Enhancing Radar Model Validation Methodology for Virtual Validation of Automated Driving Functions

Vom Fachbereich Maschinenbau an der
Technischen Universität Darmstadt
zur Erlangung des Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

Dissertation

vorgelegt von

Lukas Elster M.Sc.

aus Fulda

Berichterstatter:	Prof. Dr.-Ing. Steven Peters
Mitberichterstatter:	Prof. Dr. rer. nat. habil. Matthias Hein

Tag der Einreichung:	01.08.2024
Tag der mündlichen Prüfung:	26.11.2024

Darmstadt 2024

D 17

Lukas Elster M.Sc.: Enhancing Radar Model Validation Methodology for Virtual Validation of Automated Driving Functions

Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2025
URN: urn:nbn:de:tuda-tuprints-289623
URI: https://tuprints.ulb.tu-darmstadt.de/id/eprint/28962

Tag der mündlichen Prüfung: 26.11.2024

Kurzzusammenfassung

Automatisierte Fahrzeuge und Mobilitätsangebote werden immer mehr zum Teil des alltäglichen Straßenverkehrs. Die Sicherheit dieser Systeme ist von außerordentlicher Bedeutung. Aus diesem Grund nimmt die Validierung der Sicherheit von automatisieren Fahrsystemen eine immer größere Rolle ein. Der Nachweis der Sicherheit lässt sich erwiesenermaßen jedoch nicht mehr durch reale Testfahrten abbilden, da die Komplexität der Systeme und der sich daraus ergebenden Szenarien keine wirtschaftliche Realisierung ermöglichen. Aus diesem Grund werden Simulationen genutzt, die jedoch für diesen Zweck eine erwiesene Validität benötigen.

Neben anderen Sensoren werden Radare als wesentlicher Bestandteil zur Umfelderkennung im automatisierten Fahrzeug verwendet. Daher sind Effekte und Unsicherheiten in Radarsensorvalidierungsmessungen sowie deren Auswirkungen auf Radarsensormodellvalidierungen Kern dieser Dissertation. Dazu werden Anforderungen an Kriterien zur Akzeptanz von Radarsensormodellen aus Radarmessungen hergeleitet und anhand der im Rahmen der Dissertation weiterentwickelten Doppelvalidierungsmetrik (DVM) quantifiziert. Diese besteht aus einem Schätzwert für den Mittelwertfehler sowie den Abweichungen in der Streuung von unsicherheitsbehafteten Größen. Die Metrik wird auf Radar-Datenstrukturen in verschiedenen Abstraktionen angewendet. Dabei lassen sich Effekte durch dedizierte Messaufbauten, Objekten und Umgebungsbedingungen abbilden.
Nachfolgend wird ein Messaufbau abgeleitet, um den Einfluss von zuvor identifizierten Effekten und Effektkorrelationen auf den Radarquerschnitt (engl. RCS) isoliert zu betrachten. Es zeigt sich, dass die untersuchten Fahrzeugmodelle einen großen Einfluss auf diese Größe haben und weitere Faktoren wie die Verbauhöhe geringe Sensitivitäten im RCS zeigen. Neben dem RCS wird eine detaillierte Analyse der Reflektionszentren der untersuchten Objekte durchgeführt.
Referenzsensoren dienen in solchen dynamischen Szenarien dazu, die gefahrenen Trajektorien mit möglichst geringen zusätzlichen Unsicherheiten in eine Simulationsumgebung zu übertragen. Die vorgestellten Versuche, um die Unsicherheiten der Referenzsensorik und die Übertragung in die Simulation zu quantifizieren, werden mit Hilfe der Superreferenz qualifiziert. Die Versuche bestätigen die Genauigkeiten der genutzten Referenzsensorik.
Abschließend wird anhand eines beispielhaften Radarmodells eine metrik-basierte Validierung durchgeführt, indem ein Validierungsversuch mit einem realen Sensor aufgezeichnet wird. Dabei werden unterschiedliche Abstraktionsebenen der Radardaten unter der Nutzung der DVM analysiert. Besonders eine Kombination von Metrikergebnis mit einem Satellitenbild ermöglicht eine objektive Ursachenanalyse.

Die neu entwickelten Methoden und Versuchsdesigns bilden eine weitere Basis zur Standardisierung von Validierungsversuchen für Radarsensoren und deren Modellen.

Abstract

Automated vehicles and mobility services are increasingly becoming part of everyday road traffic. The safety of these systems is of paramount importance. For this reason, safety validation of automated driving systems is playing an increasingly major role. However, it has been shown that safety can no longer be demonstrated through real test drives, as the complexity of the systems and the resulting scenarios do not allow for an economical implementation. For this reason, simulations are used, but they need to be validated.

In addition to other sensors, radars are used as an essential component for environmental detection in automated vehicles. Therefore, the effects and uncertainties of radar sensor validation measurements and their impact on the validation of radar sensor models are the focus of this thesis. For this purpose, requirements for acceptance criteria for radar sensor models are derived from radar measurements and quantified using the further developed double validation metric (DVM) in the thesis. It consists of an estimate of the mean error and the variance in the dispersion of uncertainties. The metric is applied to radar data structures in different abstraction levels. Effects can be represented by dedicated measurement setups, objects and environmental conditions.
In the following, a measurement setup is derived to isolate the influence of previously identified effects and effect correlations on the radar cross section (RCS). It is shown that the vehicle models investigated have a large influence on this value and that other factors, such as the height of the mounting position, show little sensitivity on the RCS. In addition to the RCS, a detailed analysis of the reflection centers of the investigated objects is carried out.
Reference sensors are used in such dynamic scenarios to transfer the driven trajectories into a simulation environment with minimal additional uncertainties. The presented experiments to quantify the uncertainties of the reference sensors and the transfer to the simulation are qualified with the help of the super reference. The tests confirm the accuracy of the reference sensors used. Finally, a metrics-based validation is performed on an exemplary radar model by recording a validation test with a real sensor. Different levels of abstraction of the radar data are analyzed using the DVM. In particular, the combination of the metric results with a satellite image allows an objective root cause analysis.

The newly acquired methods and test designs form a further basis for the standardization of validation tests for radar sensors and their models.

Preface

This dissertation was written during my time at the Institute of Automotive Engineering (FZD) at the Technical University of Darmstadt. I am grateful for the financial backing from various public institutions, enabling my participation in projects such as VIVID, SET Level, Synthetic FMCW, and VVM. The enriching interactions with numerous exceptional individuals during this period not only contributed to my academic advancement but also played a pivotal role in my personal development. Thus, I extend my appreciation to all the colleagues I had the privilege to collaborate with.

Special acknowledgment is due to my doctoral supervisor, Prof. Dr.-Ing. Steven Peters, for dedicating his time and commitment to my academic journey. His enthusiasm for providing optimal guidance, characterized by personal discussions, constructive feedback, and invaluable conversations, greatly facilitated the development of this work in its current form.
I am also thankful to Prof. Dr. rer. nat. habil. Matthias Hein for serving as the second examiner. His feedback on publications was consistently helpful and appreciated. The collaboration with the Technical University of Ilmenau was highly exemplary and closely aligned with the goals of this research.
I express my gratitude to Prof. Dr. rer. nat. Hermann Winner for granting me the opportunity to join FZD.

My appreciation extends to all my friends and colleagues at FZD. The unique camaraderie and the countless shared hours outside of work transformed colleagues into friends. I anticipate many future encounters, reminiscing about our time together. During my time at FZD, certain friends stood out, and I wish to extend my gratitude to the Sensenecke. Thanks to Clemens and Philipp, the first papers were published quickly, and I felt extremely comfortable after a short time. I would also like to thank Martin for our joint research activities and our intensive discussions on the radar during various nightly hours. Special thanks also to Christoph and Timm for the delightful moments at FZD.

I am sincerely thankful to all the students whose contributions, through various master's and bachelor's theses, student projects, and work as student assistants, made this work possible.

Lastly, I express my gratitude to my family for affording me this opportunity. And finally, a special thanks to my partner, Catherine, who always lent an ear and supported me the whole time despite the considerable time I spent away from home.

Darmstadt, August 2024.

Table of Contents

List of Symbols and Indices

Latin formula symbols

Symbol	Unit	Description
d	-	Area component of the Area Validation Metric
f	Hz	Frequency
F	-	Empirical cumulative distribution function
G	dB	Antenna gain
k	-	Counting variable
n	-	Number
P	W	Power
Q	-	Logarithmic measure of variable denoted as indice
r	m	Range
t	s	Time
v	$\frac{\mathrm{m}}{\mathrm{s}}$	Velocity
V	-	Shaking factor, defined as V_{mp}^2
x	m	x dimension in cartesian frame (“longitudinal”)
y	m	y dimension in cartesian frame (“lateral”)

Greek formula symbols

Symbol	Unit	Description
ϵ	m	Deviation between measurement and super-reference
ζ	$[\zeta]$	Measurand
θ	rad	Elevation angle
ι	-	Counting index for radar cuboid bins
I	-	Upper limit of a counting index for radar cuboid bins
κ	$\frac{\mathrm{dB}}{\mathrm{km}}$	Atmospheric attenuation
λ	m	Wave Length
σ	m^2	Radar Cross Section
ϕ	rad	Azimuth (= horizontal) angle between target and sensor center
ψ	rad	Yaw angle of vehicle

Calligraphic symbols and fraktur characters

Symbol	Unit	Description
$\mathcal{F}$	-	Probability boxes based on boundaries of multiple EDFs

Indices

Symbol	Description
ϕ	Azimuth dimension
σ	RCS
ι_ϕ	Counting index for radar cuboid bins in azimuth dimnesion
ι_r	Counting index for radar cuboid bins in range dimnesion
ι_v	Counting index for radar cuboid bins in velocity dimnesion
r	Range dimension
v	Velocity dimension
AVM	Area Validation Metric
bias	Bias
c	Corrected
C	Car
CAVM	Corrected Area Validation Metric
D	Detection
DVM	Double Validation Metric
filt	Filtered
lat	Lateral
Lb	Light barrier
le	Left
long	Longitudinal
max	Maximum
Meas	Measurement
mp	Multipath
OOI	Object of interest
Ref	Reference measurement device
ri	Right
Rx	Received
S	Sample
Sim	Simulation tool
SRef	Super-reference
tb	Tow bar
Tx	Transmitted

Accents and Operators

Symbol	Description
$\widetilde{a}$	Simulated value or value of a second measurement series
$\hat{a}$	Amplitude of variable
$\dot{a}$	First time derivative of a
$\ddot{a}$	Second time derivative of a

List of Abbreviations

ADF	automated driving function
ADMA	Automotive Dynamic Motion Analyzer
AVM	area validation metric
CCR	corner cube reflector
DVM	double validation metric
EDF	empirical cumulative distribution function
FFT	Fast Fourier Transformation
GNSS	global navigation satellite system
GT	ground truth
lidar	light detection and ranging
ODD	operational design domain
OOI	object of interest
PDF	probability density function
radar	radio detection and ranging
RCS	radar cross section
RQ	research question
RTK	Real Time Kinematic
SAE	Society of Automotive Engineers

List of Figures

List of Tables

List of Papers

The thesis is based on the following papers:

I. M. Holder, **L. Elster**, H. Winner. Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation. MDPI Energies, 15, 2022

II. **L. Elster**, M. Holder, M. Rapp. A Dataset for Radar Scattering Characteristics of Vehicles Under Real-World Driving Conditions: Major Findings for Sensor Simulation. IEEE Sensors Journal, 23, 2023

III. **L. Elster**, J. Staab, S. Peters. Making Automotive Radar Sensor Validation Measurements Comparable. MDPI Applied Sciences, 13, 2023

IV. **L. Elster**, P. Rosenberger, M. Holder, K. Mori, J. Staab, S. Peters. Introducing the Double Validation Metric for Radar Sensor Models. Automotive and Engine Technology Journal, 9, 2024

Related Papers and Presentations

Papers not included in this thesis:

I. C. Linnhoff, P. Rosenberger, S. Schmidt, **L. Elster**, R. Stark, H. Winner. Towards Serious Perception Sensor Simulation for Safety Validation of Automated Driving - A Collaborative Method to Specify Sensor Models. 24th International Conference on Intelligent Transportation Systems (ITSC), Indianapolis, IN, USA, 19.–22.09.2021, 2021

II. **L. Elster**, C. Linnhoff, P. Rosenberger, S. Schmidt, R. Stark, H. Winner. Fundamental Design Criteria for Logical Scenarios in Simulation-based Safety Validation of Automated Driving Using Sensor Model Knowledge. IV Workshop on Ensuring and Validating Safety for Automated Vehicles, Nagoya, 11.07.2021, 2021

III. K. Mori, X. Liang, **L. Elster**, S. Peters. The Inadequacy of Discrete Scenarios in Assessing Deep Neural Networks. IEEE Access, 10, 2022

IV. C. Linnhoff, K. Hofrichter, **L. Elster**, P. Rosenberger, H. Winner. Measuring the Influence of Environmental Conditions on Automotive Lidar Sensors. MDPI Sensors, 22, 2022

V. A. Haider, L. Haas, S. Koyama, **L. Elster**, M. Köhler, M. Schardt, T. Zeh, H. Inoue, M. Jakobi, A. Koch. Modeling of Motion Distortion Effect of Scanning LiDAR Sensors for Simulation-based Testing. IEEE Access, 12, 2024

VI. P. Aust, Y. Amano, A. Araki, S. Buddappagari, **L. Elster**, L. Haas, A. Haider, N. Hiruma, H. Iqbal, T. Kaneshiro, A. Khakimov, M. Köhler, V. Lizenberg, M. Mizukoshi, W. Nakamura, M. Nerio, D. Nickel, M. Pauli, M. Schardt, S.-A. Schneider, T. Sugiyama, T. Zeh. Editors: K. Hanaya, K. Mori, R. Sawamura, M. A. Hein, H. Inoue. White Paper VIVID — German Japan Joint Virtual Validation Methodology for Intelligent Driving Systems, 2024

VII. M. Benedikt, E. Böde, A. Bossert, J. Braband, T. Brade, N. Braun, T. Braun, S. Burton, T. Dallmann, W. Damm, T. Düser, **L. Elster**, T. Fingscheidt, M. Fistler, M. Franek, M. Fränzle, J. Freyer, R. Galbas, R. Gansch, D. Geyer, L. Haas, A. Haider, P. Heidl, M. Hein, A. Heyl, J. Hiller, H. Hungar, D. Hutter, R. Jung, C. Klein, J. Krüger, T. Kuhn, J. Langner, M. Maurer, K. Mayr, A. Meyer-Vitali, E. Möhlmann, A. Molin, B. Möller, J. Niehaus, B. Nolte, M. Nolte, S. Otten, J. Peleska, S. Peters, T. Poguntke, F. Poprawa, J. Reich, P. Rosenberger, N. Salem, B. Schick, D. Schneider, S.-A. Schneider, C. Schyr, C. Thomas, M. Trapp, F. Wagner, N. Wagener, T. Woopen, T. Zeh. Controlling Risk for Highly Automated Transportation Systems Operating in Complex Open Environments. White Paper SafeTRANS Closing the Gap Initiative, 2024

VIII. D. Scheuble, C. Linnhoff, M. Bijelic, **L. Elster**, P. Rosenberger, W. Ritter, H. Winner. Simulating Road Spray Effects in Automotive Lidar Sensor Models. 2024 IEEE Intelligent Vehicles Symposium (IV), Jeju Island, South Korea, 02.–05.06.2024, 2024

IX. K. Hofrichter, C. Linnhoff, **L. Elster**, S. Peters. FMCW Lidar Simulation with Ray Tracing and Standardized Interfaces. 2024 Stuttgart International Symposium on Automotive and Engine Technology, Stuttgart, Germany, 02.–03.07.2024, 2024

Presentations:

I. **L. Elster**. Opportunities and limitations of ASAM OSI within the VIVID and DIVP projects. ASAM International Conference, 29.–30. November 2022, Dresden, 2022.

II. **L. Elster**, A. Khakimov. Making ViL, SiL and Collected Radar Data Comparable: An Application of the Novel Double Area Validation Metric. 26th European Microwave Week, 17.–22. September 2023, Berlin, 2023.

1 Introduction

Since the 1920s, humanity has been pursuing the dream of autonomous road mobility, marked by initial experimental vehicles.[1] The required technologies have advanced over the decades. In 2022, the application for the first commercial operating robotaxis was submitted by Waymo and Cruise.[2] In addition to Level 4 systems, according to the Society of Automotive Engineers (SAE) definition, automation in the private and transportation sectors is on the verge of becoming part of everyday life. Mercedes-Benz[3], Honda[4] and BMW[5] have introduced the first Level 3 systems in privately used cars in the luxury segment. In the transportation sector, Torc Robotics, in collaboration with Daimler Truck[6], as well as Aurora with Continental[7], aim for automated highway driving. Despite these advancements, there are persistent voices expressing skepticism toward automation in the mobility sector. Experts are increasingly cautious in their forecasts for the widespread adoption of automated driving functions (ADFs) and autonomous vehicles at SAE Levels 3 to 5.[8,9] Negative headlines from the United States contribute to additional skepticism regarding technological progress. In an exemplary situation from recent history a pedestrian was dragged beneath an automated vehicle during an accident.[10]

In addition to various personal, societal, and technological benefits[11], the safety of complex and automated driving systems must be guaranteed based on objective measures. To establish safety, a transition from the conventional distance-based safety validation approach to a scenario-based approach is one possible solution[12]. A combination of test environments such as simulations, laboratory test benches, proving ground tests, and real-world tests are indispensable for this purpose. Consequently, alongside real-world experiments, valid simulations at various levels of virtualization will be required in the future to perform scenarios in different variations. Initial guidelines for this shift are already being defined at the regulatory and political levels, with virtual testing based on valid simulations mandated for the approval of ADF.[13,14]

1 Mobileye: A Brief History of Autonomous Vehicles (2023).

2 Grant, N.; Ludlow, E.: Waymo, Chasing Cruise, Plans Fully Driverless Rides in San Francisco (2022).

3 Mercedes-Benz: Mercedes-Benz DRIVE PILOT (2023).

4 Honda Motor Europe Ltd: Honda Receives Type Designation for Level 3 (2020).

5 BMW Group Pressclub Deutschland: Level 3 im neuen BMW 7er (2023).

6 Clavenger, S.: Torc Autonomous Truck Launch in 2027 (2023).

7 Continental Press Release: Continental and Aurora Partner Autonomous Trucking Systems (2023).

8 Society of Automotive Engineers: SAE-J3016 (2021), pp. 24–34.

9 Robson, K.: Fully Self-Driving Cars Unlikely before 2035, Experts Predict (2023).

10 Hawkins, A.: California DMV Suspends Cruise's Robotaxi Permit 'Effective Immediately' (2023).

11 Peters, S. et al.: Opportunities of Automated Driving (2023).

12 Wachenfeld, W.; Winner, H.: The Release of Autonomous Vehicles (2016), p. 442.

13 UNECE: Validation Method for Automated Driving (2022), pp. 6–8.

14 Europäische Kommission: Durchführungsverordnung (EU) 2022/1426 (2022), pp. 50–59.

1.1 Motivation

In addition to navigation, trajectory planning, and actuator control, the perception of the vehicle's surrounding is of particular importance for ADF. Objects and obstacles must be recognized by the vehicle, and an appropriate response to the resulting scenario must be derived. Therefore, the sensing and creation of the system's environmental model hold a special role within the whole automation pipeline. Various sensor modalities exist for this task, fundamentally categorized into active and passive sensors. Passive sensors collect signals from the environment originating from sources outside the automated system. Examples of such sensors include cameras and microphones. On the other hand, active sensors emit signals using a modulation technique and, in turn, receive the signal reflected from the environment to gather information about the surroundings. Active sensors usually include ultrasonic, light detection and ranging (lidar), and radio detection and ranging (radar). In addition to object position, radar sensors are capable of determining the speed of objects using the Doppler effect. Furthermore, radar sensors are known for their robustness in adverse environmental conditions such as rain or fog[15].

Synthetic sensor data have the potential to accelerate developments in sensor signal processing and downstream algorithms, such as trajectory planning, facilitating early error identification.[16] Furthermore, synthetic augmentation of real sensor data can yield performance gains for machine learning approaches.[17] Moreover, the employed models are part of the overall automated system, necessitating a safety validation for these models as well. Consequently, it follows that these simulation models must be valid with respect to their intended purpose. While initial validation methods for radar sensors exist in the literature[18], there are currently no standardized experiments and evaluation procedures to provide the proof of validity. A preliminary guideline with experiments addressing performance exists for lidar sensors.[19] However, such a catalog is missing for radar sensors. The challenge lies not only in defining requirements for radar simulation models but also in conducting and evaluating corresponding validation experiments. The comparison of synthetic and real data relies on metrics, which have to be suitable for radar and its characteristics. For this reason, this dissertation systematically identifies gaps in the state of research that need to be addressed for future standardization of radar validation campaigns.

15 Ryde, J.; Hillier, N.: Laser and Radar Ranging in Adverse Conditions (2009), p. 718.

16 Ahmann, M. et al.: Towards Continuous Simulation Credibility Assessment (2022), p. 171.

17 Linnhoff, C. et al.: Simulating Road Spray Effects in Automotive Lidar Sensor Models (2022), p. 10.

18 Chapter 3 contains detailed information about the validation methods.

19 DIN: DIN SAE SPEC 91471 Assessment Methodology Automotive LiDAR (2023).

1.2 Structure

The structure of the thesis is described below. Fig. 1-1 illustrates the organization based on chapters and their corresponding main contents. The dissertation begins with Chapter 2, where the theoretical foundations necessary for understanding subsequent publications are summarized. It includes definitions of key terms, an explanation of radar measurement principles and their interfaces, and an overview of radar simulation models, concluding with the introduction of the so-called DVM. Chapter 3 analyses the state of the art in radar validation methods and leads to the derivation of the main research question (RQ). The topic of radar sensor model validation is then further analyzed using a validation methodology from literature. Subsequently, four RQs are identified in the respective sub-chapters 3.2.1 and 3.2.4. These RQ address requirements for simulation models, experimental design for the isolation of effects, uncertainties in reference sensor data as well as their impact on validation results, and a metric-based radar model validation. Chapters 4 to 7 represent the content of the four peer-reviewed journal publications used in this cumulative dissertation, as highlighted in gray in Fig. 1-1. Each publication is associated with a RQ that is answered within the core results of the respective publication. Finally, Chapter 8 presents the findings related to the initial RQ and offers an outlook on further necessary research activities regarding radar model validation and standardization.

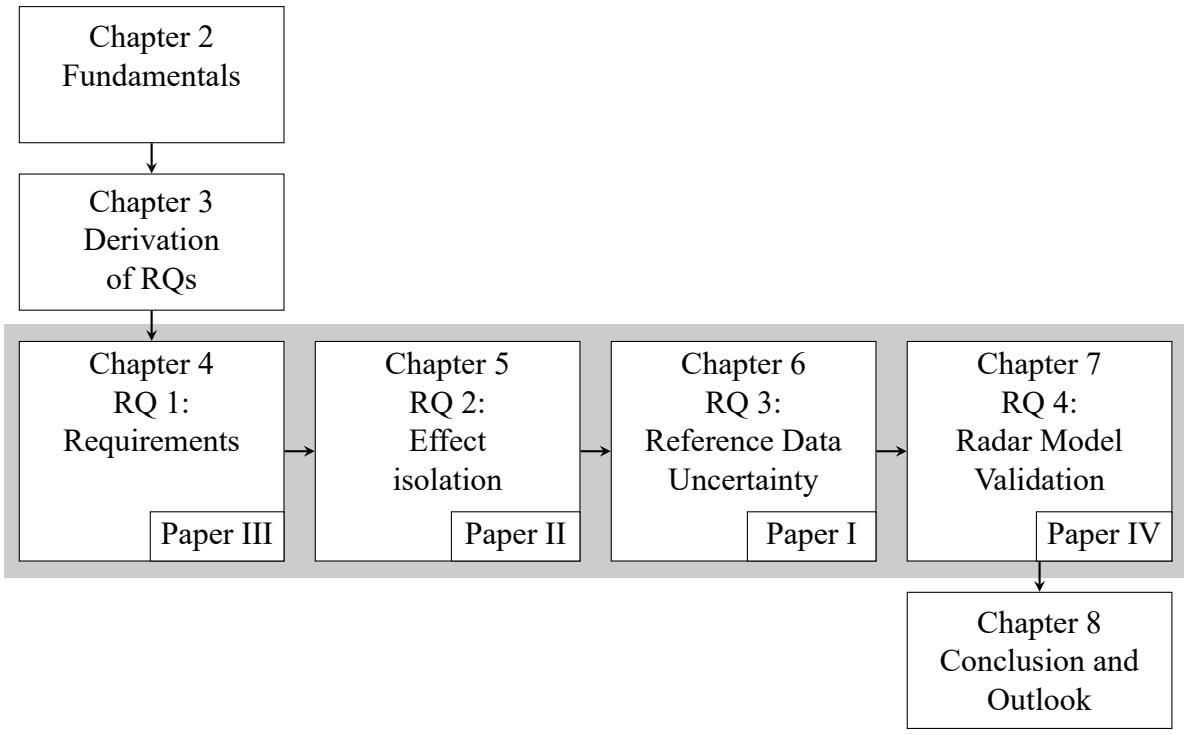


Figure 1-1: The thesis structure is based on the main research question and the resulting research questions with the associated publications highlighted in gray.

2 Fundamentals

In this chapter, the foundations are laid for a deeper understanding of the thesis. The first part provides definitions for all necessary terminologies. It is followed by an overview of the general functioning principle of automotive radar sensors and the radar data interfaces resulting from signal processing. Explanations of the RCS properties of objects conclude the introduction to the radar sensor. Subsequently, fundamental information about radar models is presented, with detailed descriptions of the model used in this thesis. The chapter concludes with validation metrics and transitions into addressing the previously raised research questions through the DVM.

2.1 Terms and Definitions

This chapter defines basic terms that will be used in the rest of the document. Initially, terminologies pertaining to simulation and modeling are elucidated.

Sensor simulation: The term sensor simulation denotes the virtual replication of a sensor and its measurement principles, incorporating selected sensor effects. The outcome of a sensor simulation comprises one or multiple timeframes of synthetic sensor data, which corresponds to the output data of the modeled sensor at various levels of abstraction. The implementation of a sensor simulation takes place within a simulation environment provided by a simulation tool.

Sensor model: A sensor model is a mathematical representation providing an approximation of the measurement principles inherent to the sensor.[20] For this purpose, the model is executed within a simulation tool that, in addition to determining the sensor's position in the virtual world, defines its relative position to other objects. These objects, depending on the type of sensor model, can be 3D entities with material properties. Wave propagation models are then employed to simulate the interaction between object surfaces and the electromagnetic waves. In the context of this dissertation, a rendering-based radar model is utilized as a sensor model.

Sensor effect: A sensor effect is any condition that induces a deviation from the original information due to interaction with the surrounding of the sensor. Causes contributing to an effect may stem from various sources, such as inherent properties of the emission unit's hardware (e.g., antenna design in radar). The manifestation of an effect requires the presence of a signal, defined as a quantified form of energy influencing the sensor

20 Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021), p. 7.

in accordance with its measurement principle. An effect can have multiple causes, and conversely, a single cause may lead to multiple effects. Consequently, a cause-effect chain can evoke subsequent effects within the signal or data.[21] In the context of radar sensor models, different implementation approaches are employed to replicate such effects.

Requirement: A requirement is defined as a capability or performance that a product (model) must fulfill or possess.[22] These model requirements are verified and validated within the V-model. Specifically, sensor model requirements describe the capability to replicate sensor effects within defined limits and abstract their impact on a model output.

Validation: Validation involves assessing how well the predictions made by a theory or model align with real-world empirical data. This process is distinct from verification, which entails checking whether a model's implementation adheres to its intended specifications. Recent suggestions for validation methods encompass straightforward comparisons of trends in means, more sophisticated hypothesis testing, and comprehensive Bayesian approaches.[23]

The focus in the following part is on metrological definitions. Thereby, reproducibility, repeatability calibration and bias are defined in accordance to ISO 3534–2.[24]

Reproducibility: Reproducibility refers to precision under consistent conditions, where precision is the agreement of independent measurement results under agreed-upon conditions. Reproducible conditions are defined as independent measurement results recorded by different experimenters with different equipment based on the same method and identical measurement setups. Reproducibility is thus the quantitative measure of result variability.

Repeatability: Repeatability refers to precision under redoable conditions, where precision is defined as in reproducibility. Repeatability conditions are the observation conditions recorded by the same experimenter with the same equipment based on the same method and identical measurement setups. The measurements must be conducted within a short time interval for the conditions to be considered repeatable. In the context of this work, the time interval is defined as 15 min, to ensure comparable environment conditions within a measurement series.

Calibration: Calibration refers to the set of activities that determine a calibration function under reference states from values of a measured quantity. Measurements based on the same calibration are defined as a measurement series.

21 Linnhoff, C. et al.: Towards Sensor Simulation for Safety Validation (2021), p. 2689.

22 VDI/VDE: VDI/VDE 2206: Development of Mechatronic and Cyber-Physical Systems (2021), p. 3.

23 Ferson, S.; Oberkampf, W.: Validation of Imprecise Probability Models (2009), p. 4.

24 DIN: DIN ISO 3534-2 Statistics - Vocabulary and Symbols (2013), pp. 55–63.

Bias: Bias refers to the discrepancy between the expected or measured result of a test or measurement and the true value. It represents the cumulative systematic error, distinguishing it from random error. Bias may comprise one or more systematic error components, contributing to the overall systematic difference from the true value. A larger bias value indicates a more significant systematic deviation from the true value. The true value is exclusively available within the simulation as the GT, but remains beyond the capability of measurement systems. In the subsequent chapters, the term bias will be employed in the context of comparing a measurement to a simulation and comparing a measurement against another. Therefore, the true value of a measurand is excluded.

Experiment: An experiment is defined by a specific setup. In the context of sensor validation measurement campaigns the position of the sensor and corresponding objects is essential. The experiment consists of at least one measurement.

Measurement: A measurement is defined as a single time series of measurement data collected within the same measurement setup and the corresponding reference data.

In addition to the definitions provided so far, the subsequent terms are crucial for understanding the scope of this work.

Reference sensor: In measurements, reference sensors are used to determine the true value of the measured quantity with an order of magnitude greater accuracy, reliability, and a higher update rate compared to the sensor under test.[25] In this work, for stationary experiments, a Real Time Kinematic (RTK)-based global navigation satellite system (GNSS) sensor is used, and for dynamic situations, an Automotive Dynamic Motion Analyzer (ADMA) is employed.[26]

Uncertainty: Uncertainty in measurements is a characteristic associated with a measurement result, describing the dispersion of values that can also arise from systematic influences.[24] Furthermore, uncertainty can be classified into aleatory and epistemic uncertainties, as illustrated in Fig. 2-1. Aleatory uncertainty, also known as irreducible uncertainty, stochastic uncertainty, or variability, refers to the inherent variation or randomness within a given system. This uncertainty results from spatial and temporal variations and is typically characterized by a probability density function (PDF) or an EDF.[27] Epistemic uncertainty, also referred to as reducible uncertainty, arises from a lack of knowledge held by the analyst involved in modeling and simulation efforts. This uncertainty is reducible through the incorporation of knowledge, achieved through methods such as experimentation and enhanced numerical approximations. With sufficient investment in time and resources, it is theoretically possible to eliminate epistemic uncertainty.[27] Epistemic uncertainty

25 Brahmi, M.: Bewertung der Umfeldwahrnehmung für Fahrerassistenzsysteme (2020), p. 33.

26 GeneSys: Datasheet ADMA-G-Pro+ V35 (2023).

27 Roy, C. J.; Oberkampf, W. L.: Framework for Verification, Validation, and Uncertainty (2011), p. 2132.

is further categorized into recognized (known) and blind (unknown) uncertainty. In the case of recognized epistemic uncertainty, a conscious decision is made to address it in some manner during the validation process or to disregard it for practical considerations. Conversely, blind epistemic uncertainty arises when there is a failure to acknowledge the incomplete nature of knowledge about the system being modeled.[28]

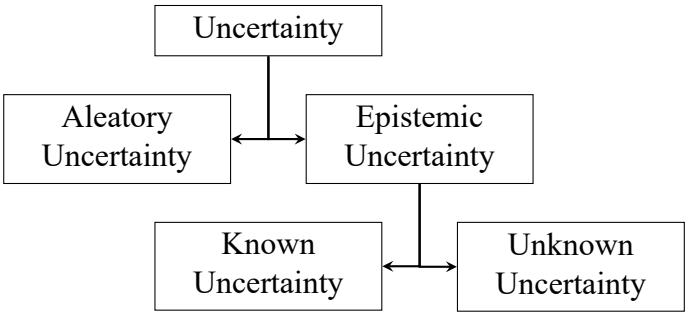


Figure 2-1: Types of uncertainity which are present in data.[28]

2.2 Radar

Automotive radar sensors emit electromagnetic waves through transmitting antennas at various frequency bands, with the most widely used frequency band for long-range radars currently ranging from 76 GHz to 77 GHz.[29a] The emitted signal is typically modulated using the frequency modulated continuous wave chirp-sequence method[29b], and the modulated frequency ideally follows a saw tooth waveform.[29c] The electromagnetic wave is absorbed, transmitted or reflected by objects and surfaces. The portion of the refelcted radiation being captured by the sensor through receiving antennas results in a measurebale signal intensity.

The transmitted power P_{Tx} and received power P_{Rx} are related according to (2-1).[29d] The factor κ takes into account the atmospheric attenuation of the signal and, together with the radial distance r, exhibits an exponential dependence. Here, G represents the antenna gain, which results from the antenna's characteristics compared to an ideal isotropic radiator. σ denotes the RCS, a measure of an object's reflectivity properties influenced by materials, surface conditions, geometric features, and the wavelength λ of the electromagnetic radiation itself.[29e,30a] The radial distance r affects the received power with a fourth-power dependency. The shaking factor V_{mp}^2 considers the effects of constructive and destructive interference of radiation due to multipath propagation.

$$\frac{P_{\mathrm{Rx}}}{P_{\mathrm{Tx}}} = \frac{10^{-2\kappa r} G_{\mathrm{Tx}} G_{\mathrm{Rx}} \sigma \lambda^2}{(4\pi)^3 r^4} V_{\mathrm{mp}}^2 \tag{2-1}$$

28 Schaermann, A.: Systematische Bewertung umfelderfassender Sensormodelle (2020), p. 19.

29 Winner, H.: Automotive RADAR (2016), a: p. 327; b: pp. 352–353; c: p. 352; d: p. 331; e: p. 328; f: p. 330; g: p. 382.

30 Buddappagari, S. J. G. et al.: Monostatic RCS Measurements of a Passenger Car (2019), a: p. 998; b: p. 999.

2.2.1 Radar Data Interfaces

After reception, the signal is mixed with the transmitted signal and low-pass filtered. The result represents the baseband signal of each antenna.[31a] Subsequent Fourier transformations on the baseband signal's frequency spectrum, obtained through multiple consecutive ramps of the mixed saw tooth waveform, identify peaks corresponding to distance, velocity and azimuth angle.[32] In recent years, so-called „4D Imaging Radar" provide elevation measurements, thanks to additional vertically distributed antenna elements. The „imaging" capability is due to increased resolution in the angle and range dimension, which is nowadays in the decimeter magnitude. In Tab. 2-1 typical state-of-the-art radar sensor parameters are listed. The elevation angle is not further considered due to the radar sensor used, which does not measure an elevation angle.

