to train the model GPR_{dif} for diffuse reflectors. Specifically, for lasers $n \in RoI_{db}$, we calculate their mean of intensities I_{db} and mean of detection ranges R_{db} under clear weather conditions (i.e. when visibility is at its maximum), mean of disappear visibility V_{db} , and the standard deviation σ_{db} of disappear visibility. The vector $\mathbf{x}_{db} = [R_{db}, I_{db}]$ forms the input of a sample to train GPR_{dif} . V_{db} and σ_{db} are used as the output ground-truth of the sample. Lasers hitting on the traffic signs are used to train

the model GPR_{ret} for retro-reflectors. Due to the special surface material of traffic signs, the laser has high intensity reflection under clear weather conditions. Under fog, even low visibility is enough for LiDAR to detect the true distance of such objects.

Fig. 5(a) shows the regression results of GPR_{dif} , where the dots represent V_{db} and the vertical lines indicate σ_{db} . We give one standard deviation bounds of GPR predictions. It can be seen that the GPR model fits well to the curves of different boards. Fig. 5(b) and 5(c) visualize the training results of GPR_{dif} and GPR_{ret} as 3D surfaces. In each figure, X-axis and Y-axis represent respectively intensity I and detection range R of a laser under clear weather conditions, and Z-axis indicates the mean of predicted distribution of disappear visibility V_{dis} .

Due to the material and color, different parts of the human model (clothing) as well as the car show different reflection features. We therefore use the K-means algorithm [16] to first segment lasers hitting on these two objects into different point clusters for each of their placements. For a laser point, the input feature vector used for clustering is $[I, V_{dis}]$, where I is the intensity under clear weather conditions, and V_{dis} is the disappear visibility in real measurements. Fig. 5(d) shows an example of the clustering results when both the human model and the car are placed at 15m from the LiDAR. It can be seen that the human model points are segmented into two clusters (denoted by *human model 1* and *human model 2*, respectively), in which the red points (higher intensity) correspond to the white coat, and the blue points (lower intensity) correspond to the navy blue pants. The car are segmented into three clusters (denoted by *car 1*, *car 2*, and *car 3*, respectively), in which the red, blue, and green points correspond to the car body, window, and front grille and tire, respectively.

Then, we apply the trained model GPR_{dif} to the human model as well as the car for each of their placements. For the former, the predicted V_{dis} of a certain point cluster of an object is the mean of the sampled values from GPR outputs of the laser points, which is shown as dashed line in Fig. 5(e). The means of the real observations of V_{dis} and the corresponding bounds of one standard deviation are also shown. It can be seen that the predictions of GPR_{dif} are generally consistent with the real observations for both parts (i.e. coat and pants) of the human model, while overall, the error increases with the object placement distance, especially for *human model 1* (i.e. coat). Fig. 5(f) shows the results of applying GPR_{dif} to the car. It can be seen that, the GPR performs better on the car body (*car 1*) than the window, front grille and tire (*car 2* and *car 3*). This is mainly because GPR_{dif} is trained with the data of calibrated boards, while the latter's reflection features (especially the material) are closer to the car body than other parts.

B. Ranging Process Estimation

Lasers hitting on calibrated boards are also used to train the model MDN_{dif} for diffuse reflectors. For a laser hitting on a certain calibrated board at a certain distance, at time

t , $\mathbf{x} = [R, I, V]$ is the input of a sample, where R and I are respectively the detection range and the intensity of the laser under clear weather conditions, and V is the current visibility. The output ground-truth is the current detection range r . All the samples are split into the training and validation sets in 90% and 10%. We stop training when the loss on the validation set no longer drops. Lasers hitting on traffic signs are used to train the model MDN_{ret} for retro-reflectors. All the samples are split into the training, validation, and test sets in 70%, 10%, and 20%, respectively. As examples, each subgraph in Fig. 6 shows the measured and predicted results of a laser hitting on an object at a certain distance. It can be seen that, on the one hand, as the visibility increases, the real measurement range rises to the true object distance after being around a small value at the beginning. On the other hand, MDN is able to predict the changes of detection range with respect to visibility, but its predicted variance cannot converge to a smaller value in the end. When the car is placed at 25m, there is a relatively large difference between the prediction and real observations. And in some cases, the final predicted distribution means of MDN and true values have small gaps.

