Universität Stuttgart

A Methodology for Validation of a Radar Simulation for Virtual Testing of Autonomous Driving

Von der Fakultät für Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Ingenieurwissenschaften (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Anthony Ngo

aus Ludwigshafen am Rhein

Hauptberichter: Prof. Dr.-Ing Dr. h.c. (DonNTU) Dr. h.c. (RAS/Sib) Hon. Prof. (RAS/Sib) Michael M. Resch

Mitberichter: Prof. Dr.-Ing Hans-Christian Reuss

Tag der mündlichen Prüfung: 18.01.2023

Institut für Höchstleistungsrechnen der Universität Stuttgart

2023

Any sufficiently advanced technology
is indistinguishable from magic.

ARTHUR C. CLARKE (1917 - 2008),
Clarke's three laws

Abstract

Autonomous driving offers great potential for reducing the number of accidents as well as optimizing traffic flow. The safety validation of such an autonomous system is an extremely difficult problem and new approaches are needed because the conventional statistical safety proof based on field testing is not feasible. The combination of real-world and simulation-based tests is a promising approach to significantly reduce the validation effort of autonomous driving.

As environment sensors such as lidar, camera, and radar are key technologies for a self-driving vehicle, they have to be validated to be able to rely on virtual tests using synthetically generated sensor data. In particular, radar has traditionally been one of the most complex sensor to model. Since a sensor simulation is an approximation of the real sensor, a discrepancy between real sensor measurements and synthetic data can be assumed. However, there exists no systematic and sound method for validating a sensor model, especially for radar models.

Therefore, this work makes several contributions to address this problem with the objective to gain an understanding of the capabilities and limitations of sensor simulation for virtual testing of autonomous driving.

Considering that high fidelity radar simulations face challenges regarding the required execution time, a sensitivity analysis approach is introduced with the goal to identify the sensor effects that has the biggest impact on a downstream sensor data processing algorithm. In this way, the modeling effort can be focused on the most important components in terms of fidelity, while minimizing the overall computation time required.

Furthermore, a novel machine learning-based metric is proposed for evaluating the accuracy of synthetic radar data. By learning the latent features that distinguish real and simulated radar point clouds, it can be demonstrated that the developed metric outperforms conventional metrics in terms of its capability to measure characteristic differences. Additionally, after training, this removes the need for real radar measurements as a reference to evaluate the fidelity of a sensor simulation.

Moreover, a multi-layered evaluation approach is developed to measure the gap between radar simulation and reality, consisting of an explicit and an implicit sensor model evaluation. The former directly assesses the realism of the simulated data, whereas the latter refers to an evaluation of a subsequent perception application. It can be shown that by introducing multiple levels of evaluation, the existing discrepancies can be revealed in detail and the sensor model fidelity can be accurately measured across different scenarios in a holistic manner.

Zusammenfassung

Das autonome Fahren bietet ein großes Potenzial sowohl zur Verringerung der Unfallzahlen als auch zur Optimierung des Verkehrsflusses. Der Sicherheitsnachweis eines solchen autonomen Systems ist ein äußerst komplexes Problem und es werden neue Ansätze benötigt, da der konventionelle statistische Nachweis auf Basis von Feldversuchen nicht wirtschaftlich ist. Hierzu bietet die Kombination aus realen und simulationsbasierten Tests einen vielversprechenden Ansatz, um den Validierungsaufwand für autonomes Fahren deutlich zu reduzieren.

Umgebungssensoren wie Lidar, Kamera und Radar sind Schlüsseltechnologien für ein selbstfahrendes Fahrzeug. Zudem müssen sie validiert werden, um sich auf virtuelle Tests mit synthetisch erzeugten Sensordaten verlassen zu können. Insbesondere Radar ist traditionell einer der am komplexesten zu modellierenden Sensoren. Da eine Sensorsimulation eine Annäherung an den realen Sensor darstellt, kann eine Diskrepanz zwischen realen Sensormessungen und synthetischen Daten angenommen werden. Es gibt jedoch keine systematische und fundierte Methode zur Validierung eines Sensormodells, insbesondere für Radarmodelle.

Daher leistet diese Arbeit mehrere Beiträge zur Lösung dieses Problems mit dem Ziel, ein Verständnis für die Möglichkeiten und Grenzen der Sensorsimulation für virtuelle Tests des autonomen Fahrens zu erlangen.

In Anbetracht der Tatsache, dass hochakkurate Radarsimulationen im Hinblick auf die erforderliche Ausführungszeit eine Herausforderung darstellen, wird eine Sensitivitätsanalyse verwendet, mit dem Ziel, die Sensoreffekte zu identifizieren, die den größten Einfluss auf einen nachgelagerten Sensordatenverarbeitungsalgorithmus haben. Auf diese Art und Weise kann der Modellierungsaufwand auf die Schlüsselkomponenten in Bezug auf die Wiedergabetreue fokussiert werden, wodurch die erforderliche Gesamtrechenzeit minimiert werden kann.

Darüber hinaus wird eine neuartige, auf maschinellem Lernen basierende

Metrik zur Bewertung der Genauigkeit synthetischer Radardaten präsentiert. Durch das Erlernen der latenten Merkmale, die reale und simulierte Radar-Punktwolken unterscheiden, kann gezeigt werden, dass die entwickelte Metrik herkömmliche Metriken in Bezug auf ihre Fähigkeit, charakteristische Unterschiede zu messen, übertrifft. Außerdem werden nach dem Training des neuronalen Netzes keine weiteren realen Radarmessungen mehr benötigt, um die Genauigkeit einer Sensorsimulation zu bewerten.

Zur Messung der Diskrepanz zwischen Radarsimulation und Realität wird ein mehrstufiger Bewertungsansatz entwickelt, der aus einer expliziten und einer impliziten Sensormodellbewertung besteht. Erstere bewertet direkt die Exaktheit der simulierten Daten, während letztere sich auf die Bewertung eines nachfolgenden Perzeptionsalgorithmus bezieht. Es kann gezeigt werden, dass durch die Einführung mehrerer Bewertungsebenen die bestehenden Unterschiede im Detail aufgedeckt werden können und die Fidelität über verschiedene Szenarien hinweg auf eine holistische Weise genau gemessen werden kann.

Acknowledgments

The present thesis and the corresponding research were completed with the High Performance Computing Center of the University of Stuttgart in cooperation with Robert Bosch GmbH.

First and foremost, I would like to thank my supervisor Michael Resch, for agreeing to supervise me as an external Ph.D. student. This allowed me to gain experience in academic research as well as to work at a company at the same time. Most of all, thank you for your high-quality feedback and for giving me the freedom to implement my ideas in this work. In addition, I would like to thank Prof. Reuss for taking over the role as the second examiner.

I want to express my gratitude to Frank Lattemann for giving me the opportunity to do research within Bosch. Thank you for always placing your trust in me and shielding my research from the distractions of a corporate environment, which cannot be taken for granted.

Furthermore, I would like to thank all the members of the simulation team for always giving me valuable feedback and creating a great, challenging, friendly, and fun place at work, which ultimately was one of the main reasons to stay at Bosch after my thesis. Special thanks to my Ph.D. sparring partner, Max Bauer. We benefited a great amount from each other and were always able to push ourselves and brainstorm ideas, not only about work-related things. Moreover, I would like to thank my students. It was always a great pleasure to work on problems with you and develop ideas together.

Finally, I want to thank my family and friends for the constant support. In particular, my partner Thiep for her encouragement, trust, and patience during all the evenings and weekends I spent writing this thesis.

Contents

List of Figures

List of Tables

Acronyms

ACC adapted cruise control.

AD autonomous driving.

ADAS advanced driver assistance systems.

AWGN additive white Gaussian noise.

CARLA car learning to act.

DDM data-driven model.

DEM deep evaluation metric.

DGPS differential global positioning system.

DP detection probabilities.

EMD earth mover's distance.

ESME explicit sensor model evaluation.

FAST fourier amplitude sensitivity testing.

FL fidelity level.

FOV field of view.

IMU inertial measurement unit.

IoU intersection over union.

IRM ideal radar model.

ISME implicit sensor model evaluation.

MOT multiple object tracking.

OSPA optimal subpattern assignment.

PNE point number error.

RCS radar cross section.

RMSE root mean squared error.

ROC receiver operating characteristic.

RQ research question.

RTM ray tracing-based model.

SAE society of automotive engineers.

SNR signal-to-noise ratio.

ViL vehicle-in-the-loop.

WD wasserstein distance.

1 Introduction

Contents

Besides the vehicle electrification, autonomous driving (AD) is currently one of the main trends in the automotive industry [7]. Fully automated driving offers the greatest potential for minimizing the likelihood of accidents and optimizing traffic flow [8]. However, the continued postponement of the market launch shows that besides the legal situation especially the testing and validation process of the AD system is highly complex and not yet solved. Although, in 2016 various companies announced fully autonomous driving for 2025 [9,10], as of today only a few dare to announce a concrete year for the release and some even believe that it could still take decades [11].

Therefore, as the industry moves towards full automation as defined by the Society of Automotive Engineers (SAE) International [12], it becomes more and more necessary to develop not only advanced safety systems but also the tools for their accurate analysis in order to validate the complex system [13]. The safety validation of such an autonomous system is an extremely challenging problem and new approaches are needed since a statistical proof of safety based on real-world testing does not scale [14]. The combination of field tests and simulation-based testing is a promising approach to substantially reduce the validation effort of autonomous driving [15].

Both a reliable perception of the vehicle environment and its precise evaluation are essential requirements for safe autonomous driving functions. A sensor-based environment detection with subsequent computer-based situation interpretation makes it possible to support the driving task of the autonomous vehicle depending on the prevailing traffic situation or even actively influence it. The sensors thus create the basis for any further processing of the environmental information and are consequently crucial components of any autonomous vehicle. Realistic models of environment perception sensors such as radar, lidar and camera play a key role in the simulation-based testing strategy [16]. These sensor models have to be validated in order to permit any reliable prediction about the behavior of the real system through virtually testing [17]. Especially the validation of radar models is a challenge and as of today there is no generally accepted methodology to validate these models [18].

The objective of this dissertation is to develop methods to investigate the capabilities and limits of a sensor simulation in order to derive reliable predictions about a real system based on tests in a virtual environment. Furthermore, this includes techniques for sensor model validation but also different methods to model a radar simulation for virtually testing autonomous driving functions such as object detection and tracking.

The present chapter introduces the motivation and relevance for this research topic, discusses the challenges in validation of a sensor simulation and presents the research questions addressed in this thesis. The final section concludes with the approach and the thesis outline.

1.1 Motivation

Wachenfeld and Winner have shown that a vast amount of driving kilometers is necessary to statistically proof the safety of an autonomous vehicle [2]. For instance to ensure that an autonomous vehicle can handle 95% of the driven kilometers safely, it would be necessary to drive a total of 10 million kilometers. Even with all these kilometers it is not guaranteed that the right scenarios are considered and all crucial situations tested, since the number of surprises declines with increasing testing distance, as illustrated in Figure 1.1 [1]. Hereby, an event is considered a surprise if the autonomous vehicle has reached an unwanted condition. This indicates that with real driving tests alone a statistical validation would not be economically reasonable. The combination of real driving tests and tests in an artificial environment is a promising approach to significantly reduce the validation effort of autonomous driving systems [15].

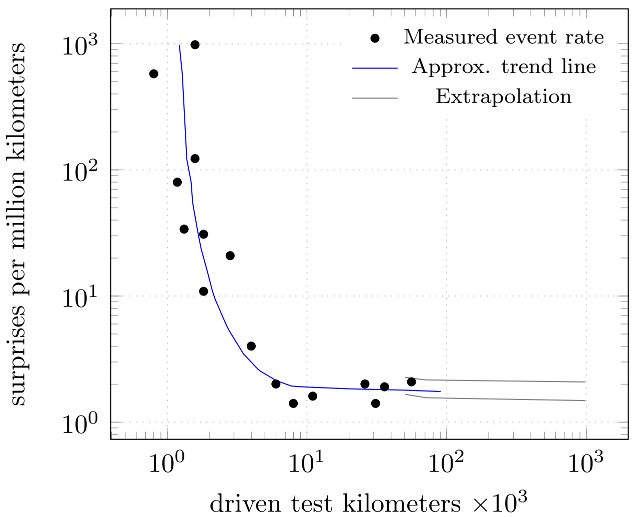

Figure 1.1: The frequency of adverse event findings with growing testing distance [1].

Virtual testing is a promising alternative, since it is less expensive, runs faster and there is no risk to human safety. Furthermore, the simulation is not exposed to the randomness in the real world. This means on the one hand that specific critical situations can be forced to happen. On the other hand, this particular inability to model the random behavior of real-life situations is a reason why the simulation alone might not be sufficient to validate the system completely in simulation, in view of the fact that a virtual model is always an approximation of the real world [19]. Nonetheless, the question remains whether this randomness is really needed. The proof that a simulation is valid is still missing [20]. For these reasons, a combination of both is a promising approach to reduce the testing and validation effort of autonomous driving. However, the question arises when to use the simulation and when driving tests, or in other words: Where are the limits of a simulation?

1.1.1 Simulation-based testing

In order to accelerate the validation process of autonomous vehicles, different approaches exist to combine real and simulation-based testing methods, as shown in Figure 1.2. It divides potential testing approaches into nine classes that differ in how they approximate the vehicle or the environment. In the following, the different methods are discussed in order to elaborate the general trade off between real and simulation-based testing.

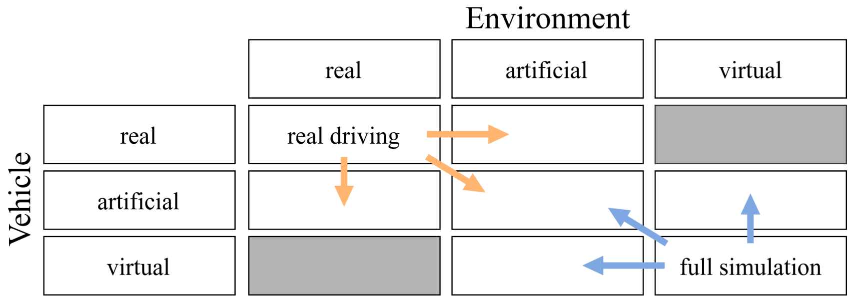

Figure 1.2: Different methods for combining simulation-based and real testing [2]. Orange arrows indicate decreasing cost, whereas blue arrows indicate increasing fidelity.

Real-world driving tests accurately represent the vehicle and the environment, because no approximation is needed. Since it involves the real vehicle, in real traffic, under real conditions, it is the only method guaranteed to produce correct results [21]. However, real driving is not used for testing safety-critical scenarios due to the risk of accidents and their consequences [1]. Furthermore, the environment can not be fully controlled, so the tested scenarios are highly dependent on the coincidences of reality. This also makes the reproducibility of observed situations impossible. Since this test approach can be deployed with the first street-legal prototypes at the earliest, other methods are needed to build trust in the system's safety prior to any real driving study or public deployment.

There are several ways to simplify reality at the cost of a loss in accuracy. In order to approximate the testing environment, real vehicles can be tested in an artificial environment. Proving grounds are an often used alternative [22]. The real environment is approximated in favor of reproducibility, safety, observability and variability. However, creating artificial test environments require additional time and money.

Instead of approximating the environment, an artificial vehicle can be used to test the system in a real environment. This artificial vehicle can be represented by a test driver who has the ability to intervene into the driving task or by a technical system that is superior to the autonomous vehicle due to a more advanced or additional sensor setup. Since the system is tested on public roads, for example, the reaction of passengers to an autonomous vehicle can be studied in order to improve the acceptance of self-driving cars [23]. The main limitations with this approach is its high cost, the risk of collision, and that it requires a deployment-ready system [21].

In Figure 1.2, approaches that combine real and virtual systems are depicted in gray, as they are technically not existent. This is due to the fact that, for example, a virtual inverter cannot generate real voltage and a real radar sensor is unable to sense a virtual environment.

Nonetheless, it is possible to combine an artificial and a real environment or vehicle. An example for this are vehicle-in-the-loop (ViL) systems [24,25]. ViL systems have been developed to test AD functions especially in safety critical traffic situations. By incorporating the real test vehicle into a traffic simulation, the advantages of both domains can be combined with this approach. Hereby, the real sensors can be artificially stimulated with synthetically generated data produced by sensor models [26].

The highest level of abstraction of both vehicle and environment is represented by the full simulation. In contrast to the previous testing approaches, the entire test domain is purely virtual, providing the greatest degree of safety, controllability, observability and repeatability [27]. Additionally, this method can also be deployed in early stages of AD development. Both the entire system as a whole and the individual components can be tested. For example, simulation can be used to find critical scenarios with the goal of validating the autonomous system [15]. On the contrary, individual modules such as the vehicle dynamic can be investigated to gain insights about behavior under different operating conditions [28].

1.1.2 Usage of synthetic sensor data

In many of the previously described methods, sensor models play a crucial role. This is because they perceive the virtual environment and can thus synthetically generate sensor data, which forms the basis for the general understanding of the prevailing situation of an autonomous vehicle.

Several areas of application can be found in the literature for using synthetically generated sensor data in the development and testing process [29]. Sligar employs an accurate physics-based radar simulation to create virtual sensor data set in order to train a machine learning-based object detection model [30]. In contrast, Hartstern et al. use probabilistic sensor models to identify the optimal sensor setup solution in early development stages since they provide a wide range of modification parameters and adjustable settings [31]. Ponn et al. utilize phenomenological sensor models to automatically create challenging and critical scenarios based on a sensor setup model of the autonomous vehicle [32].

The primary challenge of virtually testing with synthetically generated sensor data is that it requires developing and validating the employed sensor models as well as the environment models. If, however, this validity was shown, this approach allows for high reproducibility, variability and safety during testing, as the environment and the vehicle would only meet in a virtual world [1].

1.2 Problem statement

Since each sensor model is an approximation of the real sensor, a certain simulation-to-reality gap can be assumed to exist. Hence, the fidelity of a sensor model needs to be measured in order to be able to rely on virtual tests. Although there exists many approaches to model a sensor in the literature, the problem of quantitatively measure a sensor model fidelity and thus validating the model remains to be solved [29].

Whereas being considered as a key sensor for AD, the radar sensor has traditionally been one of the most complex sensors to model [33]. Many approaches exist to model a radar sensor system. Despite the fact that a lot of radar effects are understood and can be modeled today, a high fidelity simulation faces challenges in terms of the required execution speed [34]. The reason for this is that radar exhibits numerous physical characteristics, including interference, ambiguities, clutter, ghost objects and multi-path reflections [35], which leads to tremendously high demands on the computing power for a profound and comprehensive sensor simulation. However, the question arises whether such a detailed model is required for all simulation use cases. The sufficient level of detail and thus the right trade-off between model realism and computation time must be found.

Based on the aforementioned aspects the following research questions (RQ) can be derived:

RQ1 *What is missing in existing sensor model validation approaches?*

RQ2 *Which features of the radar simulation are relevant for a downstream application?*

RQ3 *How to determine the degree to which the radar simulation and experimental measurements concur?*

RQ4 *How to measure the overall simulation-to-reality gap considering a target application?*

A detailed derivation of the research questions can be found in Section 2.4. The research approach that is pursued in order to address these questions is outlined in the following section.

1.3 Approach

Virtual tests does not remove the need for large quantities of tests in a real environment. Instead, the objective of the present thesis is to investigate how much a simulation can be trusted in order to be able to define the optimal ratio between real and virtual tests. Furthermore, this work focuses on the validation of sensor models for virtually testing autonomous driving functions. The overarching goal is to develop methods to investigate the capabilities and therefore the simulation-to-reality gap of a sensor simulation for simulation-based testing.

This serves the purpose to validate a sensor model and thus provide the basis for a fundamental decision as to which test cases can be developed and tested with the aid of simulated sensor data. On the one hand, this can shorten the development time of an AD system and therefore save costs; on the other hand, testing can also become safer and more controlled compared to real world only tests.

Although this work focuses on the validation of a radar sensor simulation in particular, the methods developed for this are not exclusively designed for this type of sensor, but rather allow further abstractions to other sensor modalities such as lidar or camera as well as simulation models in general.

1.3.1 Contributions

In light of the given problem statement, this thesis contributes to the state of the art in research on the validation of sensor models with multiple validation approaches of a radar simulation for virtually testing autonomous driving. To the author's knowledge, a comparable comprehensive sensor model validation has not yet been proposed. The core findings of this thesis can be found in the following publications: [36], [37] and [38].

Contributions to specific problems can be summarized as follows:

- **Validation framework** (Chapter 3, [38]):
  A modular framework is developed for sensor model validation. This framework enables the simulation of real test drives in order to generate and compare synthetically generated data with real sensor data. Furthermore, subsequent perception algorithms can be used to investigate their performance in the virtual domain.

- **Radar sensor models** (Chapter 3.2, [36,38]):
  Three typical radar sensor models are implemented and evaluated. An ideal sensor model is introduced, which emulates idealized behavior of the radar sensor. As a second radar model, a data-driven sensor model is developed that strives to approximate radar detections by learning specific characteristics from real sensor measurements. Finally, a physical model is implemented, which models the radar wave propagation based on the ray tracing approach.

- **Identifying relevant sensor effects** (Chapter 4, [36]):
  Given that high fidelity radar simulations face challenges regarding the required computation time, this proposal introduces a sensitivity analysis approach for developing and evaluating a radar simulation, with the objective to identify the sensor effects that has the biggest impact on a system under test. This allows the modeling effort to be focused on the essential components in terms of fidelity, thus reducing the total computation time required. In addition, differences between the real and virtual domain can be examined in detail as well as results and disparities can be traced back to the contribution of the individual sub-modules of the radar simulation.

- **Evaluating simulated radar data** (Chapter 5, [37]):
  A novel data-driven metric is proposed to learn the latent features that distinguish simulated and real radar data. In this work, a neural network is developed to differentiate real and synthetic radar sensor data with the objective to learn the important features of real radar point clouds. Moreover, the classifier's confidence score for the 'real radar point cloud' class is proposed as a metric to determine the degree of fidelity of synthetically generated radar data. The presented method is evaluated and it can be demonstrated that the proposed deep evaluation metric outperforms conventional metrics in terms of its capability to identify characteristic differences between real and simulated radar data. Additional, after training, this removes the necessity for real radar data as a reference to evaluate the fidelity of sensor models.

- **Measuring the simulation-to-reality gap of radar perception** (Chapter 6, [38]):
  A multi-layered approach for measuring the radar simulation-to-reality gap of radar perception for autonomous driving is introduced. This method consists of a combination of an explicit and an implicit sensor model evaluation. The former directly evaluates the realism of the synthetically generated sensor data, while the latter refers to an evaluation of a downstream desired application. The effectiveness of the proposal is examined in terms of a sound sensor model assessment that reveals existing discrepancies and provides an accurate estimation of the overall sensor model fidelity across different scenarios.

1.3.2 Thesis outline

The present dissertation is structured in seven chapters, which are illustrated in Figure 1.3 along with the stated research questions.

Following this introductory chapter, the technical background including the radar principles, existing modeling approaches and the state of the art of sensor model validation approaches are provided in Chapter 2 in order to answer the first research question.

Chapter 3 introduces the developed framework to generate both real and simulated sensor data as well as the implemented radar perception, including the developed radar models. This framework forms the basis for the following developed methods to deal with the remaining research problems.

Thereafter, Chapters 4-6 entail the core contributions of this work and discuss the proposed methods.

Finally, Chapter 7 provides an in-depth discussion of the presented work and concludes the dissertation with an outlook.

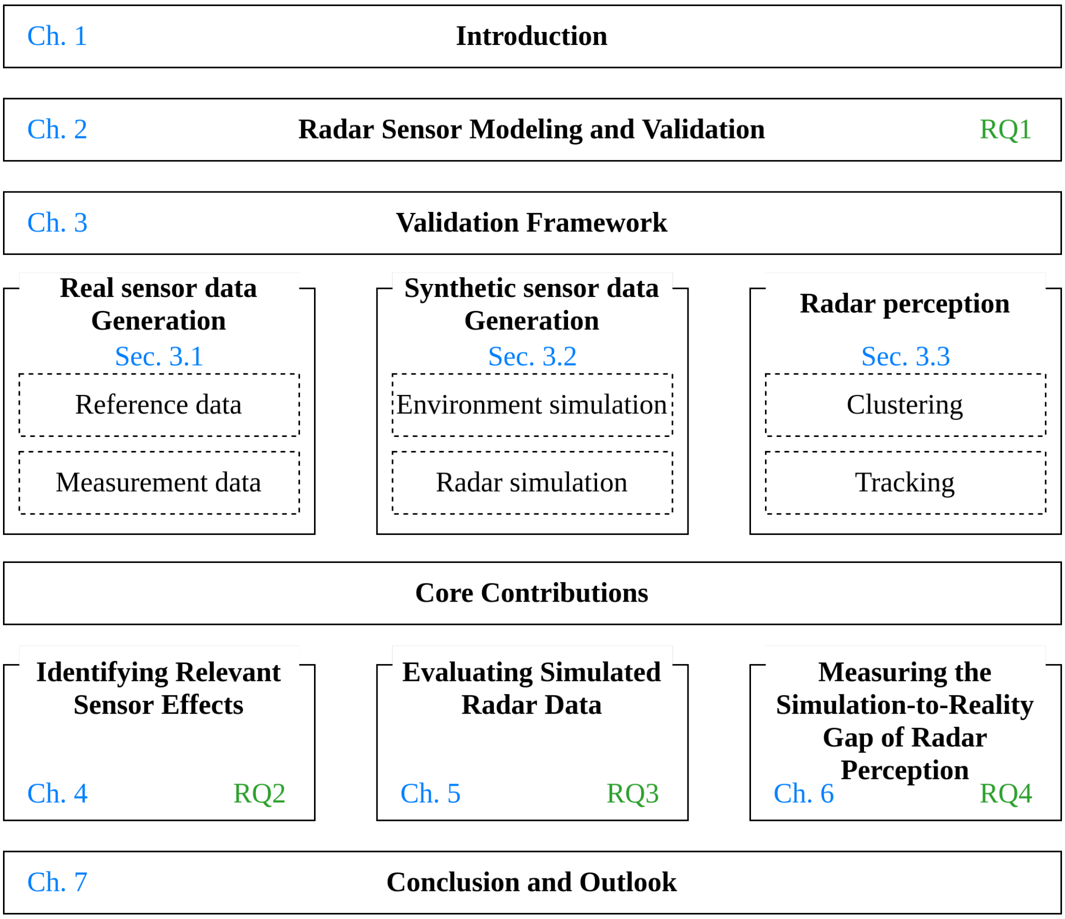

Figure 1.3: Thesis outline.