Table 2-1: Typical state-of-the-art automotive radar sensor parameters of the different dimensions.[29g,33,34]

Parameter		Value
Frequency	f	76 GHz to 77 GHz
Range	r	0.2 m to 250 m
Range resolution	Δr	0.1 m to 0.4 m
Velocity	v	−400 km/h to 280 km/h
Velocity resolution	Δv	0.1 m/s
Azimuth	ϕ	±60°
Azimuth resolution	$\Delta\phi$	3°
Elevation	θ	±15°
Elevation resolution	$\Delta\theta$	6°
Sample frequency	f_{S}	14 Hz

Fig. 2-2 illustrates the processing of the baseband signal. Initially, a FFT is performed over each ramp of the mixed signal, resulting in the frequency spectrum of distance bins as shown in Fig. 2-2 (a). Subsequently, a FFT is conducted over all ramps of the individual range bins, extracting the Doppler component in the frequency spectrum for the corresponding range bin as illustrated in Fig. 2-2 (b). Finally, a FFT is performed over all antenna elements, analyzing the angle-specific frequency components. The result is the so-called radar cuboid, containing the power ratio values of transmitted and received power in each cell. Here, ι denotes the index variable across the bins in the different dimensions of range r, velocity v, and azimuth ϕ as shown in Fig. 2-2 (c). The number of cells n in one of the dimensions is denoted by I. The cell at the maximum of all dimensions is indicated in orange with power P. All teal bins of the radar cuboid represent bins with relative velocity 0 m/s and are particularly relevant for static objects. Depending on the configuration and parameterization of the radar sensor, ambiguities may arise

31 Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021), a: p. 13; b: p. 130; c: pp. 48–52; d: pp. 130–131.

32 Meteer, O.; Bekooij, M. J. G.: Low-Power Sign-Magnitude FFT Design for Radar (2021), p. 53.

33 Waldschmidt, C. et al.: Automotive Radar — From First Efforts to Future Systems (2021), p. 138.

34 Continental Engineering Services: Datasheet Continental ARS408 (2024)

when the signal's phase deviation exceeds π.[35]

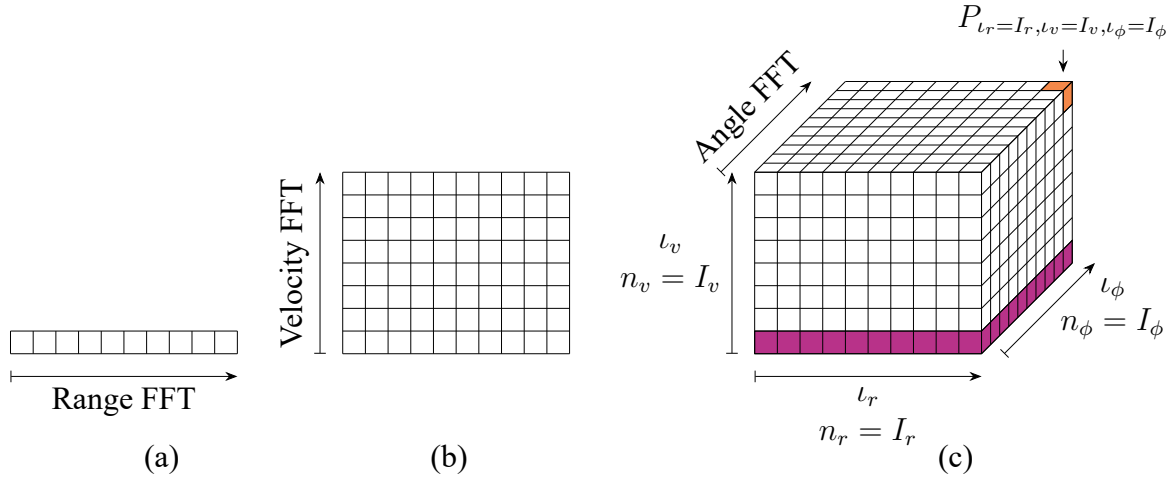


Figure 2-2: Radar data processing with FFTs in the different dimensions resulting in the radar cuboid. One cell is marked in orange and the plane with no relatve velocity is visualized in teal.

Based on the radar cuboid data in the frequency domain, dynamic thresholding is performed to extract detections from the frequency spectrum. These detections include, among other information, the parameters RCS, radial distance, azimuth, elevation, and relative velocity.[36] Objects are generated based on these detections, with an increasing application of data-driven approaches in this process.[37] For further understanding of the thesis, only the radar cuboid and the detection interface are of significance.

2.2.2 Radar Cross Section

The RCS is a measure for the detection probability representing the power level. This value is reconstructed based on internal radar sensor parameters as antenna diagram and measured distance. As previously described, there are several influencing factors affecting the value given in m^2 or $dB\,m^2$. Typically, RCS profiles are generated in radar chambers under ideal conditions by measuring the RCS under different azimuth and elevation angles.[30b,38a] Additionally, setups on proving grounds exist in the literature, utilizing either laboratory setups[39] or sensors installed in vehicles[31b]. The literature highlights the following main influencing factors on the RCS characteristics of objects in the automotive context.

Elevation angle changes due to height diversity: Due to changes in elevation, different viewing angles are obtained concerning the elevation direction to the target and multipath

35 Holder, M. F. et al.: Modeling and Simulation of Radar Sensor Artifacts (2019), p. 2.

36 ISO: ISO 23150:2021(E) Data between Sensors and Fusion (2021), pp. 78–79.

37 Zhou, Y. et al.: Towards Deep Radar Perception for Autonomous Driving (2022), p. 20.

38 Schipper, T. et al.: RCS Measurement for Automotive Objects (2011), a: p. 684; b: pp. 684–686.

39 Toss, H.; Karlsson, K.: Radar Reflectivity Spatial Profile of 3D Surrogate Targets and Real Vehicles (2019), p. 2.

propagation is directly affected. As a result of the altered viewing angle, different reflectivity properties emerge.[40a,31c,41]

Material properties: The material properties directly affect the mechanisms of reflection, transmission, and absorption. Additionally, the surface characteristics of materials, aside from the material itself, have a direct impact on the characteristics of the reflective properties.[40a,42,43,44]

Aspect angle: The aspect angle is delineated as the angular measure between the major axis of the measured object, representing its heading direction, and the line of sight corresponding to $\phi = 0°$ of the radar.[45] Due to the various viewing angles in the horizontal sensor plane, surfaces reflect differently. Measurements have shown that the RCS maxima occur typically on the sides (90° and 270°), the front (0°), and the rear (180°) of the vehicle.[46,38b,47,31d]

Object shape: The object shape influences the back scattering of radar waves.[46,38a,47,48] In theory, the RCS, which is also wavelength-dependent, can be analytically determined for simple geometries.[49a] However, the wavelength is not examined in the following, given that the frequency modulation of the used radar sensor is predefined and non-adjustable. A commonly used target in radar experiments is a trihedral corner cube reflector (CCR), which reflects incident radiation back into the same spatial direction. Despite their small size, these ideal targets achieve high RCS values. A typical vehicle, measuring 4 m in length and 2 m in width, exhibits a RCS ranging from 10 dB m^2 to 15 dB m^2, corresponding to an edge length of a CCR of 20 cm.[49b]

Polarization: For completeness, the polarization of radar waves is also listed as a factor influencing the RCS characteristics of objects. It is preset in automotive production radar sensors and lacks direct adjustability by end-users.[46,38a] Consequently, polarization is not subject to further analysis within the scope of this thesis.

40 Karlsson, K. et al.: Reducing Ground Reflection during RCS Characterization (2019), a: p. 4; b: p. 2.

41 Diewald, F.: Objektklassifikation und Freiraumdetektion auf Basis bildgebender Radarsensori (2013), pp. 49–59.

42 Abadpour, S. et al.: Dielectric Material Characterization of Traffic Objects (2023).

43 Landron, O. et al.: A Comparison of Theoretical and Empirical Reflection Coefficients (1996), pp. 346–349.

44 Langen, B. et al.: Reflection and Transmission Behaviour of Building Materials at 60 GHz (1994), p. 507.

45 Wei, Y. et al.: Extended Target Recognition in Cognitive Radar Networks (2010), p. 10.

46 Buddappagari, S. J. G. et al.: RCS Measurements of Road Traffic Objects (2020), p. 4.

47 Karlsson, K. et al.: HiFi Radar Target (2018), p. 21.

48 Aust, P. et al.: Fingerprints of the Automotive Radar Scattering of Passenger Cars and Vans (2023), pp. 7–10.

49 Winner, H.: Automotive RADAR (2016), a: pp. 4; b: p. 329.

2.3 Radar Models

In literature, various radar models with associated interfaces exist, yielding spectral radar cuboid data, detections, or object lists. The input ranges from object lists generated through simulation to ray tracing results in combination with a simulated sensor signal processing. Magosi et al. provide an overview of radar models[50], categorizing them according to Schöner. According to him ground-truth models encompass all objects within a defined search radius around the ego vehicle. Geometric models include additional information, such as the sensor's field of view, and comprise the object list filtered by object position. Stochastic models also output an object list but introduce uncertainties and noise processes to the generated list. Phenomenological and physics-based models represent a complex depiction of radar wave propagation and processing. Phenomenological models stand apart from physics-based models as they replicate effects using mathematical and statistical approximations. In addition to object lists, detections are implemented as outputs. Physical models simulate the physical properties of objects and wave propagation based on absorption, reflection, and transmission. Thus, physical models have the potential to achieve the highest degree of realism in reproducing effects, provided that the environmental simulation in conjunction with the simulation approach is valid. However, this increases the complexity of the model and the calculations, which enhances the computing time.[51]

This thesis employs an open-source phenomenological radar signal processing model for simulation purposes.[52] The model is based on Holder's Fourier tracing approach.[53] Input data consist of the output from a ray tracing algorithm, incorporating ray length, incident angle on the sensor in azimuth and elevation directions, relative radial velocity of objects intersected by the ray with respect to the sensor, and the associated ray power. The environment model and ray tracing algorithm are part of IPG CarMaker as a simulation tool. The ray tracing data, transmitted through the Open Simulation Interface[54], are conveyed to the radar model. Within the model, this data are interpreted as δ-peaks in the frequency domain. Subsequent convolution employing window functions in each dimension in the frequency domain generates spectral radar cuboid data. Detection results are obtained through dynamic thresholding across all dimensions of the radar cuboid using an ordered-statistic-constant false alarm rate algorithm[55].

50 Magosi, Z. F. et al.: A Survey on Modelling of Automotive Radar Sensors (2022), pp. 8–14.

51 Schöner, H.-P.: Automotive Needs and Expectations towards Driving Simulation (2018), pp. 11–16.

52 ENVITED OpenMSL Github: SL 1-1 Reflection Based Radar Object Model (2023).

53 Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021), pp. 92–93.

54 ASAM e.V.: ASAM OSI® (Open Simulation Interface) - Official Documentation (2023).

55 Scharf, L. L.; Demeure, C.: Statistical Signal Processing (1991).

2.4 Validation Metrics for Uncertainty Quantification

For the comparison of simulated and real sensor data, various metrics are employed. The literature identifies specific criteria that must be satisfied for the mathematical formulation of a validation metric. According to Schaermann[56], seven criteria exist, derived from a fusion of Liu et al.[57] and Oberkampf and Barone[58].

1. Metrics are intuitive, and their results derived from the metric should be easily interpretable, facilitating clear understanding.
2. Metrics exhibit broad applicability, accommodating both deterministic and non- deterministic data.
3. Metrics are characterized by their quantitative and objective nature, precluding the inclusion of user-defined parameters.
4. Metrics are designed without the incorporation of explicit acceptance criteria, yielding non-Boolean output.
5. Metrics incorporate considerations of uncertainties, encompassing both epistemic and aleatory uncertainties.
6. Metrics establish a confidence interval in relation to the quantity of measurement data.
7. Metrics adhere to the mathematical properties[59a] inherent to the definition of a metric, yielding results that might be unbounded.[59a]

Utilizing the aforementioned criteria, Rosenberger conducts an evaluation that scrutinizes validation metrics employed in the literature. In addition to the criteria mentioned earlier, he assesses interfaces, considered scenarios, measurement scales, and the consideration of both aleatory and epistemic uncertainties.[59b] Concerning interfaces, Rosenberger exclusively takes into account definitions outlined in ISO 23150, which standardizes data interfaces of automotive sensors, focusing on detections, features, and objects.[60] Notably, the radar cuboid interface is not considered in his evaluation.

Rosenberger's analysis identifies the AVM as the most favorable candidate for utilization in sensor model validation. A notable advantage lies in its inherent intuitiveness, as the metric is expressed in the unit of the measured quantity, enhancing its utility in defining acceptance thresholds for model validation.[59c] Furthermore, the AVM demonstrates the capability to account

56 Schaermann, A.: Systematische Bewertung umfelderfassender Sensormodelle (2020), pp. 20–21.

57 Liu, Y. et al.: Toward a Better Understanding of Model Validation Metrics (2011), p. 2.

58 Oberkampf, W. L.; Barone, M. F.: Measure Agreement between Computation and Experiment (2006), pp. 11–12.

59 Rosenberger, P.: Metrics for Simulating Sensors (2022), a: p. 99; b: pp. 100–101; c: p. 102; d: pp. 105–108.

60 ISO: ISO 23150:2021(E) Data between Sensors and Fusion (2021), p. 1.

for both epistemic and aleatory uncertainties and incorporates them into the metric result.[61] Given these considerations and additional analyses by Rosenberger in comparison to other metric candidates, such as Kolmogorov-Smirnov divergence, Jensen-Shannon distance/divergence and Kolmogorov-Smirnov tests[59d], this work directs its focus on the AVM.

2.4.1 Area Validation Metric

The AVM is based on calculating the area between two EDFs. These EDFs, denoted as F and $\tilde{F}$ for the distributions under comparison, may generate areas above F (referred to as d^-) as well as below F (referred to as d^+) due to the EDFs respective shapes. Thus, for F representing a quantity ζ and $\tilde{F}$ representing an identical quantity from another measurement series or a corresponding simulated value, the mathematical relationship is given by

$$d_{\mathrm{AVM}}(F, \tilde{F}) = \int_{-\infty}^{\infty} |F(\zeta) - \tilde{F}(\zeta)| d\zeta. \tag{2-2}$$

Taking aleatory and epistemic uncertainties in the data into account, probability boxes $\boldsymbol{\mathcal{F}}$ and $\widetilde{\boldsymbol{\mathcal{F}}}$ emerge. This is due to the presence of multiple EDFs of measurement or simulation data. Probability boxes are obtained based on the maximum and minimum values at each measurand value of the multiple EDFs. Thus, the probability box is composed of the two enveloping EDFs.[62] This leads to the following formula based on Ferson and Oberkampf[63]

$$d_{\mathrm{AVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}) = d^- + d^+ \,. \tag{2-3}$$

The mathematical concept is visualized in the following. Fig. 2-3 (a) represents two distinct datasets in black and teal, where variations in brightness indicate different EDFs within the two datasets. In the context of validation, these could be measurements or simulation results. In Fig. 2-3 (b), the probability boxes of the EDFs from Fig. 2-3 (a) are depicted. The areas corresponding to d^+ and d^-, necessary for the AVM calculation, are marked with different colors.

According to Voyles and Roy, it is also possible to calculate the bias d_{bias} of the probability boxes.[64] For this purpose, the two areas are subtracted from each other to equalize the shapes of the functions. Thus, the expression for d_{bias} is given by

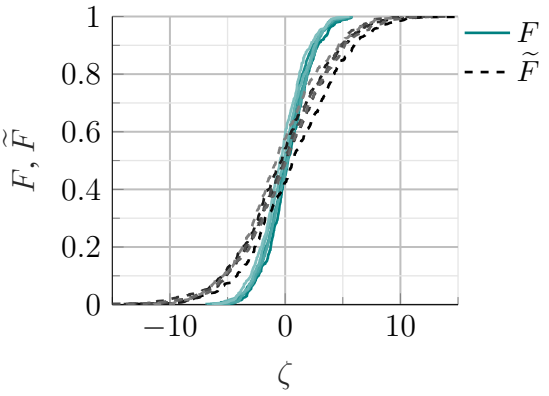
$$d_{\mathrm{bias}} = d^- - d^+ \,. \tag{2-4}$$

61 Ferson, S.; Oberkampf, W.: Validation of Imprecise Probability Models (2009), p. 13.

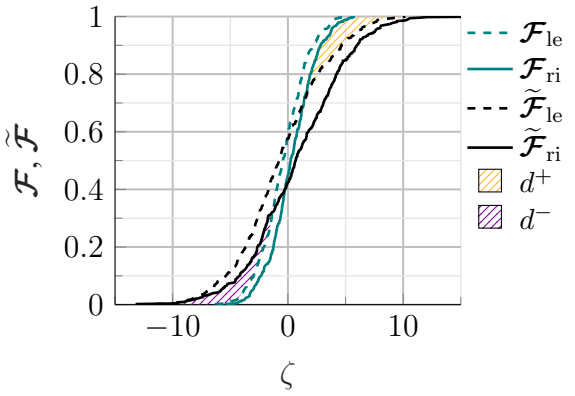
62 Williamson, R. C.; Downs, T.: Probabilistic Arithmetic (1990), pp. 107–113.

63 Ferson, S.; Oberkampf, W.: Validation of Imprecise Probability Models (2009), p. 13.

64 Voyles, I. T.; Roy, C. J.: Model Validation in the Presence of Uncertainty (2014), p. 4.



(a) Various EDFs due to e.g. multiple repetitions of a data set 1 F and a data set 2 $\widetilde{F}$. The color brightness denotes the different repetitions.



(b) Probability boxes of data set 1 $\boldsymbol{\mathcal{F}}$ and data set 2 $\widetilde{\boldsymbol{\mathcal{F}}}$ based on the EDFs in Fig. 2-3 (a). The areas d^+ and d^- of the AVM between the probability boxes are hatched in different colors.

Figure 2-3: Exemplary EDFs and probability boxes for two datasets, along with the corresponding areas d^+ and d^- for the calculation of the AVM.

In the context of radar model validation, multiple measurements give rise to a probability box for the measured data, and due to uncertainties with their variations in simulation, a probability box is generated for the simulated data. Therefore, the AVM serves as a validation metric that accommodates the specific requirements of the intended application in this work.

2.4.2 Double Validation Metric

The AVM defined in (2.4.1) exhibits no sensitivity to deviations in the data scatter, provided there is no overlap between the EDFs or probability boxes.[65a,66] This limitation is addressed by extending the conventional definition of d_{bias} to introduce the so-called corrected AVM. In this context, the AVM is computed after adjusting the EDF $\widetilde{F}$ or probability box $\widetilde{\boldsymbol{\mathcal{F}}}$ by d_{bias}, resulting in $\widetilde{F}_{\mathrm{c}}$ or $\widetilde{\boldsymbol{\mathcal{F}}}_{\mathrm{c}}$. In mathematical terms, this is expressed as[65b,65c]

$$\widetilde{\boldsymbol{\mathcal{F}}}_{\mathrm{c}}(\zeta) = \boldsymbol{\mathcal{F}}(\zeta - d_{\mathrm{bias}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}})) = \widetilde{\boldsymbol{\mathcal{F}}}(\zeta - d^- + d^+)\,, \tag{2-5}$$

$$d_{\mathrm{CAVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}) = d_{\mathrm{AVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}_{\mathrm{c}}) = d_{\mathrm{c}}^- + d_{\mathrm{c}}^+\,. \tag{2-6}$$

The two values, d_{bias} for the bias and d_{CAVM} for the scatter deviation, ultimately yield the two-dimensional DVM.[65c]

$$d_{\mathrm{DVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}) = \left(d_{\mathrm{bias}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}), d_{\mathrm{CAVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}})\right)\,. \tag{2-7}$$

65 Rosenberger, P.: Metrics for Simulating Sensors (2022), a: p. 111; b: p. 117; c: p. 119.

66 Oberkampf, W. L.; Ferson, S.: Validation Under Aleatory and Epistemic Uncertainty (2007), p. 22.

The mathematical procedure is illustrated in Fig. 2-4 for two EDFs. Subsequently, $\widetilde{F}$ is shifted by d_{bias}, resulting in $\widetilde{F}_{\text{c}}$ with d_{c}^{+} and d_{c}^{-}.

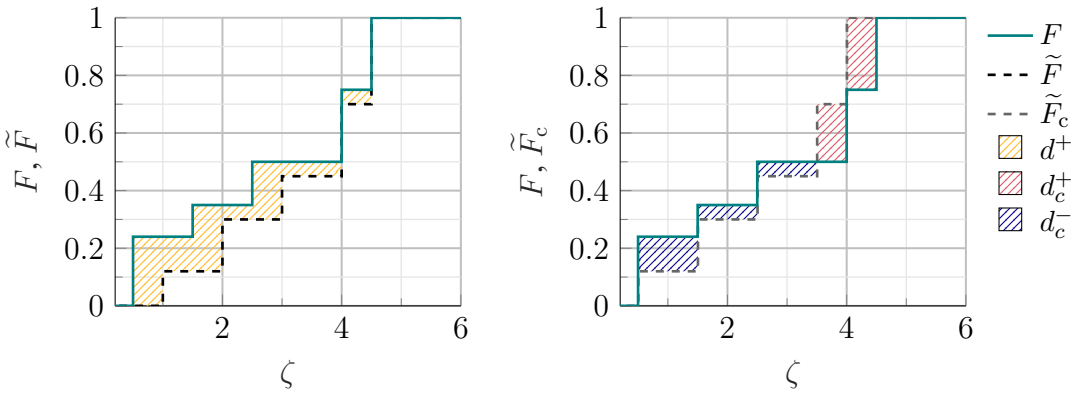


Figure 2-4: Exemplary calculation of the DVM based on two constructed EDFs. On the left, the original distributions are depicted, and on the right, the EDF corrected by d_{bias}.

The current definition of the DVM is based on probability boxes and the resulting values for d_{bias} and d_{CAVM}. However, there are conceivable scenarios that could lead to a misinterpretable metric result of the current DVM definition with probability boxes. Using the method of manufactured universes[67], various EDFs are conceived, illustrating the problem arising from using probability boxes. In Fig. 2-5, these scenarios are depicted and categorized into four edge cases.

For situation (a), the EDFs for two datasets are shown. Here, one probability box is significantly larger than the other. Two EDFs from different measurements coincide directly, thus having the same distribution function. The probability boxes do not overlap at all, but the DVM yields values of d^{+} and d^{-} as 0, resulting in d_{bias} and d_{CAVM} also being 0. In Fig. 2-5 (b), two datasets are visualized, where the EDF of $\widetilde{F}$ shows a concentration at the right edge of the probability box. This is not covered by the metric, and only the shaded area is considered. Therefore, identification of underlying distributions within the probability box based on the DVM is not possible. The distributions in (c) exhibit the same issue as in (b), but here, the other dataset is affected by the concentration. The metric also fails to capture this case. In Fig. 2-5 (d), a different shape of EDFs is present within the probability box. However, due to the two EDFs at the edge of the probability box, this is not reflected in the metric. From these four situations, the necessity arises to adapt the probability boxes to enable a meaningful and possible evaluation of the underlying distribution functions. The showed edge cases are addressed in Chapter 4.

67 Stripling, H. F. et al.: The Method of Manufactured Universes (2011), p. 1243.

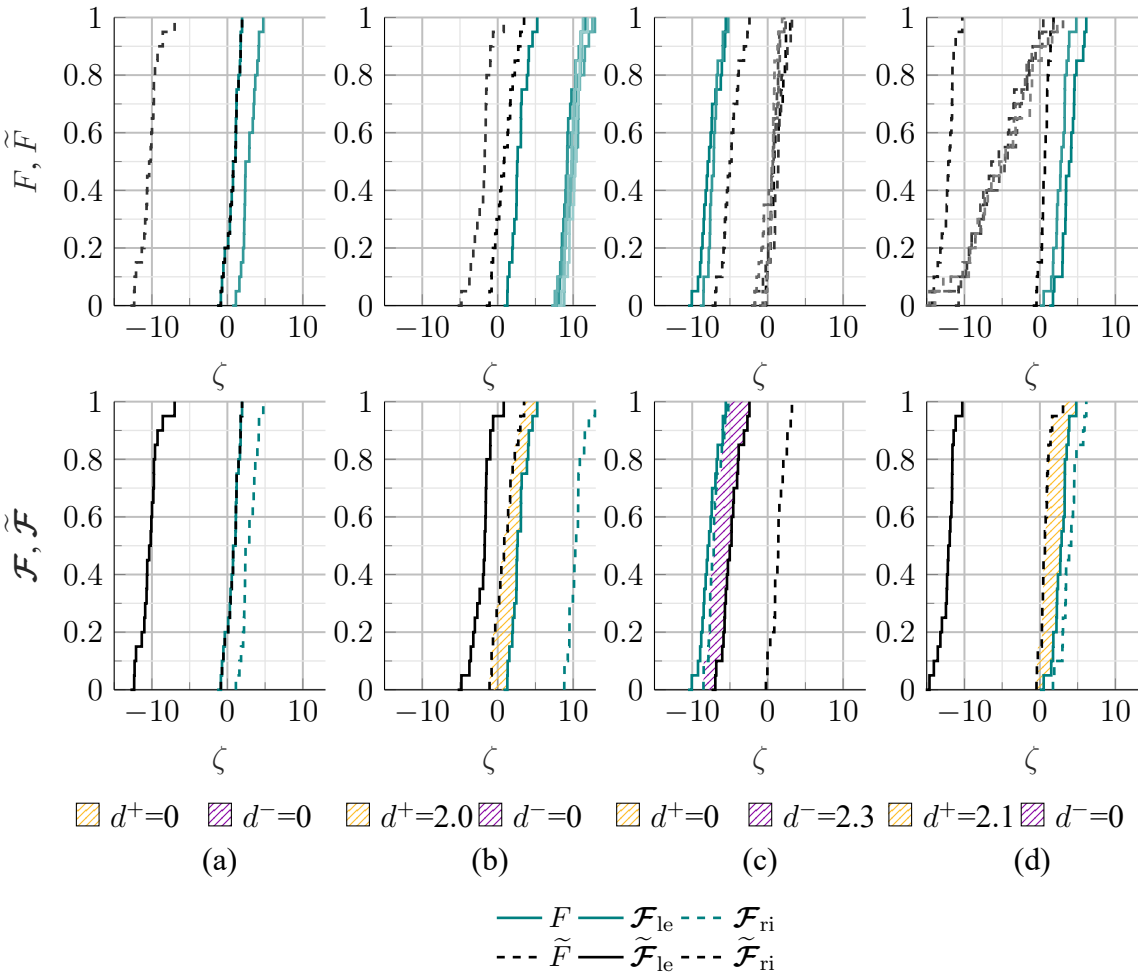


Figure 2-5: Edge cases regarding the DVM in combination with probability boxes, which are misleading in the two components of the metric.[68]

68 Elster, L. et al.: Introducing the DVM for Radar Sensor Models (2024), pp. 8–9.

3 Derivation of Research Questions

To derive the RQs, an evaluation of the state of research regarding methods for radar model validation in the automotive context is conducted. Various validation methods for active perception sensor models, including radar models, are available.

3.1 Validation Strategies for Radar Sensor Models

In his methodology for validating sensor models, Schaermann outlines a four-stage process. In the first step, real sensor data is recorded and transformed into an environment model for the ADF through fusion algorithms. This model then stimulates the driving function, influencing the vehicle dynamics through the actuators. In addition to comparing synthetic and measured sensor data directly, a comparison of subsequent processing stages is envisaged. Furthermore, necessary quantities for transferring validation experiments into the simulation are determined using reference sensors. In the simulation step, the trajectories recorded in the previous step are synthesized. Schaermann emphasizes the importance of the similarity of the real and simulated environmental simulation models. Subsequently, in the validation step, an assessment of the similarity between the captured real and synthetic value pairs is carried out. The quantitative evaluation is based on metrics, which are mathematical operators. Finally, based on the metric results and the comparison with the requirements, a decision regarding the validity is made. In the negative case, additional measurement data should be recorded, or the model should be re-parameterized. If the positive case occurs, and the requirements match the metric results, the model can be deployed.[69a]

Holder proposes a five-stage methodology for radar model validation[70], with the selection of artifacts to be implemented based on expert knowledge. In the first step, artifact allocation within the radar processing pipeline is defined, stating that the artifact or effect must be measurable in signal processing. Subsequently, artifacts are qualified, requiring the artifact to be distinguishable based on measurements of noise processes, with assignments to specific driving maneuvers or scenarios. The significance for the operational design domain (ODD) is also emphasized. A falsification experiment is defined in the third step to verify the theoretical model behind the artifact. The fourth step involves quantification through measurement databases, isolating artifacts to prevent a false assumption of the superposition of the effects of multiple artifacts. In the next step, a description model is synthesized, utilizing the simplest theory until validity is not

69 Schaermann, A.: Systematische Bewertung umfelderfassender Sensormodelle (2020), a: pp. 31–34; b: p. 33.

70 Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021), pp. 39–40.

achieved anymore. Finally, falsification experiments are derived by comparing measurement and synthetic sensor data.

Rosenberger et al.[71a] expand and adapt the existing validation methodology proposed by Viehof[72a]. This method involves an objective quality assessment of simulation models through statistical validation, applied within the context of vehicle dynamics simulation validation. Rosenberger et al. emphasize the crucial role of using appropriate reference data and ensuring trajectory reproducibility in developing an objective and widely accepted validation methodology. Their statistical approach entails identifying phenomena in sensor models for sample validity. In this context, sample validity refers to the constraint that only a specific number of experiments can be conducted, thereby limiting the comprehensive coverage of the simulation model's entire parameter space. Through sensitivity analysis, relevant configurations and parameters are determined, enabling the identification of likely scenarios for model falsification. Acknowledging the impracticality of comprehensive validation, the method focuses on essential scenarios. The process concludes with quantifying actual accuracies, asserting that synthetic data is only valid if the simulation reproduces a similar or the same PDF[73a]. The authors highlight the emerging nature of formulating model requirements and metrics in sensor data generation due to limited prior experience in the field.[75d] Rosenberger provides a detailed analysis of different metrics in his dissertation addressing this challenge.[73b] Furthermore, validation is performed for lidar sensors on the detection or point cloud level using the described methodology.[73c] The focus of the validation assessment is on distance uncertainty due to the reference sensor and its transfer into simulation as well as the impact on the metric result.[73d]

Ngo outlines a process involving measurements and simulations, where the two sets of data are directly compared using multiple metrics.[74a] Furthermore, post-processing algorithms are stimulated with both measurement and synthetic data, and the output of these algorithms is analyzed using specific metrics.[74b] The combination of these metrics results in the simulation-to-reality gap, defined as a measure of the deviation of the simulation compared to the real measurement.[74c]

Eder follows a similar process to Ngo, although the metrics for evaluation differ. The evaluation is based on statistical hypothesis testing with the Kolmogorov-Smirnov test.[75a] The research primarily focuses on various modeling approaches for radar sensors.[75b] For validation, a scenario characterized by a high number of repetitions is employed.[75c] The overarching validation process plays a less significant role.

71 Rosenberger, P. et al.: Towards Generally Accepted Validation Methodology (2019), a: pp. 9–11; b: p. 10.

72 Viehof, M.: Objektive Qualitätsbewertung von Fahrdynamiksimulationen (2018), a: p. 47; b: pp. 52–54; c: p. 54; d: p. 60; e: p. 68; f: p. 77; g: p. 82; h: p. 99.

73 Rosenberger, P.: Metrics for Simulating Sensors (2022), a: p. 46; b: p. 101; c: pp. 127–128, 134–137; d: p. 124.

74 Ngo, A.: Methodology for Validation of Radar Simulation (2023), a: p. 25; b: p. 27; c: pp. 31–33.

75 Eder, T.: Simulation of Automotive Radar (2021), a: pp. 79–82; b: pp. 33–75; c: p. 83; d: p. 27.

Magosi et al. also compare real and synthetic radar data, summarizing various scenarios. For the assessment of validity, they employ probability density functions with associated metrics.[76]

In all three mentioned publications by Ngo, Eder, and Magosi et al., no detailed considerations regarding uncertainties and how they impact the individual components of the validation chain is considered. Furthermore, the methodology lacks a process for defining requirements for a radar simulation model and obtaining validation scenarios.

In the global context, a research project exists on the Japanese side, which also addresses the modeling and validation of radar sensors.[77] In the Driving Intelligence Validation Platform project, rendering-based radar models are employed and equipped with standardized interfaces to be compatible with various simulation tools. However, according to the author's knowledge, aside from smaller presentations, there are no written publications demonstrating exact validation methods, measurements and metrics.

On the American side, NVIDIA has released an initial validation attempt with their simulation platform DRIVE Sim.[78] They validate their rendering-based radar simulation models using highly idealized objects in the form of a CCR and apply the comparison of mean and standard deviation as a metric. The focus of the experimental design is particularly on effect isolation to minimize the complexity of effect superposition. However, the described approach neglects uncertainties of the reference sensor and their impact on the validation result. In addition, the comparison of mean and standard deviation does not take into account the underlying distribution in the real and synthetic data.

Each of the authors and projects possesses an individualized model, distinct sensors, and a unique dataset acquired through measurements and scenarios, none of which are shared by others. Consequently, a lack of comparability of all approaches arises. In the current state of the art of radar validation methodologies, aleatory and epistemic uncertainties are inadequately represented, along with their impacts on both simulation and radar validation measurements. Particularly, the quality and precision of reference data require consideration for a validation campaign, as these data constitute the input to the model alongside the environmental model. From the examination of the existing validation methods concerning radar sensors, the main RQ arises:

How can effects and uncertainties in radar validation measurements, along with their associated impacts on radar sensor model validation, be identified and quantified?

76 Magosi, Z. F. et al.: Evaluation of Physical Radar Perception Sensor Models (2022).

77 DIVP Project: Driving Intelligence Validation Platform Homepage (2024).

78 Lehnen, M. et al.: Validating NVIDIA DRIVE Sim Radar Models (2023).

3.2 Validation Methodology by Viehof & Rosenberger

The validation process according to Viehof and Rosenberger encompasses all necessary steps from requirement definition to a statement of validity, considering uncertainties. This criterion distinguishes the validation process from others, prompting a more detailed examination of this process in the context of the main RQ. Therefore, the following RQs are derived based on the validation process, forming the core of this cumulative dissertation.

Fig. 3-1 illustrates the six-staged validation process by Viehof, represented with corresponding labels on the left side. On the right side, the substeps of each process stage are visualized. The shaded blue fields represent the modifications introduced by Rosenberger et al. and the adaptation for sensor models and their validation[71b] aligned with the V-model[79].