Two metrics commonly used in uncertainty estimation, i.e. Mean Squared Error (MSE) and Negative Log-Likelihood (NLL), are used to evaluate our trained models. Specifically, MSE is to measure the distance between the true values and prediction means, while NLL is to measure the uncertainty:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

$$NLL = \frac{1}{n} \sum_{i=1}^n (-\log(p(y_i|\mathbf{x}_i))) \quad (14)$$

where \mathbf{x} is the input vector; y and \hat{y} are the ground-truth and prediction mean of output. For a point cluster of an object at a certain distance, we first calculate MSE and NLL of all the lasers then take the means. The overall MSE and NLL for this cluster are the averages of the means over all the distances. The smaller these two metrics, the better the model prediction.

Table III shows the overall performance (optimal results shown in bold) of different models. "MDN comp n " means the MDN model with n mixture Gaussian models. For the combination model "GPR + MDN comp 1", the predicted detection range is the actual object distance after the measured visibility exceeds the predicted disappear visibility, and NLL is calculated only using the real observations of a laser before the predicted disappear visibility. It can be seen that increasing the number of mixture Gaussian models fails on most of objects.

For diffuse reflectors (i.e. human model and car), "MDN comp 1" shows the best results on MSE. The large MSE for *car 1* is caused by the data of car at 25m, in which the real measurements need higher visibility than the predictions of our models to reach the real object distance. "GPR + MDN comp 1" always achieves the best performance on NLL, because MDN model has a better uncertainty estimation

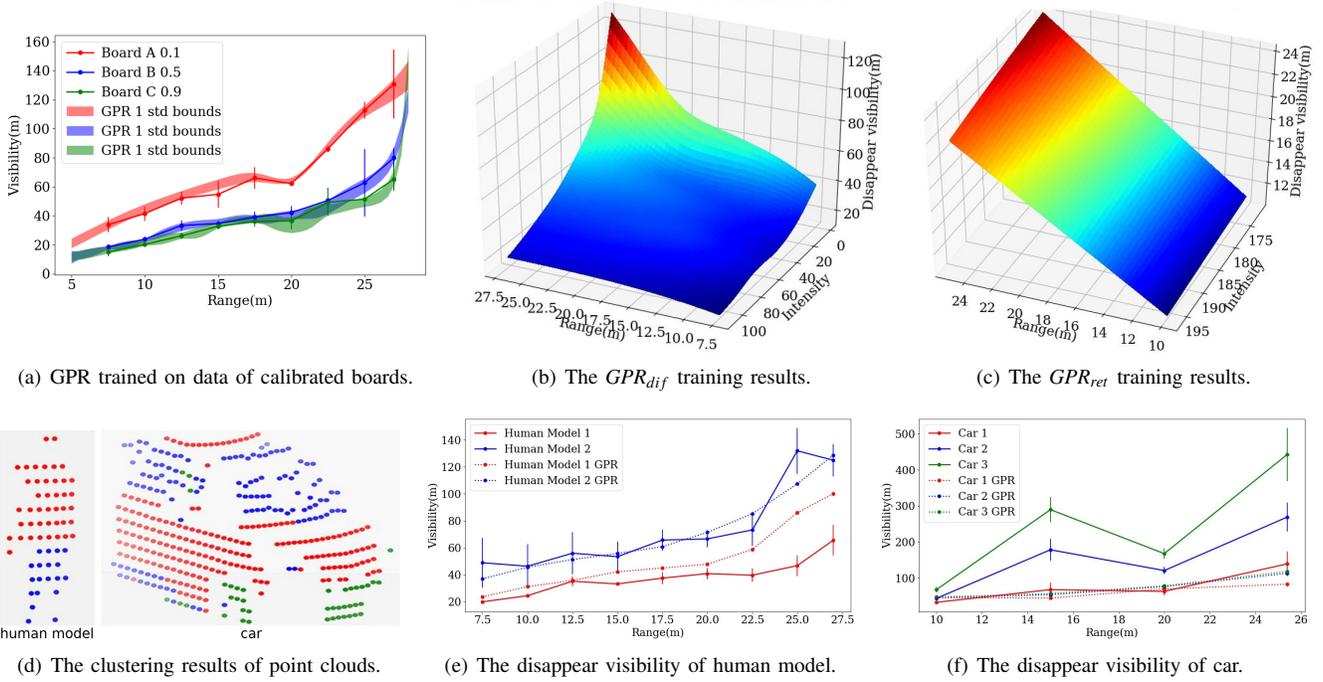


Fig. 5. GPR training and testing results. Best viewed in color. The legend colors in (e) and (f) correspond to the point colors in (d).