2 Radar Sensor Modeling and Validation

Contents

This chapter introduces the fundamental concepts, covers the technical foundations of the present dissertation, and summarizes existing work that is strongly related to the problems addressed in this work. First, the radar's physical measurement principles as well as an overview of automotive radar environment perception is provided. Subsequently, existing radar sensor modeling approaches in addition to model validation methods are introduced. Based on this, this chapter concludes by addressing the first RQ1: *What is missing in existing sensor model validation approaches?*

2.1 Automotive radar

Radar is an acronym for **ra**dio **d**etection **a**nd **r**anging. In 1904, Christian Huelsemayer patented a method for detecting distant metal objects through electromagnetic waves [39]. The foundation for this was laid by James C. Maxwell with his theory of electromagnetic waves and the proof of existence of these waves by Heinrich Hertz in 1886 [40]. Radar sensors have their origins in the military technology of World War II and also remained tied to military applications for a long time [41]. The first time a vehicle was equipped with a radar sensor was in 1998 to enable the adaptive cruise control (ACC) [42]. Since then, many other applications have been developed based on the detection of radar data, such as the automatic emergency brake [43] or the lane change assistance [44]. There are currently different frequency bands available for use in road traffic, while the 76–77 GHz and 24.0–24.25 GHz bands are mainly used. The former is used for long-range sensors, whereas the latter band focuses on mid range [45].

The major reason for the importance of radar in the automotive context is its physical principle that offers unique performance features at reasonable costs [46]. Contrary to video cameras and lidar sensors, radar is almost completely unaffected by adverse weather and light conditions, and it remains functional even in total darkness and snowfall, as illustrated in Figure 2.1. Furthermore, by exploiting the reflections of electromagnetic waves between vehicle underbody and road surfaces, even occluded objects can be detected [47]. Radar is a key technology for AD systems due to its robustness against all weather conditions and the direct acquisition of distance and velocity of targets via the Doppler effect [48].

In the following, the principles of radar sensors are briefly elaborated, which lay the foundation for environment perception and thus the scene understanding based on radar data.

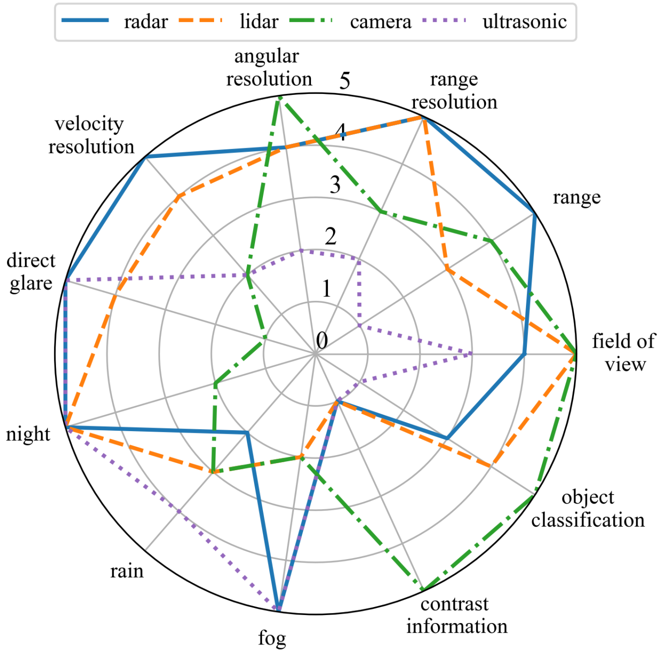

Figure 2.1: Comparison of main environment sensor modalities in the AD context (adapted from [3–5]). A larger number indicates a relatively better capability.

2.1.1 Fundamentals

Radar is an active sensor, since it operates by radiating electromagnetic energy and detecting the received echo from a reflected object (target). Based on the characteristic of the echo signal, information about the target can be inferred such as distance, radial velocity and angular location [35].

The transmission and reception of electromagnetic waves forms the functional basis for the operation of a radar. However, this creates only a carrier for information. In order to measure the distance, the information itself has to be modulated to this carrier during transmission and demodulated again when received. With modulation, the distance can be obtained from the time it takes for the radiated energy to travel to an object and back [49].

An electromagnetic wave will undergo a frequency shift due to the Doppler effect if the observer and transmitter move relative to each other [41]. This effect occurs when the radar beam is reflected by an object moving relative to the sensor. By exploiting this effect, the radial velocity of targets can be directly measured.

By using antenna systems, it is possible to furthermore determine both the yaw and pitch angle of the received signal. For example, mechanical scanning methods, where only one antenna is sufficient, or multi-antenna devices are used [42]. With sufficiently high resolution, additional features can be estimated such as the target's size, shape, and even the object type can be predicted [50].

In many applications, the radar sensor is designed to meet specific performance requirements, such as maximum distance, distance resolution, maximum velocity, velocity resolution, and covered angular field of view. A simple model to estimate the performance is the radar range equation (also simply known as radar equation) [35]. The radar equation expresses the relationship between the transmitted and received signal power, the distance to the reflected target, the characteristic of the reflected object, and the antenna properties [49]. Furthermore, it is not only useful for estimating the maximum range at which a particular radar can detect an object, but it can serve as a means for understanding the different factors affecting the performance and can be calculated as follows:

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 L_{sys}}. \tag{2.1}$$

In Equation 2.1, P_r and P_t denote the received and transmitted power respectively. The antenna properties are represented by the antenna gain G, which can be further divided into transmit and receive gain. λ is the wavelength and the size of the reflective object is governed by its radar cross section (RCS) σ. The RCS further describes the intensity of the echo received from an object exposed to an electromagnetic wave and depends on the object properties such as material, shape, size [51]. L_{sys} takes into account the overall system loss, including the atmospheric attenuation in fog or rain among others [49]. A detailed derivation of the Equation 2.1 can be found in [41].

2.1.2 Environment perception

In radar applications, the sensor measures the reflection from objects by using a single or an array of sensors. As mentioned before, by further processing these echos, information like range, velocity, and angle of the objects can be estimated. Radar sensors have already been used to develop various applications for driver assistance functions, for instance blind spot detection, cut-in warning, collision warning, and adaptive cruise control [49]. These functions are well established in the automotive industry today.

However, the requirements in the AD context are much higher. The challenges for sensors and associated perception algorithms for autonomous vehicles are enormous [14]. Compared to the conventional driver assistance systems, they must create a model of the whole static and dynamic environment around the self-driving vehicle in order to build an understanding of the prevailing situation [52]. The main applications for radar in the AD context are briefly elaborated in the following, which can be divided into two different domains, namely the detection of dynamic objects and the scene understanding of non-moving or static objects.

Detection of dynamic objects

The typical use case for radar sensors are the detection of relevant dynamic objects in the environment of the ego vehicle. Whereby, the question of relevance depends on the environment or operational domain under consideration. Whereas on the highway the number and types of objects is relatively limited, especially in the urban context a great number of different objects can be relevant. In addition to cars, these include for example cyclists, pedestrians, buses and trams. This means that radar has to provide the dimension and the motion state as well as class information of objects [52]. To meet this resulting high requirements, high-resolution radar sensors are used that can measure multiple reflections per target to be able to distinguish different types of targets [47, 50]. Hereby, it is not only necessary to detect and distinguish different objects, but also to track them over time [53].

The typical procedure for radar detection is illustrated in Figure 2.2, which is the tracking-by-detection approach. Starting with the measured reflections from an object, reflections with similar properties are clustered together in order to identify relevant objects [54]. For this clustering task, the previously described information of the measurements such as velocity, range and angular location are used and many approaches can be found in the literature [55–61]. Moreover, the size and dimension of a target can be estimated [62], which is depicted as a bounding box in Figure 2.2.

However, up to this point the nature of the detected object, its semantic class, remains unknown. A classification algorithm strives to address this problem by estimating whether a reflection or cluster belongs to a pedestrian, a car, a cyclist, a truck or another class. For this purpose, additional features such as the RCS of the reflection are used. Similar to computer vision and lidar-based perception, data-driven or machine or deep learning-based approaches outperform traditional rule-based algorithms in radar classification and various methods exist today [52, 63–69]. The semantic understanding of the surroundings of the driverless system can be used in many use cases: classification assisted tracking algorithms can outperform general tracking approaches [48, 70, 71], path planning has to take information about the object classes into account, and identifying traffic participants can reduce or even avoid critical situations early on [52].

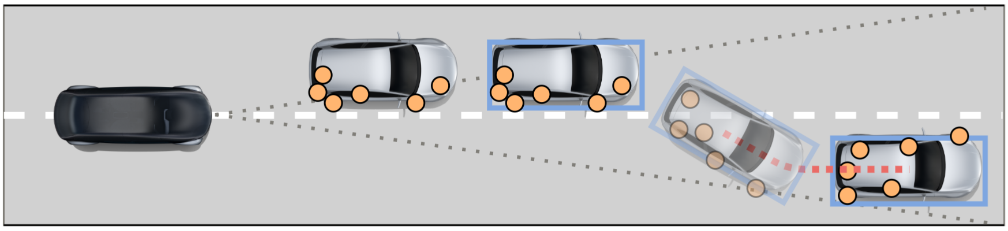

Figure 2.2: Typical Radar detection pipeline. From left to right: measurement of individual reflections, clustering of reflections that belong to an object, predicting the object size, tracking of object over time.

Multiple target tracking or multi-object tracking (MOT) describes the process of consecutively determining the number and states of multiple dynamic objects based on noisy sensor measurements [72]. Tracking is also a key technology in many other fields such as robotics, automation and surveillance. In autonomous driving, tracking algorithms play an essential role in fusion and behavior planning.

The objective of a tracking algorithm is to recursively determine the shape of the object in addition to its kinematic parameters, while the shape is usually unknown and can even vary over time [73]. Since the sensor measurements and thus the point cloud is often very sparse, it is nearly impossible to extract a shape only based on the reflections from one point in time [74]. However, by associating current object data to already known objects, a temporal data track is filtered from which the object state for the next time step can be predicted.

Scene understanding of static objects

In typical driver assistance systems, reflections from non-moving objects, the static world, were solely considered as obstacles [68]. But with the demand for more advanced assistance systems, a more sophisticated perception of the whole environment is needed and thus a semantic understanding of the static surroundings must also be built, especially for self-driving vehicles.

Although radar sensor data is more sparse in comparison to camera or lidar data, a lot of information can be inferred if it is accumulated over multiple points in time and radar physics is exploited [52]. In this respect, occupancy grid maps specifically designed for radar data can be used.

The scene understanding of static objects allow different important AD functions to be developed. For self-localization, the precise knowledge of the ego-car's own motion and position is important. Landmarks are often used in that regard, by recognizing distinctive and strongly reflective objects in the environment [47]. This information can also be used for the road course estimation. Prominent objects like reflector posts or guardrails can be used to draw conclusions on the most likely road course [48]. By additionally detecting objects like parking cars, it is possible to estimate the free drivable space [75, 76].

2.2 Radar simulation

In order to virtually develop and test perception functions, sensor models are needed to synthetically generate the sensor data. The following sections give an overview of the main elements of a sensor simulation and common modeling approaches.

2.2.1 Main components of a sensor simulation

Although the focus in this dissertation is on the sensor model itself, additional components are essential for a generation of synthetic data (see Figure 2.3). Since a sensor perceives its environment, the surroundings needs to be modeled in the simulation as well as the propagation of the energy transmitted and received by the sensor.

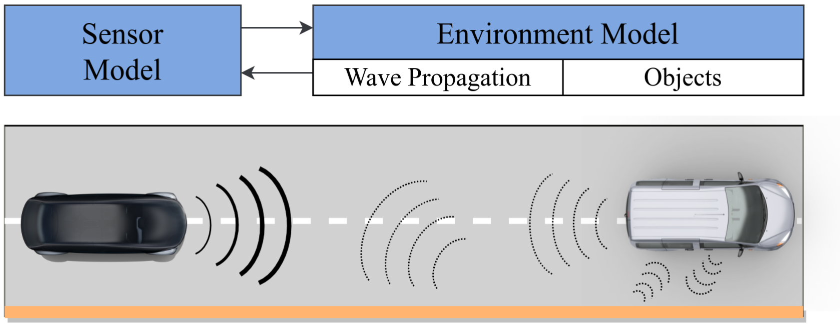

Figure 2.3: Components for a sensor system simulation. The guardrail is indicated in orange.

As illustrated in Figure 2.3, different effects such as multi-path propagation can occur in real situations. Hereby, the transmitted radio wave can be reflected from the ground or a guardrail before reaching the receiving antenna of the radar. This can cause multi-path interference, phase shifting of the signal, and ghost objects which impede the detection task and are therefore necessary to be considered in a simulation [77]. The task of accurately modeling such effects is highly complex, since it requires multiple models to be developed such as the transmitting and receiving procedure of the sensor, the propagation of the waves in the environment including effects such as the reflection from objects as well as the proper modeling of the objects in the environment and their physical properties, like the RCS. This is just one example of a physical phenomenon that shows the various components and their interactions required to accurately model this effect.

The main components for generating synthetic sensor data can be divided into the environment model, the radio wave propagation model, and the actual sensor model [78]. Various methods exist for modeling each component that together represent the principal effect chain of the sensor data generation and the typical modeling approaches are further elaborated in the following.

2.2.2 Common modeling approaches

There exist various approaches for modeling a radar sensor in the literature, but the nomenclature to define the different types of sensor models is not consistent. These definitions usually relate to the level of detail with which a sensor model approximates the real sensor. The models are often categorized by the used modeling approach [4,79,80], the input and output interface [81–83] or a combined version of both [29].

The common attributes used to describe the modeling approach include ideal, data-driven or stochastic, phenomenological, physical, white box, gray box, and black box. Furthermore, the interfaces denote the input and output format to and from the sensor simulation. The former can consist of object data like bounding boxes or the rendered environment, whereas the latter of raw data, detections (radar point cloud) or also object lists. In this context, raw data defines the data before a radar detection is generated. However, it should be noted that all mixed forms of the modeling approach as well as the input and output interfaces are possible. The used approach strongly depends on the desired target application.

In this work, the different existing radar simulations are categorized by their used modeling approach. The following modeling approaches are presented below, which are also shown in the Figure 2.4: ideal radar model, data-driven model, physical model.

Ideal radar model

The simplest form of a radar simulation is represented by an ideal sensor model, which approximates the real sensor without errors. It considers merely the optical field of view without measurement errors, i.e. objects are detected any time they are within the sensor's FOV. Therefore, radar specific physical effects are neglected and these type of models are also known as ground truth models. Since they are relatively simple and fast to compute, they are especially suitable for early testing of AD algorithms in either ideal conditions or under the assumption that sensor errors are neglectable [18].

Data-driven model

Data-driven sensor models rely on collected data from test drives in the real world. They strive to address the modeling task by learning from the recorded sensor data, since it inherently holds information about the perceived surroundings. Thus, this approach eliminates both the need to model the radar effects in detail and to have all the details about the environment. For this reason, these type of models are also termed black box models [78]. Data-driven models can exhibit essential radar specific characteristics while remaining real-time capable [33]. However, the drawback of these models is that they utterly depend on the available training data, which in most cases has been recorded on a proving ground in order to simplify the ground truth determination. Consequently, the measured scenarios are usually restricted in regard to the scenery of the environment. The generalization to more complex environments is therefore difficult, because radar data are prone to vary strongly depending on changes in the environment.

Physical model

Even though the electromagnetic radiation is governed by Maxwell's equations, it is generally not feasible to compute an analytical solution in a realistic environment [84]. That is why alternative methods are needed to accurately model a radar sensor.

Time-domain electromagnetic simulation techniques such as the finite integration technique [85], the finite element method [86], the method of moments [87], and the finite-difference time-domain method [88] are based on the spatiotemporal discretization of Maxwell's equations. These techniques can be used for an in-depth simulation of the electromagnetic phenomena observed in radar systems [89]. Although they are very accurate in principle, assuming that the analysis space is large relative to the wavelength, numerical methods based on the discretization of integral or differential equations face the challenge of extremely large memory needs and very slow computational speed [84]. Moreover, if the simulation frequency is around 77 GHz, which is within the range of automotive radar senors, it is not practicable to simulate the space enclosing an entire vehicle as a consequence of an exorbitantly long computation time [90].

A broadly utilized approach to overcome this problem is the ray tracing approach, which is based on the geometric optics diffraction theory. Hereby, the radar waves are considered as a bundle of rays [91]. As a result, ray tracing approaches enable to simulate various radar effects including reflection, multi-path propagation, or diffraction [84]. Despite the fact that this approach requires less computation power than numerical methods, they are still computationally expensive, which limits their use in real-time applications like hardware in the loop setups [18]. Besides the limitation in execution time, these methods demand a high level of detail for the simulation of the environment. Particularly material and geometry properties of all surrounding objects are a prerequisite for a high fidelity radio wave propagation simulation [92].

2.2.3 Radar simulation use cases

The different modules of the radar processing chain as well as the main radar modeling approaches are depicted in Figure 2.4.

As described before, the radar sensor emits radio waves which interact with the surroundings before the reflections are received at the sensor again. The echo signal is then further processed in several steps, which include filtering, the signal conversion from analogue to digital, fast Fourier transformation and thresholding before radar detection is identified [35]. However, these steps are not described in-depth here, since it is not the focus of this work. The output of radar depends on the subsequent perception module. For object detection usually radar detection are processed [65]. However, spectral data can also serve as the basis for algorithms such as object detection [93]. Low resolution radar sensors in particular often output measurements in form of radar object lists, which can be directly processed by downstream modules such as object level fusion or behavior planning. Additionally, typical perception modules are illustrated in the Figure 2.4, which consists of clustering, classification and object tracking.

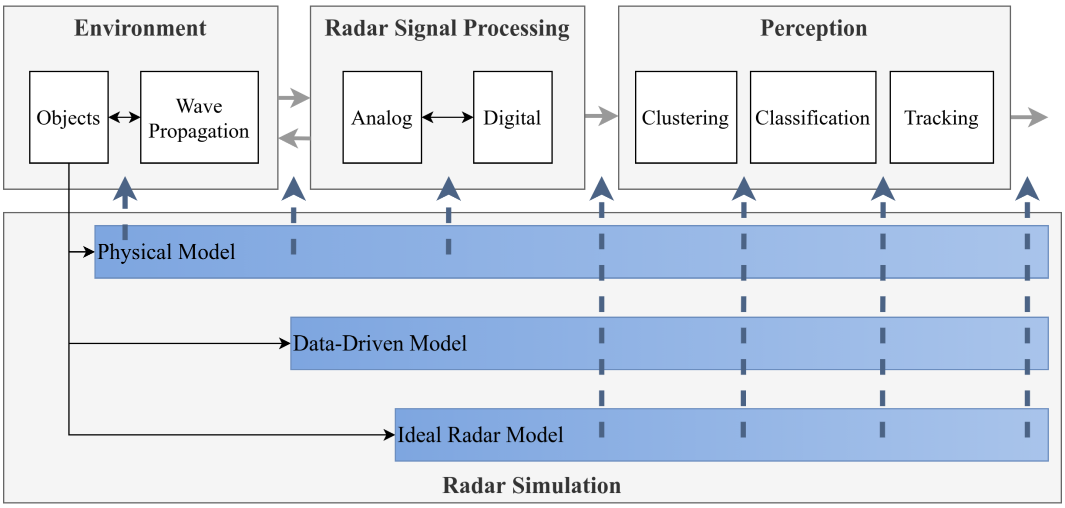

Figure 2.4: Sensor simulation use cases and common modeling approaches.

A sensor simulation strives to translate ground truth information, as available from simulation, to sensor data. Depending on the intended use case, different output formats can be generated from a sensor simulation as indicated exemplary in the Figure 2.4. It should be emphasized that variations exist in the illustrated output interfaces of the different sensor model types. Bypassing components, i.e. using their respective virtual model, allows a certain abstraction to be introduced. Thus, individual modules can be designed and tested in an early phase, reducing the overall development time.

As mentioned above, the result of a tracking algorithm can be simulated directly, which then serves as input for a subsequent behavior planning algorithm. Hereby, the goal is to test the behavior planning individually by generating radar specific object lists instead of radar detections [94]. Errors can be inserted intentionally, such as an object suddenly being lost, i.e. no longer being detected, in order to see how well the planner reacts to this. Therefore, it is possible to test module specific critical scenarios.

This procedure can also be applied to other modules in the AD stack. The processing chain up to the processing of the data by the perception module can be modeled in simulation by generating sensor data to test individual components in the perception system. For example, the ground truth information in the simulation is used to generate radar data specifically for a classification algorithm [30]. In contrast to real test drive measurements, it is relatively simple to determine the corresponding object class of a certain radar detection in simulation. However, this is a very complex task in real test drives [95], among other things due to the multi-path reflections. This simplifies the simulation-based evaluation of the perception algorithm and thus can speed up its development. Another advantage of the simulation relates to the generation of critical scenarios that are rarely measured in real drives or are too dangerous and expensive to test on a proving ground. These include, for example, highly dynamic driving maneuvers, which are of specific interest for an evaluation of object tracking approaches.

There exist different approaches to develop and test a sensor in a virtual environment. As mentioned in the previous section, the sensor simulation consists of two main components: the sensor model itself and the environment model. Consequently, either only the environment of the sensor is represented in the simulation or the sensor is modeled as well. The former using the real radar sensor hardware is referred to as the sensor stimulation, while the latter is considered as the sensor simulation, which is the focus of this work. Sensor stimulation describes the process of manipulating an entity in such a way that its state matches an environmental scenario, even though it is not physically in such a scenario, but for instance on a test bench [26]. In this way, the real radar sensor is stimulated by modeling only the environment around the radar in simulation. This allows the whole signal processing of the sensor to be tested. In order to test specific parts of the radar signal processing chain, modules of the sensor are approximated by models in simulation. This is useful, for example, if the radar sensor does not yet exist.

However, the more components are replaced by their virtual counterparts, the more complex it will be to generate accurate data, since each model introduces additional uncertainties and has to be assessed in terms of its capability to approximate the corresponding real component. Furthermore, the required computation time for a high-resolution and accurate simulation also increases [36]. This shows that there exists a certain trade-off between accuracy and complexity or required computing time, which has to be solved for the respective application purpose, because the requirements on the fidelity of a radar simulation can vary substantially depending on the intended use [38].

2.3 Validation of a sensor simulation

In the sense of the well-known aphorism of George Box "All models are wrong" [19] – which is often expanded with "..., but some are useful." – it is not possible to create a simulation model that generates valid data under all conditions. Accordingly, a certain deviation between real and synthetically generated sensor data can be assumed to exist and the question arises what accuracy is sufficient for which use case. At the same time, especially high fidelity radar simulations face the challenge of large demands regarding the simulation execution time [36]. This is why the appropriate trade-off between computation time and model realism must be found.

When virtually testing AD functions, valid sensor data is a prerequisite for meaningful results. Therefore, in addition to the development of sensor simulations, the models employed need to be validated in order to rely on simulation-based tests. Oberkampf et al. define model validation as the assessment of the error due to the assumptions and approximations made in the formulation of the model [17]. In consequence, it is crucial to meticulously test the sensor model to validate the model assumptions and simplifications with respect to the intended use of the synthetic data. However, even though it is relatively straightforward to measure the computation time of a simulation run, the error evaluation or fidelity estimation of the synthetically generated sensor data is a substantially challenging problem.

Since, as elaborated in the previous sections, a sensor simulation consists of the different sub-components which each has to be modeled, the validation of such a sensor simulation is of high complexity. In comparison to video data generated by cameras, sensor data from lidar and in particular radar data are significantly more sparse, which makes an evaluation based on visual matching difficult. Radar data is difficult to interpret due to the presence of noise, ambiguities, measurement artifacts caused by multi-path propagation, and other influences. This complicates the comparison between simulated and real radar sensor data.

In the following, the existing sensor simulation approaches are elaborated and grouped by their validation scope. Depending on the level of detail of the evaluation, four different categories are defined. The first two categories include a direct or explicit assessment of the radar data. The first one considers merely the simulation domain, whereas the second category additionally compares the simulated data with real radar data as a reference. In the third category, an indirect or implicit evaluation of the synthetic data is performed. Hereby, a target application is used that processes the radar data with the aim of investigating how the result of an application differs when fed with real and simulated data. The last class evaluates both the synthetically generated radar data as well as the result of a subsequent application. The existing approaches are listed in Table 2.1 and Table 2.2.

2.3.1 Explicit evaluation: simulation domain

The simplest method is to assess the generated sensor data only in the simulation domain. Stolz and Nestlinger [96] and Muckenhuber et al. [82] propose generic object-based sensor models. The sensor models convert an object list into a sensor specific object list. However, both approaches provide concepts for simple generic sensor models, but do not perform any evaluation of the model fidelity.

Chipengo et al. [34,92,97] are using ray tracing to model a high fidelity physical radar simulation. In order to verify the implemented radar phenomena, different test case scenarios are defined which use three steel plates as obstacles with a known reflectivity characteristic, i.e. RCS, to assess the resolutions in different dimensions [34]. The three plates are placed at a range of 15 meters relative to the sensor, while the radar itself is stationary. Since all obstacles are located at the same range, only one target should be detected. However, the estimated Doppler velocities of the targets are expected to vary, since two plates fall into the same Doppler bin. Considering the third dimension, the azimuth angle, the estimation of the direction of arrival should resolve all three obstacles in the angle. The simulation results are compared in a qualitative way concerning the assumed hypotheses.