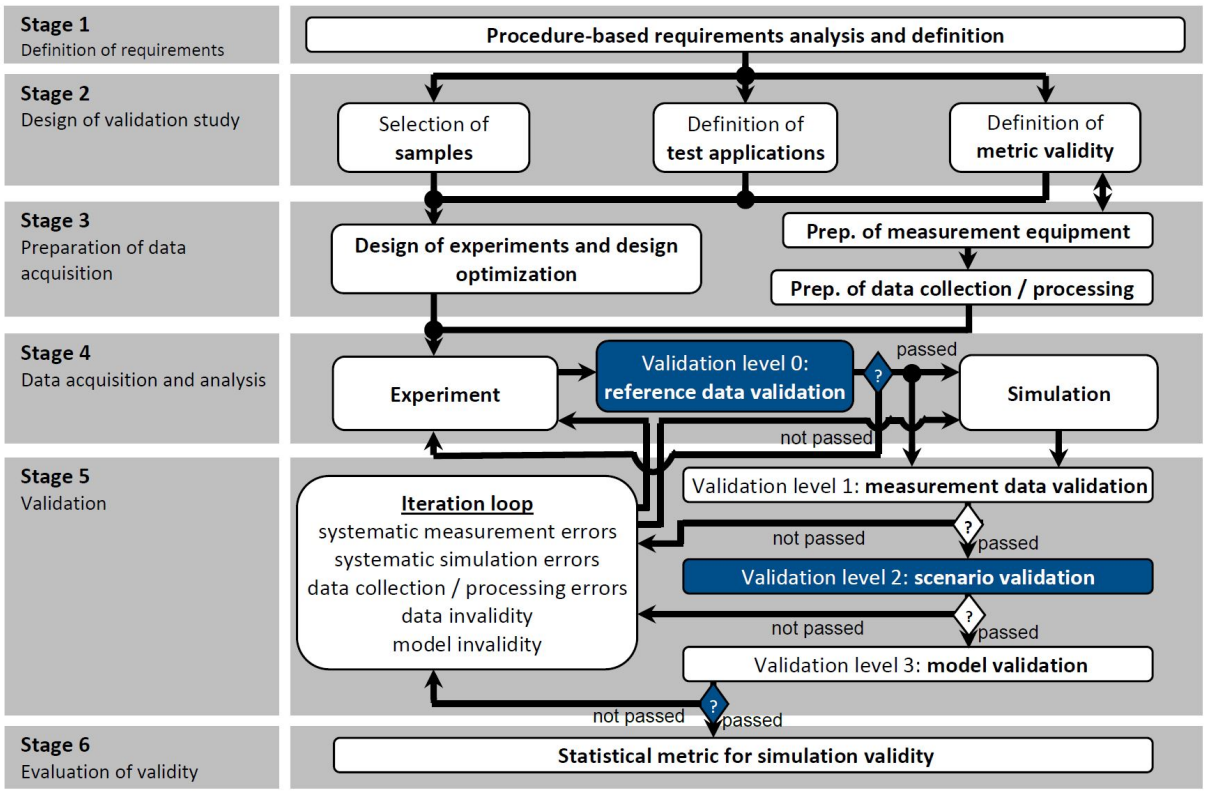


Figure 3-1: Validation process according to Viehof[72a] and the adaptations in blue by Rosenberger et al.[71b], arising from the application to sensor models.

In the first stage, requirements are derived through a methodical approach. In addition to requirements for the simulation model, supplementary requirements such as real-time capability need to be defined. Furthermore, the ODD and the associated application scope of the simulation model must be known. For instance, if an ADF is only deployed under dry conditions, then effects caused by rain need not be further considered in the simulation model and the environment. The requirements must be quantifiable for validation purposes.[72b]

In stage 2, the validation study is designed. Therefore, configuration samples are selected, test

79 VDI/VDE: VDI/VDE 2206: Development of Mechatronic and Cyber-Physical Systems (2021), pp. 22, 38.

applications are defined, and a metric is determined that allows a requirements based validity analysis. The selection of the configuration sample involves identifying simulation parameters that react particularly sensitively to changes at a specific operating point. This step is exclusively simulated through a sensitivity analysis, posing open questions regarding the discretization of the parameter space.[72c] In defining the test application, real-world experiments are introduced. Laboratory or test bench experiments serve as a substitute or complement to real experiments for validation purposes.[72d] The definition of metric validity criteria pertains to the output variables of the simulation model and subsystem interfaces. In the case of a radar model, these are the detection interface and the radar cuboid as a subsystem interface. Particularly concerning the metrics to be used, there is no consensus within the sensor model validation community.[73b,69b]
In stage 3, the experiment repetitions and the coverage of the parameter space are optimized. This is done using methods such as the design of experiments.[72e] Additionally, the measurement equipment with corresponding reference sensors is prepared, and data processing is arranged to ensure that measurement and simulation data are in an identical format.[72f]
Subsequently, in stage 4, data acquisition and simulation are carried out. An intermediate step is introduced that verifies the validity of the reference data. This involves checking the compliance with the required accuracy specifications from the requirement definition for the reference sensors. The exact details in the context of trajectory recording are not methodically elaborated.
In stage 5, it is ultimately decided based on three levels of validation whether an iteration in the execution of the experiment or simulation is necessary through an adaptation of the modeling approach or the parameterization of the model. In data validation, potential formal and systematic errors are identified[72g]. Furthermore, this point aims at the transfer of real data into the simulation.[71b] The synthesizing of trajectories into a simulation tool and their re-simulation are widely used in sensor model validation as an established procedure.[80] However, a check of the simulated trajectory is necessary since some simulation tools use the trajectory as a set point, thereby stimulating their trajectory control. This results in a deviation between measured and simulated trajectories. This uncertainty must be minimized in the validation process. Validation level 2 refers to scenario validation, comparing the sensitivity of the real scenario to a specific parameter compared to the sensitivity analysis from stage 2.[71b] Subsequently, the evaluation of the validation study is performed in validation level 3 based on the metrics defined in stage 2.
In stage 6, all results from various validation scenarios and simulations are collected and structured. The final validation report is compiled with the aid of a so-called validity assessment card. In addition to an absolute comparison of real and synthetic data, a relative comparison is also conducted. This is achieved by comparing the relative deviation between two real experiments with the relative deviation of the virtualized experiments.[72h]
The described process for radar sensor model validation reveals stages that undergo a more in-depth analysis in the following. These include stages 1, 2, 4, and 5. Detailed explanations with justifications follow in chapters 3.2.1 to 3.2.4, resulting in the RQs that are analyzed.

80 Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021), p. 130.

3.2.1 Requirements for Radar Sensor Models

At the outset of model development, requirements need to be defined. One potential process for definition is the Credible Modeling Process from the publicly funded SET Level research project. This involves an initial analysis of the modeling tasks and their objectives, including the definition of „*key performance indicators and other criteria*“. However, the description lacks a precise specification on how to determine these criteria. In the subsequent phase of requirement definition, relevant assumptions and model requirements are established.[81] A more detailed elaboration is provided by Ahmann et al. In the implementation phase and the necessary formal validation, a quantitative determination of simulation validity is to be performed, considering uncertainties.[82a] However, the selection of specific metrics is left unaddressed.[82b]

Linnhoff et al. propose a process for requirement definition resembling a fault tree analysis structure based on sensor effects and their interactions.[83] This involves an assessment of resulting chains based on the frequency of occurrence in the ODD and the relevance for the system under test.[84] Rosenberger applies this approach exemplarily.[85] However, defining values and boundaries for each effect, which are allowed to exist, poses a challenge for making the requirements quantifiable and assessable.

Aust et al. address this issue by conducting measurements and deriving quantifiable boundaries[86], following a similar approach to Viehof in the context of vehicle dynamics simulation[87]. They employ two metrics for comparing point clouds resulting in significant fluctuations in the metric. Therefore, the reliability of these metrics for requirement definition is questionable. To address this, they analyze histogram distributions and the relative probability distributions of detection positions using two other metrics. This allows for the determination of metric boundaries, which are then utilized as requirements, along with the mean value. However, interpreting metric results remains challenging, as a relationship is defined where the optimal value corresponds to 0 and the worst value to 1. The bin discretization of histograms in Cartesian coordinates relies hereby on expert knowledge and neglects the polar or spherical nature of radar cuboid bins. This leads to RQ 1:

How can radar measurements be used to establish quantifiable requirements for a radar sensor model?

81 Heinkel, H.-M.; Steinkirchner, K.: Credible Simulation Process Framework (2023).

82 Ahmann, M. et al.: Towards Continuous Simulation Credibility Assessment (2022), a: p. 176; b: p. 180.

83 Linnhoff, C. et al.: PerCollECT - LidarLimbs (2022).

84 Linnhoff, C. et al.: Towards Sensor Simulation for Safety Validation (2021), pp. 4–5.

85 Rosenberger, P.: Metrics for Simulating Sensors (2022), pp. 148–165.

86 Aust, P. et al.: Numerical Synthesis of Radar Target Detections (2023), pp. 27–29.

87 Viehof, M.: Objektive Qualitätsbewertung von Fahrdynamiksimulationen (2018), p. 51.

3.2.2 Effect Isolation and Separation

After identifying relevant effects and defining corresponding requirements for the considered interfaces, validation experiments are required. Therefore, it is crucial to consider various effects and their causes. For the experimental design, it is essential that effects are measured under repeatable or reproducible conditions with as much isolation as possible. The ideal effect isolation occurs under laboratory conditions, where, for example, multipath propagation for radar is suppressed using radiation-absorbent materials in radar chambers. This test environment is suitable for examining individual effects and using them for validation. However, the behavior of sensors deviates in real-world scenarios. Therefore, in addition to these laboratory measurements, real-world experiments are also necessary to identify and demonstrate correlations or causality between effects. For this reason, experiments on test sites are required alongside laboratory experiments.

Such attempts, where effect isolation is the primary focus, are presented by NVIDIA.[88] They utilize three different scenarios to validate the effects of sensor field of view, separation capability, and Doppler velocity. For this purpose, CCRs are employed as objects, and the driven trajectories are transferred into the simulation using corresponding reference sensor data. However, there is a lack of additional assessment of the returned power of the objects.

Holder follows the approach of effect-based validation in his dissertation, designing falsification experiments for multipath propagation effects with object occlusion[89a], interference caused by multipath propagation[89b], intrinsic sensor uncertainties, noise, clutter[89c], and extrinsic uncertainties[89d]. One of the significant parameters in radar at the detection interface, comparable to a point cloud, is the previously described RCS.[89e] Thereby, the RCS is a measure of the reflection intensity compensated by range and the antenna diagram. The RCS is particularly sensitive to changes in angle as mentioned in Chapter 2.2.2. Holder uses a slalom setup for the falsification of RCS properties, employing only one object for validating the radar models used. The experimental setup is challenging, mainly due to the low reproducibility of radar experiments. Another issue is the difficulty in isolating the effects from various influencing factors related to RCS. Especially, analyzing correlations of various effects is necessary. To illustrate the complexity of effect isolation and still consider correlations and causality due to the complexity of real environmental conditions, RQ 2 arises:

How is it possible to isolate effects related to the RCS of road vehicles and make them measurable under real world conditions?

88 Lehnen, M. et al.: Validating NVIDIA DRIVE Sim Radar Models (2023).

89 Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021), a: pp. 43–47, 51–60; b: pp. 48–50, 61–69; c: pp. 70–75; d: pp. 76–78; e: pp. 145–147.

3.2.3 Re-Simulation of Reference Data

Reference data from, for example, RTK-based GNSS systems are fundamental for transferring driven trajectories from reality to simulation. Thus, the simulation and, consequently, the validation result depend directly on the accuracy of this transformation. Rosenberger et al.[90] explain that re-simulation of trajectories is meaningful only if the accuracy of the reference data in time and space meets the requirements.[91a] Especially in urban environments, fluctuations in reference data accuracy can be expected due to building occlusions and therefore multipath propagation of the satellite signals.[92] For these reasons, a qualification of reference data accuracy is necessary beforehand, as outlined in the validation methodology in Fig. 3-1.

Furthermore, not only aleatory uncertainties of the reference sensor system need to be considered but also the transfer of the trajectories into simulation. Deviations due to temporal effects, such as interpolation or extrapolation of timestamps, and the implementation in the simulation tool, including how a trajectory is recreated, are sources of epistemic uncertainties.

Roth et al. use the Virtual Test Drive software by Vires and transform the corresponding reference data into a format readable by the simulation tool. The recreation of trajectories within the accuracy of the reference sensor is ensured based on the imported data.[93] However, Roth et al. do not provide experimental evidence to support their claim. Schaermann also emphasizes the need for an exact reproduction of the real trajectory to avoid introducing additional uncertainties, but there is no evidence provided for the precise translation into the simulation.[94a] Similarly, Magosi et al. adapt the CarMaker simulation tool to eliminate epistemic uncertainties, but no evidence is presented to support this statement.[95]

Particularly, aleatory uncertainties from reference data, such as RTK-based GNSS systems, are not considered in existing validation methods. There is a lack of experiments and procedures to capture the aleatory uncertainties of reference measurement systems. Additionally, a quantitative determination of epistemic uncertainties from the simulation tool is required. This leads to the third research question regarding uncertainties in radar model validation, RQ 3:

How are radar model validation campaigns affected by the accuracy of the reference sensor system?

90 Rosenberger, P. et al.: Towards Generally Accepted Validation Methodology (2019), pp. 7–8.

91 Schaermann, A.: Systematische Bewertung umfelderfassender Sensormodelle (2020), a: p. 44; b: pp. 33.

92 Gottschalg, G.: Data Fusion Architecture for State Estimation (2022), p.103.

93 Roth, E. et al.: Analysis and Validation of Perception Sensor Models (2011), p. 5.

94 Rosenberger, P.: Metrics for Simulating Sensors (2022), a: p. 32; b: p. 68–72; c: pp. 120–128.

95 Magosi, Z. F. et al.: Evaluation of Physical Radar Perception Sensor Models (2022), p. 8.

3.2.4 Validation with Uncertainties of Radar Sensor Models

After identifying and quantifying sources of uncertainties in reference sensors and the fundamental uncertainties of validation measurements, the validation process involves comparing real radar data with synthetic radar data. The quantitative comparison of data is carried out using metrics. Metrics in the automotive sensor validation context are categorized into hypothesis tests, confidence intervals, comparison of PDF, and correlation measurements. Hypothesis tests provide information on whether a hypothesis is accepted or rejected, where the determination of the acceptance level is subjective. Confidence intervals estimate the difference between the experiment and the model. The comparison of PDF is based on the similarity analysis of probability distributions. Correlation measurements provide a measure of the similarity of several statistical quantities.[91b] Generally, the application of which metric is used in different situations is not yet standardized. Rosenberger provides an overview of metrics in all named categories.[94b]

The metric must additionally be capable of incorporating uncertainties from reference sensors into the validation process, as the GT position of sensors or objects is unknown. Therefore, it is necessary to transfer uncertainties into the simulation by simulating both the measured value and the value affected by uncertainties. As an example, consider the object's x-position relative to the sensor, determined with a reference sensor to be 10 m. The uncertainty of the measurement system is 0.02 m. Consequently, multiple simulations are conducted with the measured value and values between 9.98 m and 10.02 m. This procedure is already established for the distance uncertainty only in lidar.[94c]

In addition to the uncertainties of the reference sensor, the measurement uncertainties of the radar sensor itself must also be taken into account in the validation result. A sensor with a very high reproducibility of the validation measurements requires completely different properties of the simulation model than a sensor with very large aleatory uncertainties. For this reason, a series of measurements consisting of several measurements is required for a plausible validation.

For the reasons mentioned above the extension of the described procedure to other uncertainty-affected quantities and the radar sensor with its interfaces is lacking. The final research question, RQ 4, is therefore:

How can reference sensor and radar sensor measurement uncertainties be taken into account in a metric-based radar model validation across various interfaces?

4 Requirements

Prior research emphasizes methodologies and defining active perception sensor model requirements based on effects as described in Chap. 3.2.1. However, a quantifiable measure is missing to define acceptance criteria for radar sensor model requirements on different interfaces in the radar signal processing. Also, the influence of different, yet mutually independent effects and the uncertainties due to the measurement setup itself are still an open point in the context of radar sensor model validation. From the mentioned points RQ 1 arises addressing stage 1 in the overall methodology in Fig. 3-1:

How can radar measurements be used to establish quantifiable requirements for a radar simulation model?

Especially in the context of validation measurements, it is essential to ensure that measurements at different locations under comparable conditions yield acceptable deviations, and resulting maximal allowable deviations need to be defined. The deviations resulting from overlapping influences constitute the central focus of this chapter and the research findings presented herein. Therefore, in a first step, a methodology is derived to evaluate the various interfaces using the DVM. The content is based on **paper III**.[96a]

4.1 Methodology

A maximum reachable simulation accuracy is derivable based on real world measurements.[97] Therefore, along with a carefully chosen metric, effects and uncertainties can be determined from the measurement data. Hence, the necessity arises for a suitable metric that is particularly easy to interpret to associate potential deviations to the corresponding underlying effects. From the description in Sec. 2.4.2, it is evident that the DVM can be utilized as a metric to statistically analyze various measurement quantities. However, a method is lacking to apply it to the respective interfaces of radar processing considering the edge cases in Fig. 2-5.

Fig. 4-1 illustrates the methodology used for evaluating radar cuboid and detection data. The recorded measurement data is processed and filtered according to the position of the target object. Subsequently, the number of data points in the resulting measurement set n is compared for different measurements. If a deviation under 10 % is present, the EDF is formed for all sets of comparable measurement data. This value should be checked for future work with hypothesis

96 Elster, L. et al.: Making Automotive Radar Sensor Validation Measurements Comparable (2023), a: -; b: p. 4; c: p. 11; d: p. 20–21.

97 Roache, P. J.: Validation in Fluid Dynamics and Related Fields (2019), p. 665.

testing to ensure the comparability between the data. For each EDF combination, d_{bias} is then calculated, and the EDF is shifted by this value. Subsequently, d_{CAVM} is computed, and both values are presented as box plots for all comparisons for evaluation. This approach differs from Rosenberger in that the p-boxes are resolved, and the distribution of deviations in the EDFs are examined. This enables dissolving the edge cases previously defined.

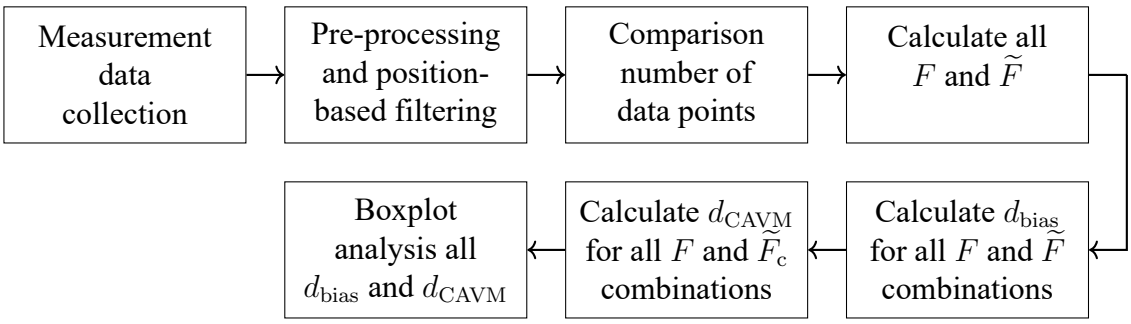


Figure 4-1: Methodology for the comparison of different experiment setups, measurements and input data on radar cuboid as well as detection level.[96b]

To quantify measurement offset d_{bias} as well as scattering deviations d_{CAVM}, multiple static measurements without and with objects are conducted. The focus in the following is on analyzing the dismantling and setup of the measurement configuration, as well as the influence of rain as an exemplary environment condition on radar measurements. In the first scenario, no objects are present in the measurements, and in the second setup, a Volvo XC90 serves as an exemplary vehicle on the August-Euler Airfield proving ground in Griesheim. The evaluation is performed for the radar cuboid and for each individual cell of the radar cuboid. Details on the experimental setups and their evaluations, including the detection level, can be found in **paper III**.

4.2 Disassembly and Assembly Measurement Setup

In this section, the impacts on validation measurements arising from the disassembly and rebuild up of the measurement setup are analyzed. Eight experiments without any specific object in the sensor's field of view are conducted on both Day 1 and Day 2. Fig. 4-2 illustrates the evaluation for the entire radar cuboid. The diagram presents experiment repetitions in teal (Day 1) and black (Day2) with corresponding brightness gradients. The visual impression generally indicates a high agreement in the offset and curve shapes. The variation between -87 dB and -79 dB arises from noise in the more distant bins, while higher power values result from vegetation and asphalt in the sensor's proximity. The tabulated counts of recorded data show only a one percent difference, attributed to the data recording process itself. In a detailed analysis, it is observed that despite the predefined 60 s measurement duration, some timeframes are missing in the measurement data, resulting in the mentioned discrepancy. The maximum deviation in the number of bins is 10 %. Thus, it is imperative to ensure a reduction in this deviation during data recording to maintain the

comparability of the metric outcome. On the right side of the figure, the box plots of the DVM are presented, revealing a median difference of only 0.3 dB for d_{bias} and 0.19 dB for d_{CAVM}.

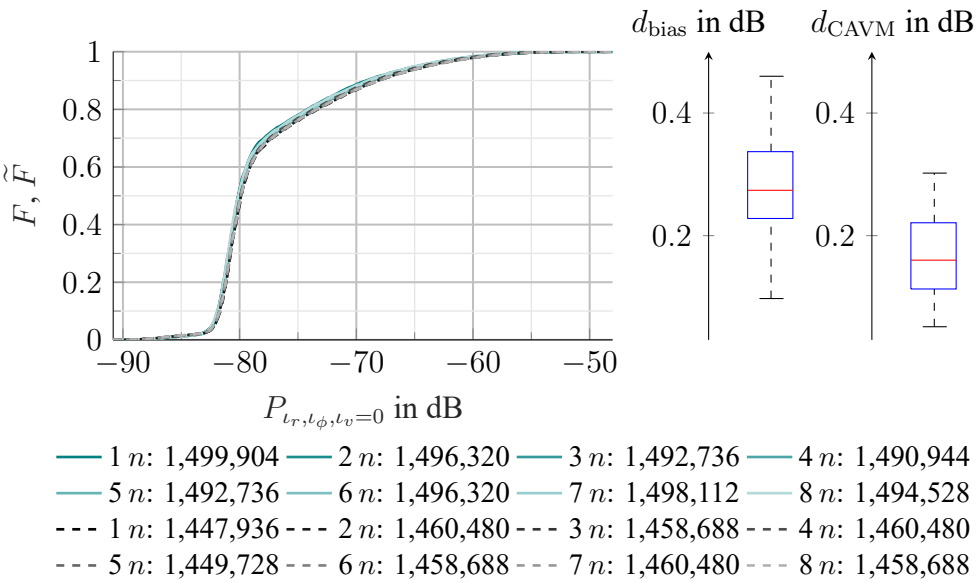


Figure 4-2: EDFs of all eight measurements for Day 1 (—) and Day 2 (- -) each on the radar cuboid level without a placed object. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.[96c]

A detailed examination of the impact of assembly and disassembly, as well as the general scattering of the sensor, is presented in Fig. 4-3. Once again, the DVM values over day 1, day 2 and the direct comparison between both days are visualized. The comparison reveals that the variations between individual measurements on day 1 and day 2 are smaller compared to both days. The assembly and disassembly result in an increase in the median deviation by 0.25 dB. Thus, the reproducibility of individual measurements is greater than that of the assembly and disassembly. The deviations are primarily attributed to the angular alignment of the sensor.

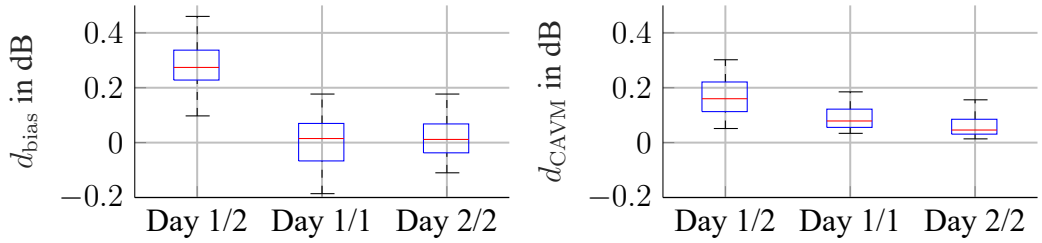


Figure 4-3: Comparison of d_{bias} and d_{CAVM} based on the whole radar cuboid for a comparison of Day 1/2, Day 1/1, and Day 2/2. The repeatability between the measurements of each day is higher than the reproduciblity of the measurement setup.[96c]

4.3 Influence of Rain on Radar Validation Measurements

Fundamentally, environmental conditions such as temperature[98] and rain[99] have an impact on radar measurements and, consequently, on comparability. For this reason, these conditions must also be taken into account in the validation campaign. This is evident in Fig. 4-4 between dry and rainy conditions. Due to the rain, there are more cells in the radar cuboid with higher power, leading to a shift towards larger values in the power dimension. In principle, while atmospheric attenuation increases due to rain, raindrops reflect a portion of the electromagnetic radiation, resulting in a relative increase in power at short distances. This can also be observed in the DVM values. A shift in the median of d_{bias} by 1 dB indicates an increase in power. A 1 dB deviation in d_{CAVM} reveals increased scattering in radar measurements due to rain. For completeness, the number of data points is again presented in tabular form, showing negligible deviations.

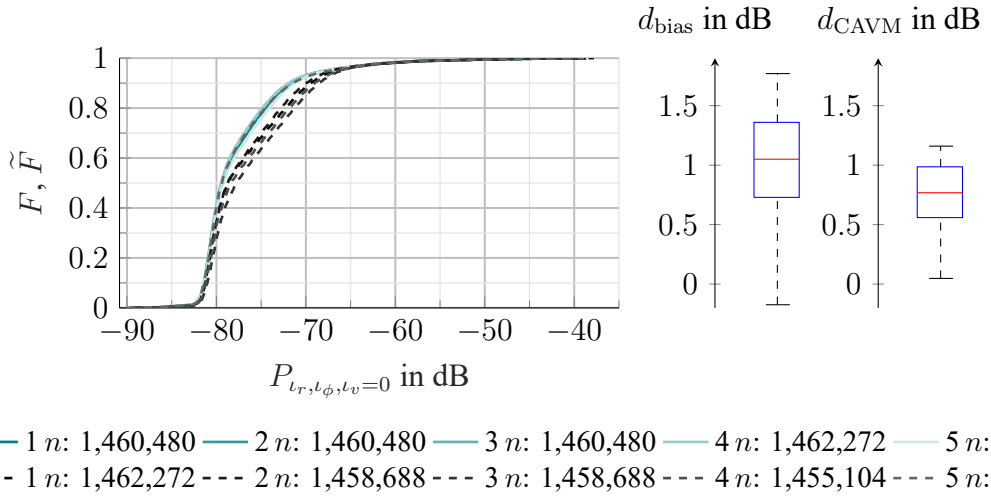


Figure 4-4: EDFs of five measurements of the XC90 in dry (—) and rainy (- -) conditions on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.[96d]

The analysis of the entire radar cuboid demonstrates the fundamental influence. A detailed examination follows to identify basic influence areas within the sensor's field of view. For each individual cell of the radar cuboid, the DVM is calculated and combined with a satellite image. The result of this approach is shown in Fig. 4-5. The two images display the maximum value of the box plot for d_{bias} and d_{CAVM} for each radar cuboid cell in the comparison of rainy and non-rainy conditions. The red circle in the figure visualizes the position of the center of the bounding box of the Volvo XC90. The increased power is evident in both d_{bias} and d_{CAVM} represented by the big green areas in the satellite image. Notably, d_{bias} experiences an increase in proximity to the sensor, extending up to approximately half of the maximum sensor range.

98 Arage Hassen, A.: Signal Degradation and Optimization Radar Sensors (2006), pp. 14–18.

99 Li, H. et al.: The Effect of Rainfall and Illumination on Automotive Sensors Detection Performance (2023), p. 11.

This phenomenon exerts a more pronounced impact on the road surface due to rain and the resulting change in material properties. In the case of d_{CAVM}, the most significant deviations manifest within the mid-range of the field of view. Additionally, it is observed that the road is more susceptible to these effects than the vegetation, a characteristic that can be attributed to the stochastic properties of the vegetation movements due to wind.

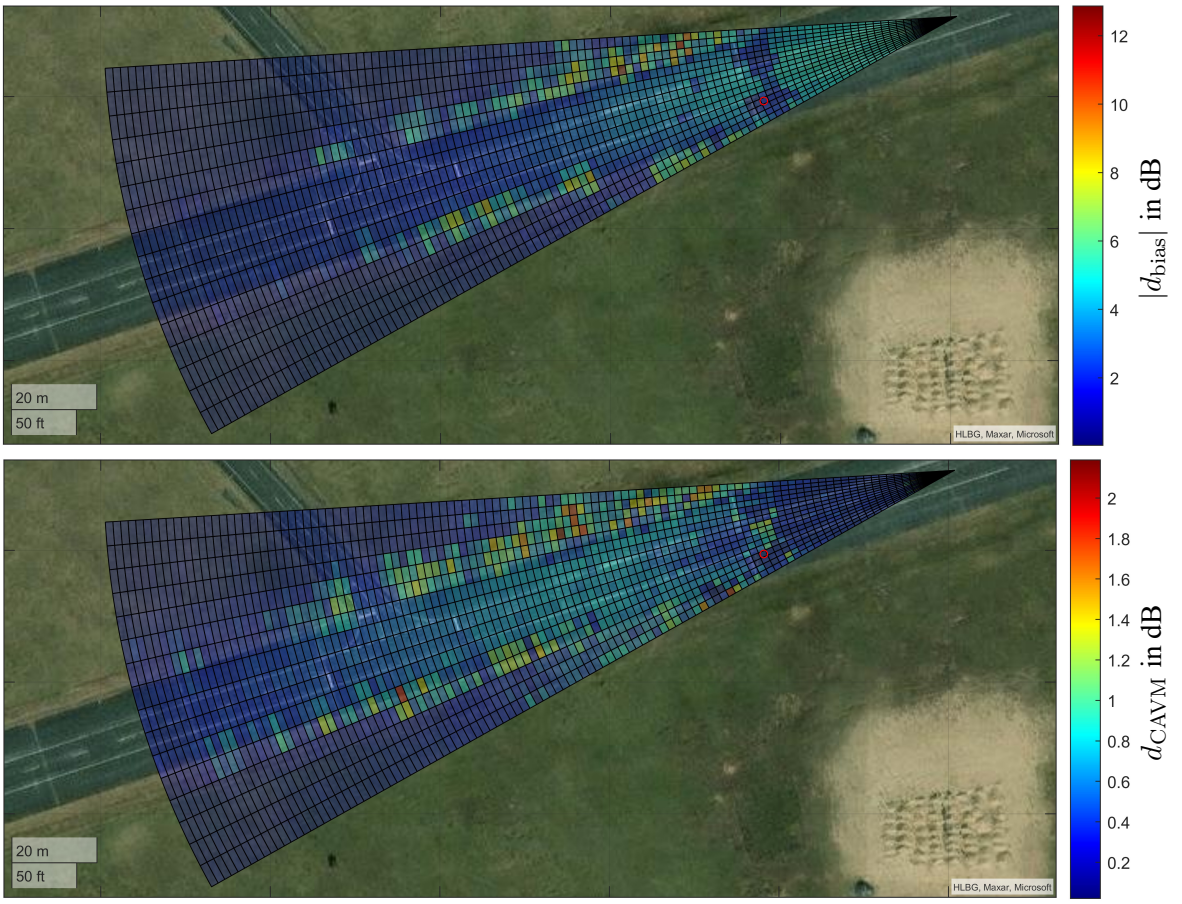


Figure 4-5: Satellite image fused with $|d_{\mathrm{bias}}|$ and d_{CAVM} coded in colors for each range azimuth bin. The object under examination is a Volvo XC90 in rainy and dry conditions. The red circle visualizes the center of the vehicle bounding box.[96d]

The procedure for deriving requirements through measurements is applicable exclusively to sensors that currently exist in the physical domain. However, effects in signal propagation are verifiable independently of signal processing in the sensor, making these effects generally verifiable. For this purpose, consideration of wavelength, objects, the environment, and environmental conditions is necessary. Thus, the presented methodology is also suitable for developing and testing sensors that exist purely as a digital twin in the future.

5 Design of Validation Study

This chapter presents the content of **paper II**.[100a] The subsequent discussions align with stage 2 of the overall validation methodology, as depicted in Fig. 3-1. As previously discussed, it is necessary to isolate effects in signal propagation and signal processing as effectively as possible to enable a sensor model validation. Therefore, in the experimental design, efforts must be made to reduce or even eliminate potential mutual interference between effects. Therefore, from Chapter 2.2.2, the following effects arise, which occur in real measurements in radar sensors:

Multipath propagation: The electromagnetic radiation of the radar sensor is reflected over the asphalt and under the vehicle's chassis. Multipath propagation is dependent on the mounting height. Therefore, measurements at different sensor heights must be taken simultaneously.

Yaw angle of the object of interest (OOI): The reflection of radar radiation depends on the orientation of the object relative to the sensor. Therefore, it is necessary to vary the relative yaw angle between the OOI and radar sensor during the experiment.

Geometry and material of the OOI: The RCS as well as the distribution of detections over the bounding box are significantly dependent on the geometry and material of the vehicle components. Therefore, in the measurement campaign, 14 objects are examined and analyzed.

How is it possible to isolate effects related to the RCS of road vehicles and make them measurable under real world conditions?

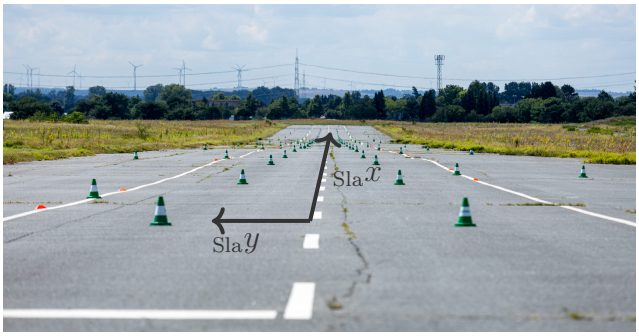
5.1 Measurement Setup

With the defined requirements, the experimental setup is depicted in Fig. 5-1. To analyze the impact of multipath propagation on the RCS, six radar sensors of the same type are mounted on top of each other on the ego vehicle as shown in Fig. 5-1 (a). The calibration of radar sensors is individually conducted for each sensor using a CCR. This approach circumvents issues related to the reproducibility of trajectories and ensures direct comparability of the measurement results from the different radar sensors. Due to the interference suppression implemented in the radar sensors, crosstalk effects will only affect the Doppler component and the signal-to-noise ratio of

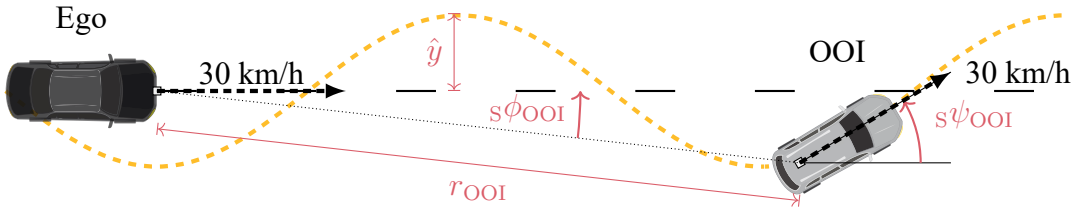
100 Elster, L. et al.: Dataset Radar Scattering Characteristics (2023), a: -; b: p. 4875; c: p. 4876; d: p. 4877; e: p. 4878.

the detections and not the RCS.[101] This property is checked before evaluating the measurements. To vary the other mentioned parameters from the requirements, a slalom trajectory for the OOI and a straight-line motion of the ego vehicle are suitable. Fig. 5-1 (b) and Fig. 5-1 (c) depict the setup of the slalom on the August-Euler proving ground in Griesheim, along with the corresponding parameters. The indices for the local coordinate system Sla of the slalom and the sensor coordinate system S are placed in front of the variable as a subscript. Both the ego vehicle and the OOI are equipped with a RTK-based GNSS system ADMA, and the measurement data is synchronized using GNSS time. In Fig. 5-1 (c) $\hat{y}$ is denoted as the slalom's amplitude, ${}_{\mathrm{S}}\phi_{\mathrm{OOI}}$ is denoted as azimuth angle, r_{OOI} is denoted as range in the polar sensor coordinate system and ${}_{\mathrm{S}}\psi_{\mathrm{OOI}}$ is denoted as the yaw angle difference between ego vehicle and OOI. For each vehicle, a slalom with 10 periods was repeated 10 times. The measurement data[102] and the evaluation code on GitLab[103] are publicly available.