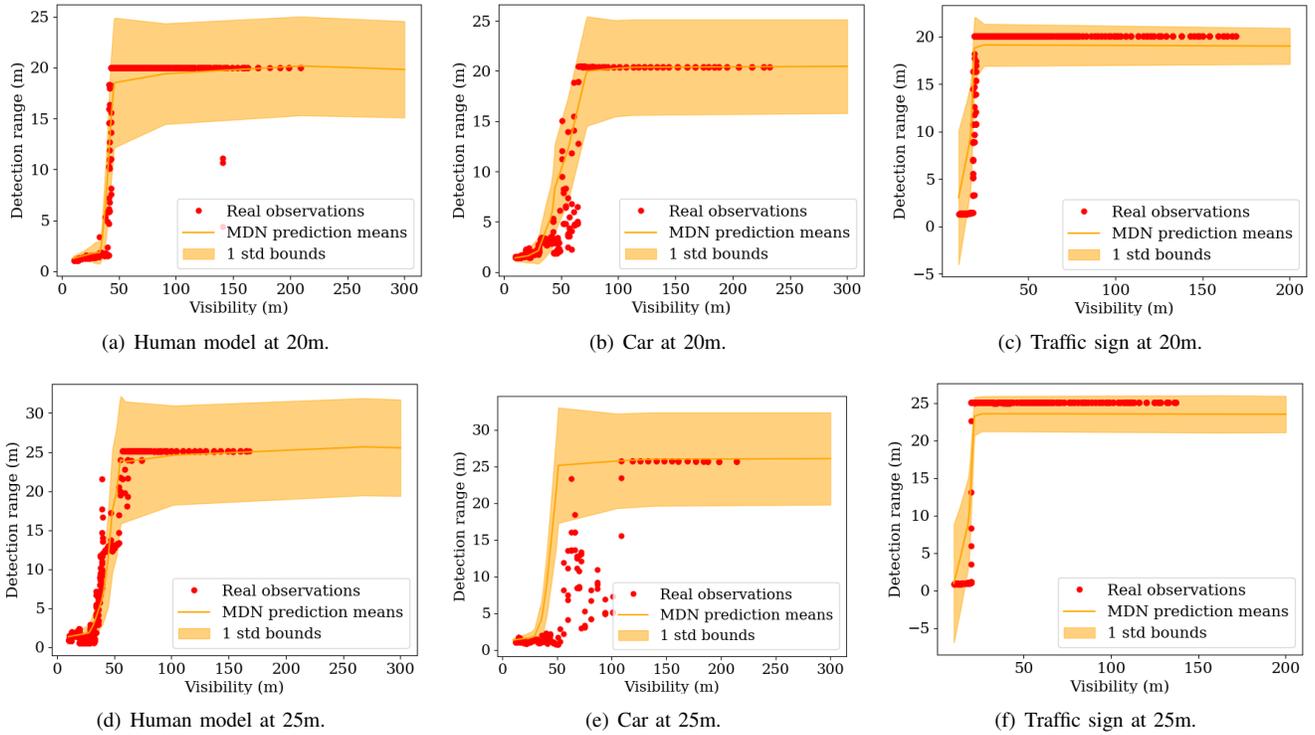


Fig. 6. Measured and predicted results of a laser hitting on an object placed at a certain distance. The orange curve represents the predicted distribution means, and only one Gaussian model is used in MDN.

TABLE III
OVERALL MSE AND NLL ON DIFFERENT OBJECTS

MSE				
	Human model 1	Human model 2	Car 1	Traffic signs
MDN comp 1	4.42	7.54	26.32	8.39
MDN comp 3	7.23	22.66	35.09	35.74
MDN comp 5	6.41	27.54	42.56	36.60
GPR + MDN comp 1	4.44	8.24	26.71	7.35

NLL				
	Human model 1	Human model 2	Car 1	Traffic signs
MDN comp 1	1.78	1.61	2.21	1.97
MDN comp 3	1.49	2.91	3.06	2.59
MDN comp 5	1.64	2.63	3.05	3.01
GPR + MDN comp 1	1.40	1.25	1.84	2.98

before the predicted distribution mean reaches the true object distance, and the model combination ultimately avoids the large variance of MDN model.