The simulated data is often compared with expected results in a qualitative way, i.e. by visual matching. The synthetically generated data can thus be quickly inspected. However, the reliability of such a method is highly dependent on the expertise of the examiner and is therefore prone to subjective findings and erroneous assessments. Furthermore, this makes the reproducibility as well as the scalability of the evaluation difficult, which are essential advantages of simulation-based testing.

2.3.2 Explicit evaluation: simulation and real domain

In this section, the existing approaches are presented that evaluate the synthetically generated sensor data not only in simulation, but also in comparison with real sensor measurements. Nevertheless, there exists no generally accepted evaluation criteria or requirements for assessing the reliability of simulated radar data [18]. One problem with the analysis of radar data is that there is no standard interface like lidar with point clouds or camera with RGB images. This complicates both the evaluation itself as well as the comparison between different evaluation approaches. The data type used usually depends on the downstream algorithm which processes the sensor data. Spectral data [71, 75] or radar detections [62, 65] often serve as the basis for environmental understanding. For this reason, several distinct methods have been presented in the literature, which can be distinguished by the considered model output interface.

Table 2.1: Overview of existing radar simulations and sensor model validation approaches.

Validation scope	Authors	Validation method
I. Explicit evaluation: simulation domain	Stolz [96]	none
	Muckenhuber [82]	none
	Chipengo [34, 92, 97]	qualitative assessment
	Haider [98]	qualitative; one dynamic scenario
	Dudek [99–101]	qualitative; two scenarios
	Buehren [102]	qualitative; one static scenario
	Thieling [103]	qualitative; two static scenario
	Hanke [104]	qualitative; two static scenario
	Gubelli [105]	qualitative; one scenario
	Azodi [106]	qualitative; one scenario
	Sturm [107]	qualitative; one scenario
	Kim [108]	qualitative; one scenario
	Ouza [109]	qualitative; one scenario
II. Explicit evaluation: simulation and real domain	Machida [89]	qualitative; and via correlation
	Holder [110]	qualitative; two scenarios
	Schuler [111]	qualitative; static in the laboratory
	Holder [18]	qualitative; specific phenomena
	Hirsenkorn [83, 112, 113]	qualitative; one static scenario
	Hirsenkorn [114]	qualitative; lidar
	Wheeler [33]	qualitative; three static scenarios
	Bernstein [94]	qualitative; one scenario
	Eder [115]	qualitative; one dynamic scenario
	Cao [78]	qualitative; static scenarios
	Roth [116]	qualitative; two dynamic scenarios
	Slavik [79]	qualitative; one scenario
	Nathaniel [117]	qualitative; one scenario
	Martowicz [118]	qualitative; one scenario
	Li [119]	qualitative;
	Hanke [120]	quantitative; static scenario; lidar
	Deep [121]	quantitative
	Eder [122]	quantitative; one dynamic scenario

In the approach by Schuler et al. [111], a model for complex targets for radar simulation is developed. A technique is presented to generate a simplified RCS model of a vehicle with a limited number of virtual scattering centers. The work is furthermore based on ray tracings simulations. The model is verified by comparing the generated data with measurement data recorded in a laboratory environment, a radar chamber. The evaluation is performed in a qualitative manner.

Hirsenkorn et al. [83] perform tests on a small road besides fields with a stationary target vehicle at three different distances. The sensor as well as the target object remain static throughout the experiments. By comparing the real and simulated received signal power, the elementary functionality of the radar simulation is verified. Holder et al. [18] analyze their developed sensor model regarding specific radar phenomena such as occlusion, separability, and the sensitivity of RCS to the aspect angle. Different scenarios are defined to specifically examine these effects whether the simulation outcomes match the expectations.

In contrast to these qualitative approaches, Hanke et al. [120] use measurable criteria for the model validation, including correlation coefficients and occupancy grids. However, they did not examine radar data in particular, but point clouds from lidar sensors. Although lidar point clouds are similar to radar detections, they generally have a denser spatial distribution. In the simulation model by Eder et al. [122], a hybrid model is proposed combining ray tracing with a data-driven approach. Furthermore, a quantitative evaluation is conducted to assess the simulated data with recorded measurements. Three different criteria are identified for that: the mean distance of all radar detections to the enclosing bounding box, the number of detections and their mean deviation, and the Kullback-Leibler divergence.

2.3.3 Implicit evaluation

Instead of assessing the generated sensor data directly, it is possible to evaluate the prediction results of a subsequent algorithm which processes the data. In the automotive context, these algorithms include object detection or classification approaches. The typical validation procedure is to feed the algorithm simulated data in addition to real measurements and compare the differences in prediction. In this way, implications for the radar model can be implicitly derived without directly evaluating the sensor data, which are difficult to interpret. Another advantage is that the evaluations of the algorithms employed can be used, which are usually more mature because they have a longer history and there are many approaches in the literature.

Jasinski [123] introduces a validation scheme to measure the reliability of radar models. By calculating the intersection over union (IoU), the prediction of a tracking method is evaluated. Reway et al. [124] propose a test method for measuring the simulation-to-reality gap of camera-based object detection techniques for AD. The test experiments are conducted on a proving ground and represented in a virtual environment. Additionally, different weather conditions are considered: day, night, rain and fog. With the purpose to measure the simulation gap, different performance metrics are computed across the real and virtual domains. These include typical object detection metrics like precision, recall, and object tracking accuracy.

2.3.4 Explicit and implicit evaluation

Up to this point, the presented approaches have either explicitly or implicitly evaluated a sensor simulation according to its reliability. Despite that a direct comparison is essential, it is alone not sufficient for sensor simulation validation [16]. Although an ideal sensor model might lack accuracy in a direct comparison, the results from a downstream algorithm can still show a large consensus. Accordingly, sensor simulations in the context of AD should not be treated as stand-alone applications. The target application which processes the sensor data has to be considered [80]. Nevertheless, particularly the example of an ideal sensor model shows that it is important to consider both evaluation dimensions. Otherwise, a pure implicit evaluation could lead to the conclusion that an ideal model has a higher fidelity than a more sophisticated physical model. Since it is easier to process ideal sensor data for object detection, for example, the gap between reality and simulation would be smaller with the ideal model. However, this implication would be incorrect.

To the best of the author's knowledge, there are no approaches that consider both evaluation dimensions for radar models. For lidar simulations, Scharmann et al. [16] and Rosenberger et al. [125] introduce similar methods. They examine lidar point clouds as well as the result of a subsequent algorithm. For the former metrics such as point clouds distances and occupancy grids are used, while for the latter the optimal sub-pattern assignment (OSPA) metric [126] is computed.

Table 2.2: Overview of existing radar simulations and sensor model validation approaches (continued).

Validation scope	Authors	Validation method
III. Implicit evaluation	Holder [127]	qualitative; two scenarios
	Jasinski [123]	conceptual quantitative evaluation
	Sligar [30]	none
	Reway [124]	quantitative; one scenario; camera
	Hartstern [31]	none
	Schouten [128]	qualitative; simulation only
	Hoeber [129]	qualitative; camera
	Ponn [4]	none
IV. Explicit and implicit evaluation	Rosenberger [80]	validation concept
	Rosenberger [125]	quantitative; two scenarios; lidar
	Schaermann [16]	quantitative; lidar

2.4 Discussion

In light of the presented state of the art in validation of radar simulation, this section discusses the existing research gaps. The first research question of the present dissertation (RQ1: *What is missing in existing sensor model validation approaches?*) is addressed in this chapter. Based on this, the main problems that need to be solved for the validation of a sensor model are deduced and the scope of this work is further presented.

2.4.1 What is missing in existing sensor model validation approaches?

The existing radar simulations as well as the model validation approaches are listed in Table 2.1 and Table 2.2. They are structured regarding their validation scope and are further analyzed in Figure 2.5 in terms of the applied validation method. The respective validation methods are used as meta-data to grasp the level of detail of the assessments. Hereby, the methods are distinguished by the following classes: no validation (none), a qualitative assessment (qualitative), and a quantitative evaluation of the simulated sensor data (quantitative).

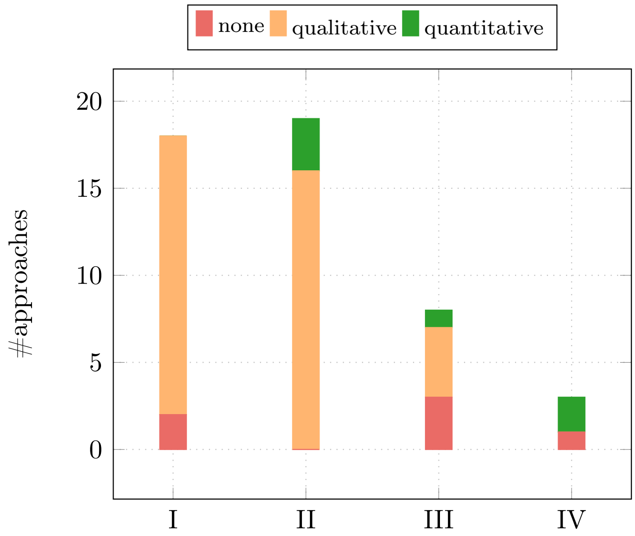

Figure 2.5: Approaches found in the literature for modeling radar systems and validating sensor simulations, categorized by the evaluation method used.

From Figure 2.5, it is evident that most approaches focus on the assessment of the radar data directly, the explicit evaluation. However, merely three authors can back up an evaluation of their model with objective numbers. The majority of the listed proposals involve a qualitative assessment, often by visually comparing the generated results with expectations. A few also do not perform any evaluation of the simulated data at all. This can be especially observed for the validation scope III and IV, the implicit evaluation and the explicit and implicit evaluation.

To this day, no systematic approach could be identified that provides an objective and quantitative method for the validation of a radar simulation for AD. Most methods use only a qualitative model assessment of the sensor data. There are, however, some approaches that include both the sensor data directly and the downstream application in the evaluation. However, these are mainly related to lidar sensors, which raises the question whether these approaches are also suitable for radar data, since radar data have different characteristics, such as much more sparse and stochastic data.

2.4.2 Formulation of research problems

The general procedure of a sensor simulation validation is illustrated in Figure 2.6. At first, the test scenarios are defined which serve as the basis for conducting real test drives to record the sensor measurements as well as for the corresponding tests in the simulation. With the generation of both the synthetic radar data and the radar measurements, the explicit evaluation can be performed. By furthermore including a target application, the radar model is implicitly assessed. Hereby, the target application is typically represented by a radar perception algorithm such as clustering, object classification, or tracking. In the following, the remaining research problems in the validation of a radar simulation are derived and formulated.

RQ2: *Which features of the radar simulation are relevant for a downstream application?*
As previously described, a radar simulation consists of different modules including an environment model, a wave propagation model, and the sensor model itself. Each model approximates various radar characteristics in order to generate sensor data in a simulation. Given that a high fidelity radar simulation often face the challenge of demanding computation times, the question arises whether every effect must be modeled in a high level of detail or whether certain properties can be simplified. Since it can further be assumed that this consideration depends on the intended use of the sensor data, a method is needed to identify relevant effects for a particular use case.

RQ3: *How to determine the degree to which the radar simulation and experimental measurements concur?*
In comparison to camera and lidar data, radar measurements are much more sparse and of stochastic nature. This one of the reasons why special expert knowledge is required for the visual interpretation of radar sensor data. Thus, the reliability of such a qualitative evaluation approach is highly dependent on the expertise of the assessor and is therefore prone to subjective results and incorrect assessments. However, as seen in Figure 2.5, most of the existing approaches evaluate the sensor model in a qualitative way. Another disadvantage of a manual and qualitative assessment of simulated sensor data is the lack of scalability, which is one of the main advantages of simulation-based

testing. This leads to the conclusion that a quantitative evaluation methodology is needed that uses metrics which are in particular suitable to evaluate radar data.

RQ4: *How to measure the overall simulation-to-reality gap considering a target application?*
Although an evaluation of the simulated sensor data is essential, it alone is not sufficient for a complete validation of a sensor simulation. For example, an ideal sensor model might lack accuracy in an explicit evaluation, but the outcome from a subsequent algorithm can still show a high consensus. The intended use of the simulated sensor data should therefore be considered. Moreover, there exists no method that combines the results of both evaluation domains, explicit and implicit. Consequently, in addition to a quantitative evaluation of the synthetically generated radar data, a validation method should also include the intended use case. In order to scale the simulation and to be able to evaluate multiple scenarios quickly, a validation method should enable a holistic evaluation on the one hand, but should also allow an in-depth assessment of specific situations on the other hand.

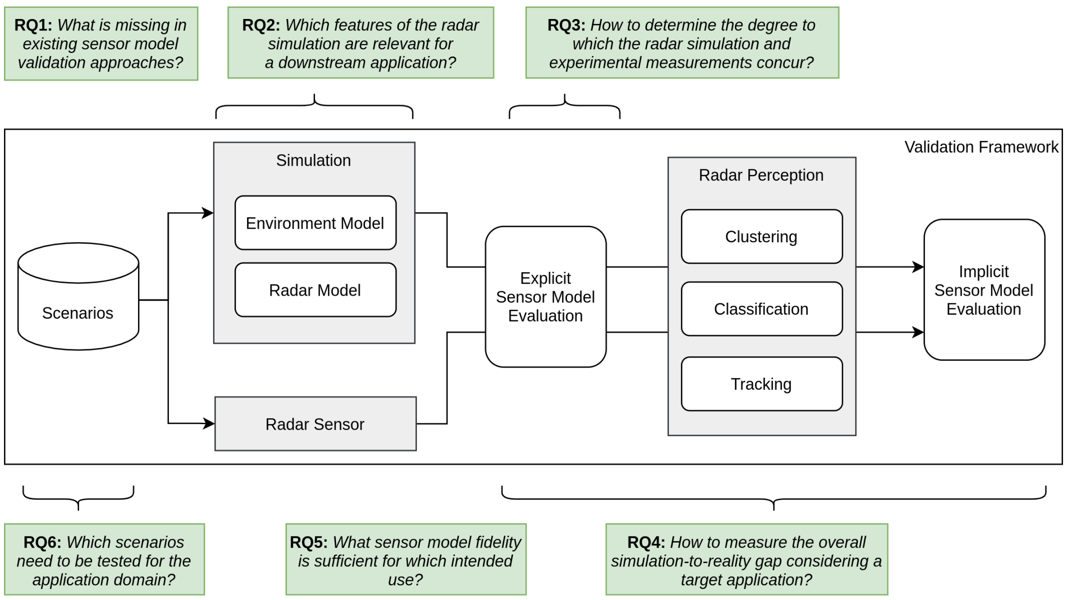

Figure 2.6: Existing research gaps in the validation of a sensor simulation.

RQ5: *What sensor model fidelity is sufficient for which intended use?*

Besides the problem of estimating the sensor model fidelity, i.e. the measurement of the discrepancy between real and simulated test, the question arises what fidelity is needed for which use case. Due to demanding requirements in terms of execution speed, furthermore, the sufficient degree of realism must be found. Since a certain simulation-to-reality gap can be assumed to exist, the problem remains what error is an acceptable disparity. In spite of that, there currently exist no method to solve this problem.

RQ6: *Which scenarios need to be tested for the application domain?*
When validating a sensor simulation, it is reasonable to use real sensor data for comparison. However, since especially the ground truth generation of the measurement runs is very difficult in complex scenarios, such as urban environments with a lot of different traffic participants, the measurement runs are conducted on a simplified test site. As a consequence, since the reference data such as sensor measurements are then also recorded on the proving ground, the problem of transferability of results emerges. A methodology is needed to overcome this gap between validation and the application domain.

2.4.3 Scope of work

Simulation-based tests do not remove the need for large quantities of tests in the real world. Rather, the objective of this work is to investigate how much a simulation can be trusted in order to be able to determine the ideal ratio between real and virtual tests.

In particular, the present dissertation focuses on the development of models and methods to answer the first four research questions (RQ1-RQ4). This provides the foundation for the remaining problems (RQ5-RQ6) and thus leading to a complete validation of a sensor simulation for AD. These models and methods can then be leveraged to reduce the need for future real-world testing.

3 Validation Framework

Contents

This chapter is structured into three main sections and introduces the validation framework for generating both real (Section 3.1) and synthetic radar data (Section 3.2) in addition to the developed radar perception algorithms (Section 3.3). The framework is illustrated in Figure 3.1 and serves as the basis for the investigations in the following chapters.

Since sensor modeling itself is not the focus of this work, the developed framework serves as a modular basis for further research. Furthermore, it is not exclusively designed for radar models and perception algorithms, but allows further abstractions to other sensor modalities and algorithms. For example, data from a lidar or camera sensor can also be examined and evaluated.

3.1 Real sensor data generation

This section is divided into two parts: the reference data and the measurement data generation. The former describes the state data of the ego vehicle and the

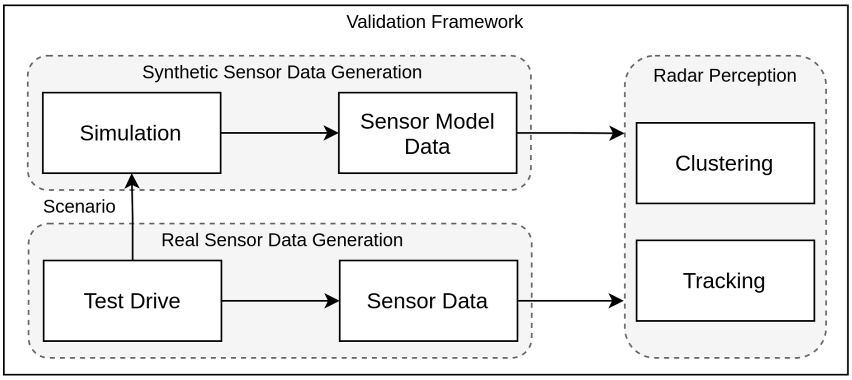

Figure 3.1: Validation framework overview.

target vehicles. The state data includes the positions, orientations and velocities of the objects and is used as ground truth, whereas the measurement data represents the data recorded from the radar sensor. The generation procedure of both is elaborated in the following.

3.1.1 Reference data

The generation of real sensor measurements as a reference for comparison is an crucial element for the sensor model validation. For the reason that a simulation approximates real effects and thus has certain errors, it is necessary to find out how large these errors are. Data from the system to be modeled can be used as a reference to investigate the existing deviations. This means that a real test drive is reproduced in the simulation to compare the synthetically generated sensor data with real measured data.

In addition to the actual radar sensor recordings, further measurements are necessary for the simulation of tests. The so-called ground truth data of the environment is needed, which refers to the true or correct value of a quantity. This supplementary information serves as a reference and can be obtained by using a dedicated sensor system, which, however, introduces additional uncertainties. In contrast to the determination of ground truth data from real recordings, there is usually no residual uncertainty in the case of a simulation. This is due to the fact that no reference sensors are needed for ground truth measurement in simulation, since the data is directly accessible. The ground truth data of the surroundings is primarily used to simulate the tests in simulation. The information required can vary widely depending on the modeling approach used. Generally, the state information of objects in the vicinity is used to represent the test scenario. This data includes quantities such as position, orientation, velocity, object class, object size or the reflection characteristic. For modeling the wave propagation as well as the environment, additional quantities can be utilized including for example temperature or atmospheric attenuation values to consider the influence of weather.

Figure 3.2: Proving ground in Immendingen, Germany [6].

However, the complexity of the ground truth data determination of a recording is highly related to the variety of the prevailing environment. According to this, the ground truth estimation in an urban environment is particularly complex, as the states of many different objects, such as pedestrians, cyclists, and cars, have to be determined. By simplifying the testing area to a dedicated proving ground, certain traits can be achieved, including reproducibility, safety, controllability and the relatively simple measurement of the ground truth information previously described [22].

In this work, the test drives are conducted on a specific proving ground from Daimler AG in Immendingen Germany [6], which is depicted in Figure 3.2. On this test site various scenarios can be tested, ranging from simple maneuvers like overtaking to more complex situations including the encounter of multiple vehicles at a junction. Additionally, different road signs are located on the site with the purpose to test for example a traffic light detection. Apart from the street, the rest of the surrounding terrain is relatively simple and lacks typical infrastructure elements such as sidewalks or buildings [130].

In order to determine the correct position and orientation of the surrounding objects, a dedicated reference system is used. The true state of the objects in the environment are referred to as reference data in the remainder of this dissertation. Since the reference data serve as the basis for the simulation, a high degree of accuracy is crucial. This can be achieved by using a differential global positioning system (DGPS) with an inertial measurement unit (IMU). Each object of a scenario in this work is equipped with this reference system for a precise acquisition of position, orientation and velocity.

3.1.2 Measurement data

Besides the information of all the surrounding objects, the actual sensor measurements are required for the validation of a sensor simulation. Since the sensors used in this work are proprietary, their technical properties are not described in detail. For the reason that the focus is on model validation rather than the development of sensor simulations and this methodology is independent of the specific sensor used, this limitation is reasonable. In the remainder of this work, measurement data is referring to the real radar sensor detections. Furthermore, the sensor recordings are synchronized with the reference data in order to enable a continuous evaluation between real and synthetically generated data.

3.2 Synthetic sensor data generation

In this thesis, the generation of synthetic sensor data can be divided in two main steps: the environment simulation and the actual sensor simulation. The former refers to the modeling of the environment including infrastructure and traffic participants as well as the control and movement of the objects based on the defined test scenario. Whereas the latter denotes the generation of a virtual scene of the surroundings as perceived by the radar. The implementation of both modules is elaborated in the following. In the course of this, three different typical radar models are introduced, which are schematically illustrated in Figure 3.3: an ideal sensor model, a data-driven or stochastic sensor model, and a ray tracing-based model.

3.2.1 Environment simulation

A simulator or simulation platform serves as a tool to create a virtual environment as well as to control the movements of the objects in the virtual world. It also calculates the optics, renders the visual output, and calculates the object physics, including collisions. Objects include, for example, vehicles, buildings, traffic signs, and the map.

There are various simulation tools in the automotive context such as Vires Virtual Test Drive [131], IPG CarMaker [132], LG SVL [133]. However, a lot of them are specifically designed for particular use cases. In this thesis, the simulator **Car** **L**earning to **A**ct (CARLA) [134] is used as the environment simulation tool, because it is open-source, has a variety of assets built in, and also already provide implementations of basic sensors.

Based on the recorded reference data, the scenario is extracted and serves as a basis for simulation-based tests. Here, the ground truth states of the road users are utilized to place their virtual counterparts in the simulation. Moreover, the testing site was virtually reproduced in the simulation and perceived by the radar model to generate simulated radar detections.

3.2.2 Ideal radar model (IRM)

An ideal sensor model is an ideal model that imitates the idealized behavior of the radar sensor. Such models simply take into account the geometric field of view (FOV) without measurement errors or sensor-specific physical effects, i.e. objects are always detectable if they are within the sensor's sensing range. Due to their simplicity and fast performance, they are suitable for early tests of perception algorithms either under ideal conditions or under the assumption that sensor errors are inconsiderable [18].

In order to assure a consistent sensor model output format across the sensor models used in the form of a radar point cloud, the radar detections of the IRM are distributed along the shell of a detected vehicle according to a simplified scattering center model [111, 135] (see Figure 3.3).

3.2.3 Data-driven radar model (DDM)

Stochastic or data-driven sensor models seek to approximate radar results by learning certain features from real sensor measurements without having to

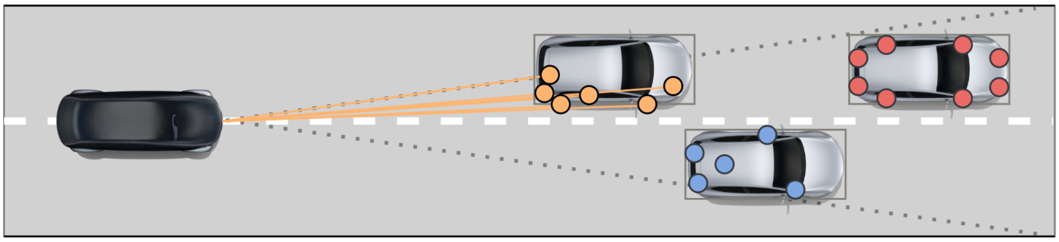

Figure 3.3: A schematic illustration of the radar simulations developed: ideal radar model (●), data-driven model (●), ray tracing-based model (●).

explicitly model the radar phenomena. Since these measurements inherently contain information about the observed surroundings, the need for extensive details about the environment, such as the material properties of objects, is omitted.

As a data-driven or stochastic model, a modified implementation of an available radar sensor measurement model designed and trained by Scheel et al. [136]. Here, the measurement model was learned from real datasets using variational Gaussian mixtures. It is used to process multiple radar detections of an object to perform measurement-to-object mappings in extended object tracking. In this thesis, it is build on this approach in reverse by predicting the radar measurements from a given object state. Moreover, samples are drawn from the learned marginal density, which is conditioned on the aspect angle of a detected vehicle, with the goal to generate radar detections. The number of samples is determined by a distribution that depends on the radial distance defined by the real radar measurements.

3.2.4 Ray tracing-based radar model (RTM)

Radar is an electromagnetic system for detecting and locating reflective objects and works by radiating energy into space and detecting the echo signal reflected from a target. The propagation of radio waves is modeled using a ray tracing or rather a ray casting approach based on the diffraction theory of geometrical optics [91]. In this concept, the radio waves are approximated as a bundle of rays, and each ray that hits an object in the sensor's FOV results in a reflection.

Furthermore, the radar range equation, which is often simply referred to as the radar equation, relates the range of a radar to the effects of the antenna, transmitter, environment, target, and the receiver. Accordingly, it is not only useful for estimating the maximum range at which a radar can still detect an object, but it can also serve as a means toward understanding the various factors that contribute to the radar performance [41]. The radar equation is defined as follows:

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 L_{sys}} \tag{3.1}$$

For this very reason, the radar equation is used in this work to estimate the signal power received P_r. A detailed derivation of this formula can be found in the work of Skolnik [35]. The symbols used are listed in Table 3.1 along with their units and a brief description.

Table 3.1: List of symbols, their units and description.