(a) Picture of the ego vehicle equipped with six identical radar sensors.[100b]



(b) Slalom course marked with traffic cones and slalom-centered coordinate system.[100c]



(c) Measurement setup of slalom with corresponding parameters where ${}_{\mathrm{S}}\phi_{\mathrm{OOI}}$ is denoted as azimuth angle and r_{OOI} denoted as range in the polar sensor coordinate system, ${}_{\mathrm{S}}\psi_{\mathrm{OOI}}$ is denoted as the yaw angle difference between ego vehicle and OOI as well as $\hat{y}$ is denoted as the slalom's amplitude.[100d]

Figure 5-1: Experimental setup for the examination of isolated effects on the RCS of different vehicles. © IEEE 2023.

To analyze the introduced effects, the entire detections are initially filtered based on the reference data from the ADMA. In addition to the ego and OOI poses, the relative radial velocity of the OOI is used as a filter. The RCS value of the filtered detections are de-logarithmized and summed. In this process, the phase of the signal resulting from the path difference and corresponding reflections is neglected. This is because signal processing already involves interpolations that

101 Norouzian, F. et al.: Phenomenology of Automotive Radar Interference (2021), p. 1057.

102 Elster, L. et al.: RCS Measurement Dataset (2022).

103 Elster, L.: RCS Measurement (2022)

do not allow precise conclusions about the actual signal phase at the object on the detection interface. Furthermore, the phase changes due to vehicle movements over a measurement cycle of the Chirp-Sequence cycle, making the determination of the exact phase impossible. Therefore, for RCS analysis, the following relationship is employed, where Q denotes the logarithmic size of σ, $n_{\mathrm{D,max}}$ denotes all detections per measurement cycle and $\sigma_{k,\mathrm{filt}}$ denotes detections filtered by position and velocity. To obtain the resulting RCS, the sum of all $\sigma_{k,\mathrm{filt}}$ is calculated by the antilogarithm of the corresponding value. Subsequently, this sum is transformed back into the logarithmic scale to obtain the overall RCS in dB m^2:

$$Q_{(\sigma)} = 10 \log_{10} \sum_{k=1}^{n_{\mathrm{D,max}}} \sigma_{k,\mathrm{filt}} \tag{5-1}$$

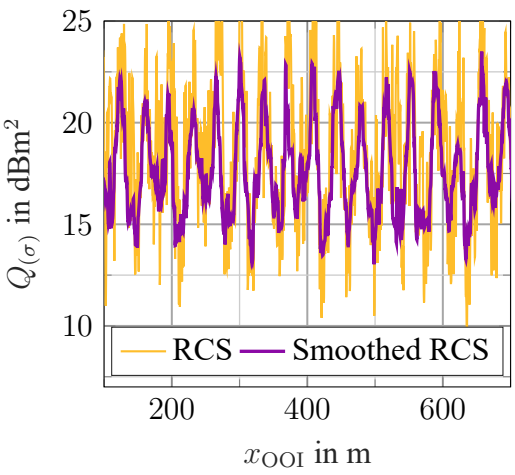
To examine the various effects and conduct separate analyses, different evaluation methods are employed. Due to the sinusoidal setup, in addition to temporal or spatial analysis, a frequency-based examination is suitable. Furthermore, the RCS data is statistically processed, and the spatial distribution is considered.

5.2 Spectral Analysis of RCS

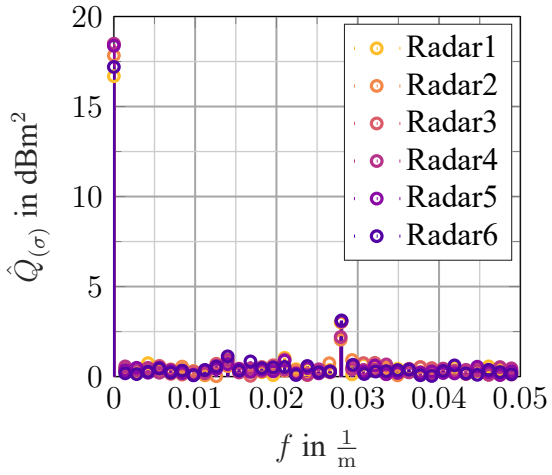
In Fig. 5-2, the periodic RCS profiles resulting from the slalom are analyzed. Fig. 5-2 (a) displays the RCS across the x-position of the OOI for one radar sensor. Upon processing the periodic profiles of distinct radar sensors through FFT analysis, Fig. 5-2 (b) is generated with the frequency on the abscissa. The two amplitude peaks at zero-frequency and the double harmonic frequency of the slalom reveal the effects of the sensor mounting position on the RCS. Slight variations of approximately 3 dB m^2 and 1 dB m^2 are noticeable, but they do not yet provide any generalizable insights into the mounting position in comparison to the overall slalom RCS amplitude.

For this reason, in a subsequent step, the analysis is conducted across all test repetitions and vehicles. The result is depicted in Fig. 5-3. Concerning the two frequency components, a high degree of repeatability is observed among different test repetitions, with negligible differences between the various radar sensor mounting positions in the context of automotive applications. Only in the case of radar sensor 6 is a greater dispersion evident among different test repetitions. Hence, it is apparent that multipath propagation influences the RCS within the vehicle primarily at low mounting positions. The vehicle geometry is ruled out as the cause, as the increasing dispersion in radar 6 is evident across all vehicles.

105 Elster, L. et al.: Dataset Radar Scattering Characteristics (2023), a: p. 4879; b: p. 4878; c: p. 4880.



(a) RCS of Toyota Auris observed while it is driving through the slalom. A periodic pattern is visible after moving median smoothing.[100d]



(b) Spectral analysis of RCS during the slalom in Fig. 5-2 (a). $\hat{Q}_{(\sigma)}$ denotes the amplitude of the oscillating RCS pattern.[100e]

Figure 5-2: Periodic characterization of the RCS for a Toyota Auris. © IEEE 2023.

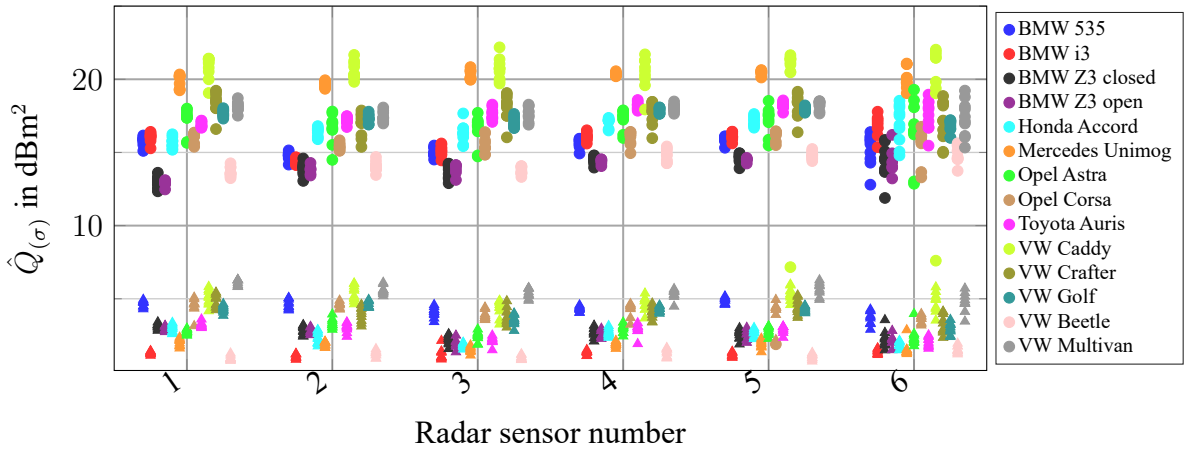


Figure 5-3: Spread of spectral amplitudes obtained by FFT over all studied cars and trials. Circles denote zero-frequency amplitude, triangles indicate amplitude at double natural frequency. The figure shows all radar sensors and vehicles.[104a105b] © IEEE 2023.

5.3 Analysis of Detection Distribution

In addition to frequency analysis, examining the RCS distribution is meaningful for comparing different sensors, test repetitions, and vehicles. To illustrate this, the Toyota Auris is examined in detail in a first step. In Fig. 5-4 (a), the EDF of RCS values for various radar sensors and the three most similar[106] test repetitions are visualized.

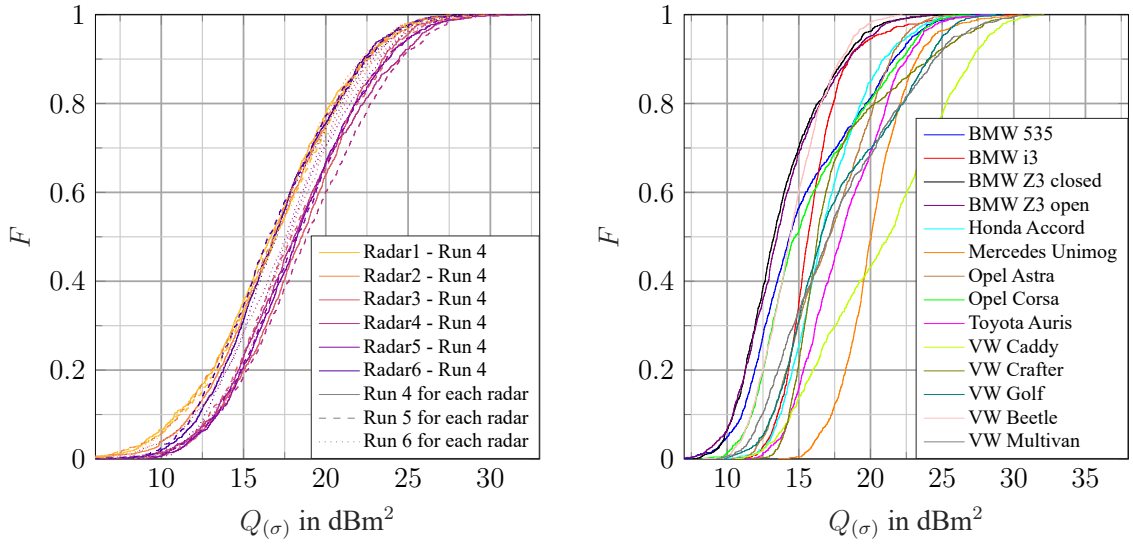
In general, it can be observed that deviations for a radar position deviate by approximately 2.5 dB m^2. This confirms the findings from Fig. 5-3 and additionally shows that the outliers are also reproducible measurable. Radars 1, 2, and 6 deviate by only 1 dB m^2 from each other, with

106 The evaluation is based on the Hilbert criterion, which is explained in detail in **paper II**.

the lowest RCS. Radars 3, 4, and 5 exhibit the highest RCS values.

In the various test repetitions, the distance between the OOI and ego is also different, which is noticeable in the measurement results regardless of the distance. Due to the varying distances, there is also a difference in the azimuth angle distribution, suggesting that the antenna pattern and its angular dependence have no impact on the RCS under the given conditions. Further in-depth analyses of different oscillations confirm this hypothesis.[107]

In Fig. 5-4 (b), the RCS distributions for different vehicles are presented. In addition to the offset of the different vehicles, there is also a difference in the shape of the EDFs. The RCS characteristics of different objects deviate by up to 7 dB m^2 from each other. Thus, in terms of RCS, it can be concluded that the RCS characteristic of each vehicle on the road is different, and there is no readily generalizable RCS profile. The presented measurement setup provides the opportunity to quantify these resulting differences.



(a) EDF of Toyota Auris' RCS of all six radar sensors and the three best runs in terms of a defined criterion.[105b]

(b) EDF of all OOI's summed RCS of all radar sensors of the best run in terms of the Hilbert criterion.[105b]

Figure 5-4: Analysis of the different RCS distribution functions for one vehicle as well as all vehicles. © IEEE 2023.

To conduct a more in-depth analysis, an examination how the detections are distributed across the bounding boxes of various vehicles. Fig. 5-5 illustrates the locations of the largest RCS detections for each vehicle and radar sensor.[108] Therefore, this analysis provides insights into the primary scattering areas of each vehicle. The color coding in the figure represents the yaw angle difference ${}_{\mathrm{S}}\psi_{\mathrm{OOI}}$ between the ego vehicle and the OOI, while black rectangles delineate the bounding boxes of the vehicles.

107 Rapp, M. L.: Messkampagne für winkelabhängige RCS-Profile (2021), pp. 75–82.

108 The figure with all vehicles is visible in **paper II**.

Detections outside the bounding box can be attributed to timing effects resulting from different measurement sampling rates and the inherent resolution of the radar sensor itself. In general, these detections tend to cluster at the rear center of the vehicles, consistent with findings in previous studies.[109,110] However, there are subtle distinctions between radar 1/6 and radar 2/3/4/5 in the case of BMW 535i, Honda Accord, Opel Corsa, Toyota Auris, and VW Caddy.

Notably, certain unique characteristics are observed in the distribution of the detection for specific vehicles. The BMW i3, with its carbon and aluminum body, exhibits distinctive features. The Mercedes Unimog, equipped with various rear and side add-on parts, presents a different detection distribution. The VW Crafter displays detections on the wheel arches at the front, setting it apart from others. Lastly, the VW Beetle has CCR-shaped side sills, leading to a unique pattern.[111] Consequently, it can be inferred that the position distribution of RCS detections is influenced by the vehicle's body shape and the yaw angle between the radar sensor and the OOI.

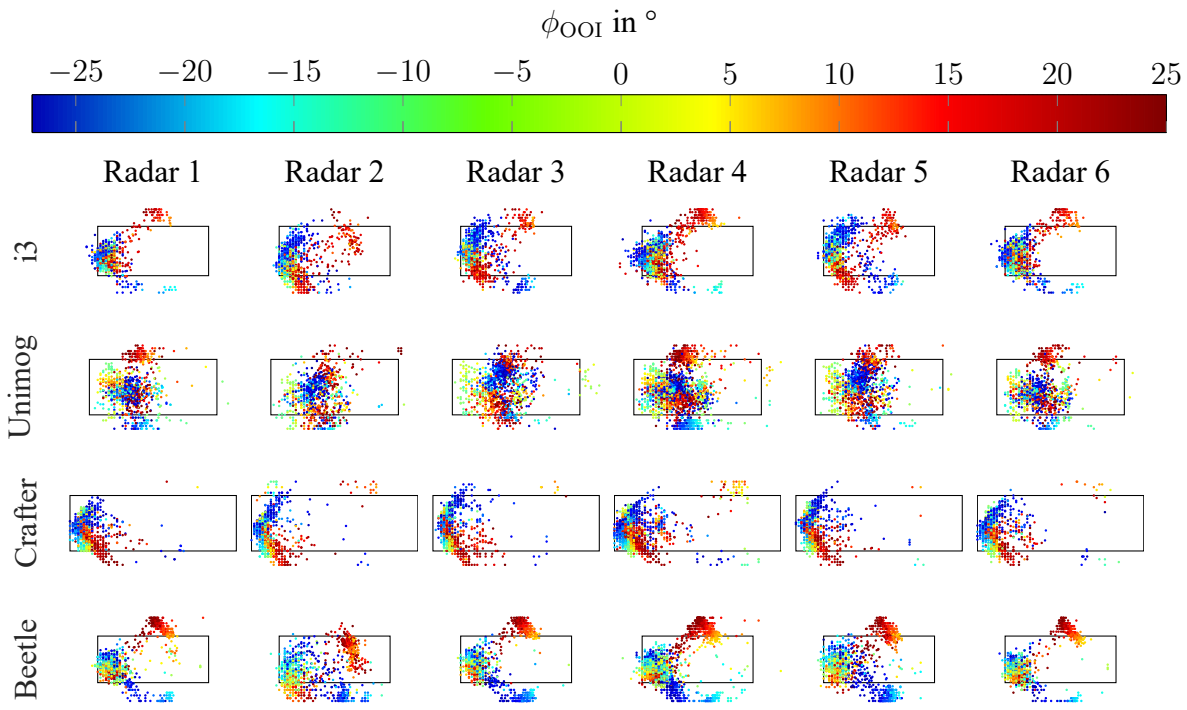


Figure 5-5: Location of detections with the highest RCS per measurement cycle around and inside the bounding box for specific vehicles and each radar sensor.[105f] © IEEE 2023.

In conclusion, the research results demonstrate the feasibility of effect isolation, albeit requiring extensive preparation and effort. The initial stage of the validation study necessitates the consideration and incorporation of potential effects. Overall, the stochastic properties of RCS characteristics exhibited by diverse vehicles present a significant challenge to the experimental setup.

109 Andres, M. et al.: 3D-scattering Center Detection of Automotive Targets Using Radar Sensors (2012).

110 Aust, P. et al.: Fingerprints of the Automotive Radar Scattering of Passenger Cars and Vans (2023).

111 In the readme of the evaluation code repository on GitLab pcitures of the corresponding vehicles are available.[103]

6 Data Acquisition and Analysis

Following the establishment of a foundational experimental setup designed to enable the isolation of effects, the experiment and simulation are subsequently conducted. This corresponds to the „Data acquisition and analysis“ stage of the validation methodology from Fig. 3-1. To accurately recreate the real measurement in the virtual simulation environment, precise transfer of reference data into the simulation is essential.[112a] On the one hand, for ray tracing-based simulation models, the 3D environment containing all objects and the associated discretization of geometries, along with their materials, are crucial factors influencing the simulation quality. On the other hand, recorded trajectories constitute a fundamental component of validation measurements for re-simulation of test drives. However, in addition to the aleatory uncertainties inherent in the measurement system itself, the transfer of references into the simulation environment introduces further uncertainties. Not only the accuracy of various data types but also the interpretation of the corresponding data affects the reference data transferred into the simulation. For example, considering a trajectory recorded with the ADMA, instead of directly implementing the measured pose data of the experiment, the transfer process into the simulation involves interpolation, extrapolation, or closed-loop control within the simulation tool. As a result, additional sources of uncertainties arise that impact the validation result. Consequently, the research question is derived from these considerations and is the subject of investigation outlined in **paper I**.[113a]

How are radar model validation campaigns affected by the accuracy of the reference sensor system?

6.1 Super-Reference for Trust in Reference Data

The preceding motivation leads to the necessity of verifying the accuracy of the reference sensors and, consequently, the quality of the reference sensor data. Furthermore, the reference sensor must exhibit higher accuracy than the sensor under test that requires validation.[112b] Additionally, in measurement campaigns and datasets, there is often reference to a GT which denotes the correct value of a measurement at a specific time. However, determining the GT with a measuring instrument is impossible, and it is referred to as the true value in the literature.[114] To identify the uncertainties in the reference data, a method is required to quantify and validate the respective deviations. This procedure is introduced as “super-referencing” and is defined as follows:

112 Rosenberger, P. et al.: Towards Generally Accepted Validation Methodology (2019), a: p. 10; b: pp. 7–8.

113 Holder, M. F. et al.: Digitalize the Twin (2022), a: -; b: p. 5; c: p. 2; d: p. 8; e: p. 10; f: p. 9; g: p. 12; h: p. 13.

114 DIN: DIN ISO 3534-2 Statistics - Vocabulary and Symbols (2013), p. 53.

"Comparing the result ξ obtained by device $\mathcal{A}$ to that of device $\mathcal{B}$. The underlying measurement principle of $\mathcal{B}$ is fundamentally different to $\mathcal{A}$, i.e., $\mathcal{B}$ is invariant to error sources of $\mathcal{A}$. Measuring ξ by means of $\mathcal{B}$ is characterized by high fidelity, accuracy, repeatability, and intuition. $\mathcal{B}$ is thereby seen as a super-reference for obtaining ξ".[113b]

Fig. 6-1 displays the introduced terminology along with the deviation of the respective data from the GT. The fundamental principle of this method is not limited to GNSS but can be generalized to other reference quantities, such as temperature or humidity.

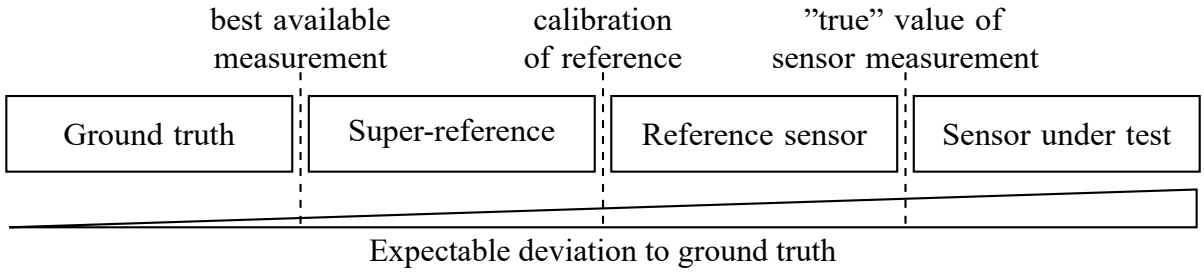


Figure 6-1: Relationship between GT, super-reference, reference sensor, and sensor under test.[113c]

Of particular significance in the transfer of trajectories with the aleatory uncertainties of the measurement system into the simulation is, in addition to the accuracy of velocity for radar sensors, the positional accuracy of the object during motion. To minimize the impact of uncertainties in reference data on the validation result, it is essential to reduce aleatory uncertainties as much as possible. Prior to conducting the validation experiment, a thorough analysis of the reference data employed in the simulation is necessary. Exemplifying this approach, lateral and longitudinal positions, as well as yaw angle, are examined under static conditions. Furthermore, the investigation aims to apply the super-referencing method to dynamic measurements of the absolute position of the ADMA and the relative position of two measurement systems. The selection of these measurement parameters is guided by the accuracy specifications of the ADMA and the conditions present on the proving ground. Subsequently, a more detailed examination is conducted on the absolute position of the reference system during motion, along with a focus on the relative positions in scenarios involving multiple vehicles.

6.2 Absolute Positioning in Dynamic Case

For the first of the two described investigations, the experimental setup from Fig. 6-2 is utilized. In this scenario, the ego vehicle equipped with the ADMA decelerates with various initial velocities (25 m/s, 20 m/s, 10 m/s) through three light barriers Lb_1, Lb_2, and Lb_3 and comes to a stop after passing the last light barrier. By knowing the moment of passing through the light barrier and the corresponding position, a comparison between the ADMA and the super-reference of the light barrier becomes feasible. In light blue the different local light barrier coordinate systems

are shown and in green the car coordinate system is visualized. To enable the differentiation of effects between high accelerations and constant-speed driving, the experiment is repeated with a cruise control drive at 30 km/h. While passing through, foam lines applied to the road surface before each trial are crossed. This allows the determination of not only the longitudinal position by triggering the light barrier but also the lateral position of the vehicle. The distance from the foam line to the corresponding light barrier corresponds to the longitudinal distance between the vehicle's front and the center of the front axle, mitigating the influence of vehicle yaw.

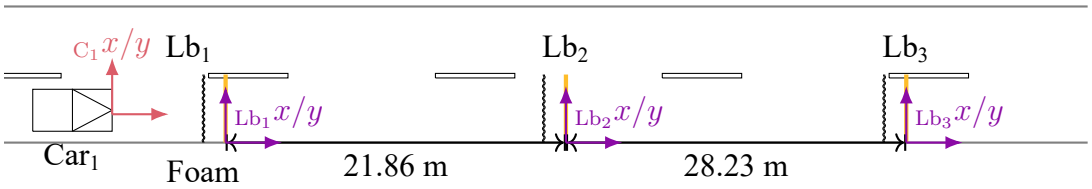


Figure 6-2: Measurement setup for super-referencing absolute positioning in the dynamic case.[113d]

To analyze the measurement results, Equ. 6-1 is employed for the longitudinal deviation ϵ_{long}, and Equ. 6-2 is used for the lateral deviation ϵ_{lat}. For this purpose, the longitudinal and lateral positions, denoted as x and y, of the reference sensor (Ref) are used with the positions of the super-reference (SRef), where $t_{\text{Lb}_{1\ldots3}}$ denotes the times at which the light barriers are crossed.

$$\epsilon_{\text{long}}(t_{\text{Lb}_{1\ldots3}}) = {}_{\text{C}_1}x_{\text{Ref}}(t_{\text{Lb}_{1\ldots3}}) - {}_{\text{Lb}_{1\ldots3}}x_{\text{SRef}}(t_{\text{Lb}_{1\ldots3}}) \tag{6-1}$$

$$\epsilon_{\text{lat}}(t_{\text{Lb}_{1\ldots3}}) = {}_{\text{C}_1}y_{\text{Ref}}(t_{\text{Lb}_{1\ldots3}}) - {}_{\text{Lb}_{1\ldots3}}y_{\text{SRef}}(t_{\text{Lb}_{1\ldots3}}) \tag{6-2}$$

The evaluation of these results yields Fig. 6-3. The lateral and longitudinal error for each light barrier depicted as different color is visualized. Additionally, the triangles denote the measurements with deceleration and the circles denote the measurements with constant velocity. Across different measurements and light barrier positions, a high level of precision is achieved in both longitudinal and lateral measurements. However, it is observed that the longitudinal error exhibits increasing deviation as the vehicle's speed becomes greater during its passage through the light barrier. As a result, it is recommended to use lower velocities to maintain an acceptable level of measurement accuracy. Alternatively, the use of devices with higher sampling frequencies is necessary. This phenomenon is attributed to measurement errors arising from the time delay associated with the operation of the light barrier. The impact of this delay is evident in the reduction of longitudinal deviation as the vehicle's speed decreases, as indicated by the triangular markers. Notably, at a speed of 25 m/s, the time delay introduces a worst-case error of 8 mm.

Furthermore, the remaining discrepancies observed in the results can be attributed to the errors inherent in the ADMA device itself, as well as the inaccuracies arising from the positioning of the experimental setup. The lateral error produced by the super-referencing method consistently

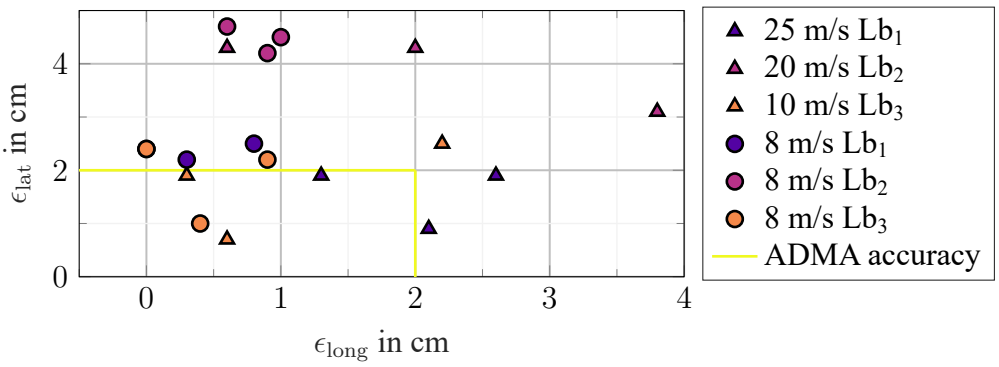


Figure 6-3: Lateral ϵ_{lat} over longitudinal error ϵ_{long} obtained by light barriers and foam.[113e]

demonstrates deviations exceeding the declared accuracy of the ADMA device, particularly at the second light barrier. These deviations are characterized by values of approximately 2.5 cm in excess of the proclaimed accuracy, thus suggesting the presence of an error inherent to the experimental setup. Hence, it is highly recommended to use precision instruments such as tachymeters for a more accurate absolute positioning of the light barriers and the marks within the foam.

Lastly, when examining the ADMA's performance under absolute dynamic conditions with low vehicle velocities, it is evident that the device consistently exhibits a positive error. This error varies between 0 cm and 3.8 cm in the longitudinal dimension and between 0.5 cm and 4.5 cm in the lateral dimension. These discrepancies can be attributed to the device's performance and highlight the need for careful consideration when utilizing a GNSS-based reference system.

6.3 Relative Positioning in Dynamic Case

In addition to the findings on absolute positional accuracy, a further investigation delves into the relative dynamic positional accuracy of two identical GNSS systems. For this purpose, the two vehicles are connected by a tow bar, and the positions of the tow bar's attachment points are used as reference points for evaluation. It is assumed that the clearance in the joint and stretch of the tow bar are smaller than the accuracy of the reference system under analysis. The driving maneuver follows the sequence 0 km/h $\rightarrow$ 30 km/h $\rightarrow$ maintaining $\rightarrow$ 60 km/h $\rightarrow$ maintaining $\rightarrow$ 30 km/h $\rightarrow$ maintaining $\rightarrow$ 0 km/h. This experimental setup is designed to include portions of constant-speed driving and accelerations. Furthermore, the experiment is repeated five times. Fig. 6-4 illustrates the corresponding experimental setup. The connection points of the tow bar are measured in the local Cartesian coordinate system L for Car_1 and Car_2.

For the evaluation of the measurement data, Equ. 6-3 is employed, with the length of the installed tow bar $l_{\mathrm{tb,SRef}}$ precisely determined with an accuracy of 0.5 mm as a super-reference.

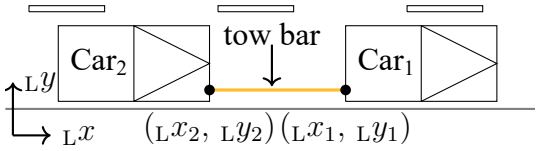


Figure 6-4: Measurement setup for dynamic dual super-reference with a tow bar in a local coordinate system L.[113f]

$$\epsilon_{\mathrm{tb}} = l_{\mathrm{tb,Ref}} - l_{\mathrm{tb,SRef}} = \left\| \begin{pmatrix} {}_{\mathrm{L}}x_2 - {}_{\mathrm{L}}x_1 \\ {}_{\mathrm{L}}y_2 - {}_{\mathrm{L}}y_1 \end{pmatrix} \right\|_2 - l_{\mathrm{tb,SRef}} \tag{6-3}$$

The deviation of the tow bar with respect to velocity $\dot{x}_{\mathrm{Ref}}$ and acceleration $\ddot{x}_{\mathrm{Ref}}$ is illustrated in Fig. 6-5. The temporal aspect of the velocity or acceleration profile is indicated by the color gradient from black to light brown, depicting all trials of the experiment. Analyzing the sensitivity of ϵ_{tb} to velocity reveals three consistent characteristics for all tests. The error in the tow bar fluctuates by 2 cm during quasi-stationary driving, aligning with the device specifications. When the vehicle is in the process of acceleration or deceleration, the measurement error remains relatively constant and falls within the acceptable range of variation. As the vehicle comes to a complete stop, the measurement error stabilizes at a specific value that remains within the permissible limits of the dual measurement setup.

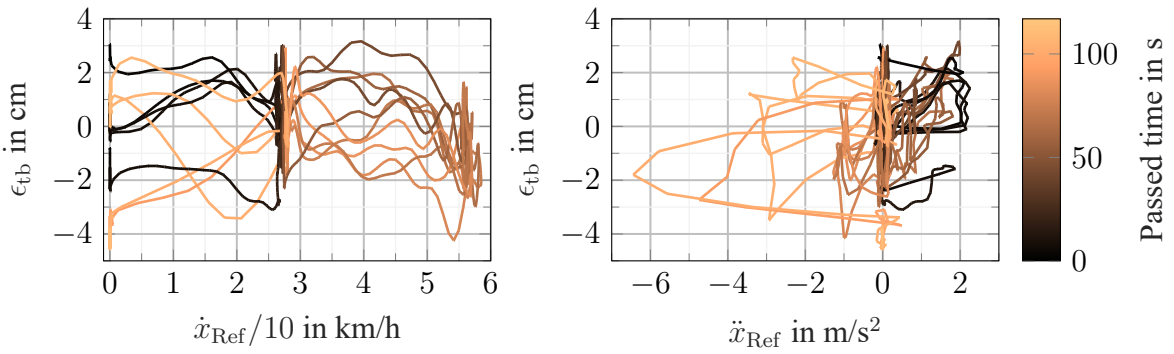


Figure 6-5: Tow bar deviation ϵ_{tb} over $\dot{x}_{\mathrm{Ref}}$ and $\ddot{x}_{\mathrm{Ref}}$. During areas of "constant" speed, the distance error settles within the accuracy of the measurement devices and the distance error only shows low sensitivity to acceleration. All trials are depicted and running time is denoted by the line's color gradient.[113g]

After confirming the accuracy of the ADMA, the impact of trajectory transfer into the simulation is analyzed. This is achieved by transferring the recorded trajectory into two commercial simulation tools using the OpenScenario[115] format.[116] Once again, the deviation from the originally measured length of the tow bar is utilized to evaluate the transfer into the simulation.

115 ASAM e.V.: ASAM OpenSCENARIO® - User Guide (2021).

116 Holder, M. F. et al.: Source Code Xosc-Converter (2021).

$$\epsilon_{\mathrm{Sim1/2}} = (l_{\mathrm{tb,Sim1/2}} - l_{\mathrm{tb,SRef}}) \tag{6-4}$$

Fig. 6-6 illustrates the deviations in both simulation tools. The experimental data reveals that variations in measurement accuracy differ across multiple test runs. These variations are notably influenced by the vehicle's speed, with lower speeds resulting in reduced measurement errors, and the highest errors occurring during acceleration or deceleration phases. The resulting disparities between the reference measurement data and the data from simulation tools stem from the software-based interpretation of the reference data. Specifically, the temporal interpolation or extrapolation of trajectory data and control algorithms prior to vehicle dynamics simulation are the causes of the divergent results. In the first simulation, during phases of near-stationary velocities, the measurement errors occasionally surpass the specified accuracy of the ADMA. In contrast, the second simulation consistently falls within the specified accuracy range. The findings indicate that the best performance in replicating test drives is achieved with the first simulation tool when the vehicle maintains relatively constant speeds and low accelerations below 2 m/s^2. This disparity complicates the comparison between simulated and measured data, as the reference data itself displays variations.

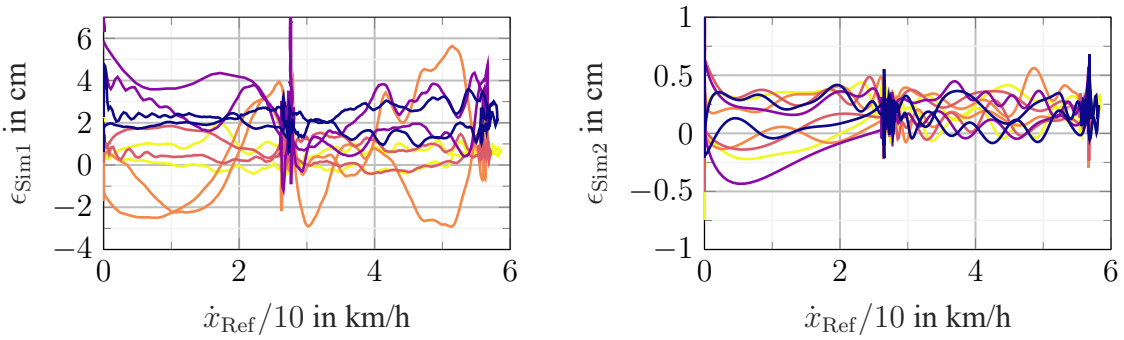


Figure 6-6: On the left side, ϵ_{Sim1} over $\dot{x}_{\mathrm{Ref}}$ is shown in case of the first simulation. On the right side, ϵ_{Sim2} over $\dot{x}_{\mathrm{Ref}}$ is shown within the second simulation. Each measurement trial is visualized with a different color.[113h]

7 Validation

After reducing, determining, and validating the uncertainties of the reference sensors used in the experiment following the validation process by Viehof and Rosenberger, the simulation is conducted. Subsequently, validation is performed based on the previously defined requirements and the isolated effects. Therefore, evaluating the validity of a radar sensor model is only possible by considering the uncertainties of both the reference sensors and the radar measurements themselves. This leads to the overarching research question of this chapter:

How can reference sensor and radar sensor measurement uncertainties be taken into account in a metric-based radar model validation across various interfaces?