For retro-reflectors (i.e. traffic signs), most of observations of a laser are at the true object distance. The variance of the MDN prediction is able to achieve a relatively small value when reaching the true detection range. Combining the GPR gets the best MSE but lower NLL. Considering that we want to sample from the predicted range distribution before a certain visibility and ensure that the predicted detection range is stable at the true object distance after the certain visibility, we use the ‘‘GPR + MDN comp 1’’ models.

C. ROS-based Simulation

The simulation has been fully implemented in ROS. All experiments were carried out with Ubuntu 16.04 LTS (64-bit) and ROS Kinetic, with an Intel i7-7700HQ processor, 16 GB memory, and a NVIDIA GTX 1060 GPU (only used for neural network processing). A ROS package is developed, which first loads the pre-trained noise models, then reads the LiDAR data (i.e. *rosbag* file) recorded under clear weather conditions, and finally outputs the point cloud with noises.

Fig. 7 shows some visualization results of noise simulation using our method on CEREMA chamber data. It can be seen that, as the visibility increases, in both real measurements and simulation, noises are reduced and become more dispersed, while objects become more complete and visible. However, there are still some differences, especially the real noise points appear a more random and concentrated (in the vicinity of LiDAR) distribution in the space. The main reasons are that, during the CEREMA data collection, there is stagnant water on the ground and the distribution of the artificial fog is not absolutely homogeneous, which affect the reflection of the lasers. Fig. 8 gives some examples of noise simulation on the point cloud in real traffic scenario (EU long-term dataset [17]), recorded with two Velodyne HDL-32E LiDARs (only one is used in our experiments). We change the simulated visibility under fog from 60m to 20m. It can be seen that, as the visibility decreases, the effective

detection range gradually decreases while the noises increase and gather around the LiDAR.

VI. CONCLUSIONS

In this paper, we proposed a two-stage data-driven method to model the 903nm ToF LiDAR performance under fog. As the first stage, the GPR models predict the distribution of the disappear visibility, i.e. the minimum visibility that allows the corresponding ranging process to return the true distance measurements. In the second stage, the MDN models output the detection range distribution, in the case where the true object distance cannot be displayed through the first stage. The experimental results based on a dataset collected in CEREMA chamber show that our method is capable of providing a promising description of the 903nm ToF LiDAR performance in fog environment.

Our approach is currently experimented with point clouds generated by the Velodyne LiDARs. Since the intensity is used as an input in our method while different LiDAR manufacturers perform different methods to calculate this value, experiments with other products are expected in the future. Other work will include extending the proposed approach (i.e. used for data augmentation) to efficient 3D LiDAR data denoising [18], as well as robust object detection and tracking under adverse weather conditions [19]. The release of the developed ROS package will also take place in accordance with the project regulations.

REFERENCES

- [1] R. H. Rasshofer, M. Spies, and H. Spies, ‘‘Influences of weather phenomena on automotive laser radar systems,’’ in *Advances in Radio Science*, vol. 9, no. B.2. Copernicus GmbH, Jul. 2011, pp. 49–60.
- [2] J. Lekner and M. C. Dorf, ‘‘Why some things are darker when wet,’’ *Applied Optics*, vol. 27, no. 7, pp. 1278–1280, Apr. 1988.
- [3] T. Peynot, J. Underwood, and S. Scheding, ‘‘Towards reliable perception for unmanned ground vehicles in challenging conditions,’’ in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Oct. 2009, pp. 1170–1176.
- [4] A. Filgueira, H. Gonzalez-Jorge, S. Laguela, L. Dıaz-Vilarino, and P. Arias, ‘‘Quantifying the influence of rain in lidar performance,’’ *Measurement*, vol. 95, pp. 143–148, 2017.
- [5] S. Hasirlioglu, A. Riener, W. Huber, and P. Wintersberger, ‘‘Effects of exhaust gases on laser scanner data quality at low ambient temperatures,’’ in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2017, pp. 1708–1713.
- [6] I. Ashraf and Y. Park, ‘‘Effects of fog attenuation on lidar data in urban environment,’’ in *Smart Photonic and Optoelectronic Integrated Circuits XX*, vol. 10536. International Society for Optics and Photonics, 2018, p. 23.
- [7] M. Colomb, K. Hirech, P. Andre, J. J. Boreux, P. Lacote, and J. Dufour, ‘‘An innovative artificial fog production device improved in the European project ‘‘FOG’’,’’ *Atmospheric Research*, vol. 87, no. 3, pp. 242–251, Mar. 2008.
- [8] P. Duthon, F. Bernardin, F. Chausse, and M. Colomb, ‘‘Methodology Used to Evaluate Computer Vision Algorithms in Adverse Weather Conditions,’’ *Transportation Research Procedia*, vol. 14, pp. 2178–2187, Jan. 2016.
- [9] M. Bijelic, T. Gruber, and W. Ritter, ‘‘A Benchmark for Lidar Sensors in Fog: Is Detection Breaking Down?’’ in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2018, pp. 760–767.
- [10] M. Kutila, P. Pyykonen, H. Holzhuter, M. Colomb, and P. Duthon, ‘‘Automotive lidar performance verification in fog and rain,’’ in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1695–1701.