Symbol	Unit	Description
B_n	$1/s$	noise bandwidth
G	-	transmitting & receiving antenna gain
k_B	J/K	Boltzmann constant
F_n	-	noise figure
L_{sys}	-	overall system loss
P_t	W	transmitting power
P_n	W	noise power
P_r	W	receiving power
R	m	radial distance
SNR	$-$	signal-to-noise ratio
T_0	K	standard temperature
λ	m	wavelength of transmitted signal
σ	m^2	radar cross-section

The radar cross-section σ is assumed to be ideal depending on the aspect angle to the object and is illustrated in Figure 3.4. In this work, the focus is on vehicles as objects and the corresponding radar cross-section (RCS) values are derived from the work of Abadpour et al. [137] and Matsunami et al. [138].

A simplified generic antenna gain G is employed for both transmitting and receiving with a maximum gain of 20 dB and a side lobe suppression of -13 dB, derived from Gamba [49]. In addition, the antenna diagram is modeled

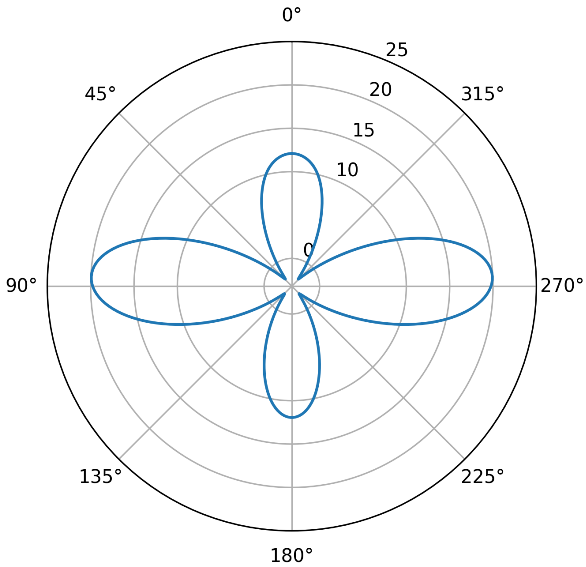

Figure 3.4: Simplified radar cross-section.

with a simple sinc filter resulting in Figure 3.5).

In addition, the ability of a radar to detect an echo signal is limited by the noise that is always present and occupies the same portion of the frequency spectrum as the radio signal. This is denoted by the noise power P_n, which depends on the Boltzmann constant k, the noise bandwidth B_n, the noise figure F_n, and the standard temperature T_0 and is defined as in Equation 3.2. The noise power is approximated as additive white Gaussian noise (AWGN).

$$P_n = k_B F_n B_n T_0 \tag{3.2}$$

In general, the performance of a radar sensor is indicated by the ratio between received signal power P_r and noise power P_n, which results in the signal-to-noise ratio SNR and defined in Equation 3.3).

$$SNR = \frac{P_r}{P_n} \tag{3.3}$$

Combining Equation 3.1 and Equation 3.2, the SNR of a radar sensor can be computed as defined by the Equation 3.4.

$$SNR = \frac{P_t G^2 \lambda^2 \sigma}{k_B F_n B_n T_0 (4\pi)^3 R^4 L_{sys}} \tag{3.4}$$

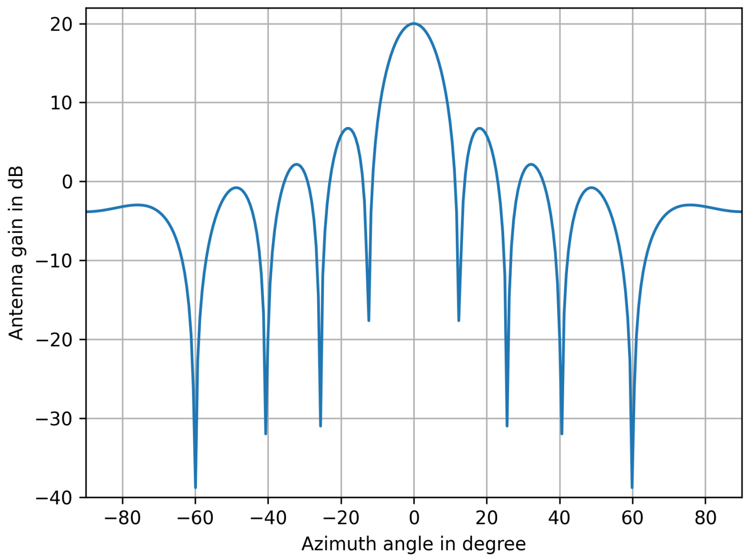

Figure 3.5: Simplified antenna diagram

For example, in rainy weather conditions the minimum SNR to generate a detection is usually increased to minimize false positives. To generate a detection from the reflected power and the SNR, a threshold detection is used. Thus, detection of a radar signal is based on setting a threshold at the output of the receiver. This threshold determines whether the output of the receiver is perceived as a signal that is present or as noise.

Since noise is a random phenomenon, the detection of signals in the presence of noise is a random phenomenon as well [41]. In this thesis, the probabilistic behavior is considered with detection probabilities (DP). These probabilities can be obtained in a simplified way by converting the signal-to-noise ratio using receiver operating characteristic (ROC) curves [4]. ROC curves are affected by the prevailing weather situation and can be dynamically adjusted in real radar systems [94], resulting in a shift in the conversion from SNR to detection probability. This is useful because, for example, in rainy weather, the minimum SNR value for detection is usually increased to minimize false alarms.

3.3 Radar perception

After both real radar measurements and synthetically generated radar data have been collected in the previous steps, both serve as input for a target application. This application represents the desired use case and represents the next stage in the perception processing chain after the sensor. As illustrated in Figure 3.1, a tracking-by-detection approach is used in this work as the desired target application. However, the framework is not exclusively designed for these modules and they serve only as exemplary placeholders in this work, as they can also be replaced by other algorithms, such as a classification.

This is due to the fact that the perception algorithm is not the focus in the present research work and the sensor simulation evaluation should be generically applicable and not developed for a specific perception approach. For this reason, the perception module is considered as a black box in this research.

4 Identifying Relevant Sensor Effects

Contents

4.1 Introduction

Realistic models of environment sensors such as lidar, camera, and radar play a crucial role in a simulation-based testing strategy [16]. Furthermore, radar is considered as a key sensor for autonomous driving [48] and has traditionally been one of the most complex sensors to model [33].

Many different approaches to simulate a radar system have been reported in the literature. Even though a lot of radar effects are understood and can be modeled today, a high fidelity simulation faces challenges regarding the required execution time [92]. This stems from the fact that radar exhibits several properties such as interference, multi-path reflections, ghost objects, clutter, and attenuation [35]. Consequently, this leads to extremely high requirements on the computation power for an accurate sensor simulation. The question arises whether a detailed radar sensor model is required in all simulation use cases. The problem to find the sufficient level of detail remains unsolved and the right trade-off between computation speed and model realism must be found. In order to answer the question of whether the fidelity of a sensor model is suitable for a particular application as well as to permit reliable predictions about the behavior of the real system through simulation-based tests, the sensor model must first be validated [17]. However, although there exist many different approaches to simulate a sensor, the problem of validating a model still remains. In light of Chapter 2 which derived the different problems need to be solved for a sensor model validation, this chapter addresses the second research question RQ2: *Which features of a radar simulation are relevant for a downstream application?*

Therefore, a sensitivity analysis approach for developing and validating a radar simulation is introduced in this chapter. The objective is to identify the radar sensor effects with the greatest impact on a target application. By focusing on the most important effects, a high model accuracy can be achieved while reducing the required computation time. In addition, a proof-of-concept implementation of the sensitivity analysis method is presented to analyze a clustering algorithm as an exemplary target application.

The remainder of this chapter is structured as follows. Section 4.2 elaborates the developed method in-depth and the experiments are conducted and evaluated in Section 4.3. Finally, the effectiveness of the proposed approach is discussed in Section 4.4.

4.2 Method

The present section elaborates the developed approach, starting with an overview of the proposed method, followed by an in-depth explanation of the different components needed.

4.2.1 Sensitivity analysis approach

In order to address the second research question (RQ2), this section introduces the developed method, which focuses on improving the existing approaches by incorporating a quantitative and objective assessment of the relevance of the implemented radar sensor effects. This is accomplished by performing a sensitivity analysis with the objective to measure the impact of each effect on a target application.

Sensitivity analysis is the study of how uncertainty in the result of a model can be attributed to different sources of uncertainty in the model input [139]. In contrast, an uncertainty analysis concentrates on measuring the uncertainty in model output. Model simplification in the context of computationally intensive and complex models, quality assurance, or robust evaluations are examples of applications of sensitivity analysis [140]. Furthermore, the sensitivity analysis approaches can be inter alia divided in quantitative and qualitative methods, while generally qualitative are more efficient, but less accurate. Fourier amplitude sensitivity testing (FAST) was introduced by Cukier et al. [141] which is an effective variance-based quantitative sensitivity analysis method and used in this work. This is due to the fact that variance-based methods provide quantitative measures of how much each parameter attribute to the overall variance of the model response in addition to quantifying the effect of the interaction between the parameters. Additionally, these methods can be applied to complex non-linear and non-monotonic models such as sensor models [142].

A radar simulation often consists of multiple sub-modules in order to model the radar characteristics. This includes for instance a model for the radar cross section (RCS), an antenna model and many others. For each sub-model, the question arises whether it can be simplified, for example, by using constant values, or whether a detailed approximation is necessary. This stems from the fact that the resulting computation time is no longer feasible if every effect is modeled with the highest possible accuracy. The modeling of the RCS for an existing vehicle alone is extremely complex and many approaches exist in the literature [90, 111, 121], ranging from simple models to highly accurate ray tracing based models. Therefore, for the selection of a suitable modeling approach it is necessary to find out which relevance a certain radar characteristic has on the final result of an target application such as object detection. This has the goal to simplify the radar simulation with respect to the complexity of the sub-modules and thus to minimize the required computation time while maintaining a sufficient accuracy.

The developed method consists of mainly four different steps, which are illustrated in Figure 4.1 and explained in the following.

1. Firstly, the sensor characteristics or effects to be investigated and their bounds are specified. Furthermore, a distinction is made between scenario-dependent and sensor-dependent parameters, while the latter is focused in the present work.

2. As a second step, the samples of the defined parameters are generated. This results in a matrix which dimensions are defined by the number of parameters and the number of samples used.

3. Subsequently, the samples are used to parameterize the model under test in order to run the model. Thereby, this step is composed of several steps. The real test drives are conducted and are simulated to generate both the real sensor measurements as well as the synthetic sensor data. Each is processed by the target application and the results are evaluated to quantify the discrepancy between both.

4. In a final step, the actual sensitivity analysis is performed, which uses the specified parameters with their bounds and the model response as inputs, resulting in the sensitivity indices that describe the relevance of the effects investigated.

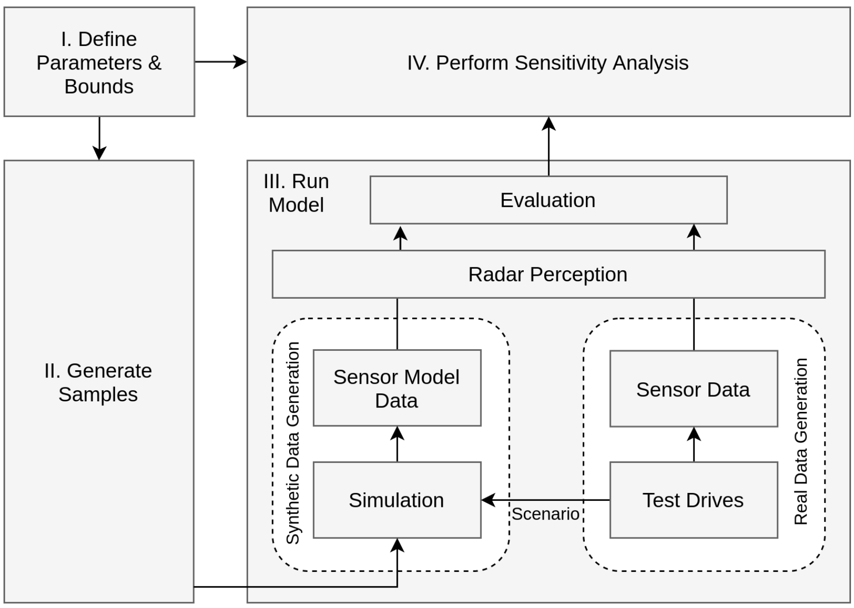

Figure 4.1: Overview of the sensitivity analysis approach to measure the relevance of sensor characteristics regarding a target application.

In the following sections, the different modules needed for the sensitivity analysis are elaborated.

4.2.2 Real sensor data generation

The generation of real radar sensor measurements as a reference for comparison is an crucial part of the sensor model validation. In this regard, it is necessary to determine the correct position and orientation of the surrounding objects, i.e. the ground truth data, which is not a trivial problem. A high degree of accuracy is essential, since the ground truth data is used to perform the test drives in a virtual environment. In this work, simple scenarios on a proving ground are tested along with a DGPS and an IMU as the reference system in order to enable a precise ground truth acquisition. The sensor measurements are recorded in a scenario in which the ego vehicle equipped with the radar is stationary and a target vehicle drives a path in form of an eight in front of it. This scenario is suitable for investigating the influence of the spatial change of the observed vehicle on the generation of radar detections.

4.2.3 Synthetic sensor data generation

The generation of simulated sensor data consists of mainly two steps, the reproduction of the test drives in the simulation and the actual rendering of the virtual environment to generate the radar detections. In contrast to the method presented, it would be conceivable to test all possible scenarios not only those observed in the real test drives. However, since a comparison to real sensor measurements is essential, only the observed tests are simulated.

The generated samples as well as the tested scenarios are used to perform the test drives in the simulation. Thereby, the number of samples determines how often and with which parametrization a specific scenario is simulated. The open-source simulator CARLA is used to model the environment and run the virtual tests.

Given the characteristics set out in Chapter 3, the ray tracing-based model is utilized in the present approach to model the radar sensor. Since the radar effects are modeled physically, they can be varied and analyzed in detail. In particular with this modeling approach, it is necessary to identify the sufficient level of detail of the individual sub-modules or effects in order to overcome the restrictions in computation time.

Radar is an electromagnetic sensor system for the detection and location of reflecting objects. It operates by radiating energy into space and detecting the received echo signals from an object in the environment. The radar range equation describes the range of a radar in relation to the characteristics of the target object, environment, transmitter, antenna, and the receiver. Therefore, it is not only useful for estimating the maximum range at which a radar sensor can detect an object, but it can also be used to understand the factors affecting the radar performance [41]. For this very reason, the radar equation is utilized in this method to investigate the impact of the different radar effects and characteristics on a target application.

4.2.4 Radar perception and evaluation

Subsequently, after generating real and simulated radar data in the preceding steps, both are fed to the radar perception algorithm and thus form the basis for an evaluation. The perception represents the downstream stage after the sensor in the AD processing chain and is realized by a spatial object clustering algorithm using radar detections in this work. Furthermore, k-means clustering is a well known unsupervised learning algorithm and is used here. The discrepancy of the prediction performance between real and simulated data is evaluated by computing the difference between the euclidean distances of the predicted cluster centers. In this way, the spatial distribution of the simulated radar detections are investigated.

4.2.5 Parameters, generation of samples and sensitivity analysis

In the final step of the proposed method, the sensitivity analysis is performed in order to identify the effects of individual radar characteristics in driving the output and its uncertainty.

The specified parameters are listed in Table 4.1 and are described in detail in Chapter 3. They serve as the input for the FAST technique with the objective to measure the impact of the different sensor effects on the results of the perception algorithm. Thereby, the bounds of the parameters are defined according to typical automotive radar values reported by Gamba [49] and Skolnik [35]. The parameters are further varied within these defined bounds in order to investigate the corresponding model response variance.

Table 4.1: Sensitivity analysis parameters with their specified bounds.

Symbol	Description	Unit	Min/Max
$AWGNoise$	AWGN standard deviation	dB	0/8
DP_{offset}	detection probability offset	–	−5./5
G_{max}	maximal antenna gain	dB	10/25
F_n	noise figure	dB	10/20
L_{sys}	overall system loss	dB	0/20
RCS_{mean}	mean radar cross-section	$dBsm$	−10/10

The result of the FAST method are the following sensitivity indices: the first-order sensitivity index S_i and total-order sensitivity index S_{Ti}. The former measures the main effect of a parameter, while the latter relates to the overall impact of a certain parameter, i.e. the total effect. Furthermore, the difference $S_{Ti} - S_i$ between both indices is a measure of the strength of the interactions [140].

4.3 Experiments and results

In this section, the results of the developed method are evaluated. First, the difference between both real and synthetic sensor data is analyzed by comparing the clustering predictions. In this way, implications for the sensor model can be derived purely subjectively and qualitatively, based mainly on the expert knowledge about the radar and the simulation of the investigator. In order to enable an objective and measurable assessment, the sensitivity analysis approach is used for a deeper evaluation of the simulated data.

4.3.1 Clustering evaluation

A target object drives in this experiment a path in form of an eight in front of the ego vehicle which itself is stationary and equipped with the radar sensor. The scenario is indicated in Figure 4.2. Hereby, the goal of this scenario is to assess the influence of the orientation as well as the range of the object on the generation of radar detections. It is assumed that the detections change in distribution and density over the distance.

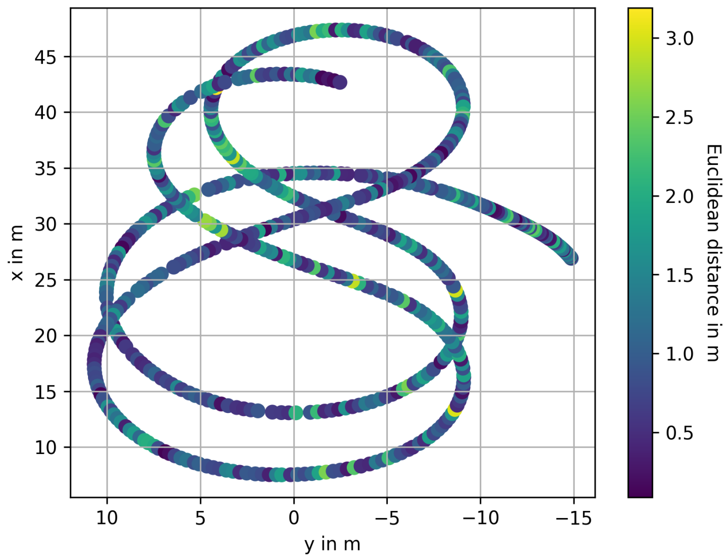

Figure 4.2: The evaluation result of the clustering algorithm fed with real and synthetically generated radar data.

In this section, it is investigated whether and to what degree the radar simulation can model this behavior. Therefore, the target application, which is represented by the clustering algorithm in this work, is fed with real and simulated radar data. The difference between the resulting predictions is measured by the euclidean distance between both predicted cluster centroids. The outcome of this evaluation is illustrated in Figure 4.2 as well, with the degree of error is emphasized with different colors. Moreover, a higher euclidean distance indicates a larger spatial discrepancy between both sensor data sources.

From the figure, it can be observed that deviations between measurement and simulation exist. However, in a holistic assessment the discrepancies are relatively small over the whole scenario. The larger error are noticeable in particular when the target object turns. This is probably due to an inaccurate RCS model and/or model of the radar antenna, as these models strongly influence the longitudinal and lateral resolution of the sensor data generation. Additionally, in consideration of the fact that the target vehicle starts from the negative y-axis, it can be noticed that larger deviations occur when the vehicle is perceived from the rear view. This finding indicated that the simplified symmetric RCS model is not sufficient, at least in the longitudinal vehicle axis.

4.3.2 Sensitivity analysis results

In light of the clustering evaluation, it is apparent that since the radar simulation is comprised of several sub-modules, it is difficult to determine which effect causes the error by qualitatively evaluating the result. Furthermore, such a subjective evaluation is strongly dependent on the expert knowledge of the examiner and is thus prone to errors. In order to address this problem, a sensitivity analysis is performed with the goal of providing a measurable and objective assessment of the relevance of the radar properties to the perception algorithm. The results of the sensitivity analysis are covered in this section.

The parameters of the radar simulation are parameterized according to Table 4.1 in order to identify the impact of each characteristic regarding the clustering algorithm as the target application. The parameters are varied within the defined bounds to generate the samples, resulting in a total of 390 samples used in this experiment. Each sample represents a specific radar model configuration, which is fed to the simulation to generate the synthetic radar data.

Both sensor data sources are evaluated according to the procedure previously elaborated. This evaluation result serves as the input for the sensitivity analysis, i.e. for the FAST method. Since this method requires a scalar value per simulation run as the input, the analysis is performed with three different approaches to calculate the evaluation value: the minimum, mean and maximum of the euclidean distance over all points in time of a simulation run. As a result, this leads to three different sensitivity analysis as illustrated in the Figures 4.3, 4.4 and 4.5.

It is apparent that no effect alone is solely responsible for the model response, which can be inferred from the first-order sensitivity indices S_i of each evaluation. Furthermore, since the interaction coefficient is relatively high, no parameter can be neglected. The interaction coefficient is computed by subtracting both indices $S_{Ti} - S_i$ [142].

The main effect of both the standard deviation of the additive white Gaussian noise (AWGN) as well as the noise figure remain in the same range of $10 - 20\%$. The total loss is relatively large when the mean and maximum values are evaluated, suggesting that the upper limit of this parameter is set

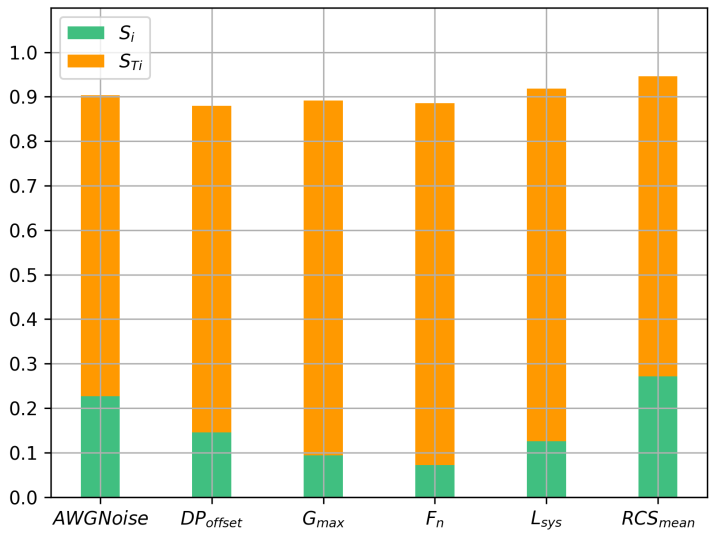

Figure 4.3: Sensitivity analysis results using minimum evaluation value.

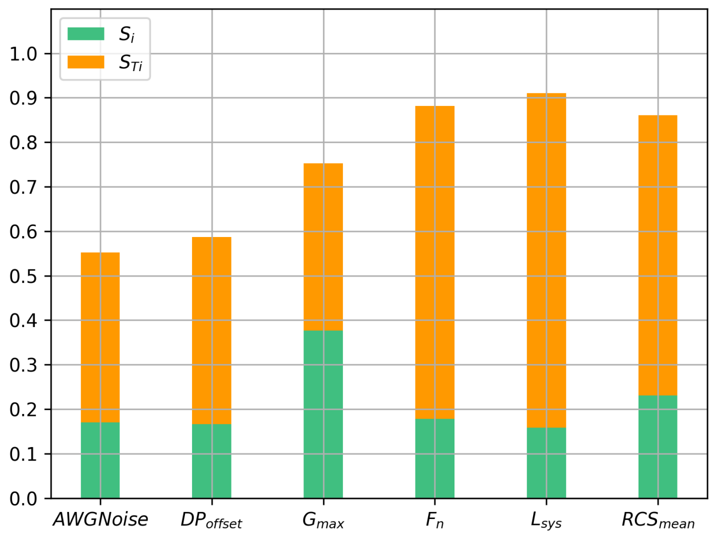

Figure 4.4: Sensitivity analysis results using mean evaluation value.

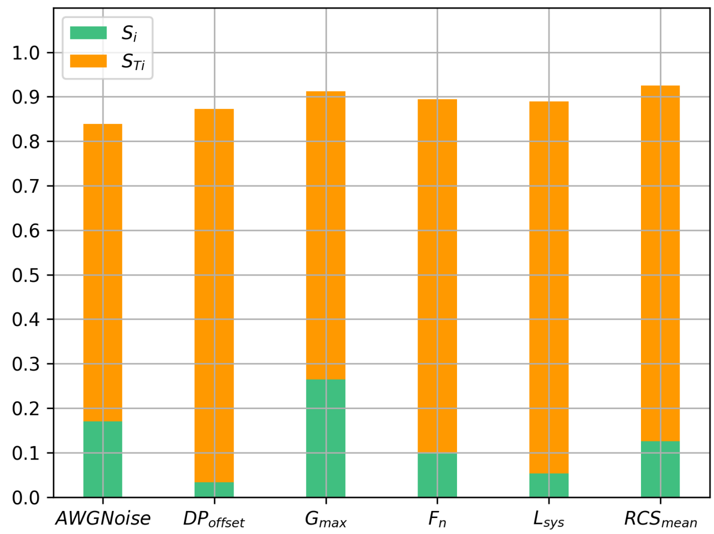

Figure 4.5: Sensitivity analysis results using maximum evaluation value.

too high. Based on the clustering evaluation of the previous section, it can be assumed that only minimal discrepancies occur in the near field. Additionally, large deviations between synthetically generated and measured sensor data are expected in the far field. Given these assumptions, the high first-order sensitivity of the detection probability for the minimum evaluation case is due to a higher number of radar detections in the radar near field, which is caused by the used ray tracing method. Based on the investigated scenario, a high impact of the antenna model can be observed, which is plausible since the antenna gain varies strongly with the azimuth angle to the object. Contrarily, the impact of the RCS model has the least effect in the maximum case, i.e. in the far field. This result indicates that a RCS model solely depending on the aspect angle might not be sufficient.