The reference sensor data recorded during the measurements, along with their uncertainties, are propagated through the simulation for the analysis of the research question. Finally, the measurement and simulation data is compared using the DVM, enabling a validity assessment based on bias and scattering deviation. This chapter is based on the content of **paper IV**.[117a]

7.1 Measurement Setup and Uncertainties

To address the research question, a simple validation scenario is conducted with a CCR to analyze the metric-based radar model validation. The measurement setup is illustrated in Fig. 7-1. In addition to the schematic representation on the left, a photo of the setup is shown on the right. CCR Position 1 (Pos 1) is used for calibrating the simulated power, while Position 3 (Pos 3) is employed for validation. The two positions differ in azimuth angle ${}_{\mathrm{S}}\phi$ in sensor coordinates while maintaining a constant radial distance ${}_{\mathrm{S}}r$. Five repetitions are done for each position to account for effects of successive measurements in the validation. Each measurement of the static experiment lasts 60 s resulting in approximately 224 timeframes based on a sample frequency of 14 Hz.

The measurement setup induces the uncertainties listed in Tab. 7-1 for the reference data. All mentioned reference data are determined using RTK-based GNSS, from which both the measured quantity and uncertainties arise. Only the sensor height and the edge length of the CCR are determined using a tape measure. Theoretically, additional influencing factors such as material properties and environmental conditions like temperature or random road surface structures should be considered. However, in the conducted measurement campaign, either no suitable reference sensors are available resulting in an increased aleatory uncertainty or the parameters are invariable in the used simulation tool.

117 Elster, L. et al.: Introducing the DVM for Radar Sensor Models (2024), a: -; b: p. 6; c: p. 19; d: p. 21.

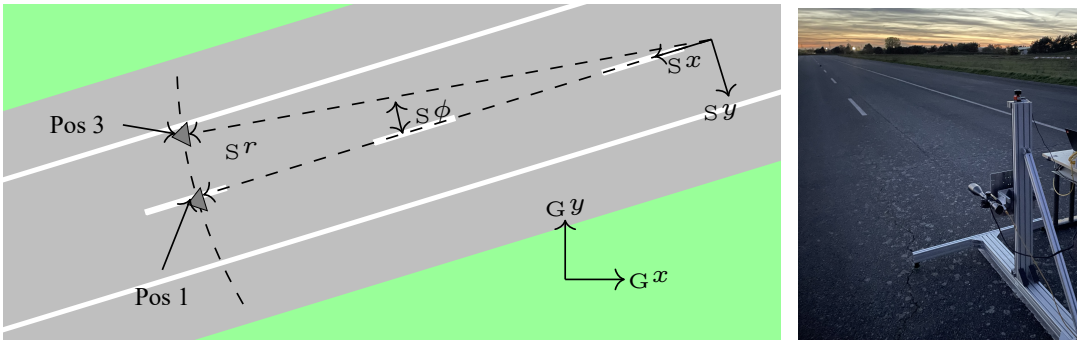


Figure 7-1: Experimental setup of the validation study. CCR position 1 ($r_{\mathrm{CCR,Pos1}} = 29.56\,\mathrm{m}$, $\phi_{\mathrm{CCR,Pos1}} = 0$) is used for simulation calibration purposes and position 3 ($r_{\mathrm{CCR,Pos3}} = 29.56\,\mathrm{m}$, $\phi_{\mathrm{CCR,Pos3}} = -8°$) is analyzed based on the presented methodology.[117b]

Table 7-1: Measured reference data uncertainties for CCR position 3 defined in Fig. 7-1. The local cartesian coordinate system G is defined in East-North-Up direction with the origin located on the August-Euler airfield in Griesheim.[117b]

Reference data	Variable	Measurement	Uncertainty
Sensor azimuth orientation in G	${}_{\mathrm{G}}\phi_{\mathrm{S}}$	197.91°	±0.07°
Sensor x position in G	${}_{\mathrm{G}}x_{\mathrm{S}}$	977.43 m	±0.02 m
Sensor y position in G	${}_{\mathrm{G}}y_{\mathrm{S}}$	241.56 m	±0.02 m
Sensor height in G	${}_{\mathrm{G}}h_{\mathrm{S}}$	12.89 m	±0.02 m
CCR x position in G	${}_{\mathrm{G}}x_{\mathrm{C}}$	948.33 m	±0.02 m
CCR y position in G	${}_{\mathrm{G}}y_{\mathrm{C}}$	216.46 m	±0.02 m
Edge length CCR	l_{C}	0.240 m	±0.005 m

7.2 Evaluation of Radar Measurement

The uncertainties defined in the previous chapter are subsequently propagated through simulation by varying both the measured value and the positive and negative uncertainties individually. A full factorial combination of uncertainties is avoided at this point due to the exploding parameter space. With seven parameters, the measured value, and two boundaries, there are 2187 possible combinations. Consequently, 15 simulation results are obtained by simulating the upper and lower boundary of each parameter and all measured values, with each simulation also conducted for 60 s. The radar model described in Chap. 2.3 in combination with the simulation tool CarMaker, is employed for the simulations.

The model is extended with a stochastic noise simulation based on additional measurements, in which the radar sensor is oriented towards the sky without any object. The resulting data from the real and virtual worlds are evaluated following the methodology outlined in Chap. 4.1 of **paper III**. In addition to evaluating the entire radar cuboid, all detections, and clustered detections detailed in **paper IV**, the evaluation of a region of interest around the CCR and each individual cell of the radar cuboid is performed. The upper limit of the propagated uncertainties is denoted as $^{+}$, and the lower limit is denoted as $^{-}$.

In Fig. 7-2, the diagrams depict the EDFs of the simulation (- -) and the measurement data (—). The left diagram illustrates all EDFs for the 15 simulations and the five measurements. The right diagram visualizes the 15 EDFs shifted based on d_{bias} relative to the first measurement, Meas 1. On the radar cuboid interface the radar data is filtered by the measured reference position of the CCR. The observed step-shaped pattern in the measurements arises from the analysis of the four filtered distinct range azimuth cells. These four cells are a consequence of the placement of the CCR on the edge of the bins in the range and azimuth dimensions. The temporal fluctuations within each cell exhibit a modest range of decibels, as discerned from the slopes evident in the EDFs. Additionally, there is notable consistency in the reproducibility of the measurements, manifesting in the substantial overlap.

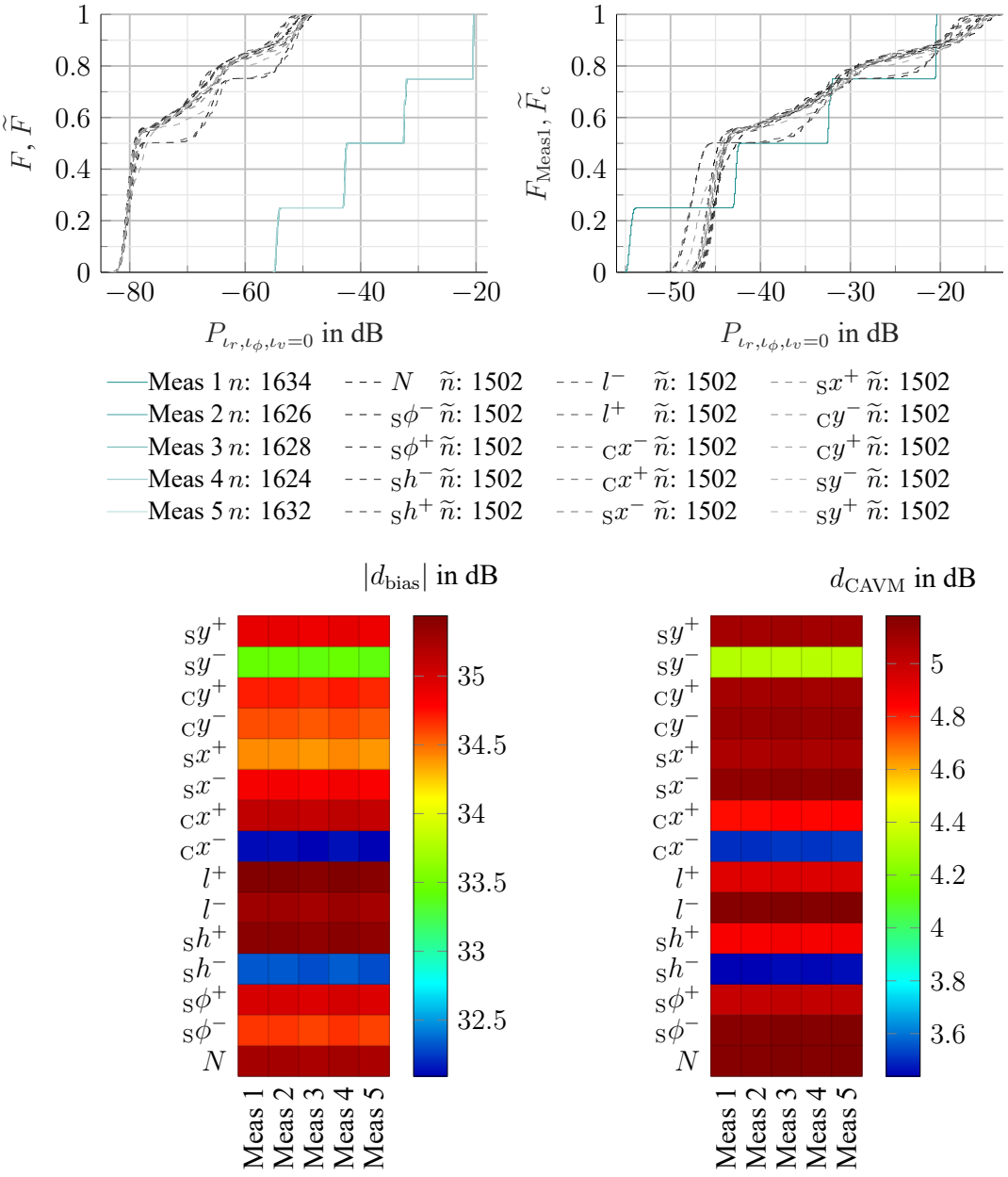


Figure 7-2: The uncorrected and corrected EDFs as well as the deviation based on the different simulations and measurements coded in colors for $|d_{\text{bias}}|$ and d_{CAVM}.[117c]

It is noteworthy that the simulations manifest a discernible model bias, primarily attributable to initial modeling inaccuracies in the signal processing and calibration. Thus, the presented experiment is suitable for substantiating the foundational maturity of a model. The model is specifically calibrated to the RCS of the CCR at position 1, resulting in disparities in RCS calculations between the radar cuboid signal strength and the detections. Moreover, the simulations are susceptible to uncertainties. The variation in length and positive height, in particular, amplifies the discrepancy between the model and the simulation. While certain simulations approximate the step-shaped pattern observed in the measurements, disparities are apparent in the slope characteristics. In particular, the initial slope in several simulations is less pronounced, and in some instances, notably smeared. This is due to a different number of simulated rays hitting the CCR in the simulation in different timeframes. The simulation tool employed utilizes sampling methods for the directional initialization of rays, resulting in a variable number of rays hitting the object depending on the time step. The effects in the simulation model on the power become more pronounced with increasing distances between the sensor origin and the object, unless appropriate algorithmic countermeasures are implemented in the ray sampling process.

The initiation of the simulation slopes is attributed to the noise simulation, where two radar cuboid cells unaffected by the CCR-induced signal strength increase are considered. This arises from the fact that the noise simulation has a mean value at the CCR position of approximately −80 dB. Deviations arise from differences in window functions between the actual sensor and the simulation model. Additionally, the discrepancy in the number of cells visualized in the table below the diagrams is a result of a simulation sampling distinction between the model and the real sensor. The previously expounded observations derived from the EDFs find corroboration in the DVM heat map visualization. Notably, simulations of ${}_{\mathrm{S}}y^{-}$, ${}_{\mathrm{C}}x^{-}$, as well as ${}_{\mathrm{S}}h^{-}$ exhibit minimal deviations in both model bias and scattering error.

In the upper image of Fig. 7-3, a discernible discrepancy of approximately 30 dB is observable in the vicinity of the CCR's location (Box 1). On one hand, the CCR's power is excessively low, a deficiency addressable through model calibration at the radar cuboid level. On the other hand, the iterative determination of the window function in the model introduces the possibility of measurement and modeling errors. The maximum deviation of 55 dB between the simulation model and actual measurement is present in the range of the CCR location but in a different azimuth angle (Box 2). Notably, the simulation model lacks input data from the ray tracing algorithm at this location, and these radar cuboid cells are solely populated by noise simulation. Despite the absence of discernible objects or asphalt irregularities during measurements justifying the observed signal strength increase, it is inferred that an unaccounted-effect in the radar sensor's signal processing, triggered by the CCR, exists. Consequently, the method is adept at identifying a systematic model error. Along the transition between runway and vegetation, persistent model bias increases ranging from 15 dB to 25 dB are evident (Box 3). In the second satellite plot, the scattering error denoted by d_{CAVM} is visually represented (Box 4). The CCR exhibits the highest deviation, and the measured distribution assumes a step

function appearance. In contrast, the simulation EDF exhibit less steep profiles, with pronounced impacts from propagated uncertainties, particularly at lower power levels. Substantial deviations are apparent along the runway and the transition to vegetation, mirroring observations in $|d_{\mathrm{bias}}|$. Of specific interest is the intersection of the taxiway with the runway (Box 5). The transition from vegetation to asphalt yields elevated values in d_{CAVM} compared to surrounding cells, as microscopic movements of the vegetation influence the shape distribution of the measurement data.

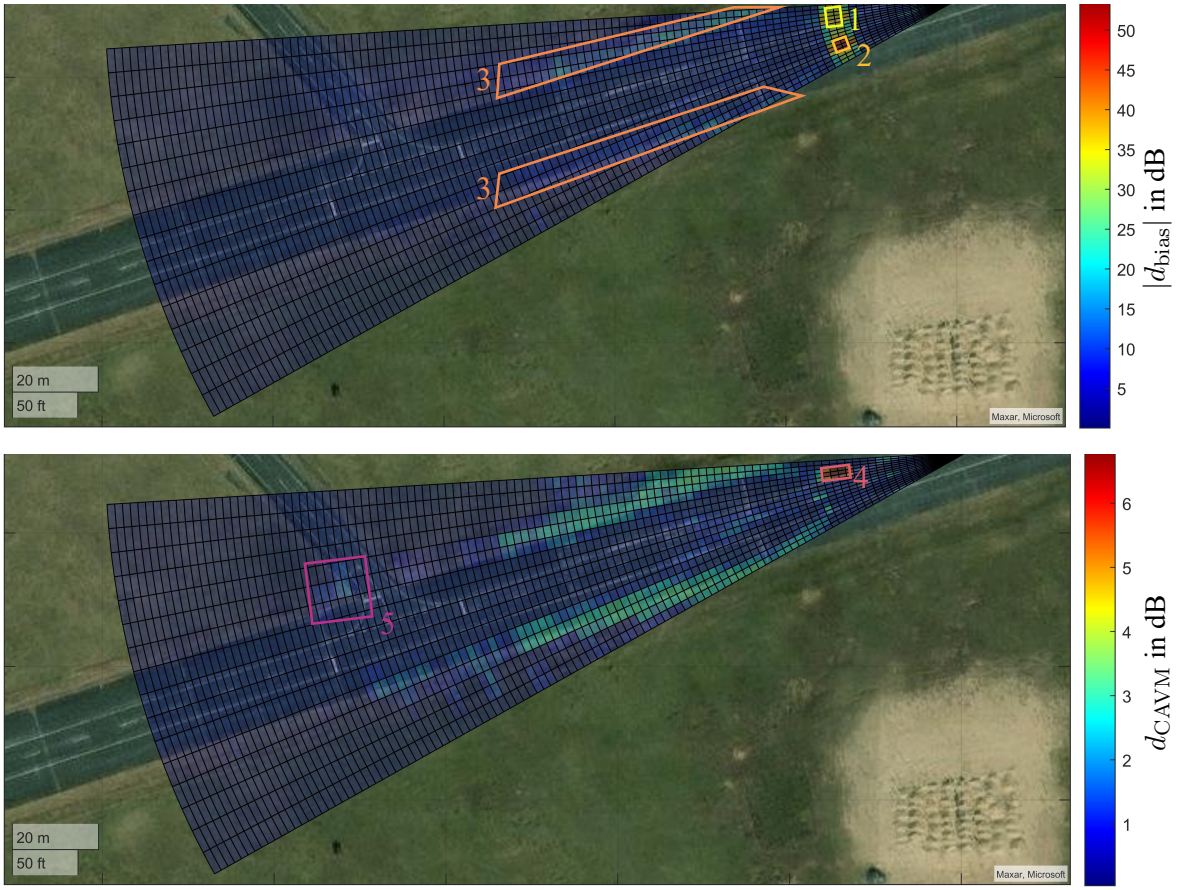


Figure 7-3: Satellite image in which the results of the DVM are shown. From top to bottom, the validation results for $|d_{\mathrm{bias}}|$ and d_{CAVM} are visualized.[117d]

8 Conclusion and Outlook

In this chapter, the findings regarding the posed RQs are summarized and recontextualized within the overall framework related to the main RQ. The guiding question of this thesis is:

How can effects and uncertainties in radar validation measurements, along with their associated impacts on radar sensor model validation, be identified and quantified?

Below, the four RQs are reiterated. Therefore, a conclusion of the RQ is provided. Afterwards the contribution to the main RQ is elaborated. Finally, an outlook on future research directions related to the validation of radar sensor models based on the insights gained is given.

RQ 1: *How can radar measurements be used to establish quantifiable requirements for a radar sensor model?*

Conclusion: To address the RQ, various measurement setups and conditions are compared. The distribution of measurement data on the radar cuboid and the detection interface is analyzed using the DVM. Different scenarios demonstrate the effects of various factors and the resulting measurement uncertainties in radar sensors. It is evident that position filtering for detections is necessary, as the number of detections between measurements fluctuates significantly. In addition to environmental influences, the complex wave propagation, particularly the multipath propagation, is responsible for this phenomenon. The detailed analysis of each azimuth range cell of the radar cuboid allows insights into the causality between measurement fluctuations and the environment, providing a new form of interpretability and comparability of measurement results. This is due to in the less processed data of the radar cuboid, leading to a higher informational content compared to the detection interface. Furthermore, repeatability and reproducibility of measurements is accessible. Since the measured uncertainties represent the maximum simulation fidelity achievable through simulation, it becomes possible to quantify the impact of various effects such as rain. Consequently, this allows the derivation of fundamental requirements for effects on radar signal processing interfaces from simple scenarios. In addition to quantifying requirements, the methodology presented facilitates the comparison of various measurement setups. This includes the evaluation of the effects of different test grounds, laboratory facilities, and environmental conditions.

Contribution to main RQ: Thus, in the context of the main RQ, it is evident that validation measurements in radar sensors have to be simple to demonstrate and validate individual effects in the sensor model. Furthermore, utilizing reference sensors for environmental conditions is crucial to quantify external factors and determine their influence. In conclusion,

the methodology presented allows for comparability of measurements across locations in radar model validation campaigns.

Outlook: This aspect plays a crucial role, especially in the potential future standardization of validation experiments, ensuring comparability. Additionally, it is essential to explore the feasibility of extending the procedure to dynamic measurements. On the detection interface, filtering based on the Doppler effect simplifies position filtering. However, the relative velocities of the wheels at the wheel contact point complicates the filtering, as the relative speed corresponds to the static environment. Therefore, a preliminary examination of this effect on radar measurements is necessary. Furthermore, the described methodology is well-suited for comparing various sensors of the same type. Consequently, it allows the identification of uncertainties stemming from the production process and their impacts on radar measurements such as the antennas.

RQ 2: *How is it possible to isolate effects related to the RCS of road vehicles and make them measurable under real world conditions?*

Conclusion: The effects on the RCS obtained from literature are investigated through a slalom setup to examine the reflectivity properties of different vehicles. Additionally, six identical radar sensors are positioned one above the other to analyze the mounting height of the sensors. A frequency analysis, considering the periodic RCS characteristics, provides amplitude and a constant component for all vehicles and radar sensors. The results indicate slight variations for all sensors between measurements, but significant differences exist between various vehicles. The lowest radar sensor exhibits the largest fluctuations in RCS, attributed to multipath propagation of radar waves. Additionally, due to the low mounting position, a greater portion of the vehicle underbody is visible, serving as another influencing factor on the fluctuations. The EDF of the vehicles confirm the findings from the frequency analysis. The uniqueness of the vehicles is ultimately confirmed through the analysis of detections relative to the position in the bounding box of the respective vehicle.

Contribution to main RQ: The demonstrated process of inferring effects on the experimental setup allows for separation and identification at the sensor output. This is demonstrated by the mounting position, and thus the multipath propagation, the aspect angle, and the various vehicle models. High reproducibility is achievable even by a human using landmarks in dynamic validation scenarios. In general, initial measurements indicate that the validation effort can be reduced by investigating the influence of specific parameters in advance, as demonstrated here, for example, with the mounting height beyond a certain threshold.

Outlook: However, the described approach lacks a methodology that systematically derives experiments from effect-based requirements. The validation experiments still rely on expert knowledge and a standardized approach is missing. Furthermore, a generalization of object models is necessary to ensure that the complexity for the virtual validation approach does

not exceed its benefits. The presented results are only valid for the specific radar sensor used. Increased range and azimuth resolutions and the elevation component can influence the presented results. Therefore, demonstrating the generalizability of the trends with another radar sensor is essential.

RQ 3: *How are radar model validation campaigns affected by the accuracy of the reference sensor system?*

Conclusion: The reference data for the re-simulation of RTK-based GNSS data is qualified through the principle of the super-reference. Methods for absolute and relative positioning of individual and multiple systems in both static and dynamic cases are presented. The evaluation of the experiments confirms the specified accuracies but also indicates the need for further optimizations for the experimental setup with the light barriers. The transfer of reference data into the simulation introduces additional epistemic uncertainties.

Contribution to main RQ: The developed methods serve to identify the aleatory uncertainties in the reference sensor data. This allows for the consideration of these uncertainties in the re-simulation by propagating them through the simulation. The experiments enhance confidence in the measurement results of the reference and demonstrate the previously specified requirements for the reference sensors.

Outlook: Furthermore, it is necessary to validate the reference data for velocity and rotational speed. This allows for the avoidance of simple errors such as incorrectly determined positions of the ADMA in the vehicle, which directly impact measurement results, as well as the absence of timestamps. Additionally, experiments are necessary in areas that do not represent ideal measurement conditions for the reference system. In the case of the sensor used, these areas include urban canyons or forests, where accuracy is reduced due to multipath effects and occlusion. Moreover, simulation environments must ensure precise repeatability of recorded trajectories. Otherwise, they are unsuitable for radar sensor model validation. Besides GNSS systems, there are emerging methods for determining object positions, such as lidar or sensor fusion. New experimental designs for the super-reference methodology are required in these cases.

RQ 4: *How can reference sensor and radar sensor measurement uncertainties be taken into account in a metric-based radar model validation across various interfaces?*

Conclusion: Using a simple validation scenario with a CCR as an object, the application of the DVM on radar data of the radar cuboid and detection interface is conducted. An adapted open-source radar model in combination with IPG CarMaker serves as the source of synthetic data. Aleatory uncertainties from radar measurements are incorporated by comparing multiple measurements taken in quick succession with simulation data. In turn, in the simulation, epistemic uncertainties from reference data are propagated through the simulation, enabling a comprehensive evaluation of the validation experiment. The

DVM, in conjunction with various evaluation methods, ensures a high interpretability of the results.

Contribution to main RQ: The presented methodology quantifies the deviations between the uncertainty-affected validation measurements and the radar simulation model. Due to the characteristic of the DVM being invariant regarding input parameters, the methodology is also applicable to the object list and other experimental setups.

Outlook: The conducted analysis is confined to examining the maximum and minimum epistemic uncertainty of the reference sensor. Currently, the combination of uncertainties is neglected. Therefore, it is recommended to reduce the parameter space of reference sensor uncertainties through pre-analysis sensitivity studies. Additionally, material properties and the shape of objects need to be incorporated into the process along with position data. Furthermore, due to the high number of necessary experiments for statistical validation, there is a need for the automation of measurement campaigns. Otherwise, validating radar models using the proposed method exceeds the affordable effort. In essence, the described approach could serve as the foundation for standardizing the evaluation process for active perception sensor models and complementing existing standards in that regard.[118]

118 DIN: DIN SAE SPEC 91471 Assessment Methodology Automotive LiDAR (2023).

A Paper I:

Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation

energies

MDPI

Article

Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation

Martin Holder [1,†,‡], Lukas Elster [2,‡] and Hermann Winner [2,*]

[1] Independent Researcher, 64287 Darmstadt, Germany; Martin.Holder@wihi.tu-darmstadt.de
[2] Institute of Automotive Engineering, Technical University of Darmstadt, 64287 Darmstadt, Germany; lukas.elster@tu-darmstadt.de
* Correspondence: hermann.winner@tu-darmstadt.de
† Martin Holder was formerly with TU Darmstadt.
‡ These authors contributed equally to this work.

Abstract: In the course of the development of automated driving, there has been increasing interest in obtaining ground truth information from sensor recordings and transferring road traffic scenarios to simulations. The quality of the "ground truth" annotation is dictated by its accuracy. This paper presents a method for calibrating the accuracy of ground truth in practical applications in the automotive context. With an exemplary measurement device, we show that the proclaimed accuracy of the device is not always reached. However, test repetitions show deviations, resulting in non-uniform reliability and limited trustworthiness of the reference measurement. A similar result can be observed when reproducing the trajectory in the simulation environment: the exact reproduction of the driven trajectory does not always succeed in the simulation environment shown as an example because deviations occur. This is particularly relevant for making sensor-specific features such as material reflectivities for lidar and radar quantifiable in dynamic cases.

Keywords: virtual validation; automated driving; ground truth; reference measurement; calibration method; simulation

Citation: Holder, M.F.; Elster, L.; Winner, H. Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation. *Energies* **2022**, *15*, 989. https://doi.org/10.3390/en15030989

Academic Editor: Arno Eichberger

Received: 29 November 2021
Accepted: 21 January 2022
Published: 28 January 2022

1. Introduction

"Ground truth (GT) data was obtained using an real time kinematic (RTK)-based global navigation satellite system (GNSS) device and provides accuracy of up to ±3 cm". Such statements are often found in research articles to justify the quality of reference data accompanying data acquisition for various tasks [1,2]. In the automotive context, RTK-aided GNSS is widely used for obtaining positions. There is no doubt that RTK-based GNSS methods can achieve accuracies in the cm range. However, this applies only to the position determination of the antenna and under favorable operating conditions of the GNSS receiver. If one is, however, interested in the position information of another reference point, e.g., the center of the vehicle's rear axle, the translational offsets between the antenna and the respective point must be determined very precisely. In complex geometries such as vehicles, further aids are needed for this. Uncertainties in the determination of these offsets can be hardly avoided. For this reason, it is unclear whether the specified precision of the device can also be achieved in its installed state.

In this work, we address the issue of the trustworthiness of reference data obtained with GNSS devices. We aim to refine the notion of GT in the context of environmental perception with different sensor modalities. It must be ensured that the reference measurement shows higher credibility against other sensors used, e.g., lidar or radar sensors. To determine this, reference measurements are required to determine the credibility of the reference, called the "super-reference". Figure 1 contextualizes the aforementioned term "super-reference" in comparison to GT and a reference sensor.

The GT can only be measured with finite accuracy, which we call the super-reference. Thereby, GT is only approximated by the super-reference, leaving a minor deviation to GT. A super-reference is typically only available in limited and controllable circumstances. A reference sensor, however, is optimized for the practical application at the cost of potentially higher GT deviation. To achieve higher trustworthiness in the accuracy of the reference, a super-reference is used for its calibration. Ultimately, when validating sensors, it is of interest to determine measurement uncertainties, which result from the difference between sensor and reference measurement.

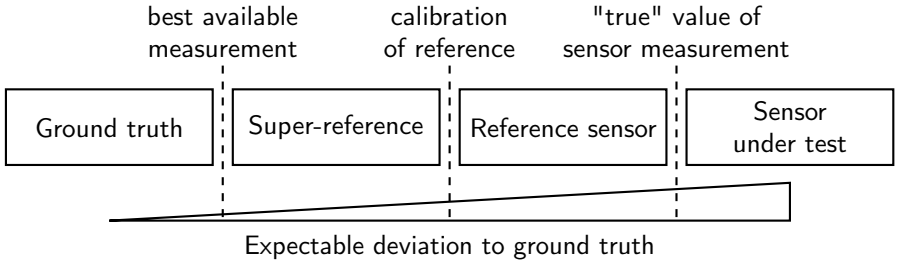


Figure 1. Relationship between GT, super-reference, reference sensor, and sensor under test.

The main interest of this article lies in increasing the trustworthiness in reference data, which enables the reenacting of real-world test drives in virtual environments. This is of particular importance in the development and validation of sensor models for the virtual validation of automated driving (AD), as reference data are required. The basic idea of transferring test drives to simulation is admittedly not new. However, our paper specifically deals with the calibration process of positioning measuring devices and discusses the achievable accuracy.

This paper is structured as follows. First, we discuss the need for the careful calibration of measurement devices that are used for collecting reference data. Next, an overview of previous research on obtaining GT data and sensor principles employed for this purpose is presented. We present stationary and dynamic calibration experiments, which serve as a reference and are thereby eligible for the calibration of measurement devices. In the practical application of our experiments, we show that the proclaimed accuracy of the positioning devices is not always met. Finally, we show the achievable precision when reenacting a real-world test drive in two simulation environments. The source code for creating scenarios with real driven trajectories based on GNSS measurements is made available.

2. A Motivational Example: Can We Trust Our Reference?

After data collection with sensors in real-world scenarios, faithful reenacting of the driven scenario in simulation is tedious, but of high interest for virtual validation aspects. There are various types of measurement phenomena that are inherent in the sensor measurement principle and can manifest as measurement artefacts. These can cause deviations between the obtained measurement result and GT. A simple yet illustrative example is the limited resolution of the (discrete) distance measurement with radar and lidar sensors, which causes quantization errors in the determination of the (continuous) distance to an object. For this reason, reference sensors are needed that are capable of measuring the movement and position of vehicles with high accuracy, precision, and reliability.

Even for simple scenarios, such as a follow-up drive with an Adaptive Cruise Control (ACC) system, one can observe non-stationary behavior when inspecting the movements of the vehicles in close detail, although the vehicle movements were subjectively perceived by the occupants as stationary. If the movement of the vehicle in front is now recorded by a sensor, further sensor-specific uncertainties are superimposed on its perception.

Figure 2 shows an example of a measurement record of an ACC drive run at 40 km/h with a medium time gap to the front vehicle. The measured variables used for the ACC

function are read out from the radar sensor, which supplies the object information for the ACC system, via the vehicle Controller Area Network (CAN).

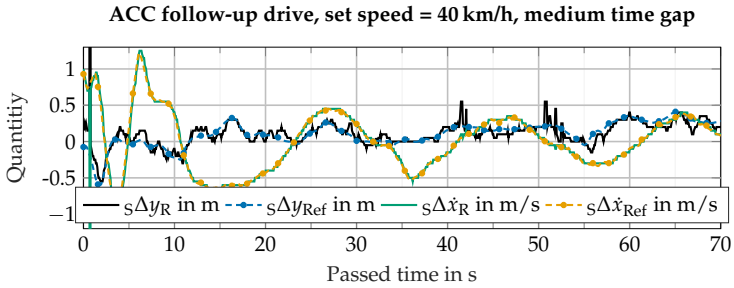


Figure 2. Variation of lateral distance and longitudinal speed between ego car and object of interest (OOI) in an ACC scenario. The left index "S" denotes the sensor coordinate system, "R" the radar sensor and "Ref" the GNSS-based reference system.

Several aspects emerge from the measurement record shown in the figure: the relative longitudinal velocity $_S\Delta\dot{x}$ shows a variation bandwidth of ± 0.5 m/s, which is around one order of magnitude above the velocity resolution of automotive radar sensors. In the lateral direction, i.e., $_S\Delta y$, the object fluctuates within its lane at around ± 0.2 m, which can hardly be noticed with visual inspection by a human driver. The object reported by the used radar sensor shows the pronounced discretization of the lateral measurement. The direct transfer of the radars' measured variables into the simulation would lead to sudden, physically implausible jumps in an object's trajectory. Although the radial velocity of objects is measured by the radar with high precision, the discretization by the sensor used in this example makes post-processing necessary in order to obtain a feasible motion profile.

There is reasonable hope that high-accuracy motion analyzers, which combine GNSS position measurements as well as accelerations and angular rates captured by inertial measurement unit (IMU) sensors, can be used to capture the motion of agents with high precision. An exemplary device is the GeneSys Automotive Dynamic Motion Analyzer (ADMA) or the RT device series by OXTS. The corresponding measurements of such a device are also shown in Figure 2 and denoted by "Ref". However, the question remains as to the actual accuracy of measuring motion and transferring the motion to the simulation. The central question of this paper is therefore as follows: How much accuracy does a reference measurement system really provide and how does one perform its calibration?

3. Related Work

Real-world traffic is a suitable data source for developing and testing automated driving functions because it is highly diverse and has random characteristics. Moreover, regarding simulation aspects, real-world data offer the highest possible quality for the validation of simulation models. Consequently, there are several previously reported approaches to transfer a real-world test drive into the simulation.

Roughly speaking, two categories can be found. These are, on the one hand, object list-based approaches. Here, the object list from sensors or a fused sensor cluster is taken as the starting point for scenario reconstruction. The goal of this method is to prepare real data for a scenario-based testing approach in the simulation. Regarding the study of absolute accuracy, previous work in the field of reference sensing deals with obtaining the accuracy that is achievable with contemporary automotive grade perception sensors.

3.1. Object List-Based Approach

In the literature, there are approaches known in which the object lists of the sensors are used to transfer the recorded scenario into a simulation. These methods aim to extract a concrete scenario from the measurement data in the sense of scenario-based testing. Logical scenarios can be abstracted from this. A typical pipeline consumes sensor data (e.g., object lists, point clouds, etc.) and compiles a standardized scenario description using the

OpenScenario/OpenDrive language format. A typical example is the framework proposed by Wagner et al. that relies on lidar sensors [3]. It is capable of sensing road objects as well as road semantics (e.g., road geometry, lane markings, etc.). There are also a number of commercial suppliers in this field that compile scenario data from sensor readings, such as [4,5].

These methods require a comprehensive object list as well as additional sensor data to facilitate the inference of the road properties or street layout. This prerequisite is not fulfilled in many everyday traffic situations, such as the occurrence of occlusions, or the cutting in/out of objects. Object detection is generally not reliable in such moments. This constraints can be gradually resolved by manual annotations of sensor data. There are no uniform quality standards, to the best of our knowledge, for the required accuracy of such methods. Based on the published information about these procedures, there is an impression that a visual inspection is performed by experts or the test engineers.