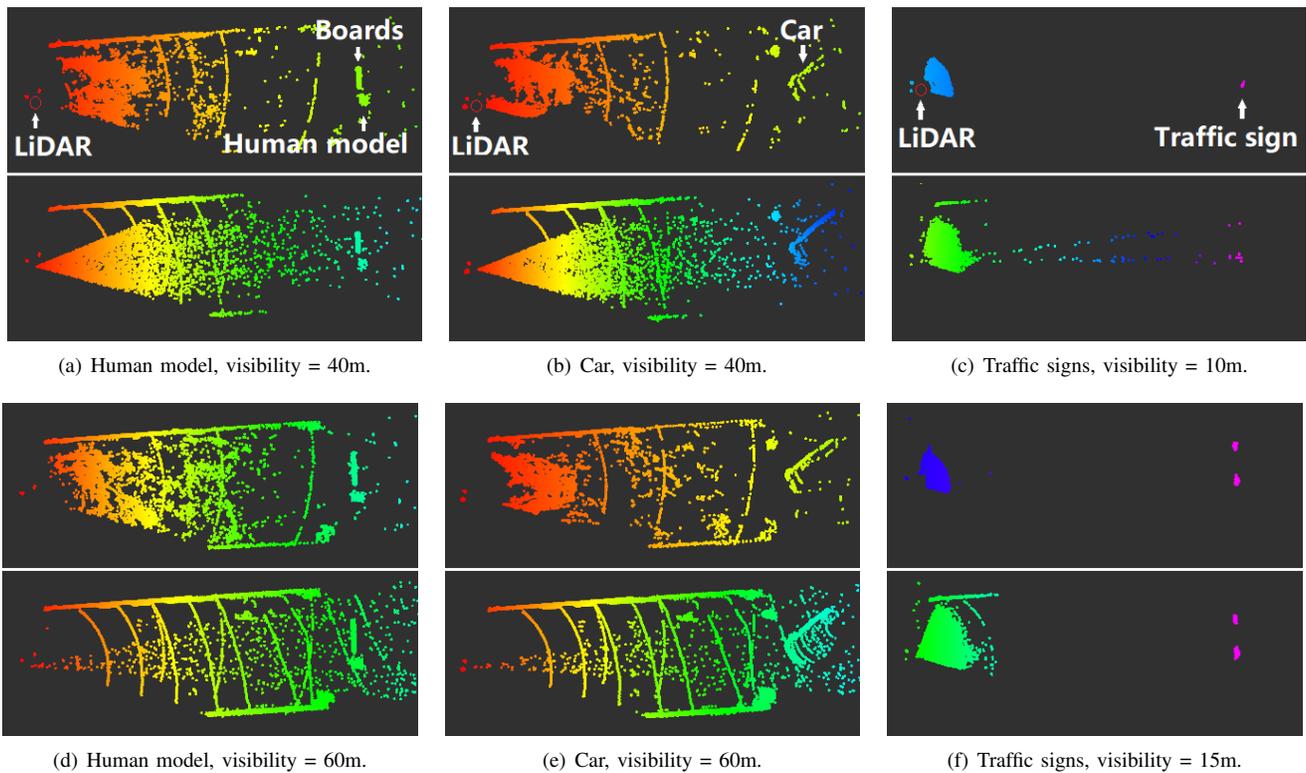


Fig. 7. Examples of noise simulation (bird's eye view). Best viewed in color. For each subgraph, the upper part is the real measurement (i.e. CEREMA chamber data) while the lower part is the simulation result. The boards, human model, car, and traffic signs are all placed at 15m under different visibility.