4.4 Discussion

This section summarizes and discusses the contributions of this chapter to the second research problem RQ2: *Which features of a radar simulation are relevant for a downstream application?* A sensitivity analysis approach is used to identify the relevance of different sensor characteristics. This chapter makes two significant contributions to the dissertation – a novel sensor model validation method as well as its implementation for automotive radar.

In the present chapter, a method was proposed for evaluating a sensor simulation in order to determine the impact of specific sensor effects regarding the prediction result of a target application. This general method can be applied to a wide variety of simulation models and was specifically used for a radar sensor model that especially has high computation requirements. A spatial clustering algorithm was further used as the target application, which represents a typical use case for processing radar data. The perception algorithm is fed with both real sensor measurements as well as synthetically generated data with the purpose of comparing both prediction outputs. To conduct the experiments and investigate the effectiveness of the developed method, the FAST algorithm was used to perform a sensitivity analysis result taking the result of the clustering evaluation as input.

It can be shown that a sensitivity analysis enables a more detailed, measurable and objective evaluation of the simulated sensor data in comparison to a qualitative assessment. The results from specific situations can be traced back to the contribution of the individual sub-modules or sensor effects of the radar simulation, resulting in an efficient analysis of the simulation. Therefore, the developed approach complements the research towards virtual validation of AD functions.

There are different conceivable extensions of the approach to further enhance the evaluation. Since this approach is dependent on the predefined radar effects to be analyzed, there is a risk that important effects are neglected or overlooked. In order to deal with this problem, analysis methods such as failure mode and effect analysis could be applied, which may not completely eliminate the problem, but at least provide a systematic approach for mitigating it. Furthermore, a radar simulation consists typically of multiple models, including also an environment model. Therefore, the method could also be extended to investigate whether or in which use cases, for example, a detailed vehicle model is required and which parameters are relevant for that. Additionally to radar effects, simulation-related parameters can be investigated, such has the number of emitted rays or the number of reflections during ray tracing. The required computing time can also be included, which increases exponentially with a large number of reflections.

5 Evaluating Simulated Radar Data

Contents

5.1 Introduction

Robust perception and sensing of the environment of a self-driving vehicle are essential tasks to build an understanding of the surrounding scene [68]. Since automotive radar is widely employed within modern advanced driver assistance systems (ADAS) and is a key technology for AD [48], radar sensor models are becoming more and more important.

To permit any implications about a real system based on simulated sensor data, the used model have to be validated [80]. In order to solve this validation problem, the fidelity of the model must first be measured to be able to determine whether it is sufficient or not. Despite the fact that many approaches exist in the literature to model a radar sensor, there is no generally accepted method to evaluate simulated radar data [18]. Additionally, an approach for measuring the level of accuracy of a sensor model does not yet exist and the problem of defining an appropriate metric remains. As elaborated in Chapter 2, the majority of approaches found in the literature provide qualitative and subjective assessments of simulated radar data. However, a qualitative evaluation relying on a visual matching of the data does not scale, which is one of the key advantages of simulation-based testing. In addition, existing quantitative evaluation approaches rely on self-defined metrics that assess specific sensor properties such as the spatial distribution between synthetic and real sensor data. This leads to the problem that some important effects may be neglected or overlooked, since the challenge of selecting the appropriate metric is tantamount to deciding which property or physical effect is relevant, which has not yet been solved. Furthermore, existing fidelity evaluation approaches rely on real sensor measurements as a reference for comparison. This assumes on the one hand that the sensor already exists and on the other hand that measurement data has been recorded. Especially the latter, as described in Chapter 3, is only practicable in simple test environments, because of the complex ground truth determination. However, no approach exists to assess the accuracy of the simulated sensor data in another environment where it is not feasible to generate reference data.

Consequently, a method is needed that allows an objective and quantitative evaluation of synthetically generated radar data, taking into account the important sensor characteristics. For this purpose, the present chapter proposes an approach to answer the third research question RQ3: *How to determine the degree to which the radar simulation and experimental measurements concur?*

A machine learning-based evaluation approach to assess a radar model is presented in this thesis. Therefore, a neural network (PointNet++ [143]) is trained in order to classify real and synthetic radar data with the objective of learning the latent and characteristic features of real radar point clouds. Moreover, the classifier's confidence score of the '*real radar point cloud*' class is proposed as a metric to measure the degree of fidelity of simulated radar data.

The remainder of this chapter is organized as follows: The developed method is introduced in detail in Section 5.2. Based on this, the experiments and the results are described in Section 5.3. Finally, the effectiveness of the proposed approach is discussed in the Section 5.4, including a concise outlook on further research.

5.2 Method

This section proposes the developed method which focuses on the enhancement of the existing approaches at the detection level by incorporating a quantitative evaluation without the need for self-defined metrics. A machine learning-based approach is introduced and in order to investigate the effectiveness of the method, the proposed deep evaluation metric (DEM) is compared with conventional metrics to assess the fidelity of simulated radar point clouds. The approach consists of the following four main steps which are illustrated in Figure 5.1: generation of real and simulated radar data, conventional metrics as well as the developed DEM.

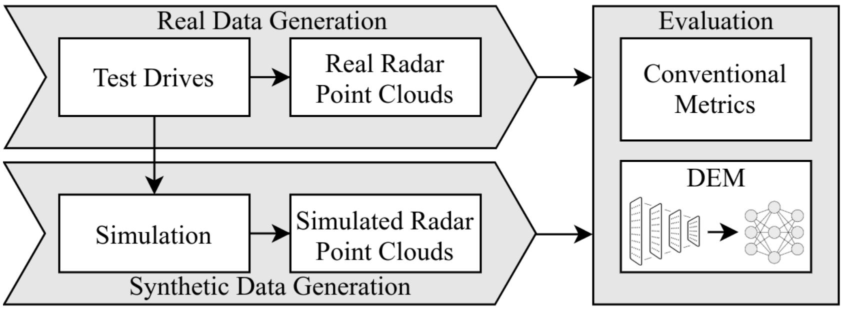

Figure 5.1: Overview of the proposed machine learning-based evaluation method.

5.2.1 Sensor data generation

This sections briefly describes the sensor data generation used in this chapter, since it was elaborated in detail in Chapter 3.

The first step comprises the generation of real radar measurements. Therefore, the required test drives are conducted on a proving ground with the ego vehicle and one target vehicle. A DGPS is used in combination with an IMU for an accurate acquisition of the position, orientation and velocity of the vehicles. A precise ground truth measurement is needed, because it serves as the basis to produce the same tests in the simulation. Furthermore, the radar data is recorded with various scenarios ranging from simple stationary scenes to highly dynamic overtaking maneuvers.

On the other hand, the generation of simulated sensor data is mainly divided in two steps: the simulation of the test drives based on the recorded ground truth data, and the generation of the actual sensor data, i.e. the synthetic radar point cloud. The procedure of the latter is indicated in Figure 5.2 and the implementation details of the underlying sub-modules and formulas used are thoroughly explained in Chapter 3.

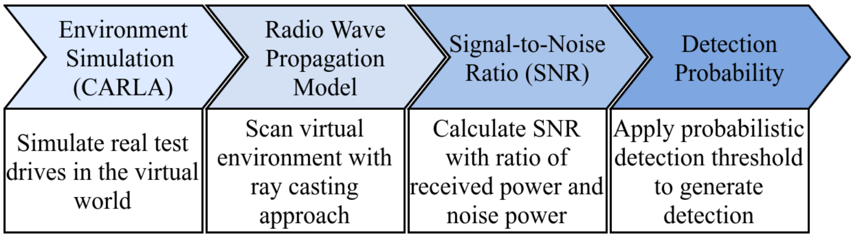

Figure 5.2: Processing pipeline of the developed radar simulation.

5.2.2 Conventional metrics

The existing conventional evaluation metrics are introduced in this section. Two different metrics are implemented to investigate and estimate the accuracy of simulated radar point clouds.

In the present thesis, each radar detection is defined by its two-dimensional location and the Doppler velocity, both dimensions are compared to evaluate the difference between real and synthetic radar data. In this regard, the normalized sum of the smallest Euclidean distance is computed from every point in the real point cloud $X = (x_1, ..., x_M)$ to the simulated point cloud $Y = (y_1, ..., y_N)$, where $x_m, y_n \in \mathbb{R}^3$ are three-dimensional points. The used point cloud to point cloud distance is first proposed by Browning et al. [144] and is defined as:

$$D'_{pp}(X, Y) := \frac{1}{M} \sum_{m=1}^{M} \min_{1 \leq n \leq N} ||x_m - y_n||. \tag{5.1}$$

The metric has the benefit that the disparity in values of each point and the difference in the number of points between both radar data sources are considered. Furthermore, it is divided by the respective number of points in order to normalize the result. Additionally, the worst-case is assumed, as D'_{pp} is a non-symmetrical range metric:

$$D_{pp}(X, Y) := \max(D'_{pp}(X, Y), D'_{pp}(Y, X)). \tag{5.2}$$

As for the second metric, the Wasserstein distance is used to compare the point distributions between both radar point clouds. This metric is also known as the earth mover's distance (EMD). It is further based on the Kantorovich-Rubinstein theorem [145], which addresses the optimal transportation problem [146]. Thus, EMD measures the disparity between two distributions by the optimal cost of rearranging one distribution into the other:

$$EMD(X,Y) := \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} f_{m,n} d_{m,n}}{\sum_{m=1}^{M} \sum_{n=1}^{N} f_{m,n}}. \tag{5.3}$$

Besides the three-dimensional point clouds X and Y, m and n represents the number of points in the point sets. In addition, the solution to the transportation problem between both radar data distributions is described by the optimal flow $f_{m,n}$. The Euclidean distance is used for the ground distance $d_{m,n}$. Therefore, EMD naturally extends the notion of a distance between individual radar points to that of a distance between distributions of points. An in-depth elaboration of the used formulas can be found in Rubner et al. [147].

5.2.3 Deep evaluation metric

Conventional existing metrics rely on self-defined metrics which assess specific characteristics like the spatial distribution between synthetic and real radar data. The problem of selecting an appropriate metric remains, which is equivalent to selecting the relevant sensor properties or physical effects.

The following section proposes a machine learning-based metric to estimate the fidelity of simulated sensor data. The aim of the developed method is to train a neural network to distinguish between real and simulated radar point clouds. Contrarily to the previously described conventional evaluation approach, the intent of the data-based method is to learn the latent features that distinguish real from synthetic radar point clouds without having to determine in advance which characteristic specifically to consider. Therefore, the classifier's predicted confidence score of the '*real radar point cloud*' class is proposed as a metric to measure the accuracy of simulated radar data.

The process of choosing and adjusting a suitable neural network architecture is elaborated in the following in addition to the used data set and the training and testing procedure.

Network architecture

The radar point clouds have to be transformed to a regular format before inserting them into a neural network, because the input of most networks follow a regular structure like a grid map representation. In this respect, Qi et al. provide with PointNet++ [143] an architecture to overcome this limitation and work directly with point clouds without the need for previous mapping. PointNet++ is a hierarchical network and is able to learn local features and to deal with point sets that vary in density. Schumann et al. [66] and Danzer et al. [62] have already shown that PointNet++ can be applied to radar data, which is why this approach is used in this thesis.

Data set

In order to simplify the data acquisition, only the radar detections around the target vehicle are considered. Since the test drives are performed on an empty proving ground, this is a reasonable simplification. As described in the previous chapters, the simulated sensor data is generated by reproducing the real test drives in the virtual domain. This is the reason why the resulting data set is rather balanced between real and synthetic measurements. 235 scenarios are tested, which corresponds to 1.59×10^5 point clouds with 3×10^6 radar detections in total. Hereby, each radar detection that is fed into the neural network, contains two spatial coordinates along with the Doppler velocity. Finally, the entire set is randomly split into a training and testing data set with a 70/30 ratio.

Training and testing

The used network is trained from scratch, using both synthetic and real radar data. In order to avoid model overfitting, the data set is augmented during the training process. For this reason, the data is perturbed using random Gaussian noise with a zero mean and a standard deviation of 0.1. The resulting random noise is used to alter each feature dimension of every data point, i.e. the spatial positions as well as the Doppler velocity of the detections are modified. For the purpose of maintaining a fixed input size for each point cloud, sampling is performed by randomly drawing (undersampling) or duplicating (oversampling) each set up to 10 detections. Furthermore, the initial learning rate of the model is set to 0.001 and the batch size for training is 32. The Adam optimizer is utilized for training and is performed for 20 epochs on two two NVIDIA GeForce RTX 2080 Ti GPUs. Throughout testing, the batch size is set to 1 to allow a variable number of points to be fed into the model. The trained network achieves a classification accuracy of 82.14% within testing.

5.3 Experiments and results

The experimental setup is introduced in this section. Subsequently, the performance of the trained network is investigated in order to analyze whether and to what extent the model has learned the latent features that differentiate both real and simulated radar point clouds and is thus able to distinguish between them. Based on this, it is further analyzed whether the result of the final network layer, i.e. the confidence score of the 'real radar point cloud' class, can be used as a metric to measure the sensor model fidelity. Therefore, the DEM is evaluated in comparison with the implemented conventional metrics and the effectiveness of both approaches is compared and discussed.

5.3.1 Experimental setup and classification performance

For the sake of comparability, the metrics are compared using the same scenario in which a target vehicle drives a path in the shape of an eight in front of the stationary ego vehicle. The ego vehicle is equipped with the radar sensor and the scenario is illustrated in Figure 5.3. It can be assumed that the real radar point clouds change in distribution and density over different positions and orientations of the target vehicle. Thus, the goal of this scenario used is to investigate to what degree the radar model is capable to model this behavior.

Given that the implemented radar simulation incorporates a random module (detection probability) in order to model the stochastic nature of real radar measurements, the evaluation results are also subject to random effects. In order to diminish these effects, the scenario is simulated 100 times and the results are averaged over these runs.

Besides the path of the target vehicle, the predictions of the classification of the trained network is color coded in Figure 5.3. Hereby, the model is fed with real radar data as well as synthetically generated sensor data to examine the capability of the model to differentiate between both data sources. This specific scenario was withheld from the training and testing set to guarantee an unbiased evaluation of the model performance.

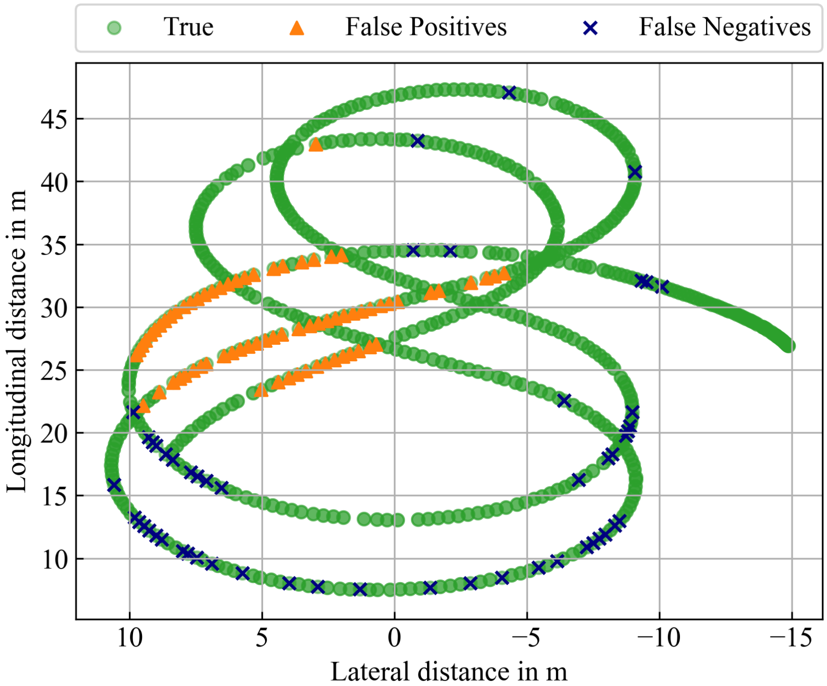

Figure 5.3: Classification results on the withheld scenario. The model is fed with real radar data as well as the corresponding synthetic radar point clouds. The green detections indicate all correct classification predictions. Additionally, the false positives (input: simulated, prediction: real) and false negatives (input: real, prediction: simulated) are illustrated.

In the examined scenario, the model achieves a classification accuracy of 88.59% with simulated and 91.99% with real radar measurements as input. It can be observed that most of the false predictions are located in specific areas for both inputs. The center of the false predictions is located at a longitudinal distance of about 27 meters and a lateral distance of approximately 5 meters. The results indicate that either the vehicle was not observed enough in the training before in this area or that this specific zone exhibits a weakness of the radar simulation. Contrarily, the larger part of the false negatives are found in the near field of the sensor.

In summation, the neural network can predict most of the radar point clouds correctly, which indicates that the model could learn the distinctive features that differentiate real and synthetic radar data. This allows a deeper investigation of the proposed deep evaluation metric in comparison with conventional metrics.

5.3.2 Results of evaluation approaches

At first, the post-processing procedure is elaborated. Based on this, the qualitative and the quantitative evaluation are performed and compared between the different metrics.

Postprocessing

Apart from the result averaging of over 100 simulation runs, the data are further post-processed in order to ensure a valid comparison between different metric results. For the reason that the resulting values can vary widely, a min-max normalization is applied, i.e. rescaling the range of data to [0, 1]. Moreover, the axes are reversed in a way that zero represents the worst (lowest sensor model fidelity) and one the best possible result, meaning high sensor model accuracy. As the final post-processing step, the data is smoothed for a clearer visualization to be able to compare the trend of the different results. For this, the Savitzky-Golay filter [148] is applied.

Qualitative evaluation

In the following, the key differences between both radar data sources are defined, which are identified by a qualitative evaluation based on visual matching of both radar measurements. Significant sample points of both the real and simulated radar point clouds are depicted in Figure 5.4. According to this, the metrics are then examined to what degree they can quantifiable reproduce the observed qualitative disparities.

Since the developed radar model uses a ray casting-based technique to approximate the radar wave propagation, it is evident that on the one hand large differences between real and synthetic radar point clouds occur, especially at close range due to an increase in the number of simulated detections. On the other hand, the number of radar detections decreases too much with increasing range in comparison to the real measurements.

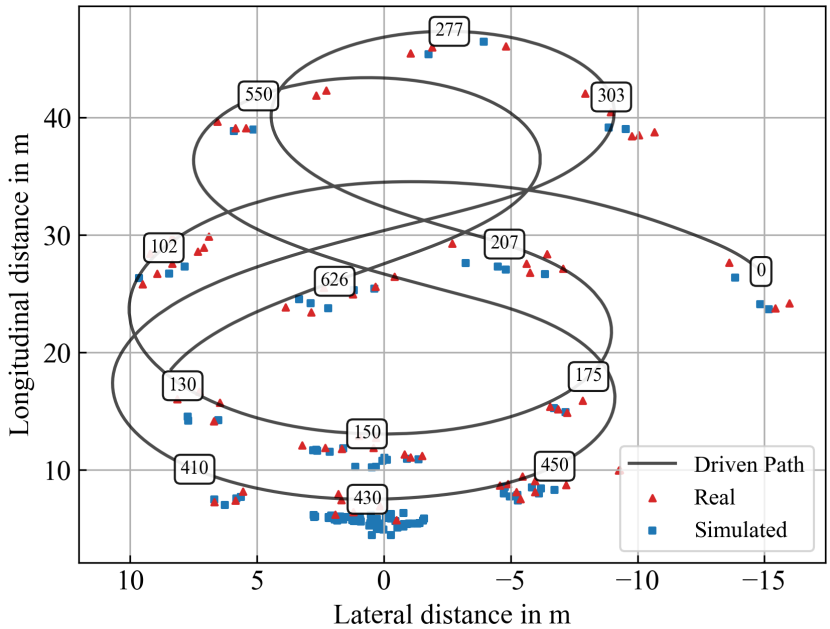

Figure 5.4: The real and simulated radar detections are illustrated and the white boxes indicate the frame number.

Apart from that, an additional effect can be observed, which occurs in particular in the near field. This effect is the formation of a L-shaped point clouds in this area, which is due to the fact that only the outer shell of the vehicle is modeled and the resulting aggregated high number of points in the close range make the edges of the shell to be clearly visible. However, this point cloud shape is rather untypical for radar, since generally detections can also be found inside the vehicle shell.

Quantitative evaluation

In order to further analyze the metrics, the results are additionally plotted over time in Figure 5.5. It is especially apparent that all three metrics indicate a relatively high model fidelity, in particular the EMD ($\mu = 0.79$, $\sigma = 0.09$) and D_{pp} (μ=0.90, σ=0.09). However, the proposed DEM predicts the lowest fidelity with a relatively high standard deviation (μ=0.72, σ=0.19).

Whereas the EMD does not imply any decline of fidelity of the simulated data in the near field, i.e. around frame 150 and 430, the D_{pp} has a strong minimum peak in the areas directly in front of the sensor. This minimum

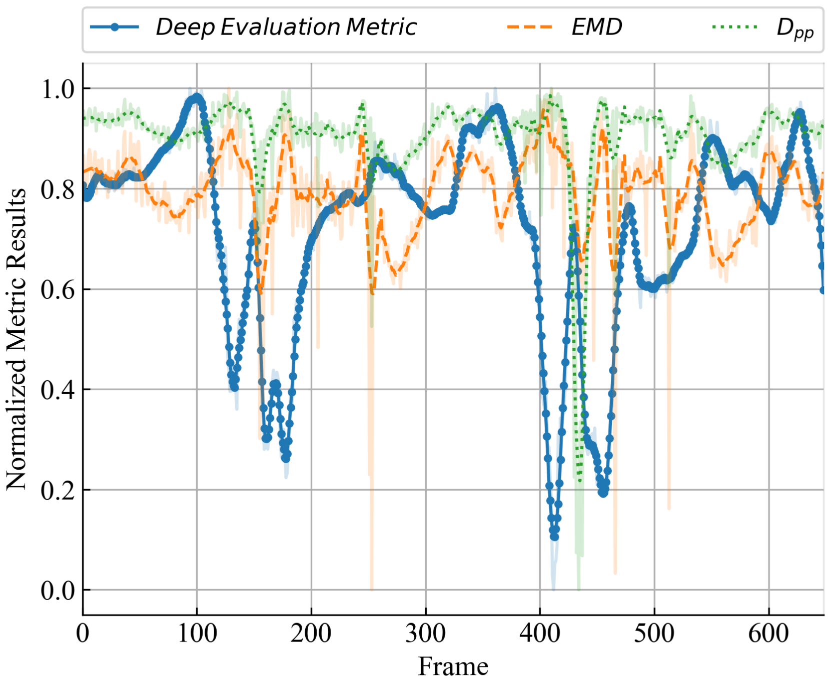

Figure 5.5: The moderately transparent and solid lines indicate the unfiltered results, whereas the dashed lines represent the smoothed point cloud metric results.

can be explained by the significant increase in the number of points in the simulated data, which causes an strong increase in the calculated sums of the individual radar points. Although DEM indicates a low fidelity in both near ranges mentioned, peaks upwards can additionally be noticed in these zones. These are probably caused by the turning point of the observed vehicle, as the object is perpendicular to the sensor in these positions and thus the previously described L-shape disappears, which could result in an sharp increase of the fidelity estimated. However, in Figure 5.4 it can be observed that the number of points differ considerably especially in the near field, which should result in a further descent of fidelity. The effect of too few simulated points appearing at longer distances is not significantly reflected by any of the metrics presented. The number of points might be too small to allow a reliable prediction of the neural network. For EMD and D_{pp}, an insufficient number obviously does not affect the estimated quality of the simulated point cloud.

5.4 Discussion

The present section recapitulates and discusses the main contributions of this chapter to the third research problem RQ3: *How to determine the degree to which the radar simulation and experimental measurements concur?* A novel data-driven metric is introduced to measure the relevant characteristics of radar measurements in order to estimate the accuracy of a radar simulation for AD.

In this chapter, a machine learning-based metric, the DEM, is presented to estimate the fidelity of simulated radar detections. To analyze the effectiveness of the proposed approach, conventional metrics are used and their ability to measure relevant differences between real and synthetic sensor data is compared.

It can be shown that the developed DEM, unlike the conventional metrics, is able to identify the weaknesses of the simulated radar detections in the near-field region. However, it was not possible to measure all effects such as the inaccurate number of detections at long ranges, which none of the metrics were able to do. Overall, the proposed metric shows great potential as it could reflect the intuitive result of a qualitative assessment much better than the other metrics.

Various extensions of the approach are conceivable to further improve the data evaluation. Future work will focus on improving the training data, for example by learning the entire scene perceived by a sensor or including classes other than cars, such as pedestrians or cyclists. In addition to enhancing the training data, it can be investigated to what extent it is beneficial to include time information to account for the temporal evolution of objects. Additionally, it can be further examined whether the network is able to generalize the predictions to other scenarios not observed before. It is also conceivable to investigate to what degree general statements can be made about the quality of radar data without using real measurement data as a reference, for example, if the real radar sensor does not yet exist.

6 Measuring the Simulation-to-Reality Gap of Radar Perception

Contents

6.1 Introduction

Since the safety validation of an AD system is an incredibly complex problem, novel approaches are needed. This is because a statistical proof of safety based on real testing in the real world does not scale [14]. The combination of field tests and tests in a virtual environment is a promising method to considerably decrease the validation effort of autonomous driving [15].

In the literature, numerous application areas for the use of simulated sensor data in the development and test process can be found [29]. Sligar uses an accurate physics-based radar simulation to create a synthetic sensor data set to train a machine learning-based object recognition model [30]. In contrast, Hartstern et al. use probabilistic sensor simulations to determine the optimal sensor setting at early stages of development, because they provide a wide range of modification parameters and customizable settings [31]. Ponn et al. use phenomenological sensor models to automatically create critical scenarios based on a sensor setup model of the autonomous vehicle [32].

Since the requirements for a sensor model can vary greatly depending on the target application, the sensor model must be validated and the right compromise between model realism and computation time must be found. Especially for highly realistic radar simulations, the computation time requirements are very high. Although it is relatively simple to measure the run time of a simulation execution, estimating the accuracy of a sensor simulation is quite complex, because not only the sensor model itself but also the virtual environment has to be assessed [80].