Services that provide so-called GT information for the annotation of sensor data (lidar point cloud, camera images), such as [6,7], are not within the scope of this paper because no reference data are used for this purpose. Instead, recorded sensor data, which may be calibrated extrinsically to each other when multiple sensor modalities are used, are annotated in a manual or (semi-)automated fashion.

3.2. Reference Sensors

With the help of reference sensors (e.g., high-precision GNSS measurement technology or laser scanners), the position of traffic participants can be obtained within the respective measurement accuracy. Data sets such as KITTI, nuScenes, WaymoOpen, etc., therefore provide GT information of traffic participants obtained from an automotive-grade laser scanner mounted on the roof of the ego vehicle. This approach provides useful results for annotating bounding boxes such as those used for labeling in machine learning methods. Minor inaccuracies in the labeling, so-called label noise, can even increase the robustness of the learning algorithm under certain circumstances. In order to use bounding boxes that are labeled in this way as a reference when transferring the scenario to the simulation, a specification of the accuracy over several time steps is required. This is not given in most data sets. The suitability of automotive-grade lidar sensors was investigated in a paper by Schalling et al. [8]. However, the limitations of lidar sensors with respect to the factors influencing their measurement result prevent their justification as GT sensors.

Thorough research on referencing the reference system ("super-referencing") has been presented by Brahmi [9]. His focus is on the evaluation of object-based advanced driver assistance system (ADAS) systems. The basic ideas presented in his thesis can essentially be applied to the problem of this paper, namely the transfer of a real test drive to a simulation.

In a paper by Steinhard, the suitability of a lidar sensor system for GT determination is investigated [10]. As with Brahmi, a high-precision laser scanner with sub-mm resolution serves as a super-reference.

3.3. Gaps in State of the Art

The determination of GT is mostly done via RTK-based GNSS or high-precision lidar sensors with mm-scale resolution e.g., Leica D5. In this context, however, there is no verification that the proclaimed accuracy is actually met under all circumstances. Previous experiments, such as the work from Brahmi [9], have indeed identified the need for a calibration procedure with reference sensors. What remains unresolved so far is to study the fidelity of "GT" in dynamic cases, as well as the stationary analysis of the yaw angle between two reference systems, which is of the utmost interest in reflectivity studies and signal drift.

The digitalizing of a test run relies on the position accuracy of the RTK-based GNSS device. However, it lacks the discussion of whether the proclaimed accuracy is maintained during dynamic situations. Modern lidar and high-resolution radar sensors have distance

resolutions in the cm range. If sensor models are to be validated, high demands are therefore made on the accuracy of the trajectory reproduction in the simulation.

4. Calibration Aspects: The Need for a Super-Reference

At this point, a discussion of the term "GT" in the context of automotive simulation is needed to obtain a common understanding of it. It is often used to describe the true state of an object, and potentially also the future state, e.g., in terms of planned actions. Thus, there is a state that can be estimated or measured. Its true value is called GT. It is initially irrelevant how GT is determined. The only relevant aspect is that the GT value serves as a reference against other methods for determining a certain value (measurement, estimation). Especially in the field of virtual environments, which consider 3D representations of objects, the term can be used in a broader sense: it covers material assignments, reflectively properties, as well as geometry detailing, and others.

When a "GT" is obtained with a prospective device, the resulting deviations can be conceptualized in terms of "accuracy" and "precision". The term "accuracy" is defined as "*the degree to which the result of a measurement or calculation matches the correct value or a standard*" [11]. Moreover, the term "precision" is defined as "*the quality of being exact, accurate and careful*" [12].

GT can hardly claim to be completely accurate. It represents rather a value that can be faithfully measured to the best of one's knowledge and belief, as well as up to the accuracy of the measurement equipment used. Prominent examples are object states, such as its longitudinal and lateral positions, as well as the object's orientation. Measurement errors of all kinds, as they are present in all measuring instruments, mean that GT can basically only be obtained with finite accuracy. Nevertheless, the measurement data obtained using the highest-precision device are considered to be a GT measurement. Consequently, a GT to the "GT" is needed. Thus, for verification of the reference sensor, a more accurate reference is needed, the so-called "super-reference". We define the term "super-reference" as follows:

"Comparing the result ξ obtained by device $\mathcal{A}$ to that of device $\mathcal{B}$. The underlying measurement principle of $\mathcal{B}$ is fundamentally different to $\mathcal{A}$, i.e., $\mathcal{B}$ is invariant to error sources of $\mathcal{A}$. Measuring ξ by means of $\mathcal{B}$ is characterized by high fidelity, accuracy, repeatability, and intuition. $\mathcal{B}$ is thereby seen as a super-reference for obtaining ξ".

In order to distinguish the term "super-reference" from the calibration of a measuring device, the definition of calibration is considered. Calibration is defined as "*to mark units of measurement on an instrument so that it can be used for measuring something accurately*" [13]. Therefore, the usability of a measuring device for determining the "GT" is qualified by a calibration procedure.

The "super-reference" principle is demonstrated using position measurements with GNSS. A GNSS device is chosen to serve as a reference measurement technique. To determine the shortest distance between two GNSS points, their Euclidean distance according to the obtained GNSS positions can be used. The result is subject to all errors affecting the GNSS measurements and can only be seen as correct within ± 2 cm. A super-reference for calibrating this method is given by a length-measuring device such as a tape measure or meter stick, which usually have an accuracy level in the sub-mm range according to EC Regulation 2004/22/EC [14]. Thereby, the demand for accuracy during the setup of the measurement to obtain these values has to be absolutely exact regarding experimental conduct.

4.1. Super-Referencing in Automotive Use Cases

The current state of an object is given by its translational and rotational degrees of freedom and the respective rates of change and accelerations, which are defined according ISO 8855 [15]. In a Cartesian frame, these would be $\mathbf{x} = [x, y, z, \phi, \theta, \psi]$ along with $\dot{\mathbf{x}} = [\dot{x}, \dot{y}, \dot{z}, \dot{\phi}, \dot{\theta}, \dot{\psi}]$ and $\ddot{\mathbf{x}} = [\ddot{x}, \ddot{y}, \ddot{z}, \ddot{\phi}, \ddot{\theta}, \ddot{\psi}]$, as well as $\dddot{\mathbf{x}}$ when also considering jerk.

For the calibration of these 24 quantities, only the longitudinal acceleration values offer a natural reference value: standard acceleration due to gravity (approx. 9.81 m/s^2 [16]) can

be calculated for different locations and altitudes [17]. An acceleration sensor measuring along the axis pointing to the center of the earth can be referenced via this value. Aids are required for calibrating the other measured variables. For example, translational distances can be referenced via auxiliary means, such as the aforementioned meter stick. With respect to manufacturing tolerances, high-precision Computerized Numerical Control (CNC) machinery would provide sufficient accuracy for calibration rotation angles [18].

Finding a super-reference is more difficult for velocities. Although the speed of sound defines a reference, it is beyond relevant velocities in the automotive domain. Furthermore, the specified velocity resolution of precision-measuring instruments such as ADMA or OXTS is in the range of less than 0.01 m/s. This is an order of magnitude above the velocity resolution of automotive radar sensors via the Doppler effect [19] (p. 272).

Technically, velocity can be determined by the change in location within a time interval. However, this requires very high sampling rates in the automotive context, as the following calculation example illustrates: let an object's longitudinal velocity $\dot{x} = 10\,\mathrm{m/s}$ and the lowest possible distance between two measurement points $\Delta x = 5\,\mathrm{cm}$ be the parameters of the measurement setup; the necessary sampling frequency f_s is calculated by the time difference between $\dot{x}$ and the sum of $\dot{x}$ and the velocity accuracy $\Delta\dot{x} = 0.01\,\mathrm{m/s}$. Then, the following consideration is valid under the assumption of constant velocity.

$$f_s = \frac{1}{\frac{\Delta x}{\dot{x}} - \frac{\Delta x}{\dot{x}+\Delta\dot{x}}} \approx 200\,\mathrm{kHz} \tag{1}$$

This sampling frequency exposes high demands on typical measurement devices and is therefore beyond the scope of our considerations.

4.2. Materials and Methods for Practical Super-Referencing

The following section is organized as follows: first, the ADMA is described. Next, the different experimental setups for super-referencing the lateral y_{SRef} and longitudinal x_{SRef} position in stationary and dynamic cases with the corresponding materials, as well as determination of the yaw angle, are described. Thereby, the index "SRef" denotes the super-reference measurement. In the automotive sensor modeling and validation context, these values are of the utmost interest.

The ADMA-G-PRO+ by Genesys Offenburg GmbH is available as a reference measurement technique in this study. Because of the high accuracy of up to ± 2 cm [20], high sampling frequencies of up to 1000 Hz and the possibility to use the device as standalone, as well as the combination of two systems, the methods and results can be generalized for comparable devices. Next to the position, the yaw angle accuracy is specified by $\pm 0.05°$ [21] and the velocity is measured with an accuracy of less than ± 0.01 m/s.

The ADMA is mounted via a rack on the vehicle. To configure the device, the mounting offset between its measuring center and the GNSS antenna is required. The ADMA is capable of outputting the poses and their derivatives in a defined point of interest (POI), provided that their positions with regard to its measuring center are known. In our case, we define and measure two POIs: the center of the rear axle and the connection point of a tow bar in the front/back of the vehicle. We use cross line lasers, a measurement tape, and meter rods to determine the described aforementioned offsets with an accuracy of ± 2 mm. Additional supporting points are obtained by photogrammetry measurement of the vehicle.

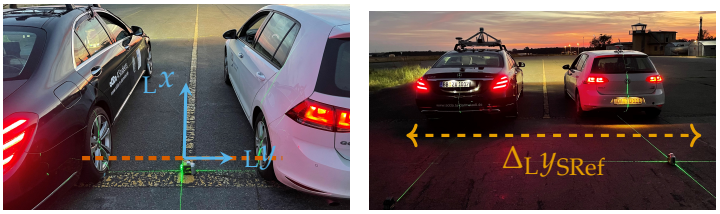
4.2.1. Calibration of Lateral and Longitudinal Position in Stationary Conditions

To determine the correct measurement procedure during the setup of the ADMA and antenna in the vehicle, a stationary calibration experiment has to be conducted to ensure lateral and longitudinal positioning correctness. The accuracy of the measurement device can be determined by two reference points. These points must be known with regard to their geodetic or Cartesian position. One of these reference points marks the origin of a local coordinate system, of which one axis spans through the second reference point. For

a positioning device, a given lateral/longitudinal displacement between the POI and the measurement origin is to be indicated. When one of them is brought to zero, the other quantity can be determined directly. The method is applicable for a single or dual car setup.

For the single-car calibration, the vehicle equipped with the positioning measurement device is placed along one of the axes of the reference coordinate system; see Figure 3a. The measured longitudinal component should now indicate zero, while the lateral component can be determined with a reliable distance measurement device such as a meter stick. The remaining errors indicate the calibration offsets of the positioning device, such as in the aforementioned mounting offsets.

For determining the position of two cars with regard to each other, the setup is fundamentally similar. The rear axles of two vehicles are placed parallel to each other, resulting in zero displacement in the longitudinal direction. The lateral distance can now be obtained in the same way with a meter stick. To ensure the correct positioning of the vehicle's POI at the position ${}_{\mathrm{L}}x = 0$ in a local coordinate system "L", a cross line laser is used. The super-reference measurement of $\Delta_{\mathrm{L}}y_{\mathrm{SRef}}$ is done by means of two cross line lasers focusing on the middle axis of the vehicles, as visualized in Figure 3b. The measured values are then compared to the output of the GNSS device.



(**a**) (**b**)

Figure 3. Dual-car calibration setup. Super-reference is provided by perpendicular cross line laser lines. (**a**) Zero longitudinal offset (i.e., $\Delta_{\mathrm{L}}x_{\mathrm{SRef}} = 0$) between the vehicles is verified by cross line laser through center of rear axles. (**b**) Lateral offset is obtained by measuring the distance between cross line laser lines focusing along the vehicle's middle axis.

4.2.2. Yaw Angle

The yaw angle and, in turn, the relative orientation between vehicles is among the relevant quantities in evaluating movement patterns in road traffic. Given the sensitivity of the reflectivity of vehicles with regard to the aspect angle for radar and lidar sensors, its accurate determination is highly desirable.

When using IMU-based systems for angle measurement, drift of the displayed angle may occur. This error is caused by the integration of the measured rotation rate and the angular acceleration by the IMU. An offset error can hardly be avoided, which results in a higher drift after a longer operating time, without correction by additional efforts. This so-called drift stability is usually provided in the sensor specification.

As a super-reference for the yaw angle, the cosine theorem is used: it determines the enclosed angles from the given side length of a triangle, i.e., $\cos(\psi_{\mathrm{SRef}}) = \frac{x_1^2+x_2^2-d^2}{2x_1x_2}$. The measurement setup for the stationary yaw angle super-reference is shown in Figure 4. This experiment is suitable as a super-reference, because the underlying measurement principle is completely different in comparison to the device under test.

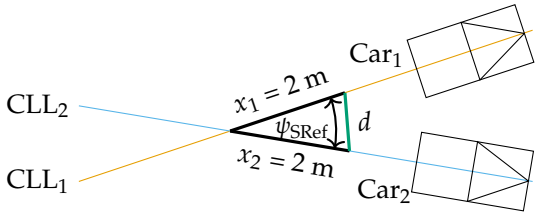


Figure 4. Measurement setup for yaw angle super-reference.

We use two cross line lasers, which are positioned in the same directions as the x-axes of the two cars, to obtain the origin of the straight x_1 and x_2. Cross line lasers are aligned so that they point exactly through the centerline of the vehicles. The manufacturer's logo on the trunk and the shark radio antenna on the roof serve as support points when aligning the lasers. Starting from the intersection of the laser lines, the side lengths of the triangle can now be determined. The edges x_1 and x_2 are determined using a 2 m long meter stick. Meter rods offer accuracy classes in the sub-mm range, which is considered adequate for the intended use here. This simplifies angle determination by means of the law of cosine because two side lengths are already fixed. The length of d is measured by a measurement tape and ψ is calculated by the three given lengths and the cosine theorem. Five measurements are made within 36 min. The accuracy of this measurement method can be calculated based on the Gaussian error propagation. The values for the error propagation are $x_1 = x_2 = 2 \pm 0.005$ m, $d_{\min} = 0.902 \pm 0.005$ m and $d_{\max} = 1.529 \pm 0.005$ m.

$$\Delta\psi_{\max,\mathrm{SRef}} = \left|\frac{\partial\psi}{\partial x_1}\right|\Delta x_1 + \left|\frac{\partial\psi}{\partial x_2}\right|\Delta x_2 + \left|\frac{\partial\psi}{\partial d}\right|\Delta d = \pm 0.085\mathrm{deg} \tag{2}$$

4.2.3. Absolute Positioning in Dynamic Case

To investigate the absolute accuracy of the reference measurement technique in the dynamic case, the following experiment is proposed: a vehicle passes through three light barriers designated as Lb_1, Lb_2, and Lb_3. These are aligned perpendicular to the roadway. The timesteps $t_{\mathrm{Lb}_{1\ldots3}}$ at which a light barrier is crossed mark the point in time with zero longitudinal offset between the light barrier and the front point of the vehicle in a light barrier-centered coordinate system. In addition, a foam line is drawn perpendicular to the road. The measurement principle of the super-reference is again completely different to the ADMA and therefore this experiment is suitable as a super-reference. The full measurement setup is illustrated in Figure 5.

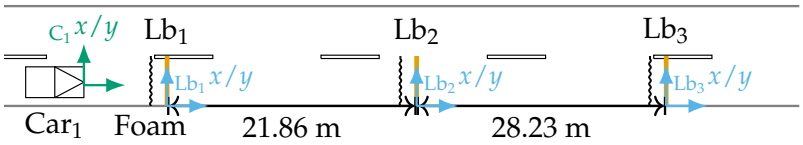


Figure 5. Measurement setup for super-referencing absolute positioning in the dynamic case.

The error of the reference system is found at each light barrier as

$$\epsilon_{\mathrm{lat}}(t_{\mathrm{Lb}_{1\ldots3}}) = {}_{\mathrm{C}_1}x_{\mathrm{Ref}}(t_{\mathrm{Lb}_{1\ldots3}}) - {}_{\mathrm{Lb}_{1\ldots3}}x_{\mathrm{SRef}}(t_{\mathrm{Lb}_{1\ldots3}}). \tag{3}$$

and when crossing the foam line, the lateral error can be determined based on the tire marks that remain on the foam. The lateral offset can only be determined at the wheels. The imprint of the tires is determined with a measurement tape and gives the lateral distance between the light barrier and the wheels. To account for the offset between the front of the vehicle and the front axle, the foam line is applied in front of the light barrier with an offset by this amount to minimize errors due to yaw angles. In other words, the longitudinal offset is known at the time at which the light barrier is crossed and should be zero. The longitudinal error is obtained at $t_{\mathrm{Lb}_{1\ldots3}}$ for each light barrier and reads:

$$\epsilon_{\mathrm{long}}(t_{\mathrm{Lb}_{1\ldots3}}) = {}_{\mathrm{C}_1}y_{\mathrm{Ref}}(t_{\mathrm{Lb}_{1\ldots3}}) - {}_{\mathrm{Lb}_{1\ldots3}}y_{\mathrm{SRef}}(t_{\mathrm{Lb}_{1\ldots3}}). \tag{4}$$

The experiment is conducted with the vehicle passing the light barriers at a constant velocity of 30 km/h and with an initial set speed of $\dot{x}_{\mathrm{C}_1}$ = 100 km/h at Lb_1 and braking. When the vehicle is decelerated while passing through the light barriers, the accuracy of the positioning in the dynamic case can be studied. Crossing the barriers with constant velocity indicates the potential sensitivity of positioning errors to velocity.

Three SICK WL 12-2 light barriers that have a specified delay time of 330 µs are chosen for use in the experiment. The light barriers are connected to a second ADMA that is placed stationary next to Lb_2. Time synchronization between both devices is given by timestamps conveyed in the GNSS signal. Both ADMAs operate with f_s = 1000 Hz to minimize the positioning error due to sampling discretization.

The position of the light barriers in GNSS coordinates is measured with the RTK-aided Piksi Multi GNSS Module by Swift Navigation. The position of the point is averaged by measurement over 60 s. The verification of these GNSS coordinates is given as it matches the distance between the light barriers, which is determined by a measuring tape with mm accuracy. The spherical GNSS coordinates are converted into an East-North-Up (ENU) coordinate system based on the WGS84 ellipsoid, which is a metric Cartesian system.

4.2.4. Relative Positioning between Vehicles in Dynamic Case

To determine the accuracy of the ADMA in the dual measurement setup under dynamic conditions, a constant distance between the two vehicles can be used. A tow bar mounted between two vehicles fulfills the requirement between the respective mounting points, also while driving. The position of the towing lugs on the vehicles relative to the ADMA is defined as a POI. By using the positioning information obtained, the calibration goal is to obtain the length of the tow bar, denoted $l_{\text{tb,Ref}}$, which is assumed constant when neglecting strain effects of materials. Then, the resulting error, i.e., $\epsilon_{\text{tb}} = l_{\text{tb,Ref}} - l_{\text{tb,SRef}}$, is obtained, which should give zero for an ideal measurement. Measured length $l_{\text{tb,SRef}}$ by a measuring tape of the tow bar is defined as the Euclidean distance of the measured mounting points in Cartesian world coordinates, i.e.,

$$l_{\text{tb,Ref}} = \left\| \begin{pmatrix} {}_{\text{L}}x_2 - {}_{\text{L}}x_1 \\ {}_{\text{L}}y_2 - {}_{\text{L}}y_1 \end{pmatrix} \right\|_2 \tag{5}$$

Car_1 accelerates from standstill to a given set speed. After a period of constant velocity, the front vehicle brakes the convoy to standstill. The velocity is controlled by Car_1's speed limiter, while Car_2 rolls behind in towing mode, i.e., neutral gear position. Three velocity profiles were studied, each with multiple repetitions.

1. $0 \rightarrow 30\,\text{km/h} \rightarrow$ maintaining $\rightarrow 60\,\text{km/h} \rightarrow$ maintaining $\rightarrow 30\,\text{km/h} \rightarrow$ maintaining $\rightarrow 0$
2. $0 \rightarrow 30\,\text{km/h} \rightarrow$ maintaining $\rightarrow 0$
3. $0 \rightarrow 80\,\text{km/h} \rightarrow$ maintaining $\rightarrow 0$

The profiles differ in the duration and intensity of acceleration or deceleration, as well as the duration of cruising at "constant" speed. In this way, the influence of these motion phases on the error can be studied. It is to be noted that the set speed of the speed limiter is the speedometer value, which is above the actual GT speed. The general scenario setup is shown in Figure 6.

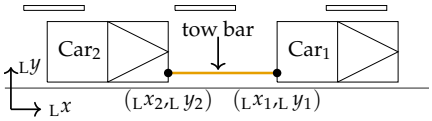


Figure 6. Measurement setup for dynamic dual super-reference with a tow bar in a local coordinate system L.

5. Super-Referencing Results Obtained in Practical Experiments

The proposed super-reference methods were performed at the August Euler airfield near Darmstadt, Germany, between April and September 2021. The ADMA devices used

were mounted, measured, and initialized according to the manufacturer's instructions. A 2015 VW Golk Mk7 and a 2018 Mercedes S Class V222 were available as test vehicles.

5.1. Yaw Angle

Figure 7 shows the results of our yaw angle referencing experiment. It compares the heading angle as calculated from the law of cosine to the measured value from the ADMA, i.e., $\epsilon_\psi = |\psi_{\text{Ref}} - \psi_{\text{SRef}}|$. The experiment was conducted five times at various positions and data were collected for around 60 s each. The vehicles were moved only for the purpose of changing position and were otherwise stationary, especially during the determination of the super-reference, which took a couple of minutes. Stationary operating conditions particularly favor the occurrence of yaw angle drift. The drift objectively shows little effect and the deviations are less than 1 deg even after 36 min. It should be noted that the IMU and GNSS fusion system utilizes the dynamic movements of the device. Such a stationary experiment over a long time is challenging for the system. Drift is therefore an expected side effect.

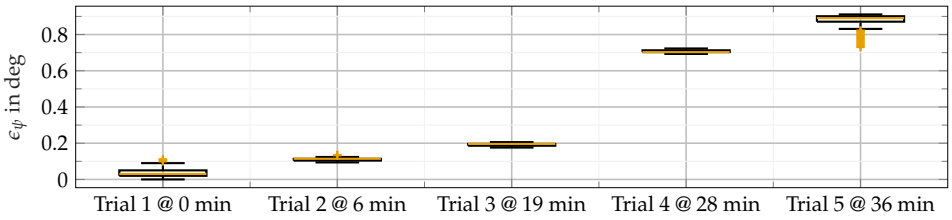


Figure 7. Statistical analysis of heading angle error ϵ_ψ.

5.2. Absolute Positioning in Dynamic Case

The results of the super-referencing absolute positioning in the dynamic case by using light barriers (see Section 4.2.3) are shown in Figure 8. The lateral and longitudinal errors are denoted by ϵ_{lat} and ϵ_{long}, respectively.

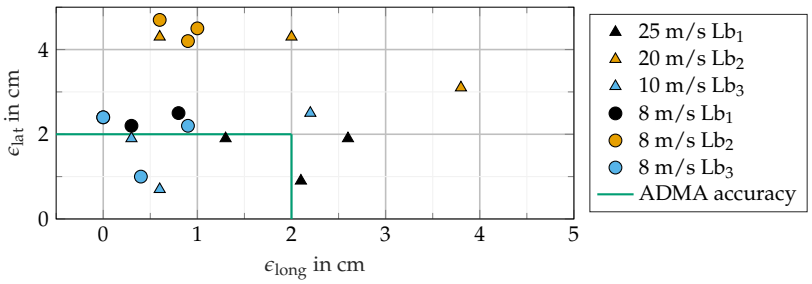


Figure 8. Lateral ϵ_{lat} vs. longitudinal error ϵ_{long} obtained by light barriers and foam.

In general, high longitudinal and lateral precision in the three trials of every experiment and light barrier position is achieved. It is to be noted that ϵ_{long} is larger with a higher speed of the vehicle as it crosses the light barrier. Therefore, low velocities should be used as target velocities to achieve sufficient accuracy or devices with higher sampling frequencies. This is explained with measurement errors due to the light barrier's time delay $\Delta t_{\text{Lb}} = 330\,\mu\text{s}$ [22]. This explains the decreasing deviation in the longitudinal direction with decreasing speed, visible by the triangle markers. The delay results in a worst-case error at 25 m/s of

$$\epsilon_{\text{long,max}} = \dot{x}_{\text{C}_1,\text{max}} \Delta t_{\text{Lb}} = 8\,\text{mm} \,@\, 25\,\text{m/s}. \tag{6}$$

The remaining deviation is the error of the ADMA and the positioning error of the experimental setup. The lateral error ϵ_{lat} of our calibration method shows deviations higher than the proclaimed accuracy of the ADMA consistently present at the second light barrier. It shows deviations of around 2.5 cm from the proclaimed accuracy and indicates the experimental setup error. The ADMA's error in the absolute dynamic case with a low

velocity is always positive and differs between 0 cm and 3.8 cm in the longitudinal and 0.5 cm and 4.5 cm in the lateral direction.

5.3. Relative Positioning in Dynamic Case

The relative positioning error in the dynamic case is obtained by estimating the length of a tow bar mounted between two vehicles while driving; see Section 4.2.4. Figure 9 shows exemplary results obtained during one trial of the experiment. It is structured as follows: the error, which is obtained when estimating the tow bar length, i.e., ϵ_{tb}, varies within ± 3 cm. Because of the dual measurement setup, the worst-case error based on (5) and $\Delta x_{1/2} = \Delta y_{1/2} = \Delta x = \pm 2$ cm is:

$$\Delta \epsilon_{\text{tb}} = \sqrt{(\Delta x_1 + \Delta x_2)^2 + (\Delta y_1 + \Delta y_2)^2} = 2\sqrt{2}\Delta x = 5.7\,\text{cm} \tag{7}$$

Therefore, the deviation of the devices is in accordance with their specification. Longitudinal acceleration in Car_1 or Car_2 with the fixed coordinate system shows little difference due to the mechanical coupling by the tow bar, which causes crabbing at the rear car. Moreover, the velocity profile is shown and does not indicate a strong correlation between error dynamics and longitudinal acceleration.

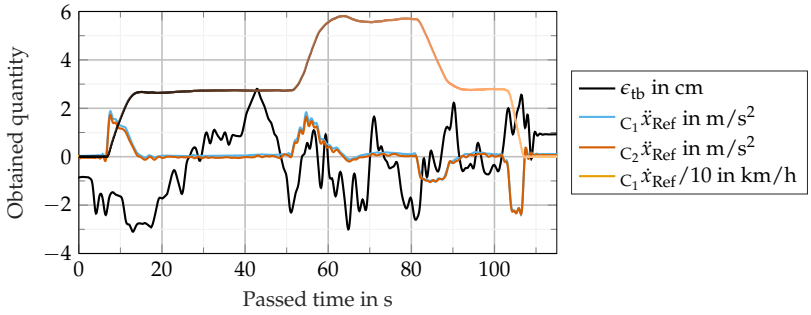


Figure 9. Exemplary measurement reading from one out of five trials. The distance error in cm, longitudinal acceleration, and velocity of the front vehicle are shown. Note that velocity is scaled for better readability. The color gradient with the velocity indicates the running time.

In Figure 10, the influence of velocity and acceleration on ϵ_{tb} is shown. The time course of the velocity or acceleration profile is coded in the color gradient from black to light brown and all trials of the experiment are shown. Studying the sensitivity of ϵ_{tb} to velocity reveals three consistent characteristics for all tests; see the left column in Figure 10.

1. The error shows a fluctuation range of approximately 2 cm during quasi-stationary driving and matches the specification.
2. During acceleration and braking phases, the error remains at a tolerable constant value within the fluctuation range.
3. When reaching standstill, the error settles at a certain value, which lies inside the specification of the dual measurement setup.

No consistent correlations follow from the acceleration profile, as shown in the right column of Figure 10. However, it can be seen that the error also changes during the acceleration phases in the range of a few cm. It is worth noting, however, that the error profile shows some consistency when the acceleration profile is similar, as shown in the portion highlighted by a light blue ellipse in the right column and the first row of Figure 10.

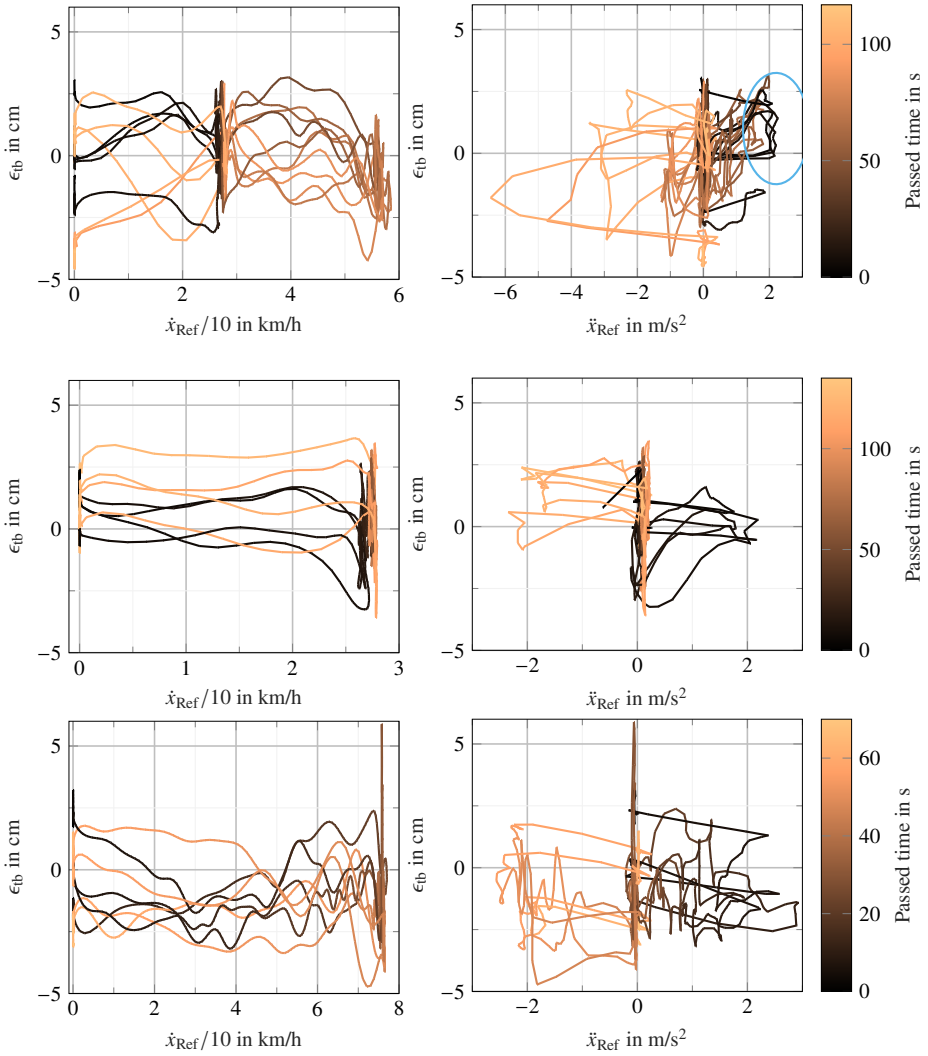


Figure 10. In the left column, ϵ_{tb} vs. velocity is shown for the different velocity profiles mentioned in Section 4.2.4. During areas of "constant" speed, the distance error settles within the accuracy of the measurement devices. In the right column, ϵ_{tb} vs. acceleration is shown. The distance error dynamics show only low sensitivity to acceleration. All trials are depicted and running time is denoted by the line's color gradient.

6. Feasibility of Transferring Real-World Test Drives to Simulation

The main interest in using reference sensors in the context of virtual validation is ultimately to transfer real-world test drives to virtual environments. Under the so-called "Measurement2Sim" method, modern simulation tools such as IPG CarMaker, Vires VTD, or CARLA are able to control an actor's position based on a given trajectory. The tow bar experiments are suitable to represent a simulation's capability to render recorded measurements in the movement of objects. For this purpose, these experiments were transferred to two different simulation environments: Sim1 and Sim2.

The results are given in Figure 11 and are organized as follows. The left column shows the error ϵ_{Sim1} of the first simulation and the right column shows the error ϵ_{Sim2} of the second simulation. The topmost figures show five trials of the experiment, where the two vehicles undergo two phases of acceleration and deceleration with semi-stationary drive in between, i.e., from 0 to 30 km/h, 30 km/h to 60 km/h, back to 30 km/h, and finally to 0. The middle figures show the experiment with 30 km/h and the bottom figures with 80 km/h. The figures visualize the error between the reference measurement, as discussed in Section 5.3, and the simulation environment. Zero error would indicate that

the measurement of the distance between vehicles, obtained either in simulation or via the reference measurement, exactly corresponds to the length of the tow bar.

$$\epsilon_{\text{Sim1/2}} = (l_{\text{tb,Sim1/2}} - l_{\text{tb,Ref}}) \tag{8}$$

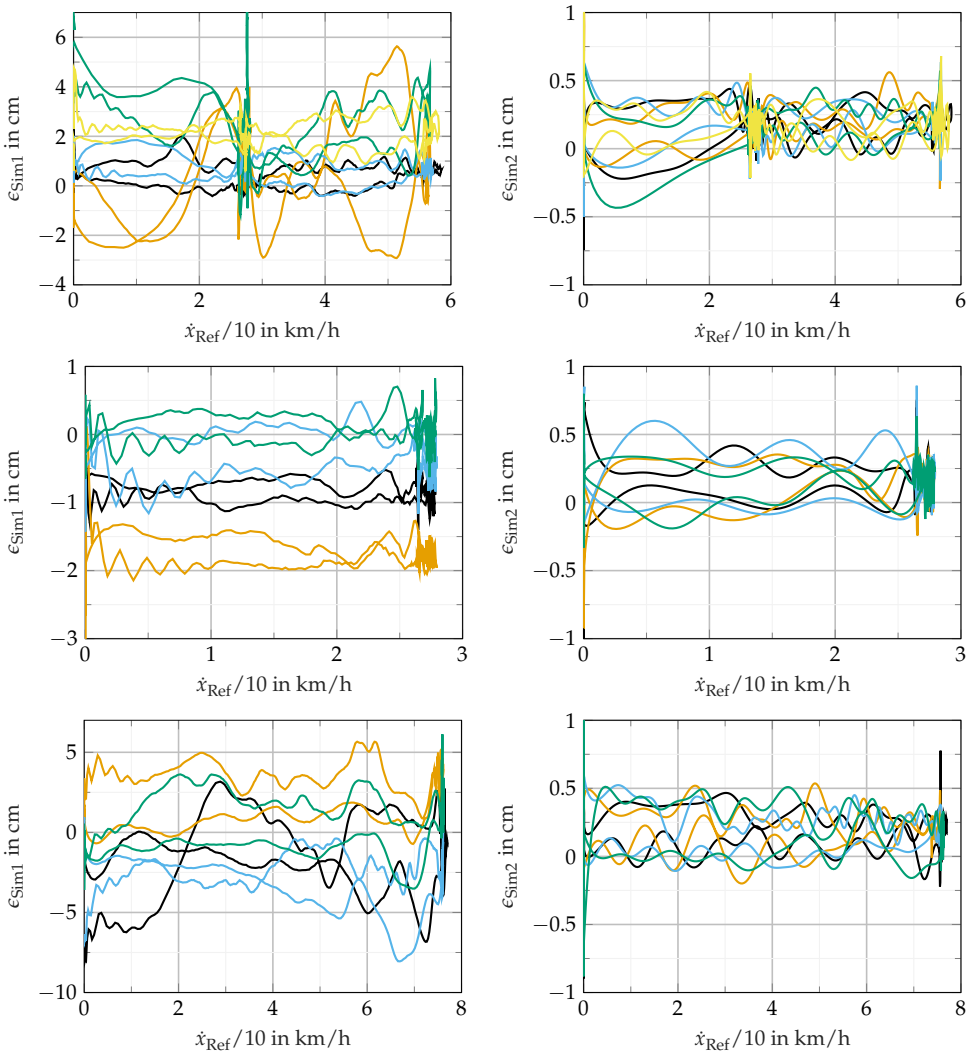


Figure 11. In the left column, ϵ_{Sim1} vs. velocity is shown for the different velocity profiles in the first simulation. In the right column, ϵ_{Sim2} vs. the same profiles is shown in the second simulation. Each trial is visualized with a different color.