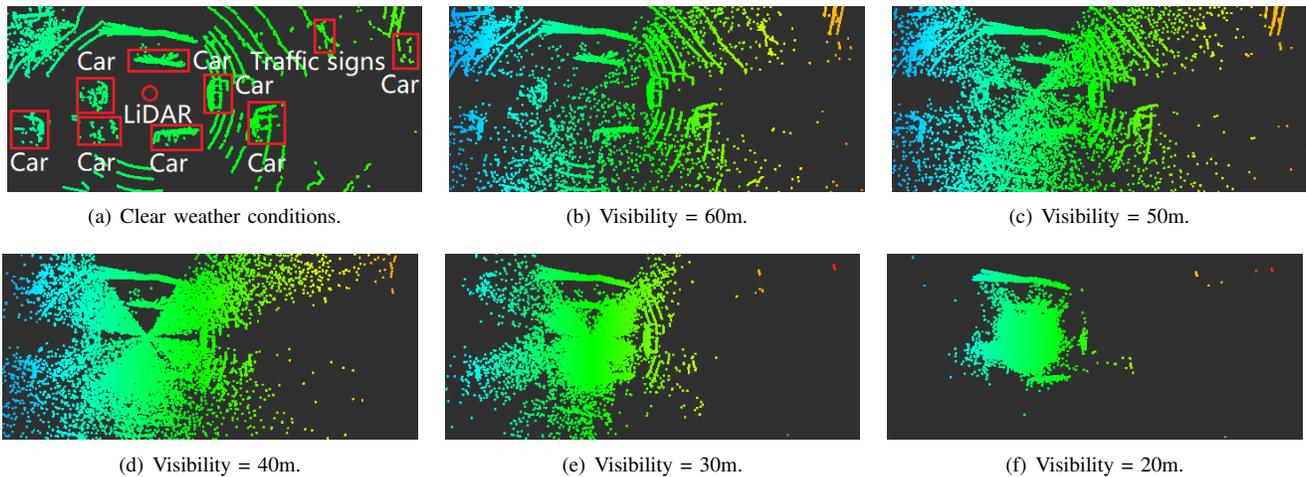


Fig. 8. Examples of noise simulation (bird's eye view) on the point cloud in real traffic scenario (EU long-term dataset [17]).

[11] M. Colomb, P. Duthon, and F. Bernardin, "Spectral reflectance characterization of the road environment to optimize the choice of autonomous vehicle sensors*," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct 2019, pp. 1085–1090.

[12] T. Ogawa and G. Wanielik, "TOF-LIDAR signal processing using the CFAR detector," in *Advances in Radio Science*, vol. 14, no. F. Copernicus GmbH, Sep. 2016, pp. 161–167.

[13] Y. Li, P. Duthon, M. Colomb, and J. Ibanez-Guzman, "What happens for a tof lidar in fog?" *IEEE Transactions on Intelligent Transportation Systems*, vol. Early Access, pp. 1–12, 2020.

[14] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[15] A. Brando, "Mixture density networks (mdn) for distribution and uncertainty estimation," 2017, report of the Master's Thesis: Mixture Density Networks for distribution and uncertainty estimation.

[16] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, no. 2, pp. 451–461, Feb. 2003.

[17] Z. Yan, L. Sun, T. Krajník, and Y. Ruichek, "EU long-term dataset with multiple sensors for autonomous driving," *CoRR*, vol. abs/1909.03330, 2019.

[18] R. Heinzler, F. Piewak, P. Schindler, and W. Stork, "Cnn-based lidar point cloud de-noising in adverse weather," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2514–2521, 2020.

[19] F. Majer, Z. Yan, G. Broughton, Y. Ruichek, and T. Krajník, "Learning to see through haze: Radar-based human detection for adverse weather conditions," in *2019 European Conference on Mobile Robots (ECMR)*. IEEE, 2019, pp. 1–7.