Schlager et al. derive the accuracy of a sensor simulation by focusing on the inputs and outputs as well as the modeling technique used [29]. If a sensor simulation relies on rendering methods such as ray tracing, the sensor model is classified as high fidelity. However, this is a rather qualitative evaluation and does not necessarily apply to radar models. Furthermore, various approaches can be found in the literature that qualitatively compare synthetic and real radar data [34, 90, 121]. Apart from a direct evaluation of virtual sensor data, the model can also be evaluated indirectly by investigating the disparity of the results from a subsequent algorithm, which processes the sensor data [94, 127].

As an alternative, there exist different evaluation methods for lidar simulations, which are based on distance metrics and occupancy grids [16, 125, 144]. However, the question arises whether these approaches can be applied to radar data, considering that they are much more stochastic and sparse in comparison to lidar point clouds [18]. As for camera data, Reway et al. introduce a testing method to measure the difference between simulation and real video data by indirectly assessing the model with an object detection algorithm fed with real and simulated data [124].

Despite the fact that many radar simulation approaches exist in literature, the problem of quantitatively evaluating the overall accuracy of a sensor model remains unsolved. However, it is crucial to validate the employed radar simulation to be able to rely on simulation-based tests [80], because it can be assumed that a discrepancy exist between a radar simulation and the actual radar sensor.

Consequently, a novel method is needed that allows to accurately estimate the simulation-to-reality gap, serving the objective to decide whether a given model is sufficient for an intended use. Therefore, the target application should be considered in a model evaluation, since the requirements on the simulation can vary depending on the desired use. Thus, this chapter introduces a methodology to address the fourth research problem RQ4: *How to measure the overall simulation-to-reality gap considering a target application?*

A multi-level testing method is proposed to measure the overall gap between simulation and reality for virtually testing perception functions. The remainder of the present chapter is structured as follows: Firstly, Section 6.2 elaborates the developed validation methodology, including the defined evaluation levels. Based on this, Section 6.3 presents the conducted experiments in order to analyze the effectiveness of the approach. Finally, Section 6.4 concludes this chapter with a discussion and a concise outlook on possible further work.

6.2 Method

The present section introduces the method which focuses on measuring the simulation-to-reality gap of a radar simulation regarding an intended use in order to render existing differences visible. A multi-level testing method is developed and the procedure is illustrated in Figure 6.1.

The approach consists of a combination of an explicit and an implicit sensor model evaluation. Whereas the former directly evaluates the synthetic data, the latter refers to an indirect assessment by analyzing the prediction result of a subsequent target application. Additionally, a multi-object tracking approach is chosen as an exemplary intended use of the radar data in this thesis, which is a typical use case of radar perception in the AD context.

Both assessment levels are further subdivided into a holistic (high level) and a detailed (low level) evaluation. This results in four separate fidelity levels

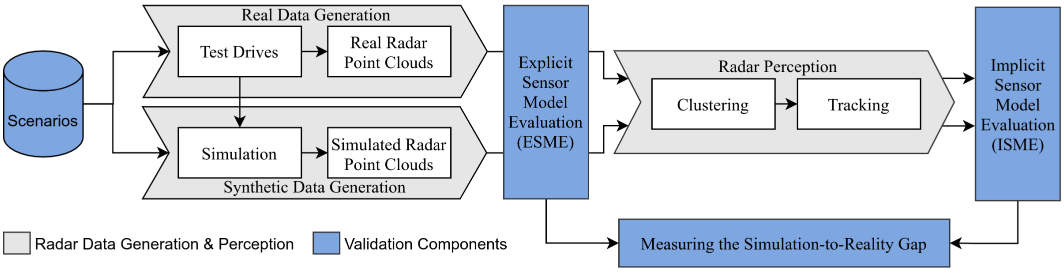

Figure 6.1: Overview of the proposed validation approach in order to measure the radar simulation gap.

(FL) illustrated in Figure 6.2. This separation into four fidelity levels allows a holistic sensor model assessment that makes existing discrepancies transparent and enables an accurate measurement of the overall model fidelity.

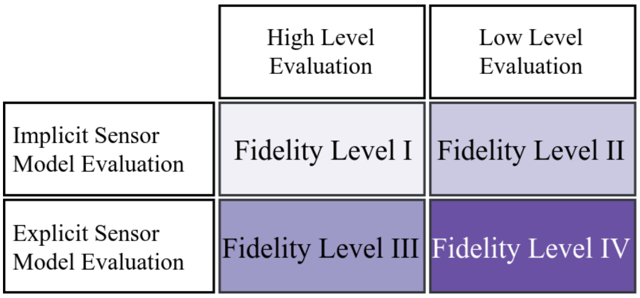

Figure 6.2: Four different fidelity levels are introduced to allow an accurate estimation of the overall radar model fidelity.

In the following, the different modules are elaborated, in particular the proposed fidelity levels along with the quantification of the overall radar simulation gap is described in detail.

6.2.1 Radar data generation and perception

The present section briefly describes the implemented generation of real and synthetic data as well as the perception algorithm used. The employed models are elaborated in-depth in Chapter 3.

The generation of the real radar data along with the ground truth acquisition is analogous to the previous chapters. In the following experiments, all three radar models developed are used to generate synthetic data. This serves the goal to examine the developed method on different use cases and to analyze to what extent the method can measure the strengths and weaknesses of the different modeling procedures. For this reason, the following typical models of a sensor are used: an ideal radar model (IRM), a data-driven or stochastic model (DDM), and a physically-based model using ray tracing (RTM).

The radar point clouds generated by these employed models are further processed by a perception module, which is represented in this thesis by a tracking-by-detection approach as the target application. This module is comprised of a clustering and a tracking method.

6.2.2 Explicit sensor model evaluation (ESME)

In this work, the direct or explicit sensor model evaluation focuses on the radar detection level, which relates to the interface after a reflection passes the detection threshold, leading to the radar point cloud. Both the real and the synthetically generated radar data are compared regarding their similarity. Thereby, the evaluation is conducted on two different levels of detail. The first FL refers to a single score metric to evaluate the fidelity of the simulated data in a holistic perspective, i.e. focusing on the point cloud as a whole (high level evaluation). In contrast, the second FL independently investigates the individual features of the point cloud, including the number of detections (low level evaluation).

Explicit - high level evaluation

Each detection in the radar point cloud is defined by its two-dimensional position as well as the Doppler velocity. The normalized sum of the smallest Euclidean distance from every point in the real $X = (x_1, ..., x_M)$ and synthetic point cloud $Y = (y_1, ..., y_N)$ is used, where $x_m, y_n \in \mathbb{R}^3$ are three-dimensional points. The point cloud to point cloud distance D_{pp} introduced in Chapter 5 is also used here and defined as follows:

$$D_{pp}(X, Y) := \frac{1}{M} \sum_{m=1}^{M} \min_{1 \leq n \leq N} ||x_m - y_n||. \tag{6.1}$$

As the second metric, the Gaussian Wasserstein distance (WD) is used, which was introduced in Chapter 5 as the earth mover's distance. It compares the point distributions of two point clouds and measures the discrepancy between two sets determined by the optimal cost of rearranging one into the other. The Wasserstein distance is defined as follows:

$$WD(X,Y) := \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} f_{m,n} d_{m,n}}{\sum_{m=1}^{M}\sum_{n=1}^{N} f_{m,n}}. \tag{6.2}$$

Besides the point clouds X and Y, m and n represent the number of points in the clouds. Furthermore, the optimal cost between both distributions is defined by the optimal flow $f_{m,n}$. As the ground distance $d_{m,n}$ the Euclidean distance is used. In consequence, WD naturally expands the notion of a range between single detections to that of a distance between distributions of points.

Explicit - low level evaluation

In order to enable a detailed assessment of the synthetic data and to render existing deviations visible, individual features of the point cloud are specifically evaluated. Since a radar point is defined by its two-dimensional position (radial distance r, azimuth ϕ) and the Doppler velocity, each dimension is assessed separately and the respective differences are measured across the real and simulated domain by the feature specific Wasserstein distance ($WD_{feature}$). This results in the following metrics: WD_r, WD_ϕ, $WD_{Doppler}$. As the last metric of this fidelity level, the difference in the number of points is considered by the absolute point number error (PNE).

6.2.3 Implicit sensor model evaluation (ISME)

The indirect or implicit sensor model evaluation investigates the output of a subsequent perception module fed with real and simulated sensor data. By feeding the data from both domains into a perception algorithm optimized for real radar data and comparing both resulting predictions, the strengths and weaknesses of the sensor simulation can be investigated. Analog to ESME, ISME is subdivided into high and low level evaluation. The former focuses on the overall prediction evaluation, whereas the latter independently measures the discrepancies of specific features of the prediction result. The metrics used are briefly described in the following, because the assessment of perception methods is generally better investigated and more mature compared to the evaluation of radar detections.

Implicit - high level evaluation

For this holistic fidelity level, two different metrics are used to measure the overall performance of the object tracking. The widely known optimal subpattern assignment (OSPA) is used as the first metric, which was proposed by Schumacher et al. [126]. OSPA measures two different characteristics, accounting for localization and cardinality errors. Furthermore, it has two adaptable parameters p and c that can be interpreted as the outlier sensitivity and cardinality penalty respectively. In this thesis, they are set to $p = 2$ and $c = 5$.

In order to measure the performance of the bounding box prediction, the intersection over union (IoU) is used as the second metric. The IoU is defined as the area of intersection between the predicted shapes based on synthetic and real radar data, divided by the area of the union of the two shapes.

Implicit - low level evaluation

With the purpose of evaluating the object tracking prediction, the root mean squared error (RMSE) of the longitudinal x-position and lateral y-position estimation are computed. Furthermore, the cardinality estimates are compared by calculating the absolute cardinality error, with the estimate based on real radar measurements as the ground truth.

6.2.4 Measuring the simulation-to-reality gap

After evaluating each level individually in the previous sections, the final step is to combine all results into an overall gap. Therefore, the simulation-to-reality gap G is proposed with the goal to measure the total disparity between the radar simulation and reality. It is calculated as follows:

1. choose a test scenario
2. perform the test drive and record ground truth data and sensor measurements
3. reproduce the tests in a virtual environment in order to generate synthetic sensor data
4. run the perception module with real and synthetic sensor data

5. perform explicit as well as the implicit sensor model evaluation
6. normalize metric results to the interval $[0, 1]$ so that zero represents the best case (no deviation)
7. aggregate evaluation results on each fidelity level
8. compute the averaged gap over all fidelity levels to obtain the combined simulation-to-reality gap G

6.3 Experiments and results

In order to analyze the effectiveness of the developed method in terms of its capability to accurately estimate the gap between simulation and reality, the performed experiments are elaborated in this section. At first, the three radar simulations (IRM, DDM, RTM) are evaluated in-depth for one exemplary scenario to analyze to what degree they can model the real sensor behavior across the different FL. Based on this, the fidelities of the radar simulations are further investigated across different scenario categories.

6.3.1 Single scenario - qualitative evaluation

The radar models are qualitatively evaluated based on their respective object tracking performance. Hereby, they are analyzed to what extent they can approximate the result of the prediction algorithm fed with real radar measurements.

In the example scenario, the sensor is stationary located in $(0, 0)$ and a target vehicle drives a path in form of an eight in front of it. The predictions of the tracking module fed with real and simulated data from each radar model are illustrated in Figure 6.3.

Since the object tracking is optimized with real radar measurements, the predictions of the object position show a relatively small deviation from the true track of the vehicle. Due to the fact that the IRM is rather simple, only small deviations can be observed, similar to the predictions based on real data. These results are expected, because the radar detections are uniformly distributed along the bounding box of the target vehicle. Therefore, the position estimation is simpler in comparison to the more stochastic detections

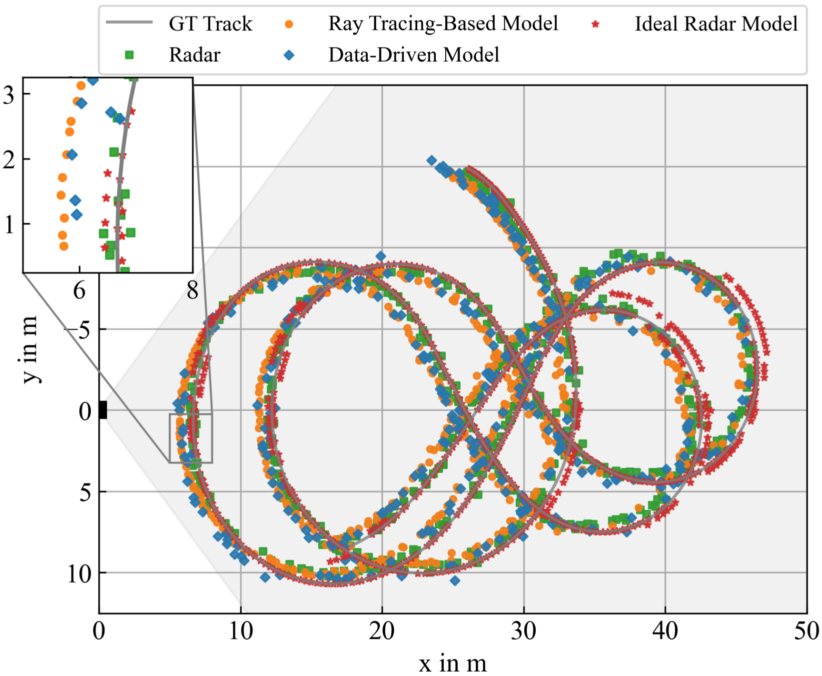

Figure 6.3: The color-coded points indicate the radar-based predicted tracks based on real and simulated radar sensor data, while the gray line represents the ground truth object track (GT Track).

generated by DDM and RTM, as they also model sensor errors, resulting in more noisy data in general. The latter model has a constant offset since the simple ray casting approach is utilized. As a result, the radar points are distributed more along the visible edge of the object, leading to a slightly shifted prediction. Despite the fact that similar results are observed with the synthetic detections generated by DDM, the observed error is relatively smaller. This can be explained by the fact that the radar points are spread over the entire vehicle and are not predominantly concentrated on the outer shell.

6.3.2 Single scenario - quantitative evaluation

Despite that IRM is an ideal radar model, it provides a more realistic estimate in terms of object tracking prediction, i.e. similar to the real radar tracking predictions. With the purpose to measure an accurate overall model fidelity (simulation-to-reality gap), not only the tracking prediction but also the direct result of the model, in this case the radar point cloud, must be considered.

The results of the implicit as well as the explicit sensor model evaluation are combined in order to obtain a comprehensive estimate of the overall gap between simulation and reality. For a holistic evaluation of the sensor model, several metrics are used at each level, each assessing different features in the data. The findings of the approach elaborated in Section 6.2 are presented in Table 6.1.

Here, the value of a metric relates to the mean value across the scenario. Since the resulting range of values can vary greatly between the different metrics, min-max normalization is applied to rescale the data range to $[0, 1]$. Aggregating the individual results of each evaluation level makes it easier to interpret the results of the metrics, as shown in Figure 6.4. In addition, the overall gap between simulation and reality G is shown, allowing a quantification of the overall fidelity with respect to the intended use under study.

Based on the results, it can be observed that the qualitative evaluation by the visual comparison is reflected in the metric results, in the sense that the IRM shows a better performance in the implicit than in the explicit evaluation compared to the other models. However, due to the relatively very large deviations in the explicit comparison, the ideal model shows the highest overall gap.

On the other hand, for the data-driven model and the ray tracing-based

Table 6.1: The assessment results of the radar models implemented are presented for the different fidelity levels. The down arrow (resp. up arrow) indicates that the performance is better if the quantity is smaller (resp. greater).

Fidelity Level	Metric	IRM	DDM	RTM
FL I	$OSPA \downarrow$	0.342	0.314	0.304
	$IoU \uparrow$	0.545	0.347	0.346
FL II	$RMSE_x \downarrow$	0.292	0.287	0.231
	$RMSE_y \downarrow$	0.243	0.258	0.189
	$Cardinality\ Error \downarrow$	0.093	0.0	0.004
FL III	$DPP \downarrow$	0.381	0.051	0.152
	$WD \downarrow$	0.398	0.055	0.173
FL IV	$PNE \downarrow$	0.435	0.304	0.064
	$WD_r \downarrow$	0.26	0.341	0.331
	$WD_\phi \downarrow$	0.089	0.102	0.112
	$WD_{Doppler} \downarrow$	0.409	0.04	0.161

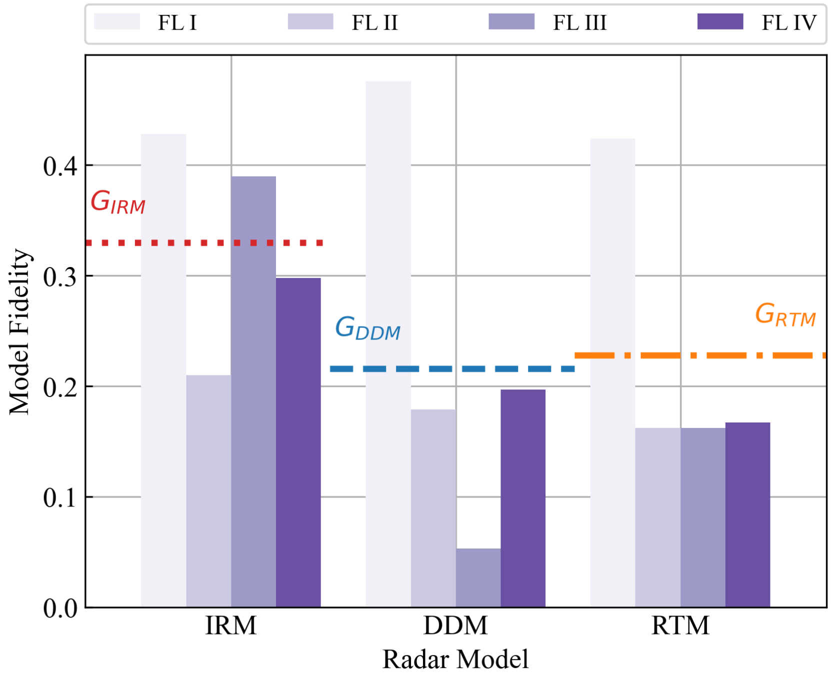

Figure 6.4: Aggregated results of the fidelity levels together with the resulting overall simulation-to-reality gap for scenario 'eight'.

model, a significant difference between FL I and the other levels is clearly visible. This can be explained by a large difference in the bounding box prediction, which translates into a poor IoU score. With the exception of FL II, the results of the other fidelity levels are very similar, which is why the total deviation of DDM and RTM is almost the same.

6.3.3 Evaluation across multiple scenarios

In contrast to the evaluation of a specific scenario, this section extends the evaluation of the accuracy of each model to different categories of scenarios. Multiple different scenario categories are evaluated, because the output of sensor data as well as the performance of object tracking can vary greatly depending on the situation tested. The scenarios are divided into single (s) and multi-object (m) scenarios, as indicated in Table 6.2. In addition, typical difficult scenarios were selected to demonstrate the proposed method. The chosen scenarios do not claim to cover all relevant scenarios, but rather are of exemplary character.

Table 6.2: List of tested scenarios and their description.

Scenario Name	Description
oncoming_s	target enters sensor FOV in far range
overtake_s	target enters FOV in near range and overtakes ego
leading_s	ego follows a leading target
eight_s	target drives an eight in front of static ego
$\text{occlusion}_\text{m}$	multiple targets are occluded
leading_m	ego follows multiple leading targets driving in parallel
overtake_m	multiple targets overtake ego
crossing_m	multiple targets cross in front of ego

The results of the respective gaps between simulation and reality are shown in a radar diagram illustrated in Figure 6.5. It can be seen that the ray tracing-based model achieves the smallest error in almost all tested scenarios. However, the exception is the crossing scenario, where both IRM and RTM perform relatively poorly, while DDM has the smallest deviation from the real data. It is also observed that the deviations are larger in the multiple vehicle scenarios than in the single object scenarios. This is probably due to the fact that the radar models used are relatively simple and therefore do not take into account effects such as multi-path reflections.

6.4 Discussion

In this section, the main contributions of the present chapter to the following research question are discussed RQ4: *How to measure the overall simulation-to-reality gap considering a target application?* A multi-layered sensor model evaluation approach is proposed in order to measure the gap between a radar simulation and the real radar sensor.

To analyze the effectiveness of the developed method, three typical radar models were implemented and their fidelity assessed. It can be shown that the different levels of evaluation introduced can reveal existing discrepancies in detail visible as well as investigate the sensor model fidelity across different scenarios. In addition, this quantitative and objective evaluation approach enables scaling of virtual tests and provides the basis for an informed decision based on simulation results, thus reducing the validation effort of AD functions.

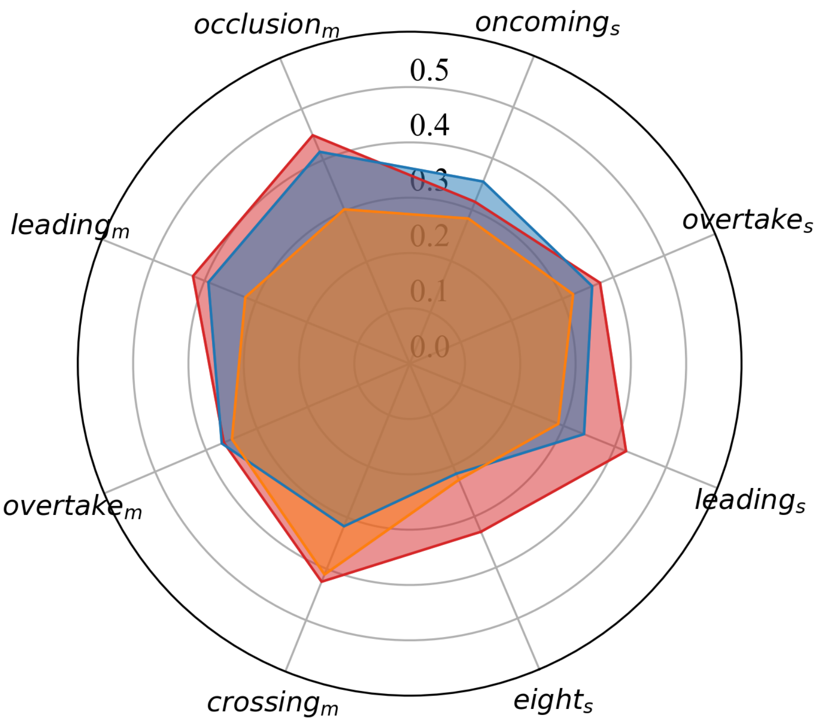

Figure 6.5: The simulation-to-reality gap of each radar model across multiple scenarios with IRM( ), DDM( ), RTM( ).

Despite the fact that this thesis studies the applicability of radar simulations for tracking multiple objects, the presented concept is not exclusively designed for this class of perception module, but also allows further abstractions for other algorithms such as classification. Still, the metrics used for evaluating the implicit perception models would need to be modified to fit the particular use case. Moreover, the multi-layered approach could be straightforwardly adapted to other sensor modalities like lidar sensors, since lidar point clouds are comparable to detections from a radar.

Apart from an extension of the method, some improvements of the approach are also conceivable. An apparent enhancement exists in the metric calculation, because some information might be lost due to the averaging of the evaluation results over a scenario run. Therefore, it needs to be investigated to what extent a deeper analysis of the time series can improve the overall model estimation. Additionally, it was assumed that the used metrics on each FL are weighted equally. Although this is a reasonable procedure, the weighting problem needs to be further explored in future work. Similar to the weighting of the metrics, it was assumed that each model fidelity level is equally important, but depending on the use case, a similar perception result may be favored over an exact radar point cloud and vice versa. Therefore, it is necessary to determine what model quality requirements are needed for the varying use cases.

7 Conclusion and Outlook

Contents

Simulation-based testing in the context of AD relies inter alia on accurate models of the environment, sensors, and their signal propagation. These models are by necessity approximations, but must be designed to adequately represent the real world. This thesis is a step towards developing sound and validated sensor simulation approaches for virtual testing of autonomous driving functions, serving the broader goal of gaining trust in simulation and synthetic data to reduce the overall validation effort and the need for real-world testing. This chapter summarizes the approach taken in this thesis, highlights the main contributions made, and discusses the limitations of the developed methods. This is complemented with an outlook of further potential research directions.

7.1 Conclusion

Having virtual tests does not eliminate the need for a significant number of tests in a real environment. Rather, the overarching objective of this dissertation was to investigate the extent to which a sensor simulation can be trusted and to measure this trust in a quantifiable way. By analyzing the capabilities and limitations of a sensor model, reliable predictions about the real system can be derived based on tests in a virtual environment.

Therefore, different approaches for the validation of a sensor simulation were provided in this thesis along with techniques to model a radar simulation for virtually testing AD functions such as object detection and tracking. This work proposed several contributions to different aspects of sensor model validation.

In the following, the developed approaches are recapitulated and discussed with respect to the research problems derived and formulated in Section 2.4.

RQ1: *What is missing in existing sensor model validation approaches?*
The first research problem is addressed in Chapter 2. Existing approaches on radar simulation and sensor model validation were structured regarding the validation method used.

The most significant observation is that there exist no systematic proof of fidelity that provides an objective and quantitative method for the validation of a radar simulation for AD. Although most approaches focus on the direct assessment of the synthetically generated radar data, merely three authors can back up their sensor model evaluation with objective numbers. Most of the proposals presented perform a qualitative evaluation, often by visually comparing the generated results with expectations. A few also do not assess the simulated data at all. Additionally, since there is currently no benchmark for objective assessment of sensor model accuracy, many approaches proclaim their validity based on a momentary observation in individual scenarios. However, no approach exist that investigates a larger number of scenarios.

Besides a direct evaluation of the sensor data, some approaches include a downstream target application in the assessment as well. Nevertheless, these methods are mainly applied to data from a lidar sensor. This raises the question of whether these approaches are also suitable for radar simulations, since radar data are typically much more sparse and stochastic in nature.

A detailed derivation of the resulting research problems can be found in Section 2.4.