The experiments show that the resulting deviations vary between trials through all trials of the experiments. The error in both simulation tools thereby shows sensitivity to velocity: from the results shown, it can be concluded that the error becomes less with lower speeds, while showing the largest error during phases of acceleration or deceleration. In the first simulation during phases of semi-stationary velocities, the error occasionally extends the proclaimed accuracy of the ADMA. In the second simulation, in turn, the errors are always in the accuracy range.

Our results show that the reenacting of test drives performs best with the first simulation tool when the velocities of the vehicles are kept fairly constant and the accelerations are low, i.e., less than $2\,\text{m/s}^2$. The absolute deviation between measurement and simulation is in orders of magnitude exceeding the distance resolution of lidar sensors or high-resolution radar sensors. This makes the comparison of simulation to measurement considerably more difficult, since the basis of comparison shows deviations.

7. Discussion

In this paper, we present four different experimental setups to obtain super-reference measurements. With the proposed methods, the confidence in GNSS- and IMU-based reference data for lateral and longitudinal positions can be strengthened in stationary and dynamic cases, as well as the drift analysis of the stationary yaw angle. We note that the highest precision is required when setting up the measurement equipment in order to achieve useful results in terms of super-referencing. During our experimental setups, we encountered the necessity of excellent measuring conditions regarding GNSS measurement devices, because the accuracy of the device's data is highly dependent on the surrounding conditions. Effects such as multipath propagation, shading by other objects, and loss of differential GNSS and RTK connection result in deviation that is an order of magnitude above ideal conditions.

Our experiments reveal the strengths and weaknesses of the reference system under study, the ADMA. The stated measurement accuracy is almost consistently met. The yaw angle measurement quantifies the expected drift of the device. The reference system confirms the proclaimed accuracy during the light barrier experiment. The experiment shows the difficulty in verifying the position accuracy by means of the super-reference, showing less deviation than the system under test. In the dynamic dual measurement setup with the tow bar, the deviation always lies within the specification.

Our comparison of the simulation and real test drive shows a new possibility of verifying the fidelity of so-called "Measurement2Sim" methods. Not only the transfer of the trajectory into the simulation is a source of deviations between measurements and simulation, but also the simulation tool itself provides errors due to the trajectory discretization. The results between the two simulation tools differ clearly. The sources of the deviations cannot be directly identified. When the "Measurement2Sim" method is used in the context of validation of sensor models, it has to be noted that the deviation must not exceed the accuracy of the sensor itself. In the case of lidar, typically, accuracy lies within in the centimeter range. The second simulation tool is better suited to reproducing sensor effects in sensor simulation models with the "Measurement2Sim" method. This simulation tool converts the trajectories very well on the basis of an OpenScenario xosc file based on the measurements of $\mathbf{x} = [x, y, z, \phi, \theta, \psi]$. For future verification and validation experiments in combination with "Measurement2Sim" methods, we highly recommend the analysis of the transfer error of the measurement into the simulation.

Regarding virtual validation by means of digital twins, our results indicate that sample validation using "reference measurement sensors" can hardly be achieved. This is of particular importance when considering the accuracy of perception sensors, which is close to the stochastic deviation margin of the reference measurement system. Rather, our findings strengthen the argumentation for stochastic validation approaches that explicitly take the measurement uncertainties of the reference system into account.

Author Contributions: M.H. and L.E.: conceptualization, methodology, formal analysis, investigation, software, visualization, writing—original draft preparation. H.W.: conceptualization, methodology, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work received funding from *VIVID* grant number 16ME0173, promoted by the German Federal Ministry for Education and Research (BMBF), based on a decision of the Deutsche Bundestag. We also acknowledge support by the Open Access Publishing Fund of the Technical University of Darmstadt.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source code for the conversion of the ADMA data into an xosc file is available on *TUdatalib* under http://https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/2993, accessed on 9 November 2021.

Acknowledgments: The authors would like to thank Dominik Huber from GeneSys Elektronik GmbH for his support during the commissioning of the ADMA measurement system in our vehicles, Jonathan Knerr and Anthony Ngo for their contributions to earlier versions of our method and Timm Ruppert for preparing the Piksi Multi GNSS Module System, and Felix Glatzki for supporting the measurement campaign.

Conflicts of Interest: The authors have no conflict of interest. The funders had no role in the writing of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

ACC	Adaptive cruise control
AD	Automated driving
ADAS	Advanced driver assistance system
ADMA	Automotive dynamic motion analyzer
CAN	Controller area network
GNSS	Global navigation satellite system
GT	Ground truth
IMU	Inertial measurement unit
POI	Point of interest
RTK	Real-time kinematics

References

1. Ramezani, M.; Wang, Y.; Camurri, M.; Wisth, D.; Mattamala, M.; Fallon, M. The Newer College Dataset: Handheld LiDAR, Inertial and Vision with Ground Truth. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 4353–4360. [CrossRef]
2. Skoglund, M.; Petig, T.; Vedder, B.; Eriksson, H.; Schiller, E.M. Static and dynamic performance evaluation of low-cost RTK GPS receivers. In Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 19–22 June 2016; pp. 16–19. [CrossRef]
3. Wagener, A.; Katz, R. Automated scenario generation for testing advanced driver assistance systems based on post-processed reference laser scanner data. In *Fahrerassistenzsysteme 2016*; Springer Fachmedien Wiesbaden: Wiesbaden, Germnay, 2018; pp. 175–190.
4. Automotive Artificial Intelligence (AAI) GmbH. Scenario Cloning and Extraction—Closing the Loop between Reality and Simulation. 2016. Available online: https://www.automotive-ai.com/technologies/scenario-cloning (accessed on 15 April 2021).
5. Atlatec Gmbh. Scenarios for Simulation. 2021. Available online: https://atlatec.de/scenarios-for-simulation/ (accessed on 15 April 2021).
6. Annotel. Annotation & Ground Truth. 2021. Available online: https://www.annotell.com/annotation (accessed on 15 April 2021).
7. Scale AI, Inc. 3D Sensor Fusion for Simulation. 2021. Available online: https://scale.com/3d-sensor-fusion (accessed on 15 April 2021).
8. Schalling, F.; Ljungberg, S.; Mohan, N. Benchmarking LiDAR Sensors for Development and Evaluation of Automotive Perception. In Proceedings of the 2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Kedah, Malaysia, 27–29 November 2019.
9. Brahmi, M. Bewertung der Objektbasierten Umfeldwahrnehmung für Fahrerassistenzsysteme Mithilfe von Referenzsystemen. Ph.D. Thesis, Technische Universität Braunschweig, Braunschweig, Germany, 2020.
10. Steinhard, N.; Kaufmann, S.; Sven, R.; Lages, U.; Goerick, C.; Noutangnin, Y. Lidar Based Object Tracking Evaluation Forautomotive Applications. 2016. Available online: https://www.honda-ri.de/pubs/pdf/3216.pdf (accessed on 15 April 2021).
11. Oxford University Press "Accuracy". Oxford Leaner's Dictionary of Academic English. 2021. Available online: https://www.oxfordlearnersdictionaries.com/definition/academic/accuracy (accessed on 18 September 2021).
12. Oxford University Press "Precision". Oxford Leaner's Dictionary of Academic English. 2021. Available online: https://www.oxfordlearnersdictionaries.com/definition/english/precision (accessed on 18 September 2021).
13. Oxford University Press "Calibrate". Oxford Leaner's Dictionary of Academic English. 2021. Available online: https://www.oxfordlearnersdictionaries.com/definition/english/calibrate (accessed on 18 September 2021).
14. European Parliament and Council. Regulation 2004/22/EG. 2004. Available online: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32004L0022 (accessed on 19 September 2021).
15. International Organization for Standardization. ISO 8855:2011 Road Vehicles—Vehicle Dynamics and Road-Holding Ability—Vocabulary. 2011. Available online: https://www.iso.org/standard/51180.html (accessed on 19 September 2021).

16. Bureau International des Poids et Mesures. Resolution 2 of the 3rd CGPM. Declaration on the Unit of Mass and on the Definition of Weight; Conventional Value of gn. 1901. Available online: https://www.bipm.org/en/committees/cg/cgpm/3-1901/resolution-2 (accessed on 19 September 2021).
17. Physikalisch-Technische Bundesanstalt. g-Extractor. 2007. Available online: http://www.ptb.de/de/org/1/11/115/g-extractor.htm (accessed on 27 September 2021).
18. Bosetti, P.; Bruschi, S. Enhancing positioning accuracy of CNC machine tools by means of direct measurement of deformation. *Int. J. Adv. Manuf. Technol.* **2012**, *58*, 651–662. [CrossRef]
19. Winner, H. Radarsensorik. In *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme fúr Aktive Sicherheit und Komfort*; Winner, H., Hakuli, S., Lotz, F., Singer, C., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2015; pp. 259–316. [CrossRef]
20. Hessische Verwaltung für Bodenmanagement und Geoinformation. 2021. SAPOS Hessen HEPS. Available online: https://sapos.hvbg.hessen.de/service.php#EPS (accessed on 27 September 2021).
21. GeneSys Elektronik GmbH. *ADMA-G-Pro+ Technical Documentation*; GeneSys Elektronik GmbH: Offenburg, Germany, 2020.
22. Sick AG. *Technical Documentation SICK SENSICK WL12-2*; Sick AG: Düsseldorf, Germany, 2007.

B Paper II:

A Dataset for Radar Scattering Characteristics of Vehicles Under Real-World Driving Conditions: Major Findings for Sensor Simulation

A Dataset for Radar Scattering Characteristics of Vehicles Under Real-World Driving Conditions: Major Findings for Sensor Simulation

Lukas Elster, Martin F. Holder, and Manuel Rapp

***Abstract*—In the virtual validation of automated driving, trustworthy simulation models of perception sensors are required. Radar sensors are particularly hard to model, as their measurements are notoriously difficult to interpret. This is due to their complex measurement principle, involving multipath propagation of mm-waves, varying backscattering characteristics of objects, and further factors such as limited measurement ranges and resolutions that introduce uncertainty to the measurements. This work presents a method for studying the backscatter characteristics of vehicles under real-world driving conditions. A slalom-like driving scenario, which is representative of road driving where the vehicle is visible under different aspect angles, has been designed. It aims at a high level of reproducibility of the trajectories driven by the vehicles, hence reducing additional sources of uncertainty that were otherwise present in the measurements. In a large-scale measurement campaign, 13 vehicles have been studied. The vehicles under test are observed by multiple radars, mounted at different heights, and carry reference sensors for obtaining their positions. In this article, we present the measurement campaign and show major findings from our measurement results. Our focus lies on drawing conclusions for trustworthy sensor simulation. Both sensor measurement data and MATLAB code for data analysis are made publicly available alongside this article.**

***Index Terms*— Millimeter-wave sensors, sensor model analysis, sensor testing and evaluation.**

I. Introduction

VIRTUAL test methods are required for the release of automated driving. Such methods use simulation models of the involved sensor modalities. Radar sensors are of the utmost importance for automated driving functions, due to their ability to measure relative velocity, robustness against adversarial weather conditions, and low costs in comparison to frequency-modulated continuous-wave (FMCW) lidar systems.

Still, the development of simulation models of radar sensors, which are ready for deployment in simulation-based testing toolchains, is the subject of current research. A research problem, that is closely linked to radar sensor simulation, is the ability to model mm-wave scattering behavior realistically in virtual environments. Radar measurements are known to have a noisy characteristic, which originates from the measurement principle. Physically interpretable quantities, such as range, velocity due to the Doppler effect, and angular positions are deduced from spectral analysis of modulated mm-wave signals.

The backscattered energy is dictating the existence and achievable accuracy of radar detection. In radar theory, the physics for obtaining backscatter is conceptually absorbed in the radar cross section (RCS) denoted as σ. It is a measure of the strength of the signal reflected by an object. Influencing factors include object size, geometry, material, wavelength, and aspect angle under which the object is visible to the

Manuscript received 20 October 2022; revised 13 January 2023; accepted 13 January 2023. Date of publication 24 January 2023; date of current version 28 February 2023. This work was supported in part by the German Federal Ministry for Education and Research (BMBF) through Virtual Validation Tool Chain for Automated and Connected Driving (VIVID) under Grant 16ME0173 and in part by the Open Access Publishing Fund of the Technical University of Darmstadt. The associate editor coordinating the review of this article and approving it for publication was Prof. Qiliang Li. *(Lukas Elster and Martin F. Holder contributed equally to this work.) (Corresponding author: Lukas Elster.)*

The authors are with the Institute of Automotive Engineering, Technical University of Darmstadt, 64287 Darmstadt, Germany (e-mail: lukas.elster@tu-darmstadt.de; martin.holder.tuda-alumnus@gmx.de; manuel.rapp@stud.tu-darmstadt.de).

Digital Object Identifier 10.1109/JSEN.2023.3238015

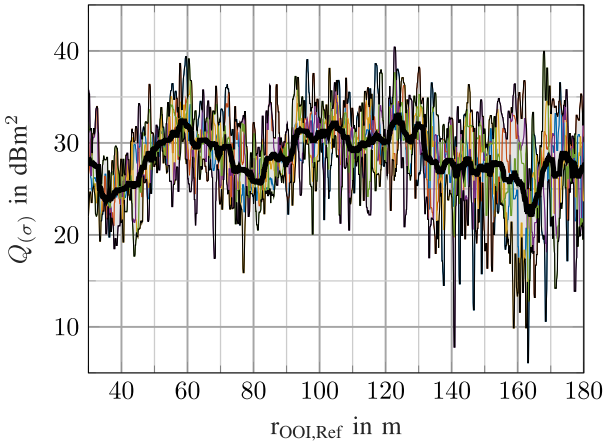


Fig. 1. Stochastic nature of RCS observed for a vehicle during five trials. The average value is indicated as a bold line and is obtained by using a moving average filter over 2.5 s. In our work, we use $Q_{(\sigma)}$ to denote RCS in the logarithmic scale, and σ for the linear scale.

sensor [1]. When inserted into the radar equation [2], as given in the following equation, the received power P_{Rx} can be obtained:

$$P_{\mathrm{Rx}} = P_{\mathrm{Tx}} \frac{G^2\, V_{\mathrm{mp}}^2\, \sigma\, \lambda^2}{(4\pi)^3\, r^4}. \tag{1}$$

Here, P_{Tx} denotes the transmitted power, and G, λ, and r denote the antenna gain, wavelength, and radial distance, respectively. This equation disregards the atmospheric attenuation and introduces V_{mp} as the so-called shaking factor ($0 \leq V_{\mathrm{mp}} \leq 2$) to account for influences by multipath propagation.

The RCS is well known, or analytically available, for simple geometries such as flat plates, spheres, or corner cube reflectors (CCRs). Analytical models are, however, not available for large objects with complex surfaces, such as vehicles.

In automotive radars, the RCS of objects is not measured directly. Instead, it is inferred from the received signal strength, the radial distance, and the angular position. Therefore, influences due to the free-space propagation loss, which is governed by r^{-4}, and the antenna gain pattern are compensated. Still, it is widely used as a quantification of the backscatter on an ordinal scale. Therefore, RCS, indicating the intensity of a detection, allows for object detection or free space estimation [3]. The noisy characteristic of the received power is also present when obtaining the corresponding RCS values in measurements. This is illustrated by the following example, obtained in a simple scenario where a vehicle is placed in front of the Ego vehicle, where "Ego" refers to the radar sensor carrier vehicle. As it is traveling at a higher speed than the Ego, the distance between the cars increases gradually. The RCS of the detections associated with the object of interest (OOI) are displayed in Fig. 1 and a distinct stochastic nature of the RCS is visible. By smoothing the noisy measurements, a moderately constant RCS value is revealed. This is, however, overlaid by additional phenomena, for example, multipath propagation, that causes a significant spread between repetitions and even adjacent measurement values. These results indicate that RCS measurement is influenced by the number of factors in a stochastic, nondeterministic fashion.

Radar models must be validated with real data to allow their usage for virtual safety validation. Therefore, the virtual environment is assessed together with the sensor model, as the real sensor is also subject to many real-world effects outside of idealized testbench conditions and anechoic radar chambers.

Although methods from the field of computational electromagnetic wave propagation allow us to compute backscatter profiles, it is debatable to what extent such simulation results may be transferred into real-world conditions. In reality, effects like multipath propagation and interference directly influence the RCS profiles and the validity of analytical models in comparison to real-world measurements, as it is highlighted in [4] and [5]. Similarly, measurement campaigns conducted in anechoic radar chambers need yet to prove their applicability to field measurements on public roads.

Therefore, we conducted an open-source available large-scale measurement campaign to obtain vehicle RCS patterns from automotive radar sensors under real driving conditions.

A. Related Work

The noisy and random nature of RCS is a well-known phenomenon in many radar applications. In radar literature, stochastic behavior is also referred to as fluctuation loss. Prominent research on this topic was carried out already in 1954 by Swerling [6]. He derived probabilistic models based on Rayleigh distributions describing the statistical properties of the RCS of objects with complexly formed surfaces, such as an aircraft. However, their application to automotive radar application, operating at mm-waves, is limited as different conditions apply here: Most notable is the presence of multipath reflections, for example, induced by reflections of the pavement, which are not present in airborne radar.

For automotive radar, a number of researchers have carried out reflectivity measurements in anechoic chambers obtained with automotive-grade radars, or vector network analyzers [7], [8]. Here, a vehicle has been placed on a rotating plate and its 360° RCS profile has been obtained from the received signal strength for a given rotation angle. Direct applicability of such results to automotive scenarios is not immediate, since the distance between radar and object is often chosen to be very small (e.g., less than 25 m), which is debatable for two reasons: First, the Fraunhofer criterion for far-field with typical aperture sizes in automotive radar is only fulfilled with 15 m onward. Second, long-range radar usually deploys high-pass filtering to avoid overloading the analog-to-digital (AD) converter for detections close to the sensor. Nevertheless, these results retain their validity and the fact that comparable results have been generated in different studies emphasizes their integrity. As the RCS profile is computationally expensive to calculate, simplified models have been derived. Based on observations and measurements, so-called "scattering center models" have been proposed [9], [10]. This theory assumes the total reflectivity of complex geometries such as vehicles is composed of individual scatterers. At the same time, there are areas on the vehicle from which a stronger backscatter is expected, such as the license plate, the wheel arches, or the exterior mirrors. The challenge with such models is the deduction of generalizability to different vehicle types and sizes, as large databases are missing.

In recent years, more radar datasets became publicly available, such as View-of-Delft [11], Radar Scenes [12], nuScenes [13], and others. They are designed for benchmarking and development of object detection and classification algorithms. It is difficult to extract reproducible scenarios for detailed RCS investigations from such datasets. In most scenes, parked (static) vehicles can be seen, or the rear of the vehicles in front while following behind. Annotations are often only available via lidar or bounding boxes, which typically do not have the precision of dedicated global navigation satellite system (GNSS) measurements.

We, therefore, design our study of vehicle backscatter such that it is representative of real road traffic conditions and at the same time generates a broad statement through many repetitions of the experiment and the use of different vehicles.

The measurement data,[1] and the evaluation source code[2] are all open-source and publicly available.

B. Research Questions and Method Outline

We state the following research questions, which we see as particularly relevant when studying the RCS characteristics in real-world traffic conditions, also in the light of radar simulation.

RQ1: What is the influence of the aspect angle of the radar to the object? Vehicles are often seen by the radar under different perspectives, for example, during lane changes. Our work addresses this research question by obtaining measurements of vehicles that follow a slalom-shaped trajectory. Therefore, they are visible under different aspect angles while conducting multiple repetitions to strengthen the underlying database for evaluation of the experiments. Utilizing a slalom course for studying reflectivity has been carried out in the author's previous work [14].

RQ2: What is the influence of the radar mounting height on RCS? In today's vehicles, radars are usually mounted at different heights. In order to study the effect of mounting height on the measured RCS, we use six identical radar sensors, mounted on a sensor rack (cf. Fig. 2).

RQ3: What is the influence of vehicle body shape on RCS? In everyday traffic, a wide variety of vehicles is encountered. Variations in their individual RCS profile are expected, due to differences in geometry, size, and material compounds. A total number of 13 vehicles of different categories (e.g., compact car, sedan, and truck) have been used in this study.

RQ4: Where is the strongest backscatter located depending on aspect angle and body shape? The position of radar detections with the highest RCS per measurement cycle is expected to depend on the aspect angle as well as the body shape of the vehicle. By comparing different vehicles scattering centers and scattering characteristics can be revealed.

II. Experimental Setup

The overall experimental setup and calibration methods for the measurement campaign as well as the quality criterion of the driven trajectory are explained in this chapter.

[1] https://www.fzd-datasets.de/rcs/

[2] https://gitlab.com/tuda-fzd/fzd-datasets/rcs-measurement

Fig. 2. Mounted sensor rack with six radar sensors at the front of the Ego vehicle. The rack has a lateral offset of 265 mm to the center. In addition to the radars and GNSS, it is also carrying lidar sensors. They are not considered in this work but the lidar data is also available within the published dataset.

TABLE I
Mounting Heights of the Radar Sensors

Sensor	Radar 1	Radar 2	Radar 3	Radar 4	Radar 5	Radar 6
h in mm	931	817	716	617	501	388

A. Sensors, Measurement Setup, and Calibration

Our research vehicle is a 2018 Mercedes-Benz S450 retrofitted with multiple Continental ARS408 radar sensors that are stacked on top of each other. Fig. 2 shows the installed measurement rack of the six radar sensors at the front of the vehicle. These radars are open-market adaptions of Continentals fourth-generation long-range radars, representing state-of-the-art automotive-grade 77-GHz radar [15]. This setup enables investigating the influence of the mounting height on the RCS profile of an OOI. The mounting heights of the sensors, measured from the ground to the center of the sensor, are specified in Table I. Due to the interference mitigation techniques of the manufacturer, interference is only recognizable in a raised signal to noise ratio (SNR) because of the six mounted sensors. As there are six same sensors used, we assume that the results focusing on RCS values are not affected based on our measurement setup.

These positions include above-average sensor heights for currently available vehicles, for example, the radar position in the Mercedes-Benz GLS series [16].

For obtaining the vehicle positions and motion, both the Ego vehicle and the OOI are equipped with the real-time kinematic (RTK)-based GNSS device automotive dynamic motion analyzer (ADMA).

Radar calibration was conducted both extrinsically and intrinsically. For extrinsics, CCRs have been placed at precisely measured positions with respect to the radar for obtaining the horizontal twist angles due to mounting. Intrinsic calibration aims at verifying the RCS value obtained by the radar. This is realized by placing a CCR with predefined RCS at a known distance from the radar (e.g., 40 m) over minimal-reflecting terrain. Both the radar and the CCR are mounted at a height of 3 m to avoid disturbances due to ground reflections.

Calibration of the relative positions between the Ego and OOI, both obtained with GNSS systems, is done by aligning

them along their rear axles, indicating zero longitudinal offsets. At standstill, the lateral offset can then be determined with a measurement tape. The known values for longitudinal and lateral offset are compared to the in the GNSS systems measured relative position data. These calibration values absorb uncertainties during obtaining mounting positions and are considered during the processing of the measurement data.

The radars report a set of detection points at a scan rate of approximately 14 Hz. Each scan contains a range, radial velocity by utilizing Doppler effect, azimuth, and RCS, and the radars are able to transmit up to 250 detections per measurement cycle. As large objects can cause multiple detections, the total RCS of an object is found by the linear sum over $n_{\mathrm{d,max}}$ detections that are linked to the object, that is,

$$Q_{(\sigma)} = 10 \log_{10} \sum_{k=1}^{n_{\mathrm{d,max}} \leq 250} \sigma_k. \tag{2}$$

The association of detections with an object is determined by multistage filtering that relies on the relative position and velocities of the OOI reported by the GNSS. The GNSS system samples at 100 Hz. While the GNSS system is reporting a precise time stamp, time referencing is made by time-stamping the individual data packages upon receiving via controller area network (CAN) bus. The radars are not synchronized in their measurement cycles. In other words, the measurement cycles of the radar sensors are independent of each other, that is, it cannot be guaranteed that the radar scans are performed simultaneously among all six radars.

B. Slalom Construction

The test setup consists of a sinusoidal slalom with ten periods. We briefly describe the setup of the slalom course, which is optimized for high repeatability for driving through it, and minimal effort for construction on straight roads. For each run, the OOI approaches the slalom setup driving straight ahead, with an offset to the center line. It is followed by the Ego vehicle which continues driving straight ahead on the center line while the OOI follows the sinusoidal trajectory of the slalom. The slalom parameters are designed to match distances and aspect angles that are similar to the real-world driving scenario of the OOI entering a curve, followed by the Ego vehicle when assuming a German EKL3-type road [17]. The geometry of the slalom is defined by the amplitude $\hat{y}$ and the frequency f_{Sla}. $\hat{y}$ is defined as 5 m to ensure a wide spectrum of aspect angles between Ego vehicle and OOI. The frequency is chosen as $f_{\mathrm{Sla}} = 0.014(1/\mathrm{m})$, which results in a slalom period of approximately 71 m. In test runs with an amplitude of 5 m, this frequency proved to be effortlessly manageable by the driver. The given geometric parameters result in a maximum yaw angle of ${}_{\mathrm{S}}\psi_{\mathrm{OOI}} = 25°$ referred to the slalom's center line. To define the relative movement between the Ego vehicle and the OOI, the corresponding speeds, v_{OOI} and v_{Ego}, as well as the initial distance between the two vehicles Δr are used. While driving through the slalom course, the OOI is visible to the Ego vehicle under different aspect angles: Its azimuth and yaw angles are changing periodically and both are opposed in phase, that is, maximum yaw is obtained at 0 azimuth, and vice versa. Fig. 3 illustrates the geometrical quantities.

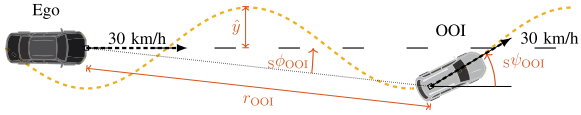


Fig. 3. Measurement setup of slalom with corresponding parameters where ${}_{\mathrm{S}}\phi_{\mathrm{OOI}}$ is denoted as the azimuth angle and r_{OOI} denoted as the range in the polar sensor coordinate system, ${}_{\mathrm{S}}\psi_{\mathrm{OOI}}$ is denoted as the yaw angle difference between the Ego vehicle and OOI as well as $\hat{y}$ is denoted as the slalom's amplitude.

The Ego vehicle's speed v_{Ego} is set to 30 km/h. This is the lowest possible speed that allows for the usage of speed control systems in both vehicles, although not all OOIs are equipped with equivalent systems. The target speed of the OOI v_{ooi} is also set to 30 km/h. By using the lowest speed possible, the duration of each experiment execution is maximized and therefore also the number of measurement samples. The sinusoidal trajectory and therefore a longer travel path in comparison to the Ego produce the most representative distance ranges with respect to the aspect angle. Lower speeds are preferred to generate as many data samples as possible during the slalom run. The usage of speed control systems causes a better reproducibility of the slalom runs, however, since they have to be manually set independent of each other, they also cause a deviation of the initial distance between Ego and OOI.

In order to guide the driver of the OOI along the ideal trajectory, traffic cones are placed along the slalom. For their placement, a sinusoidal function congruent to the ideal trajectory is calculated and eight equidistant sampling points per period are obtained. These sampling points include both vertices as well as both inflection points of each period. Each one of these characteristic positions is marked with two tall traffic cones which the OOI passes in between. The spacing between the two traffic cones is adjusted to and exceeds the width of the OOI. This is due to the fact that the reference point for the determination of the OOI's position is set as the center of the rear axle. Therefore, the front axle's center position can differ significantly from the ideal trajectory which requests additional space requirements. The four remaining sampling points are marked with flat cones over which the OOI passes. Fig. 4 shows the fully constructed slalom course. For each vehicle, ten slalom runs with five in each direction are conducted.

C. Evaluation of Trajectory

To evaluate the uniformity of the trajectory, a quality criterion based on the Hilbert transform is introduced. The Hilbert transform phase shifts a sinusoidal signal, in this case, either the trajectory of the OOI or $\hat{y}$ over x_{OOI}, by 90°. The result is an analytical signal $z(x_{\mathrm{OOI}})$, consisting of the initial signal as a real part and the phase-shifted signal as the imaginary part. This signal is referred to as $z(x)$ in the following paragraphs. To derive a quality criterion from this mathematical transformation, the absolute of the analytical signal is calculated. For an ideal trajectory, this would result

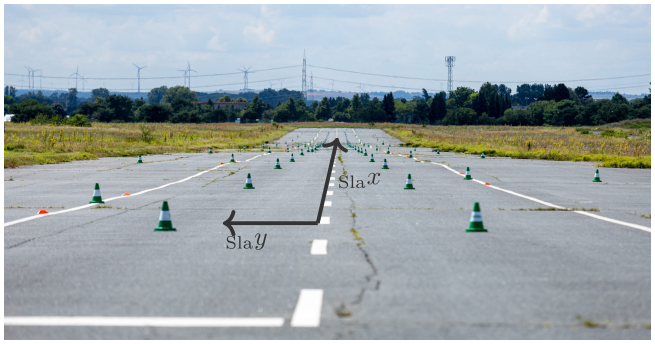


Fig. 4. Slalom course marked with traffic cones and slalom-centered coordinate system.

in a constant function with the same value as the amplitude of the slalom. However, due to human errors and environmental impacts, $z(x)$ oscillates unevenly along x_{OOI}. In the next step, this function can be easily described by its mean value $\overline{|z(x)|}$ and the according to variance $s^2(z(x))$, which is referred to as s^2 henceforth. To consider deviation from the ideal mean value and the variance, the final unitless quality criterion is calculated as shown in the following equation with the optimal value of 2:

$$k_{\mathrm{Q}} = \begin{cases} \dfrac{s^2_{\max} - s^2}{s^2_{\max} - s^2_{\mathrm{id}}} + \dfrac{\overline{|z(x)|} - \overline{|z(x)|}_{\min}}{\overline{|z(x)|}_{\mathrm{id}} - \overline{|z(x)|}_{\min}}, & \overline{|z(x)|} \leq 5\ \mathrm{m} \\ \dfrac{s^2_{\max} - s^2}{s^2_{\max} - s^2_{\mathrm{id}}} + \dfrac{\overline{|z(x)|}_{\max} - \overline{|z(x)|}}{\overline{|z(x)|}_{\max} - \overline{|z(x)|}_{\mathrm{id}}}, & \overline{|z(x)|} > 5\ \mathrm{m}. \end{cases} \tag{3}$$

Here, $s^2_{\max}$ is set as 0.05 m^2 which represents a subjectively chosen threshold based on the rough evaluation of previous test runs. Likewise, the thresholds for the mean value are defined as $\overline{|z(x)|}_{\max} = 5.2$ m and $\overline{|z(x)|}_{\min} = 4.8$ m. Furthermore, the ideal values are defined as $s^2_{\mathrm{id}} = 0$ and $\overline{|z(x)|}_{\mathrm{id}} = 5$ m, which reflect the ideal sinusoidal trajectory with an amplitude of 5 m. The quality criterion is normalized in a way that each summand equals 1 when the mean value, respectively, the variance, equals the ideal value. As soon as one of the two variables passes the minimum/maximum threshold value, the corresponding summand becomes negative. In conclusion, the ideal trajectory results in $k_{\mathrm{Q}} = 2$ and decreases steadily with increasing variance s^2 and increasing absolute difference $||\overline{|z(x)|} - 5|$.

Out of the test runs of each vehicle, the runs of the highest quality can be determined by calculating the corresponding quality criteria. These runs are of the highest available reproducibility. Fig. 5 shows the achieved precision of the driver's trajectory during all test runs. The red dashed lines represent the predefined threshold values for the mean value and the variance. On average, the achieved mean value lies below the ideal of 5 m. This is most likely due to the fact that the driver orients himself to the traffic cones on the inner side of each vertex. Factors influencing the variance include different vehicles, human factors, for example, seating positions as well as fatigue and driver experience in handling various vehicles.

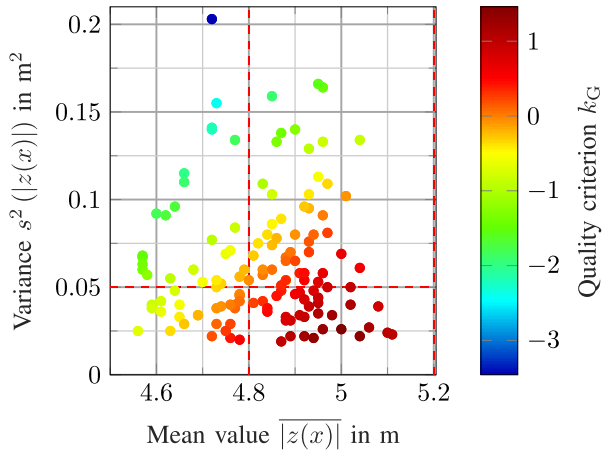


Fig. 5. Achieved precision of driver trajectory, quantified by the mean value of the Hilbert transformation's absolute, and the corresponding variance. The red dashed lines indicate the predefined thresholds for mean value and variance.

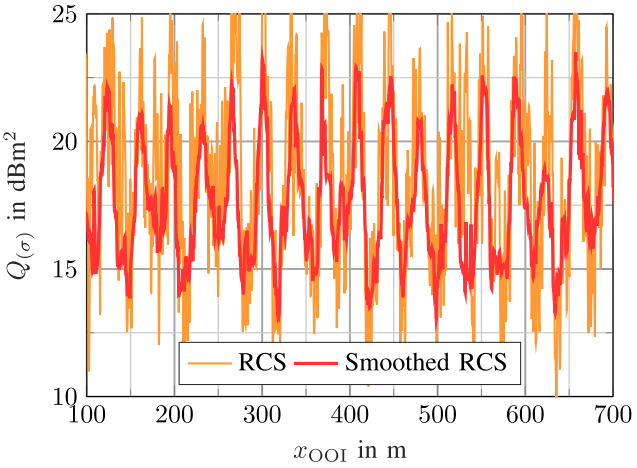


Fig. 6. RCS of Toyota Auris observed while it is driving through the slalom. A periodic pattern is visible after moving median smoothing.

III. Results

Core results, that are available from our measurement campaign, are outlined in this section and are structured according to the research questions in Section I-B. Each plot can be generated for different vehicles and sensors with the MATLAB tool that is released alongside this article. We show detailed measurement results for a Toyota Auris, representing a medium-size vehicle, and extend our findings to the full vehicle dataset.