RQ2: *Which features of the radar simulation are relevant for a downstream application?*
Chapter 4 deals with the second research question. A sensitivity analysis approach is developed to identify the relevant sensor characteristics w.r.t. to a target application.

Furthermore, the method proposed assesses a sensor simulation with the objective to determine the impact of specific sensor effects regarding a desired use case. This approach can generally be applied to various simulation models and was used in this work especially for a radar simulation, which in particular often has computationally intensive requirements. Moreover, a spatial clustering algorithm is used as a typical target application that processes radar data. The algorithm is fed with simulated as well as real radar measurements with the purpose of comparing both prediction results. In order to perform the sensitivity analysis, the FAST algorithm is used taking the result of the clustering evaluation as input. This serves the purpose of drawing conclusions about the relevance of individual parameters or features based on the measured sensitivity.

The investigation of the effectiveness of the proposed approach can show that a sensitivity analysis allows a more detailed, objective and measurable evaluation of the synthetically generated radar data compared to a qualitative assessment. Furthermore, the results from specific situations can be attributed to the contribution of each sensor property of the radar simulation, leading to an efficient analysis of the simulation result.

RQ3: *How to determine the degree to which the radar simulation and experimental measurements concur?*
This section recapitulates and discusses the main contributions of Chapter 5, which addresses the third research problem.

A novel data-driven metric is proposed to objectively measure the latent features of radar point clouds in order to estimate the accuracy of a radar simulation. The machine learning-based metric, DEM, is introduced to estimate the fidelity simulated radar detections. In order to investigate the effectiveness of the proposed metric, conventional metrics are used and their ability to measure the discrepancy between real and synthetically generated sensor data is compared.

From the conducted experiments, it is found that the developed DEM is able to detect the shortcomings of the simulated radar point clouds in the near-field region, unlike the conventional metrics. However, it was not possible to capture all radar characteristics, such as the inaccurate number of detections at long ranges, which none of the metrics could. All in all, the developed metric shows great potential as it could reproduce the expected outcome of a qualitative evaluation much better than the other conventional metrics.

RQ4: *How to measure the overall simulation-to-reality gap considering a target*

application?
The present research question is addressed in Chapter 6 and the main contributions and findings are discussed in the following.

A multi-layered approach to sensor model evaluation is developed with the goal of estimating the simulation fidelity of a radar simulation for virtual testing of perception algorithms. Four different fidelity levels are introduced to allow both a direct evaluation of the sensor data (explicit sensor model evaluation) and the consideration of a target application that further processes the sensor data (implicit sensor model evaluation). On the other hand, this multi-level validation procedure provides an in-depth evaluation of specific aspects of the simulation as well as a holistic evaluation resulting from the combination of all fidelity levels.

With the purpose of investigating the efficiency proposed validation approach, three typical radar simulations are implemented and their accuracy is assessed. From the results, it can be shown that the different levels of evaluation introduced can reveal existing differences in detail visible as well as examine the sensor model across different scenarios. Therefore, this objective and quantitative method makes the scaling of simulation-based tests possible and provides the basis for a profound decision based on tests in a virtual environment.

7.2 Further work

The present section addresses the remaining two research questions for the validation of a sensor simulation from Section 2.4. Following this, general aspects for further research directions are discussed.

RQ5: *What sensor model fidelity is sufficient for which intended use?*
Apart from the problem of measuring the accuracy of a sensor simulation, i.e. the gap between real and synthetic sensor data, the question remains what fidelity is sufficient for which use case. Due to the high demands on the execution speed, a sufficient degree of realism must be found. Since a certain discrepancy between simulation and reality can be assumed, the problem remains which error is an acceptable discrepancy.

One conceivable approach is to include safety-relevant aspects [149]. In this respect, the fidelity evaluation is extended by an additional dimension with regard to the significance for safety. Thus, differences could be neglected if they are not meaningful for the safety validation of the autonomous vehicle. An example of this would be a discrepancy that is in the far range where it is unlikely that critical situation will arise. On the other hand, a discrepancy between simulated and real sensor data in the close range directly in front of the ego-vehicle should be weighted more strongly, since an imminent threat could be present. In this way, the lowest required model fidelity could be determined at which the same criticality still prevails in the simulation as in the real reference scenario.

RQ6: *Which scenarios need to be tested for the application domain?*
As the ground truth generation of the test drives is very complex, e.g. in dense urban environments, the tests are often carried out on a simplified test site. Since the reference data such as sensor measurements are then also recorded on the test site, the problem of transferability of the results arises. Novel methods are needed to overcome this gap between validation and application domain.

There exist different approaches that proposed methods to find and generate critical scenarios. However, these often focus on the system level of an AD system [150–152] or on other sensor modalities such as lidar [153] or camera [154]. To the best of the author's knowledge, there are no methods applied to radar simulations in particular. In addition, these methods focus on finding critical scenarios related to system safety, but do not consider what tests are required to validate a sensor model. Therefore, a possible approach is to investigate how these methods can be adapted to validate a radar simulation. For this purpose, the implemented models and the developed evaluation methods from this work could be used.

General future research directions

In addition to the research problems formulated for the validation of a sensor simulation, other research directions dealing with simulation-based testing in general are discussed in the following.

The present dissertation focused on the validation of a radar simulation. However, additional novel methods are needed to compare different sensor models while ensuring comparability of results. One possible solution to this could be a common benchmark for model evaluation involving a public database of sensor recordings, similar to the well-known KITTI [155] or nuScenes [156] for perception algorithms. Apart from the actual sensor data, this still requires additional development work with regard to, for example, standardized interfaces and formats for virtual maps or scenario descriptions. Some formats already exist, but they have not yet found wide acceptance and need to be further developed and matured.

As described in Chapter 2, the environment model is an essential part of a sensor simulation. The modeling effort of new environments is very time consuming and expensive. Therefore, novel methods are needed to automate the modeling process. This is very challenging because, in addition to information about the infrastructure physical properties, such as roughness values of objects, also play an important role, especially for the generation of sensor data. Addressing this issue would also help to bridge the gap between validation and scope described earlier.

Bibliography

[1] H. Winner, W. Wachenfeld, and P. Junietz, "Validation and Introduction of Automated Driving," in *Automotive Systems Engineering II* (H. Winner, G. Prokop, and M. Maurer, eds.), pp. 177–196, Cham: Springer International Publishing, 2018. (Cited on pages xiii, 2, 3, 4, and 6.)

[2] W. Wachenfeld and H. Winner, "The Release of Autonomous Vehicles," in *Autonomous Driving: Technical, Legal and Social Aspects* (M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, eds.), pp. 425–449, Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. (Cited on pages xiii, 2, and 4.)

[3] J. Steinbaeck, C. Steger, G. Holweg, and N. Druml, "Next generation radar sensors in automotive sensor fusion systems," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, (Bonn), pp. 1–6, IEEE, Oct. 2017. (Cited on pages xiii and 13.)

[4] T. Ponn, F. Mueller, and F. Diermeyer, "Systematic Analysis of the Sensor Coverage of Automated Vehicles Using Phenomenological Sensor Models," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, (Paris, France), pp. 1000–1006, IEEE, June 2019. (Cited on pages xiii, 13, 19, 29, and 43.)

[5] J. Schrepfer, V. Picron, J. Mathes, and H. Barth, "Automatisiertes Fahren undseine Sensorik im Test," *ATZ - Automobiltechnische Zeitschrift*, vol. 120, pp. 28–37, Jan. 2018. (Cited on pages xiii and 13.)

[6] Daimler, "Test and Technology Center in Immendingen: Consolidation of worldwide vehicle testing in one location," 2019. Accessed on: Sep. 03, 2021, Available: https://media.daimler.com/marsMediaSite/en/instance/ko/Test-and-Technology-Center-in-Immendingen-Consolidation-of-worldwide-

vehicle-testing-in-one-location.xhtml?oid=44800150. (Cited on pages xiii and 37.)

[7] F. Kuhnert and C. Stuermer, "Five trends transforming the Automotive Industry," *PwC*, 2018. Accessed on: Aug. 24, 2021, Available: https://www.pwc.com/gx/en/industries/automotive/publications/eascy.html. (Cited on page 1.)

[8] B. Friedrich, "The Effect of Autonomous Vehicles on Traffic," in *Autonomous Driving: Technical, Legal and Social Aspects* (M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, eds.), pp. 317–334, Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. (Cited on page 1.)

[9] S. Buhr, "Elon Musk, Sam Altman Predict Self-Driving Cars Will Be On The Road In Just A Few Years," *TechCrunch*, 2015. Accessed on: Aug. 23, 2021, Available: https://techcrunch.com/2015/10/06/elon-musk-sam-altman-say-self-driving-cars-are-going-to-be-on-the-road-in-just-a-few-years/. (Cited on page 1.)

[10] M. Kosoff, "Lyft's President says Car Ownership will "All But End" by 2025," *Vanity Fair*, 2016. Accessed on Aug. 23, 2021, Available: https://www.vanityfair.com/news/2016/09/lyfts-president-says-car-ownership-will-all-but-end-by-2025. (Cited on page 1.)

[11] C. Mims, "Self-Driving Cars Could Be Decades Away, No Matter What Elon Musk Said," *Wall Street Journal*, 2021. Accessed on: Aug. 23, 2021, Available: https://www.wsj.com/articles/self-driving-cars-could-be-decades-away-no-matter-what-elon-musk-said-11622865615. (Cited on page 1.)

[12] SAE, "Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems," SAE Standard J3016, 2018. (Cited on page 1.)

[13] P. Koopman and M. Wagner, "Challenges in Autonomous Vehicle Testing and Validation," *SAE International Journal of Transportation Safety*, vol. 4, pp. 15–24, Apr. 2016. (Cited on page 1.)

[14] J. E. Stellet, M. Woehrle, T. Brade, A. Poddey, and W. Branz, "Validation of automated driving – a structured analysis and survey of approaches," in *13. Uni-DAS e.V. Workshop Fahrerassistenz und automatisiertes Fahren*, (Walting, Germany), p. 10, 2020. (Cited on pages 1, 15, and 71.)

[15] S. Jesenski, J. E. Stellet, W. Branz, and J. M. Zollner, "Simulation-Based Methods for Validation of Automated Driving: A Model-Based Analysis and an Overview about Methods for Implementation," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, (Auckland, New Zealand), pp. 1914–1921, IEEE, Oct. 2019. (Cited on pages 1, 3, 5, and 71.)

[16] A. Schaermann, A. Rauch, N. Hirsenkorn, T. Hanke, R. Rasshofer, and E. Biebl, "Validation of vehicle environment sensor models," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, (Los Angeles, CA, USA), pp. 405–411, IEEE, June 2017. (Cited on pages 2, 28, 29, 45, and 72.)

[17] W. L. Oberkampf, "Simulation Accuracy, Uncertainty, and Predictive Capability: A Physical Sciences Perspective," in *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives* (C. Beisbart and N. J. Saam, eds.), pp. 69–97, Cham: Springer International Publishing, 2019. (Cited on pages 2, 24, and 46.)

[18] M. Holder, P. Rosenberger, H. Winner, T. Dhondt, V. P. Makkapati, M. Maier, H. Schreiber, Z. Magosi, Z. Slavik, O. Bringmann, and W. Rosenstiel, "Measurements revealing Challenges in Radar Sensor Modeling for Virtual Validation of Autonomous Driving," in *International Conference on Intelligent Transportation Systems (ITSC)*, (Maui, Hawaii, USA), pp. 2616–2622, IEEE, Nov. 2018. (Cited on pages 2, 19, 20, 25, 26, 27, 39, 57, and 72.)

[19] G. E. P. Box, "Robustness in the Strategy of Scientific Model Building," in *Robustness in Statistics*, pp. 201–236, Academic Press, 1979. (Cited on pages 3 and 23.)

[20] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, eds., *Autonomous Driving: Technical, Legal and Social Aspects*. Berlin Heidelberg: Springer Open, 2016. (Cited on page 3.)

[21] T. A. Wheeler, *Automotive Safety Validation in Simulation*. Ph.D. dissertation, Stanford University, Stanford, California, USA, 2018. (Cited on pages 4 and 5.)

[22] N. Katzorke, M. Moosmann, R. Imdahl, and H. Lasi, "A Method to Assess and Compare Proving Grounds in the Context of Automated Driving Systems," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, (Rhodes), pp. 1–6, IEEE, Sept. 2020. (Cited on pages 4 and 37.)

[23] P. Wang, S. Sibi, B. Mok, and W. Ju, "Marionette: Enabling On-Road Wizard-of-Oz Autonomous Driving Studies," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, (Vienna Austria), pp. 234–243, ACM, Mar. 2017. (Cited on page 5.)

[24] T. Bock, M. Maurer, and G. Faerber, "Validation of the Vehicle in the Loop (VIL); A milestone for the simulation of driver assistance systems," in *2007 IEEE Intelligent Vehicles Symposium*, (Istanbul, Turkey), pp. 612–617, IEEE, 2007. (Cited on page 5.)

[25] S. Solmaz, M. Rudigier, and M. Mischinger, "A Vehicle-In-The-Loop Methodology for Evaluating Automated Driving Functions in Virtual Traffic," in *IEEE Intelligent Vehicles Symposium*, (Las Vegas, USA), p. 7, IEEE, 2020. (Cited on page 5.)

[26] F. M. Maier, V. P. Makkapati, and M. Horn, "Environment perception simulation for radar stimulation in automated driving function testing," *e & i Elektrotechnik und Informationstechnik*, vol. 135, pp. 309–315, Aug. 2018. (Cited on pages 5 and 23.)

[27] J. Ma, C. Schwarz, Z. Wang, M. Elli, G. Ros, and Y. Feng, "New Simulation Tools for Training and Testing Automated Vehicles," in *Road Vehicle Automation 7* (G. Meyer and S. Beiker, eds.), pp. 111–119, Cham: Springer International Publishing, 2020. Series Title: Lecture Notes in Mobility. (Cited on page 5.)

[28] M. Nolte, R. Schubert, C. Reisch, and M. Maurer, "Sensitivity Analysis for Vehicle Dynamics Models – An Approach to Model Quality Assessment for Automated Vehicles," in *Intelligent Vehicle Symposium*, (Las Vegas, USA), IEEE, 2020. arXiv: 2005.03872. (Cited on page 5.)

[29] B. Schlager, S. Muckenhuber, S. Schmidt, H. Holzer, R. Rott, F. M. Maier, K. Saad, M. Kirchengast, G. Stettinger, D. Watzenig, and J. Ruebsam, "State-of-the-Art Sensor Models for Virtual Testing of Advanced Driver Assistance Systems/Autonomous Driving Functions," *SAE International Journal of Connected and Automated Vehicles*, vol. 3, pp. 233–261, Oct. 2020. (Cited on pages 5, 6, 19, and 72.)

[30] A. P. Sligar, "Machine Learning-Based Radar Perception for Autonomous Vehicles Using Full Physics Simulation," *IEEE Access*, vol. 8, pp. 51470–51476, 2020. (Cited on pages 6, 22, 29, and 72.)

[31] M. Hartstern, V. Rack, M. Kaboli, and W. Stork, "Simulation-based Evaluation of Automotive Sensor Setups for Environmental Perception in Early Development Stages," p. 7, 2020. (Cited on pages 6, 29, and 72.)

[32] T. Ponn, T. Lanz, and F. Diermeyer, "Automatic Generation of Road Geometries to Create Challenging Scenarios for Automated Vehicles Based on the Sensor Setup," *arXiv:2008.11580 [cs]*, Aug. 2020. arXiv: 2008.11580. (Cited on pages 6 and 72.)

[33] T. A. Wheeler, M. Holder, H. Winner, and M. Kochenderfer, "Deep Stochastic Radar Models," *arXiv:1701.09180 [cs]*, June 2017. arXiv: 1701.09180. (Cited on pages 6, 20, 26, and 45.)

[34] U. Chipengo, A. P. Sligar, and S. Carpenter, "High Fidelity Physics Simulation of 128 Channel MIMO Sensor for 77GHz Automotive Radar," *IEEE Access*, pp. 160643 – 160652, 2020. (Cited on pages 6, 25, 26, and 72.)

[35] M. I. Skolnik, *Radar Handbook*. New York, USA: McGraw-Hill, 3rd ed., 2008. (Cited on pages 6, 13, 14, 21, 41, 45, and 50.)

[36] A. Ngo, M. P. Bauer, and M. Resch, "A Sensitivity Analysis Approach for Evaluating a Radar Simulation for Virtual Testing of Autonomous

Driving Functions," in *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, (Singapore, Singapore), pp. 122–128, IEEE, 2020. (Cited on pages 8, 23, and 114.)

[37] A. Ngo, M. P. Bauer, and M. Resch, "Deep Evaluation Metric: Learning to Evaluate Simulated Radar Point Clouds for Virtual Testing of Autonomous Driving," in *2021 IEEE Radar Conference (RadarConf21)*, (Atlanta, GA, USA), pp. 1–6, IEEE, 2021. (Cited on pages 8, 9, and 114.)

[38] A. Ngo, M. P. Bauer, and M. Resch, "A Multi-Layered Approach for Measuring the Simulation-to-Reality Gap of Radar Perception for Autonomous Driving," in *24th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, (Indianapolis, IN, USA), p. 7, IEEE, 2021. (Cited on pages 8, 9, 23, and 114.)

[39] C. Huelsemayer, "Procedure for reporting removed metallic objects to an observer using electric waves." German Patent 165546C, 1904. (Cited on page 12.)

[40] G. Galati and P. van Genderen, "History of radar: The need for further analysis and disclosure," in *2014 11th European Radar Conference*, (Italy), pp. 25–28, IEEE, 2014. (Cited on page 12.)

[41] M. I. Skolnik, *Introduction to Radar Systems*. Boston, MA, USA: McGraw-Hill, 3rd ed., 2001. (Cited on pages 12, 13, 14, 41, 43, and 50.)

[42] H. Winner, "Automotive RADAR," in *Handbook of Driver Assistance Systems* (H. Winner, S. Hakuli, F. Lotz, and C. Singer, eds.), pp. 325–403, Cham: Springer International Publishing, 2016. (Cited on pages 12 and 14.)

[43] H. Winner, "Fundamentals of Collision Protection Systems," in *Handbook of Driver Assistance Systems*, pp. 1149–1176, Cham: Springer International Publishing, 2016. (Cited on page 12.)

[44] A. Bartels, M.-M. Meinecke, and S. Steinmeyer, "Lane Change Assistance," in *Handbook of Driver Assistance Systems*, pp. 1235–1257, Cham: Springer International Publishing, 2016. (Cited on page 12.)

[45] J. Wenger, "Automotive radar - status and perspectives," in *IEEE Compound Semiconductor Integrated Circuit Symposium, 2005. CSIC '05.*, (Palm Springs, CA, USA), p. 4 pp., IEEE, 2005. (Cited on page 12.)

[46] J. Dickmann, N. Appenrodt, J. Klappstein, H.-L. Bloecher, M. Muntzinger, A. Sailer, M. Hahn, and C. Brenk, "Making Bertha See Even More: Radar Contribution," *IEEE Access*, vol. 3, pp. 1233–1247, 2015. (Cited on page 12.)

[47] F. Roos, J. Bechter, C. Knill, B. Schweizer, and C. Waldschmidt, "Radar Sensors for Autonomous Driving: Modulation Schemes and Interference Mitigation," *IEEE Microwave Magazine*, vol. 20, pp. 58–72, Aug. 2019. (Cited on pages 12, 15, and 17.)

[48] J. Dickmann, J. Klappstein, M. Hahn, N. Appenrodt, H.-L. Bloecher, K. Werber, and A. Sailer, "Automotive radar the key technology for autonomous driving: From detection and ranging to environmental understanding," in *IEEE Radar Conference (RadarConf)*, (Philadelphia, PA, USA), pp. 1–6, IEEE, May 2016. (Cited on pages 12, 16, 17, 45, and 57.)

[49] J. Gamba, *Radar Signal Processing for Autonomous Driving*. Signals and Communication Technology, Singapore: Springer Singapore, 2020. (Cited on pages 13, 14, 15, 41, and 50.)

[50] N. Scheiner, O. Schumann, F. Kraus, N. Appenrodt, J. Dickmann, and B. Sick, "Off-the-shelf sensor vs. experimental radar – How much resolution is necessary in automotive radar classification?," *arXiv:2006.05485 [cs, eess]*, June 2020. arXiv: 2006.05485. (Cited on pages 14 and 15.)

[51] E. B. Kamel, A. Peden, and P. Pajusco, "RCS modeling and measurements for automotive radar applications in the W band," in *2017 11th European Conference on Antennas and Propagation (EUCAP)*, (Paris, France), pp. 2445–2449, IEEE, Mar. 2017. (Cited on page 14.)

[52] J. Dickmann, "Radar for Autonomous Driving – Paradigm Shift from Mere Detection to Semantic Environment Understanding," p. 17, 2019. (Cited on pages 15, 16, and 17.)

[53] S. Haag, B. Duraisamy, W. Koch, and J. Dickmann, "Radar and Lidar Target Signatures of Various Object Types and Evaluation of Extended Object Tracking Methods for Autonomous Driving Applications," in *2018 21st International Conference on Information Fusion (FUSION)*, (Cambridge), pp. 1746–1755, IEEE, July 2018. (Cited on page 15.)

[54] T. S. Madhulatha, "An Overview on Clustering Methods," *IOSR Journal of Engineering*, vol. 02, pp. 719–725, Apr. 2012. (Cited on page 15.)

[55] M. Li, M. Stolz, Z. Feng, M. Kunert, R. Henze, and F. Kucukay, "An Adaptive 3D Grid-Based Clustering Algorithm for Automotive High Resolution Radar Sensor," in *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, (Madrid), pp. 1–7, IEEE, Sept. 2018. (Cited on page 15.)

[56] S. Lim, S. Lee, and S.-C. Kim, "Clustering of Detected Targets Using DBSCAN in Automotive Radar Systems," in *International Radar Symposium (IRS)*, (Bonn, Germany), pp. 1–7, IEEE, June 2018. (Cited on page 15.)

[57] A. Breloy, M. N. El Korso, A. Panahi, and H. Krim, "Robust Subspace Clustering for Radar Detection," in *European Signal Processing Conference (EUSIPCO)*, (Rome), pp. 1602–1606, IEEE, Sept. 2018. (Cited on page 15.)

[58] J. Schlichenmaier, F. Roos, P. Hugler, and C. Waldschmidt, "Clustering of Closely Adjacent Extended Objects in Radar Images using Velocity Profile Analysis," in *IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, (Detroit, MI, USA), pp. 1–4, IEEE, Apr. 2019. (Cited on page 15.)

[59] C. Ebersole, A. Buchenroth, D. Zilz, and V. Chakravarthy, "Spatiotemporal Density-Based Clustering for Dynamic Spectrum Sensing," in *2020 IEEE International Radar Conference (RADAR)*, (Washington, DC, USA), pp. 720–725, IEEE, Apr. 2020. (Cited on page 15.)

[60] T. Wagner, R. Feger, and A. Stelzer, "Modifications of the OPTICS Clustering Algorithm for Short-Range Radar Tracking Applications," in

European Radar Conference (EuRAD), (Madrid), pp. 91–94, IEEE, Sept. 2018. (Cited on page 15.)

[61] X. Feng, X. Hu, and Y. Liu, "Radar signal sorting algorithm of k-means clustering based on data field," in *IEEE International Conference on Computer and Communications (ICCC)*, (Chengdu), pp. 2262–2266, IEEE, Dec. 2017. (Cited on page 15.)

[62] A. Danzer, T. Griebel, M. Bach, and K. Dietmayer, "2D Car Detection in Radar Data with PointNets," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, (Auckland, New Zealand), pp. 61–66, IEEE, Oct. 2019. (Cited on pages 15, 25, and 62.)

[63] N. Scheiner, N. Appenrodt, J. Dickmann, and B. Sick, "Radar-based Road User Classification and Novelty Detection with Recurrent Neural Network Ensembles," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, (Paris, France), pp. 722–729, IEEE, 2019. (Cited on page 16.)

[64] K. Patel, K. Rambach, T. Visentin, D. Rusev, M. Pfeiffer, and B. Yang, "Deep Learning-based Object Classification on Automotive Radar Spectra," in *2019 IEEE Radar Conference (RadarConf)*, (Boston, MA, USA), pp. 1–6, IEEE, 2019. (Cited on page 16.)

[65] M. Ulrich, C. Glaser, and F. Timm, "DEEPREFLECS: Deep Learning for Automotive Object Classification with Radar Reflections," in *2021 IEEE International Radar Conference (RADAR)*, (Atlanta, GA, USA), p. 6, IEEE, 2021. (Cited on pages 16, 21, and 25.)

[66] O. Schumann, M. Hahn, J. Dickmann, and C. Woehler, "Semantic Segmentation on Radar Point Clouds," in *2018 21st International Conference on Information Fusion (FUSION)*, (Cambridge), pp. 2179–2186, IEEE, July 2018. (Cited on pages 16 and 62.)

[67] O. Schumann, C. Wohler, M. Hahn, and J. Dickmann, "Comparison of random forest and long short-term memory network performances in classification tasks using radar," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, (Bonn), pp. 1–6, IEEE, Oct. 2017. (Cited on page 16.)

[68] O. Schumann, J. Lombacher, M. Hahn, C. Wohler, and J. Dickmann, "Scene Understanding With Automotive Radar," *IEEE Transactions on Intelligent Vehicles*, vol. 5, pp. 188–203, June 2020. (Cited on pages 16, 17, and 57.)

[69] A. Palffy, J. Dong, J. F. P. Kooij, and D. M. Gavrila, "CNN based Road User Detection using the 3D Radar Cube," *IEEE Robotics and Automation Letters*, vol. 5, pp. 1263–1270, Apr. 2020. arXiv: 2004.12165. (Cited on page 16.)

[70] S. Haag, B. Duraisamy, W. Koch, and J. Dickmann, "Classification Assisted Tracking for Autonomous Driving Domain," in *2018 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, (Bonn), pp. 1–8, IEEE, Oct. 2018. (Cited on page 16.)

[71] T. Akita and S. Mita, "Object Tracking and Classification Using Millimeter-Wave Radar Based on LSTM," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, (Auckland, New Zealand), pp. 1110–1115, IEEE, Oct. 2019. (Cited on pages 16 and 25.)

[72] K. Granstrom, C. Lundquist, and U. Orguner, "Tracking rectangular and elliptical extended targets using laser measurements," in *14th International Conference on Information Fusion*, (Chicago, Usa), p. 8, IEEE, 2011. (Cited on page 16.)

[73] N. Scheiner, F. Kraus, F. Wei, B. Phan, F. Mannan, N. Appenrodt, W. Ritter, J. Dickmann, K. Dietmayer, B. Sick, and F. Heide, "Seeing Around Street Corners: Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar," p. 10, 2020. (Cited on page 16.)

[74] K. Granstrom, M. Baum, and S. Reuter, "Extended Object Tracking: Introduction, Overview and Applications," *Journal of Advances in Information Fusion*, vol. 12, pp. 139–174, Dec. 2017. (Cited on page 16.)

[75] F. E. Nowruzi, D. Kolhatkar, P. Kapoor, F. Al Hassanat, E. J. Heravi, R. Laganiere, J. Rebut, and W. Malik, "Deep Open Space Segmentation using Automotive Radar," in *2020 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, (Linz, Austria), pp. 1–4, IEEE, Nov. 2020. (Cited on pages 17 and 25.)

[76] M. Li, Z. Feng, M. Stolz, M. Kunert, R. Henze, and F. Küçükay, "High Resolution Radar-based Occupancy Grid Mapping and FreeSpace Detection:," in *Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems*, (Funchal, Madeira, Portugal), pp. 70–81, SCITEPRESS - Science and Technology Publications, 2018. (Cited on page 17.)

[77] M. F. Holder, C. Linnhoff, P. Rosenberger, C. Popp, and H. Winner, "Modeling and Simulation of Radar Sensor Artifacts for Virtual Testing of Autonomous Driving," in *9th Conference Driver Assistance*, (Munich, Germany), p. 9, mediaTUM, 2019. (Cited on page 18.)

[78] P. Cao, *Modeling Active Perception Sensors for Real-Time Virtual Validation of Automated Driving Systems*. Ph.D. dissertation, Technical University of Darmstadt, Darmstadt, Germany, 2017. (Cited on pages 18, 19, and 26.)

[79] Z. Slavik and K. V. Mishra, "Phenomenological Modeling of Millimeter-Wave Automotive Radar," in *URSI Asia-Pacific Radio Science Conference (AP-RASC)*, (New Delhi, India), pp. 1–4, IEEE, Mar. 2019. (Cited on pages 19 and 26.)

[80] P. Rosenberger, J. T. Wendler, M. Holder, C. Linnhoff, M. Berghoefer, H. Winner, and M. Maurer, "Towards a Generally Accepted Validation Methodology for Sensor Models - Challenges, Metrics, and First Results," in *Graz Symposium Virtual Vehicle (GSVF) 2019*, (Graz, Austria), p. 13, 2019. (Cited on pages 19, 28, 29, 57, 72, and 73.)

[81] M. F. Holder, *Synthetic Generation of Radar Sensor Data for Virtual Validation of Autonomous Driving*. Ph.D. dissertation, Technical University of Darmstadt, Darmstadt, Germany, 2021. (Cited on page 19.)

[82] S. Muckenhuber, H. Holzer, J. Rubsam, and G. Stettinger, "Object-based sensor model for virtual testing of ADAS/AD functions," in *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, (Graz, Austria), pp. 1–6, IEEE, Nov. 2019. (Cited on pages 19, 24, and 26.)

[83] N. Hirsenkorn, P. Subkowski, T. Hanke, A. Schaermann, A. Rauch, R. Rasshofer, and E. Biebl, "A ray launching approach for modeling an FMCW radar system," in *International Radar Symposium (IRS)*, (Prague, Czech Republic), pp. 1–10, IEEE, June 2017. (Cited on pages 19, 26, and 27.)

[84] Z. Yun and M. F. Iskander, "Ray Tracing for Radio Propagation Modeling: Principles and Applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015. (Cited on page 20.)

[85] M. Clemens and T. Weiland, "Discrete Electromagnetism with the Finite Integration Technique," *Progress In Electromagnetics Research*, vol. 32, pp. 65–87, 2001. (Cited on page 20.)

[86] J.-M. Jin, *The finite element method in electromagnetics*. Hoboken. New Jersey: John Wiley & Sons Inc, third edition ed., 2014. (Cited on page 20.)

[87] R. Harrington, *Field computation by moment methods*. Macmillan series in electrical science, Macmillan, 1968. (Cited on page 20.)

[88] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on Antennas and Propagation*, vol. 14, pp. 302–307, May 1966. (Cited on page 20.)

[89] T. Machida and T. Owaki, "Rapid and Precise Millimeter-wave Radar Simulation for ADAS Virtual Assessment," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, (Auckland, New Zealand), pp. 431–436, IEEE, Oct. 2019. (Cited on pages 20 and 26.)

[90] T. Owaki and T. Machida, "Hybrid Physics-Based and Data-Driven Approach to Estimate the Radar Cross-Section of Vehicles," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, (Auckland, New Zealand), pp. 673–678, IEEE, Oct. 2019. (Cited on pages 20, 47, and 72.)

[91] J. B. Keller, "Geometrical Theory of Diffraction," *Journal of the Optical Society of America*, vol. 52, pp. 116–130, Feb. 1962. Publisher: Optical Society of America. (Cited on pages 20 and 40.)

[92] U. Chipengo, P. M. Krenz, and S. Carpenter, "From Antenna Design to High Fidelity, Full Physics Automotive Radar Sensor Corner Case Simulation," *Modelling and Simulation in Engineering*, vol. 2018, pp. 1–19, Dec. 2018. (Cited on pages 21, 25, 26, and 45.)

[93] B. Major, D. Fontijne, A. Ansari, R. T. Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. Subramanian, "Vehicle Detection With Automotive Radar Using Deep Learning on Range-Azimuth-Doppler Tensors," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, (Seoul, Korea (South)), pp. 924–932, IEEE, Oct. 2019. (Cited on page 21.)

[94] S. Bernsteiner, Z. Magosi, D. Lindvai-Soos, and A. Eichberger, "Radar Sensor Model for the Virtual Development Process," *ATZelectronics worldwide*, vol. 10, pp. 46–52, Apr. 2015. (Cited on pages 22, 26, 43, and 72.)

[95] N. Scheiner, S. Haag, N. Appenrodt, B. Duraisamy, J. Dickmann, M. Fritzsche, and B. Sick, "Automated Ground Truth Estimation for Automotive Radar Tracking Applications With Portable GNSS And IMU Devices," in *International Radar Symposium (IRS)*, (Ulm, Germany), pp. 1–10, IEEE, June 2019. (Cited on page 22.)

[96] M. Stolz and G. Nestlinger, "Fast generic sensor models for testing highly automated vehicles in simulation," *e & i Elektrotechnik und Informationstechnik*, vol. 135, pp. 365–369, Aug. 2018. (Cited on pages 24 and 26.)

[97] U. Chipengo, "Full Physics Simulation Study of Guardrail Radar-Returns for 77 GHz Automotive Radar Systems," *IEEE Access*, vol. 6, pp. 70053–70060, 2018. (Cited on pages 25 and 26.)

[98] A. Haider, M. Thumann, J. Kim, A. Eryildirim, M. Gerstmair, T. Zeh, and S.-A. Schneider, "Integration of Phase Noise into a Virtual Test Driving Software to Investigate the Impact on the Radar Performance," in *2020 21st International Radar Symposium (IRS)*, (Warsaw, Poland), pp. 328–333, IEEE, Oct. 2020. (Cited on page 26.)

[99] M. Dudek, I. Nasr, D. Kissinger, R. Weigel, and G. Fischer, "The impact of antenna characteristics on target detection in FMCW-radar system

simulations for automotive applications," in *IEEE Radio and Wireless Symposium*, (Santa Clara, CA, USA), pp. 107–110, IEEE, Jan. 2012. (Cited on page 26.)

[100] M. Dudek, I. Nasr, D. Kissinger, R. Weigel, and G. Fischer, "The impact of phase noise parameters on target signal detection in FMCW-radar system simulations for automotive applications," in *Proceedings of IEEE CIE International Conference on Radar*, (Chengdu, China), pp. 494–497, IEEE, Oct. 2011. (Cited on page 26.)

[101] M. Dudek, D. Kissinger, R. Weigel, and G. Fischer, "A versatile system simulation environment for millimeter-wave phased-array FMCW-radar sensors for automotive applications," in *Proceedings of the Asia-Pacific Microwave Conference*, pp. 1478–1481, Engineers Australia, 2011. (Cited on page 26.)

[102] M. Buhren and Bin Yang, "Simulation of Automotive Radar Target Lists using a Novel Approach of Object Representation," in *IEEE Intelligent Vehicles Symposium*, (Meguro-Ku, Japan), pp. 314–319, IEEE, 2006. (Cited on page 26.)

[103] J. Thieling, S. Frese, and J. Rossmann, "Scalable and Physical Radar Sensor Simulation for Interacting Digital Twins," *IEEE Sensors Journal*, pp. 1–1, 2020. (Cited on page 26.)

[104] T. Hanke, N. Hirsenkorn, B. Dehlink, A. Rauch, R. Rasshofer, and E. Biebl, "Generic architecture for simulation of ADAS sensors," in *2015 16th International Radar Symposium (IRS)*, (Dresden, Germany), pp. 125–130, IEEE, June 2015. (Cited on page 26.)

[105] D. Gubelli, O. A. Krasnov, and O. Yarovyi, "Ray-tracing simulator for radar signals propagation in radar networks," in *Proceedings of the 10th European Radar Conference*, (Nuremberg, Germany), p. 4, IEEE, Oct. 2013. (Cited on page 26.)

[106] H. Azodi, U. Siart, and T. F. Eibert, "A fast three-dimensional deterministic ray tracing coverage simulator for a 24 GHz anti-collision radar," *Advances in Radio Science*, vol. 11, pp. 55–60, July 2013. (Cited on page 26.)

[107] C. Sturm, S. Knorzer, T. Zwick, and W. Wiesbeck, "Virtual performance evaluation of automotive radar concepts in realistic traffic environments," in *2009 International Conference on Electromagnetics in Advanced Applications*, (Torino, Italy), pp. 206–209, IEEE, Sept. 2009. (Cited on page 26.)

[108] J. Kim, "Automotive radar simulations in a real traffic scenario: Antenna design and Radar evaluation," in *Antenna Measurement Techniques Association Symposium (AMTA)*, (Newport, Usa), p. 4, IEEE, 2020. (Cited on page 26.)

[109] M. Ouza, M. Ulrich, and Bin Yang, "A simple radar simulation tool for 3D objects based on blender," in *2017 18th International Radar Symposium (IRS)*, (Prague, Czech Republic), pp. 1–10, IEEE, June 2017. (Cited on page 26.)

[110] M. Holder, C. Linnhoff, P. Rosenberger, and H. Winner, "The Fourier Tracing Approach for Modeling Automotive Radar Sensors," in *2019 20th International Radar Symposium (IRS)*, (Ulm, Germany), p. 8, IEEE, June 2019. (Cited on page 26.)

[111] K. Schuler, D. Becker, and W. Wiesbeck, "Extraction of Virtual Scattering Centers of Vehicles by Ray-Tracing Simulations," *IEEE Transactions on Antennas and Propagation*, vol. 56, pp. 3543–3551, Nov. 2008. (Cited on pages 26, 27, 39, and 47.)

[112] N. Hirsenkorn, T. Hanke, A. Rauch, B. Dehlink, R. Rasshofer, and E. Biebl, "Virtual sensor models for real-time applications," *Advances in Radio Science*, vol. 14, pp. 31–37, Sept. 2016. (Cited on page 26.)

[113] N. Hirsenkorn, T. Hanke, A. Rauch, B. Dehlink, R. Rasshofer, and E. Biebl, "A non-parametric approach for modeling sensor behavior," in *2015 16th International Radar Symposium (IRS)*, (Dresden, Germany), pp. 131–136, IEEE, June 2015. (Cited on page 26.)

[114] N. Hirsenkorn, H. Kolsi, M. Selmi, A. Schaermann, T. Hanke, A. Rauch, R. Rasshofer, and E. Biebl, "Learning Sensor Models for Virtual Test and Development," in *11th Workshop Driver Assistance and Autonomous Driving*, (Walting, Germany), p. 10, 2017. (Cited on page 26.)

[115] T. Eder, A. Prinz, L. Brabetz, and E. Biebl, "Szenarienbasierte Validierung eines hybriden Radarmodells für Test und Absicherung automatisierter Fahrfunktionen," in *Automobil-Sensorik 3*, pp. 21–43, Berlin, Heidelberg: Springer Berlin Heidelberg, 2020. (Cited on page 26.)

[116] E. Roth, T. J. Dirndorfer, A. Knoll, K. v. Neumann-Cosel, T. Ganslmeier, A. Kern, and M.-O. Fischer, "Analysis and Validation of Perception Sensor Models in an Integrated Vehicle and Environment Simulation," in *22nd Enhanced Safety of Vehicle Conference*, (Washington DC, USA), 2011. (Cited on page 26.)

[117] N. U. Nathaniel, E. C. Ashigwuike, A. U. Sadiq, and N. A. Obadiah, "Modeling, Simulation and Analysis of Automotive Radar Signal Using Wavelet Transform Technique," in *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)*, (Abuja, Nigeria), pp. 1–6, IEEE, Dec. 2019. (Cited on page 26.)

[118] A. Martowicz, A. Gallina, and G. Karpiel, "Uncertainty propagation for vehicle detections in experimentally validated radar model for automotive application," in *2019 24th International Conference on Methods and Models in Automation and Robotics (MMAR)*, (Miedzyzdroje, Poland), pp. 606–611, IEEE, Aug. 2019. (Cited on page 26.)

[119] X. Li, X. Tao, B. Zhu, and W. Deng, "Research on a Simulation Method of the Millimeter Wave Radar Virtual Test Environment for Intelligent Driving," *Sensors*, vol. 20, p. 1929, Jan. 2020. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute. (Cited on page 26.)

[120] T. Hanke, A. Schaermann, M. Geiger, K. Weiler, N. Hirsenkorn, A. Rauch, S.-A. Schneider, and E. Biebl, "Generation and validation of virtual point cloud data for automated driving systems," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, (Yokohama), pp. 1–6, IEEE, Oct. 2017. (Cited on pages 26 and 27.)

[121] Y. Deep, P. Held, S. S. Ram, D. Steinhauser, A. Gupta, F. Gruson, A. Koch, and A. Roy, "Radar cross-sections of pedestrians at automotive radar frequencies using ray tracing and point scatterer modelling," *IET*

Radar, Sonar & Navigation, vol. 14, pp. 833–844, June 2020. (Cited on pages 26, 47, and 72.)

[122] T. Eder, R. Hachicha, H. Sellami, C. van Driesten, and E. Biebl, "Data Driven Radar Detection Models: A Comparison of Artificial Neural Networks and Non Parametric Density Estimators on Synthetically Generated Radar Data," in *Kleinheubach Conference*, (Miltenberg, Germany), p. 4, IEEE, 2019. (Cited on pages 26 and 27.)

[123] M. Jasinski, "A Generic Validation Scheme for real-time capable Automotive Radar Sensor Models integrated into an Autonomous Driving Simulator," in *International Conference on Methods and Models in Automation and Robotics (MMAR)*, (Miedzyzdroje, Poland), pp. 612–617, IEEE, Aug. 2019. (Cited on pages 28 and 29.)

[124] F. Reway, A. M. K. Hoffmann, D. Wachtel, W. Huber, A. Knoll, and E. P. Ribeiro, "Test Method for Measuring the Simulation-To-Reality Gap of Camera-Based Object Detection Algorithms for Autonomous Driving," p. 8, 2020. (Cited on pages 28, 29, and 72.)

[125] P. Rosenberger, M. Holder, S. Huch, H. Winner, T. Fleck, M. R. Zofka, J. M. Zollner, T. D'hondt, and B. Wassermann, "Benchmarking and Functional Decomposition of Automotive Lidar Sensor Models," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, (Paris, France), pp. 632–639, IEEE, June 2019. (Cited on pages 28, 29, and 72.)

[126] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A Consistent Metric for Performance Evaluation of Multi-Object Filters," *IEEE Transactions on Signal Processing*, vol. 56, pp. 3447–3457, Aug. 2008. (Cited on pages 29 and 77.)

[127] M. F. Holder, J. R. Thielmann, P. Rosenberger, C. Linnhoff, and H. Winner, "How to evaluate synthetic radar data? Lessons learned from finding driveable space in virtual environments," in *13. Uni-DAS e.V. Workshop Fahrerassistenz und automatisiertes Fahren*, (Walting, Germany), p. 11, 2020. (Cited on pages 29 and 72.)

[128] G. Schouten, W. Jansen, and J. Steckel, "Simulation of Pulse-Echo Radar for Vehicle Control and SLAM," *Sensors*, vol. 21, p. 523, Jan. 2021. (Cited on page 29.)

[129] M. Hoeber, "Phenomenological Modelling of Lane Detection Sensors for Validating Performance of Lane Keeping Assist Systems," in *Intelligent Vehicle Symposium*, (Las Vegas, USA), p. 7, IEEE, 2020. (Cited on page 29.)

[130] Daimler, "Immendingen Test Center: Where Daimler works on the mobility of the future." Accessed on: Sep. 05, 2021, Available: https://www.daimler.com/innovation/specials/immendingen-2.html. (Cited on page 38.)

[131] Vires, "Vires Virtual Test Drive (VTD)." Accessed on: Sep. 30, 2021, Available: https://vires.mscsoftware.com. (Cited on page 39.)

[132] IPG, "CarMaker." Accessed on: Sep. 30, 2021, Available: https://ipg-automotive.com. (Cited on page 39.)

[133] LG, "SVL Simulator." Accessed on: Sep. 30, 2021, Available: https://www.svlsimulator.com. (Cited on page 39.)

[134] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, (Mountain View, United States), pp. 1–16, 2017. (Cited on page 39.)

[135] H. Buddendick, T. Eibert, and J. Hasch, "Bistatic scattering center models for the simulation of wave propagation in automotive radar systems," in *German Microwave Conference*, p. 4, 2010. (Cited on page 39.)

[136] A. Scheel and K. Dietmayer, "Tracking Multiple Vehicles Using a Variational Radar Model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 3721–3736, Oct. 2019. (Cited on page 40.)

[137] S. Abadpour, A. Diewald, M. Pauli, and T. Zwick, "Extraction of Scattering Centers Using a 77 GHz FMCW Radar," in *2019 12th German Microwave Conference (GeMiC)*, (Stuttgart, Germany), pp. 79–82, IEEE, Mar. 2019. (Cited on page 41.)

[138] I. Matsunami, R. Nakamura, and A. Kajiwara, "RCS measurements for vehicles and pedestrian at 26 and 79GHz," in *2012 6th International Conference on Signal Processing and Communication Systems*, (Gold Coast, Australia), pp. 1–4, IEEE, Dec. 2012. (Cited on page 41.)

[139] A. Saltelli, M. Ratto, T. Andres, and F. Campolongo, *Global Sensitivity Analysis: The Primer.* Chichester: Wiley, 2008. (Cited on page 47.)

[140] A. Saltelli and P. Annoni, "Sensitivity Analysis," in *International Encyclopedia of Statistical Science*, Springer Berlin Heidelberg, 2011. (Cited on pages 47 and 51.)

[141] R. I. Cukier, C. M. Fortuin, K. E. Schuler, A. G. Petschek, and J. H. Schaibly, "Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients," *The Journal of Chemical Physics*, vol. 59, no. 8, pp. 3873–3878, 1973. (Cited on page 47.)

[142] Y. Gan, Q. Duan, W. Gong, C. Tong, Y. Sun, W. Chu, A. Ye, C. Miao, and Z. Di, "A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model," *Environmental Modelling & Software*, vol. 51, pp. 269–285, Jan. 2014. (Cited on pages 47 and 53.)

[143] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *arXiv:1706.02413 [cs]*, June 2017. arXiv: 1706.02413. (Cited on pages 58 and 62.)

[144] B. Browning, J.-E. Deschaud, D. Prasser, and P. Rander, "3D Mapping for high-fidelity unmanned ground vehicle lidar simulation," *The International Journal of Robotics Research*, vol. 31, pp. 1349–1376, Oct. 2012. (Cited on pages 60 and 72.)

[145] L. V. Kantorovich and G. S. Rubinstein, "On the space of completely additive functions," *Vestnik Leningradskogo Universiteta*, vol. 3, no. 2, pp. 52–59, 1958. (Cited on page 61.)

[146] F. L. Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities," *Journal of Mathematics and Physics*, vol. 20, no. 1-4, pp. 224–230, 1941. (Cited on page 61.)

[147] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's Distance as a Metric for Image Retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000. (Cited on page 61.)

[148] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, pp. 1627–1637, 1964. (Cited on page 65.)

[149] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas, "Criticality Analysis for the Verification and Validation of Automated Vehicles," *IEEE Access*, vol. 9, pp. 18016–18041, 2021. (Cited on page 90.)

[150] S. Jesenski, J. E. Stellet, F. Schiegg, and J. M. Zollner, "Generation of Scenes in Intersections for the Validation of Highly Automated Driving Functions," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, (Paris, France), pp. 502–509, IEEE, June 2019. (Cited on page 91.)

[151] C. Gladisch, T. Heinz, C. Heinzemann, J. Oehlerking, A. von Vietinghoff, and T. Pfitzer, "Experience Paper: Search-Based Testing in Automated Driving Control Applications," in *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, (San Diego, CA, USA), pp. 26–37, IEEE, Nov. 2019. (Cited on page 91.)

[152] A. Lauber, M. Sommer, K. Fuchs, and E. Sax, "Evolutionary Algorithms to Generate Test Cases for Safety and IT-Security in Automotive Systems," in *2020 IEEE International Systems Conference (SysCon)*, (Montreal, QC), pp. 1–6, IEEE, Aug. 2020. (Cited on page 91.)

[153] J. Wang, A. Pun, J. Tu, S. Manivasagam, A. Sadat, S. Casas, M. Ren, and R. Urtasun, "AdvSim: Generating Safety-Critical Scenarios for Self-Driving Vehicles," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Nashville, Tennessee, USA), IEEE, 2021. (Cited on page 91.)

[154] C. Gladisch, C. Heinzemann, M. Herrmann, and M. Woehrle, "Leveraging combinatorial testing for safety-critical computer vision datasets," in

2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (Seattle, WA, USA), pp. 1314–1321, IEEE, June 2020. (Cited on page 91.)

[155] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, (Providence, RI), pp. 3354–3361, IEEE, 2012. (Cited on page 92.)

[156] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," *arXiv:1903.11027 [cs, stat]*, May 2020. arXiv: 1903.11027. (Cited on page 92.)

[157] M. P. Bauer, A. Ngo, and M. Resch, "Usage of Operating System Kernel Paradigms to Build Modern Simulation Platforms for Autonomous Driving," in *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, (Singapore, Singapore), pp. 80–85, IEEE, July 2020. (Cited on page 114.)

[158] M. P. Bauer, A. Ngo, and M. Resch, "The YASE Framework: Holistic Scenario Modeling with Behavior Trees," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Fall)*, IEEE, 2021. (Cited on page 114.)

[159] M. Staebler, "Development of a Radar System Simulation Framework for Virtual Testing of Autonomous Driving Functions," Master's thesis, University of Stuttgart, Stuttgart, Germany, 2020. (Cited on page 115.)

[160] L. Roess, "Erweiterung eines Radar-Simulations Framework für simulationsbasiertes Testen und Entwicklung eines Radar Perzeptions-Algorithmus," 2020. (Cited on page 115.)

[161] B. Li, "Evaluation of a Radar Simulation for Virtual Validation of Autonomous Driving," Master's thesis, University of Stuttgart, Stuttgart, Germany, 2021. (Cited on page 115.)

List of Publications

[36] **A. Ngo**, M. P. Bauer, and M. Resch, "A Sensitivity Analysis Approach for Evaluating a Radar Simulation for Virtual Testing of Autonomous Driving Functions," in *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, (Singapore, Singapore), pp. 122–128, IEEE, 2020.

[37] **A. Ngo**, M. P. Bauer, and M. Resch, "Deep Evaluation Metric: Learning to Evaluate Simulated Radar Point Clouds for Virtual Testing of Autonomous Driving," in *2021 IEEE Radar Conference (RadarConf21)*, (Atlanta, GA, USA), pp. 1–6, IEEE, 2021.

[38] **A. Ngo**, M. P. Bauer, and M. Resch, "A Multi-Layered Approach for Measuring the Simulation-to-Reality Gap of Radar Perception for Autonomous Driving," in *24th IEEE International Conference on Intelligent Transportation Systems (ITSC)*, (Indianapolis, IN, USA), p. 7, IEEE, 2021.

[157] M. P. Bauer, **A. Ngo**, and M. Resch, "Usage of Operating System Kernel Paradigms to Build Modern Simulation Platforms for Autonomous Driving," in *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, (Singapore, Singapore), pp. 80–85, IEEE, July 2020.

[158] M. P. Bauer, **A. Ngo**, and M. Resch, "The YASE Framework: Holistic Scenario Modeling with Behavior Trees," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Fall)*, IEEE, 2021.

List of Supervised Theses

[159] M. Staebler, “Development of a Radar System Simulation Framework for Virtual Testing of Autonomous Driving Functions,” Master’s thesis, University of Stuttgart, Stuttgart, Germany, 2020.

[160] L. Roess, "Erweiterung eines Radar-Simulations Framework für simulationsbasiertes Testen und Entwicklung eines Radar Perzeptions-Algorithmus," Bachelor’s thesis, Aalen University, Aalen, Germany, 2020

[161] B. Li, “Evaluation of a Radar Simulation for Virtual Validation of Autonomous Driving,” Master’s thesis, University of Stuttgart, Stuttgart, Germany, 2021.