A. Influence of Aspect Angle

We begin with a typical result of the RCS profile obtained during one slalom run. It is shown in Fig. 6 and reveals a periodical pattern, which is more distinct after smoothing with a moving median filter. In the smoothed signal, we can obtain a mean RCS value of approximately 18 dBm^2 with peak amplitudes of approximately 4 dBm^2. In the raw signal, the RCS shows sporadic peaks that span between 13 and 25 dBm^2. Qualitatively similar patterns can be found for all vehicles by executing the "RCS over x" plot in the MATLAB program.

Fig. 7 shows the RCS again, but now plotted over azimuth and yaw angle. Slalom driving renders as circles in this

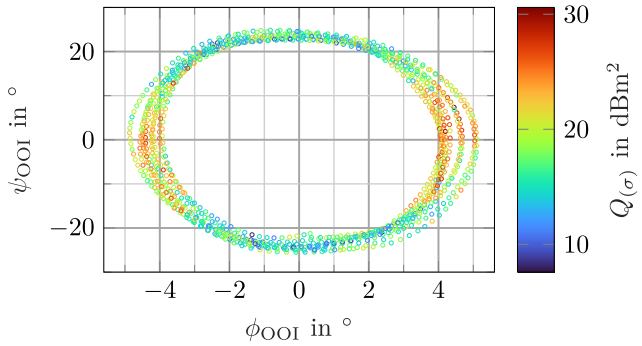


Fig. 7. RCS profile of Toyota Auris over azimuth ϕ_{OOI} and yaw ψ_{OOI}.

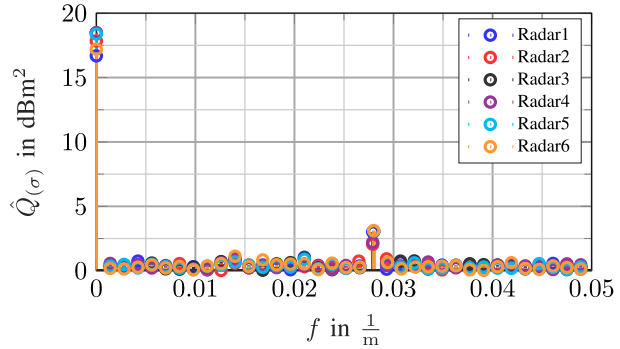


Fig. 8. Spectral analysis of RCS during the same slalom. $\hat{Q}_{(\sigma)}$ denotes the amplitude of the oscillating RCS pattern. Only minor differences in measurements between sensors at the dc-component (zero-frequency) and double harmonic frequency are visible.

representation. The RCS trends to its maximal value at close to zero yaw, that is, at maximal azimuth and minor mitigation is visible through the trial. The minimal values are found at zero azimuth but maximum yaw, that is, as the OOI is crossing through the boresight line during driving through the slalom. From both figures, we can conduct a major influence of the yaw angle on the RCS profile. This plot is generated with the "RCS over yaw and azimuth" function in the MATLAB program.

B. Influence of Mounting Height

To reveal governing frequencies visible in the harmonic course, we obtain a spectral analysis of the signal via fast Fourier transform (FFT) on the RCS over range, which was shown in Fig. 6. A typical result is shown in Fig. 8, which shows the single-sided amplitude spectrum over all radar sensors. This plot is generated with the "FFT over x" plotting function in the MATLAB program. The FFT reveals the signal amplitudes at the dominant frequencies in the signal. At zero frequency, the FFT gives the dc component, which is the amplitude at zero frequency, as the quasi-mean value of the periodic RCS signal. It shows an additional peak at double the natural frequency of the slalom, which is around $0.028(1/\mathrm{m})$. This is due to the orientation and therefore visibility of the OOIs' left- and right-hand sides. The changes in the respective magnitudes indicate the effect of the sensor mounting height and vehicle body shapes. Noise components will render minor spectral components outside these two frequencies.

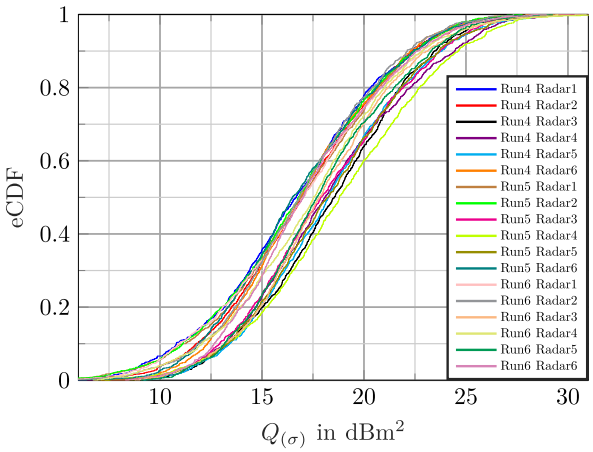


Fig. 9. eCDF plot of Toyota Auris' RCS of all six radar sensors and the three best runs in terms of the Hilbert criteria.

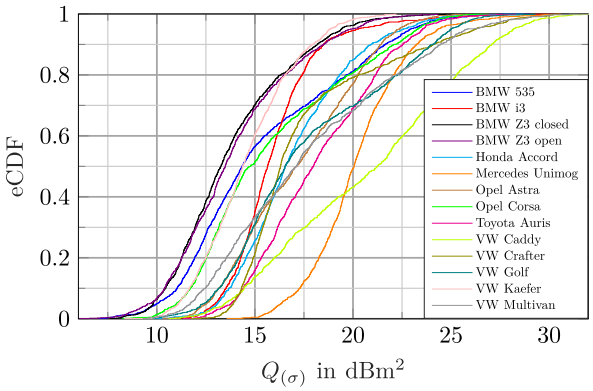


Fig. 10. eCDF plot of all OOI's summed RCS of all radar sensors of the best run in terms of the Hilbert criterion.

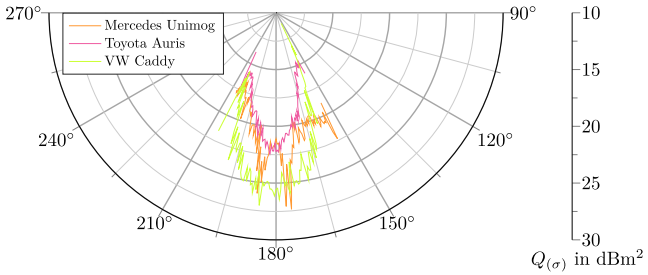


Fig. 11. RCS profiles of Mercedes-Benz Unimog, Toyota Auris, and VW Caddy.

We can only notice a little spread in RCS between sensors, which is below 3 dBm^2 and therefore small compared to its total dynamic range. It is to emphasize that these amplitudes do not represent an absolute RCS value, but the amplitudes of the oscillation of RCS during the slalom. The relation of the amplitude at the harmonic frequency to the amplitude at 0 quantifies its dynamic range. The difference between the two amplitude levels can be understood as the sensitivity of a particular vehicle body to the aspect angle. The higher the amplitude at the harmonic frequency, the higher the range between minimal and maximal observed RCS during the slalom.

So far, we have only discussed the amplitudes obtained by the FFT-based spectral analysis. The RCS measurements are characterized by considerable noise, which leads to significant

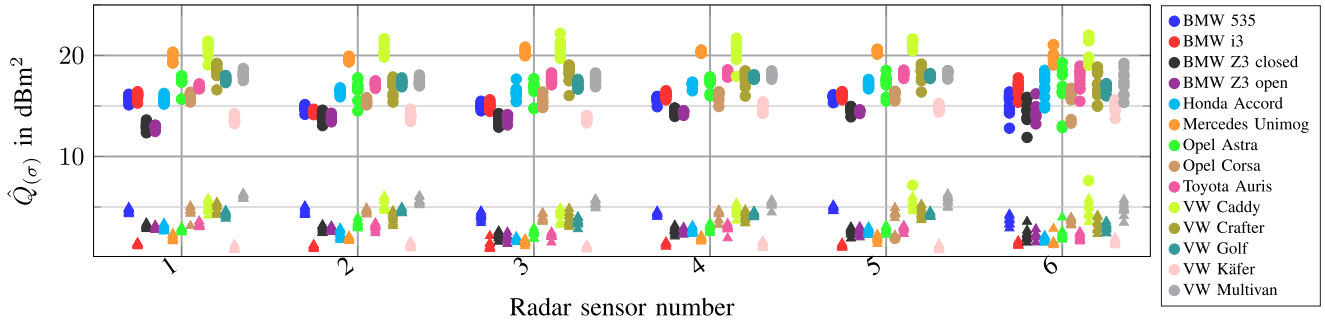


Fig. 12. Spread of spectral amplitudes obtained by FFT over all studied cars and trials. Circles denote zero-frequency amplitude, and triangles indicate amplitude at the double natural frequency. The figure shows all radar sensors and vehicles simultaneously for the best comparison.

outliers with particularly high or low RCS values. To analyze the quantitative distribution, we compute the eCDF, showing the distribution of samples and the frequency of their occurrence. Fig. 9 visualizes the eCDF of the six radar sensors and three best runs regarding the Hilbert criterion of the Toyota Auris. In general, all graphs show almost identical behavior and are lying within a 2-dBm^2 range. From this finding, we conclude that the previously mentioned noise is present in a similar manner in all six radar sensors as well as in all runs. An explanation for the recognizable spread beside the stochastic character of the RCS is the run quality, but also different distances between the Ego and the OOI. The distribution of the RCS is another indication that the influence of the sensor mounting height on the reflectivity characteristics of vehicles is lost in the signal bandwidth of the RCS itself.

C. Influence of Body Shape

For studying the behavior of different vehicle bodies, the spectral FFT analysis is now extended to the full set of vehicles and sensors. Therefore, their spread during multiple test repetitions becomes quantitatively specifiable. This gives objective insights into how well the repeatability of the experiment, that is, how well similar results can be obtained in a repeated trial.

We focus on the amplitudes at the dc component and the doubled natural frequency. Fig. 12 shows the spectral amplitudes obtained by FFT over all studied cars, trials, and mounting heights. The preliminary finding from the Toyota Auris, as presented above, can be well transferred to the other vehicles. While each vehicle differs in amplitude, we see a little spread across sensors and repetitions. Differences between vehicle bodies become more notable when taking the amplitudes at the doubled natural frequency into account. Here, we see that large amplitudes at the dc component are not necessarily followed by large amplitudes at the double natural frequency. Of particular note are the Mercedes-Benz Unimog and the BMW i3, which show strong dc component amplitudes, but only minor amplitudes at the doubled natural frequency. Contradicting examples are given by the VW Multivan and VW Caddy. From this result, we further justify minor influences of the sensor mounting position, which is less than 3 dBm^2 across all studied vehicles. The highest spread is present at radar sensor 6, which is closest to the ground. At the same time, variations between multiple trials remain at the same level.

To view the RCS distributions in addition to the FFT spectral analysis, the eCDF graphs of the different vehicles are visualized in Fig. 10. The eCDFs shows significant differences between all vehicles. No direct correlation between vehicle size and RCS can be seen. This can be especially demonstrated by comparing Honda Accord or VW Multivan and Toyota Auris. Furthermore, the eCDF of the Mercedes-Benz Unimog resembles a log-normal distribution despite its complex structures on the sides. This is also evident for BMW i3, Honda Accord, Opel Astra, Toyota Auris, and VW Käfer. BMW 535, BMW Z3, Opel Corsa, VW Caddy, VW Crafter, VW Golf, and VW Multivan exhibit a kind of kink in the distribution function, which resembles a log-logistic distribution. Therefore, not only the quantitative value of RCS differs, but also the distributions are different between all vehicles. No direct correlation between body size, shape, and material can be identified. Therefore, other radar sensors and mounting positions are the focus of future work.

The FFT analysis and the eCDF do not give information about the noise and sensitivity of the RCS at individual yaw angles. To take these aspects into account, it is recommended to display the RCS in a polar plot. From the findings in Fig. 9, we conclude that on the basis of the Hilbert criterion, an evaluation of the trajectories regarding their comparability is reasonable. Therefore, the different radar sensors as well as the three best runs of the vehicles are aggregated. The summed RCS values from each time step are assigned to the yaw angle ${}_{\mathrm{S}}\psi_{\mathrm{OOI}}$, which is discretized with 0.5°. Subsequently, the median of the RCS is formed for each discrete yaw angle and displayed in a polar plot. Fig. 11 shows RCS profile for all radar sensors and the three best runs for the Unimog, the Auris, and the Caddy. Toyota Auris and VW Caddy show a qualitatively similar RCS character, with the VW showing a significantly higher as well as more brawny shaped pattern. The many add-on parts and the complex structures of the Unimog result in an asymmetrical RCS profile, which is clearly distinguishable from all other vehicles.

Due to the fact that based on the experimental setup, the projected area of the vehicle at ${}_{\mathrm{S}}\phi_{\mathrm{OOI}} \neq 0$ is higher in comparison to a turntable in an anechoic chamber the RCS profile differs from a laboratory profile. Nevertheless, the measured characteristics of the vehicles can be used for validation purposes in sensor simulation.

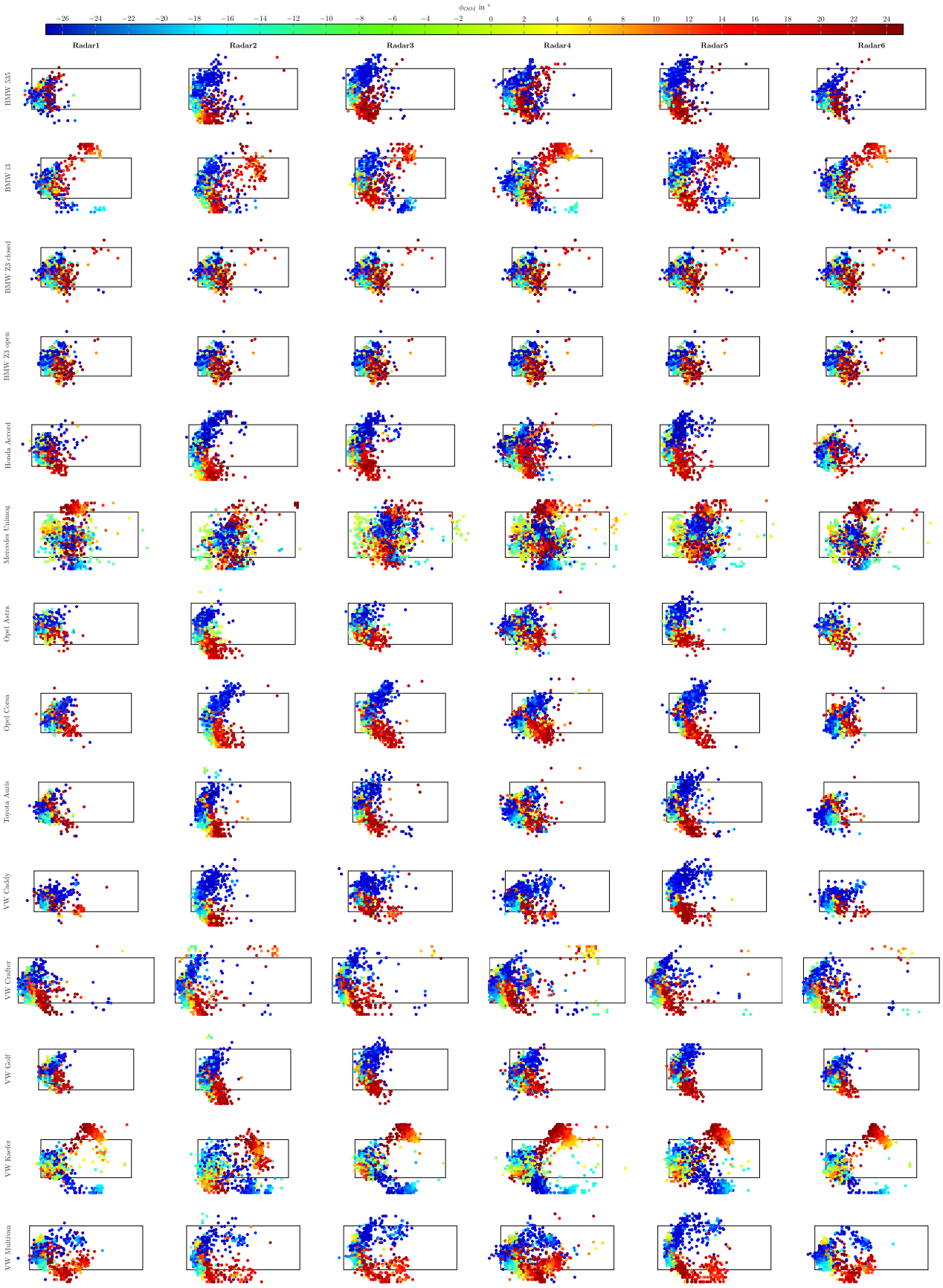


Fig. 13. Location of detections with the highest RCS per measurement cycle around and inside the bounding box for each vehicle and each radar sensor. For evaluation, we took the best run in terms of the Hilbert criterion.

D. Strongest Scatter Location

For a deeper analysis, we consider the distribution of the detections with the highest RCS over the bounding box of the different vehicles. Fig. 13 shows the detection location of the best run of each vehicle and each radar sensor. Therefore, only the detection with the highest RCS is visualized to gain insights into the main scattering parts of each vehicle. The yaw angle difference ${}_S\psi_{OOI}$ between Ego and OOI is coded as color identifier. Additionally, the bounding boxes of the different vehicles are represented in the form of black rectangles. The detections outside of the bounding box can be explained by timing effects due to different measurement frequencies as well as the resolution of the radar sensor itself. In general, the detections are at the rear center of the vehicles, as already shown in [9]. Slight differences between radar 1/6 and radar 2/3/4/5 are visible for BMW 535, Honda Accord, Opel Corsa, Toyota Auris, and VW Caddy. Special features in the distribution can be seen in the BMW i3, which is equipped with a carbon body, the Unimog, which has various add-on parts at the rear and side, the Crafter, which shows detections on the wheel arches at the front, and the Käfer, which has CCR-shaped side sills. Therefore, we conclude that the position distribution of the detections depends on the body shape as well as on the yaw angle between the sensor and OOI.

IV. Derived Findings for Sensor Simulation

We summarize the core results of our research from various points of view as follows. Here, we focus on statements that are of immediate value to the sensor simulation research community.

A. Analytical RCS Models

Sensor modeling must take into account the fact that even a reference measurement can never be made with infinite accuracy. Uncertainties remain, which are difficult to quantify but remain visible as nondeterministic and stochastic behavior. This becomes particularly clear in the deviations of the RCS measurements between the tests. Our results stress the stochastic nature of the RCS characteristics of vehicles. For each vehicle geometry, we find differences at the microscopic level that justify deviations in the RCS. Due to a large number of possible geometries of today's and future cars, we consider it difficult to maintain vehicle-specific analytical RCS models. In object-based sensor modeling, a generative model can be derived from our data that returns an RCS value for a given aspect angle. For other modeling approaches, our work gives reference measurements that can be used for simulation model falsification.

B. Spatial Distribution of Detections

We identified clear differences in the spatial distribution of detections around the vehicles. The Volkswagen Käfer is of particular note: Its body design at the bottom of the front door forms a CCR-shape and leads to a concentration of detections around this area.

These results, however, pose a challenge for traditional scattering center models. They assume "fixed" areas of significant scatter, while our results show that these areas differ significantly between vehicle shapes [6]. This also poses a challenge for object detection algorithms, relying on detection clustering. The centerpoint of clusters varies between vehicles, resulting in a systematic position offset between the true and estimated position of the vehicle.

C. Sensor Model Validation

Our results and dataset can serve as a baseline for sensor model validation. The RCS behavior obtained during reenacting the slalom drive in the virtual world should match its eCDF distribution. In "physical," or "reflection-based" sensor simulation, the detailing of 3-D models is of high importance. These models must convey material descriptions and meshing must be fine enough to preserve geometrical details. By comparing the spatial distribution of the detections, the appropriateness of 3-D models for radar simulation becomes judgeable.

Sensor models that process object information as input data do benefit from our work by having access to a publicly available dataset that allows model parameterization.

D. Transferring Results to Other Radar Sensors

The gathered RCS data and detection distribution are only valid for Continental ARS408 sensors. However, when gathering the same data with other radar sensors, the RCS profiles presented in Fig. 11 are expected to look similar in their qualitative form and the stochastic RCS behavior will have a similar dynamic range. It is also expected that the detection focal point of the spatial distribution will be identical for the various vehicles. By using radar sensors with higher resolutions, the position of the (strongest) detection can be determined more precisely. Still, our results are well aligned with previous research, such as [9], which utilizes the SAR method. In future research, radar sensors with elevation measurements are expected to lead to further insights into the 3-D location of detections.

V. Conclusion

The aim of the work was to investigate the aspect angle dependence of the RCS. For this purpose, an experimental setup was proposed in which the vehicle under investigation drives a slalom while being observed by several radar sensors.

The design of our slalom aims at minimizing uncertainties in radar perception introduced by human drivers. Hereby, we could isolate the noise that is typically present in radar measurements to the radar measurement principle. At the same time, additional influences, such as unsteady driving through a slalom course by human drivers, are isolated.

From the measurement results, we gained knowledge about the stochastic behavior of the RCS and the local distribution of the detections.

For further work, we recommend examining the SNR value in addition to the RCS value. While this will show proportionality to the RCS value, it has the advantage of taking into account the performance of the radar. Thus, the SNR value

is easier to determine than the RCS because, for example, the antenna pattern and frequency filters have no influence. When setting up the test again, we recommend considering an additional radar, which is installed with a horizontal offset. This can be used to draw conclusions about the transferability between horizontal shoring positions. In our work, we have shown that the vertical shoring position has a negligible part. The evaluation methodology can be applied straightforwardly to lidar sensors as well. In this way, similar investigations of reflectivity can also be performed for lidar sensors.

Furthermore, by providing ground truth and vehicle metadata, our dataset also provides a scientific basis for defining detection-level metrics of sensor data. This is the subject of ongoing research for the evaluation of sensors and simulations and also in the interest of industrial standardization processes.

Acknowledgment

The authors would like to thank the owners of the used vehicles for making them available for this study. They also gratefully acknowledge Prof. Hermann Winner for his guidance and advice, and Prof. Steven Peters, Clemens Linnhoff, and Philipp Rosenberger for proofreading.

References

[1] R. B. Dybdal, "Radar cross section measurements," *Proc. IEEE*, vol. 75, no. 4, pp. 498–516, Apr. 1987.

[2] H. Winner, "Automotive RADAR," in *Handbook of Driver Assistance Systems*. Cham, Switzerland: Springer, Dec. 2015, pp. 325–403, doi: 10.1007/978-3-319-12352-3_17.

[3] K. Werber et al., "Automotive radar gridmap representations," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Apr. 2015, pp. 1–4.

[4] K. Karlsson, H. Toss, J. Lang, F. Costagliola, T. Zheng, and E. Marel, "Hifi radar target: High fidelity soft targets and radar simulation for more efficient testing (real and virtual)," Res. Inst. Sweden (RISE), Saf. Transp., Electron., Tech. Rep., 2019. [Online]. Available: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1264147&dswid=-5305

[5] E. B. Kamel, A. Peden, and P. Pajusco, "RCS modeling and measurements for automotive radar applications in the W band," in *Proc. 11th Eur. Conf. Antennas Propag. (EUCAP)*, Mar. 2017, pp. 2445–2449.

[6] P. Swerling, "Probability of detection for fluctuating targets," *IRE Trans. Inf. Theory*, vol. 6, no. 2, pp. 269–308, Apr. 1960, doi: 10.1109/tit.1960.1057561.

[7] S. B. J. Gowdu, A. Schwind, R. Stephan, and M. A. Hein, "Monostatic RCS measurements of a passenger car mock-up at 77 GHz frequency in virtual environment," in *Proc. 49th Eur. Microw. Conf. (EuMC)*, Oct. 2019, pp. 996–999. [Online]. Available: https://ieeexplore.ieee.org/document/8904569

[8] I. Matsunami, R. Nakamura, and A. Kajiwara, "RCS measurements for vehicles and pedestrian at 26 and 79GHz," in *Proc. 6th Int. Conf. Signal Process. Commun. Syst.*, Dec. 2012, pp. 1–4, doi: 10.1109/icspcs.2012.6508004.

[9] M. Andres, P. Feil, and W. Menzel, "3D-scattering center detection of automotive targets using 77 GHz UWB radar sensors," in *Proc. 6th Eur. Conf. Antennas Propag. (EUCAP)*, Mar. 2012, pp. 3692–3693, doi: 10.1109/eucap.2012.6206580.

[10] S. Abadpour, A. Diewald, M. Pauli, and T. Zwick, "Extraction of scattering centers using a 77 GHz FMCW radar," in *Proc. 12th German Microw. Conf. (GeMiC)*, Mar. 2019, pp. 79–82, doi: 10.23919/gemic.2019.8698144.

[11] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+1D radar in the view-of-delft dataset," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4961–4968, Apr. 2022.

[12] O. Schumann et al., "RadarScenes: A real-world radar point cloud data set for automotive applications," in *Proc. IEEE 24th Int. Conf. Inf. Fusion (FUSION)*, Nov. 2021, pp. 1–8, doi: 10.23919/fusion49465.2021.9627037.

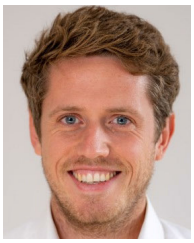
[13] H. Caesar et al., "NuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11621–11631.

[14] M. F. Holder, "Synthetic generation of radar sensor data for virtual validation of autonomous driving," Ph.D. dissertation, Inst. Automot. Eng. Darmstadt, Technische Universität Darmstadt, Darmstadt, Germany, 2021. [Online]. Available: http://tuprints.ulb.tu-darmstadt.de/17545/

[15] C. E. Services. *Data Sheet Continental ARS 408-21 Premium Long Range Radar Sensor 77 GHz*. Accessed: Jan. 13, 2023. [Online]. Available: https://www.continental-automotive.com/getattachment/8e4678e1-9358-48e1-8d5b-a0c2de942edb/ARS408-21_Datenblatt_de_170707_V07.pdf.pdf

[16] Mercedes-Benz. *GLS SUV Specification*. Accessed: Sep. 27, 2022. [Online]. Available: https://www.mercedes-benz.co.za/passengercars/mercedes-benz-cars/models/gls/suv-x167/specifications/dimensions/_jcr_content/highlightcontainer/par/highlighttile_1222582405.MQ6.0.2x.20190627112735.jpeg

[17] *Richtlinien Für die Anlage von Landstraßen*, Forschungsgesellschaft Für Straßen und Verkehrswesen e.V., Bonn, Germany, 2012. [Online]. Available: https://www.fgsv-verlag.de/pub/media/pdf/201.r.16052013.pdf



Lukas Elster was born in Fulda, Germany, in 1993. He received the B.Sc. and M.Sc. degrees in mechanical engineering from TU Darmstadt, Darmstadt, Germany, in 2016 and 2020, respectively.

Since 2020, he has been working as a Research Associate with the Institute of Automotive Engineering, TU Darmstadt. His research interests include the effects of radar and lidar sensors in complex surroundings.

Martin F. Holder was born in Heilbronn, Germany, in 1990. He received the Dr.-Ing. degree in radar sensor simulation from TU Darmstadt, Darmstadt, Germany, in 2021, the master's degree in mechanical engineering from TU Darmstadt, in 2016, and the master's degree in systems, control and mechatronics from Chalmers University, Gothenburg, Sweden.

Manuel Rapp was born in Saarbrücken, Germany, in 2000. He received the B.Sc. degree in mechanical engineering from TU Darmstadt, Darmstadt, Germany, in 2021.

Since then, he has been working as a Research Assistant with the Institute of Automotive Engineering, TU Darmstadt.

C Paper III:

Making Automotive Radar Sensor Validation Measurements Comparable

applied sciences

MDPI

Article

Making Automotive Radar Sensor Validation Measurements Comparable

Lukas Elster *, Jan Philipp Staab and Steven Peters

Institute of Automotive Engineering, Technical University of Darmstadt, Otto-Berndt-Straße 2, 64287 Darmstadt, Germany; jan.staab@tu-darmstadt.de (J.P.S.); steven.peters@tu-darmstadt.de (S.P.);
* Correspondence: lukas.elster@tu-darmstadt.de

Abstract: Virtual validation of radar sensor models is becoming increasingly important for the safety validation of Light Detection and Rangings (lidars). Therefore, methods for quantitative comparison of radar measurements in the context of model validation need to be developed. This paper presents a novel methodology for accessing and quantifying validation measurements of radar sensor models. This method uses Light Detection and Rangings (lidars) and the so-called Double Validation Metric (DVM) to effectively quantify deviations between distributions. By applying this metric, the study measures the reproducibility and repeatability of radar sensor measurements. Different interfaces and different levels of detail are investigated. By comparing the radar signals from real-world experiments where different objects are present, valuable insights are gained into the performance of the sensor. In particular, the research extends to assessing the impact of varying rain intensities on the measurement results, providing a comprehensive understanding of the sensor's behavior under these conditions. This holistic approach significantly advances the evaluation of radar sensor capabilities and enables the quantification of the maximum required quality of radar simulation models.

Keywords: automotive radar; validation measurements; virtual validation; sensor model validation; Double Validation Metric

Citation: Elster, L.; Staab, J.P.; Peters, S. Making Automotive Radar Sensor Validation Measurements Comparable. *Appl. Sci.* **2023**, *13*, 1405. https://doi.org/10.3390/app132011405

Academic Editors: Atsushi Mase, Khan Muhammad Umar, Matthias A. Hein, Christian Bornkessel and Nosherwan Shoaib

Received: 28 August 2023
Revised: 6 October 2023
Accepted: 14 October 2023
Published: 17 October 2023

1. Introduction

Automated driving is one of the biggest challenges facing the automotive industry. Advantages are expected in the areas of connection of village regions, better utilization of existing resources with accompanying minimization of vehicle downtimes, and additional technological development. However, any benefits depend on the safety of such ADFs being proven. Until it is reliable that automation is safer than humans, such systems will not be deployed in large numbers and in all areas of road transport [1]. For a variety of reasons, it is currently not possible to prove safety for such an ADF. One approach that addresses this problem is scenario-based testing [1]. These tests will be performed alongside real test drives in different simulation environments, such as software-in-the-loop, hardware-in-the-loop, or vehicle-in-the-loop. In this way, safety-critical scenarios can be performed for the vehicle and involved agents.

Along with path planning and control, the perception of the environment is an elementary component of ADFs. Therefore, in addition to vehicle dynamics and environment models, sensor models are required to represent the entire ADF pipeline in simulation. However, it is fundamental for the simulation validation process to objectively validate the models used. Both the simulation models and real-world validation measurements play a crucial role in this process. Although the first approaches to validation measurements exist, the comparability, reproducibility, and repeatability of these measurements are often problematic, especially for Radio Detection and Ranging (radar) sensors [2] (p. 68), [3] (p. 84).

In addition to radar, cameras, Light Detection and Ranging (lidar), and ultrasonic sensors are established perception sensors in production vehicles [4]. Radar sensors determine

position information, intensities, and speeds by means of frequency modulation, which is determined by the Doppler effect. There are different processing levels in the sensor that can be used for radar sensor model validation measurements. Radar cuboid as a level after a fast Fourier transform (FFT) algorithm in the radar processing chain and detections as a result of the application of dynamic thresholding algorithms [5], which are of interest for this publication. The object list will not be explored further in the course of the paper, as the identification of effects in further processed interfaces increases in complexity.

Therefore, this paper presents a method to compare radar measurements at the radar cuboid and detection levels. The application of this method is intended to provide objective, comparable, and quantifiable radar sensor validation measurements for radar sensor models as part of the safety argumentation of ADFs in the context of scenario-based testing.

1.1. Related Work

There are several measurements in the literature that intend to validate radar models. The related work is separated into radar measurement campaigns as well as lidar and general model validation measurement campaigns.

In order to show the lack of state-of-the-art in this field, the first step is to define the terms repeatability and reproducibility. We define repeatability as the variation of a measurement due to the radar sensor itself. Reproducibility is defined as the variation in the measurement data resulting from the measurement system (setup, sensor, operator) using the same sensor under the same conditions.

1.1.1. Radar

Schaerman et al. transfer real scenarios with a high accuracy of 5 cm standard deviation into the simulation to validate raw data models. However, the repeatability and reproducibility of the measurements are not directly considered [6]. Abadpour presents radar measurements of vulnerable road users and their back-scattering behavior. Nevertheless, the number of repetitions of the measurements is only one trial, and reproducibility is not further addressed [7]. Schneider analyzes different objects and their back-scattering behavior in radar sensors. He compares the characteristics using histograms to investigate the corresponding distributions in the sensor. A deeper investigation about repeatability and reproducibility on radar cuboids and detection levels is missing in his analysis [8]. Eder analyzes an overtaking maneuver with 100 repetitions and finds that the trajectory is almost identical, but the fluctuation and therefore a lack in the repeatability of the measurement result in the detection level hinder a deeper analysis [3]. Holder et al. use EDFs to study the detection distribution of occluded and unoccluded objects. The number of detections on an object is also analyzed using EDFs. Here, three repetitions of the measurement are performed, but a deeper analysis of the repeatability is not considered further [5]. Magosi et al. also aggregate the detections over the bounding box of a scenario but do not consider test repetitions. Only a comparison between simulation and measurement is made [9]. Aust et al. use 10 repetitions of a scenario by driving the radar sensor towards a static object. They aggregate the detections over the bounding box of the target object and examine the distribution of detections using the probability density function (PDF). However, they do not directly compare the measurements with each other [10]. Buddapaggari et al. also use resimulation of measured trajectories to compare detections in over-the-air/vehicle-in-the-loop test beds with real radar data. The number of experimental replicates is also limited to one, which means that measurement repeatability is not further considered [11]. Ngo uses different scenarios to determine the simulation-reality gap he defines for detections and object lists. For each scenario, only one trial repetition is considered [12] (pp. 66, 84). Holder analyzes different scenarios, running three trials for each scenario. Various variations are shown, but in the context of repeatability and reproducibility, these are not discussed further in terms of quantitative deviations and the identification of measurement outliers [2].

1.1.2. Lidar and Other Domains

In addition to radar sensors, different validation measurements of lidar sensors are also introduced in different publications. Rosenberger et al. use a relatively complex validation scenario with multiple objects but also neglect the repeatability of the experiments in their analysis [13]. Linnhoff et al. use a meteorological station to analyze lidar data based on real weather conditions. However, due to the real conditions, there is no reproducibility of the measurements, despite a high number of experimental replicates [14]. In DIN SAE SPEC 91471:2023-05 [15], different measurement setups are presented, which serve for the evaluation of lidar performance. The technical description lacks information about the number of repetitions and an evaluation metric to compare the different results with each other. Viehof uses 15 measurements in the validation of vehicle dynamics models and explicitly emphasizes the need to ensure the quality of the measurement data [16]. However, no specific procedure is given for evaluating the repeatability of the measurements and the deviation quantification.

1.2. Research Questions

The following research questions arise from the validation measurement campaigns in the literature for radar sensors.

Research Question 1: How reproducible are radar measurements and identical measurement setups?

Identical measurements in this case refer to a comparison of unchanged measurement setups in less than 10 min consecutively where the sensor is not switched off. Furthermore, it is investigated how dismantling and reconstruction of the measurement setup behave on the radar measurements at constant position with precise markings on the test track.

Research Question 2: How do different objects and object positions affect radar measurements?

Different objects here denote a corner cube reflector (CCR), which is an ideal point target, and extended targets, which differ in shape and size (Volvo XC90, Mercedes Vito). Furthermore, the poses of the vehicles to the sensor are varied to analyze this influence.

Research Question 3: How do environmental conditions such as rain affect radar validation measurements?

Due to the stochastic behavior of rain, it is challenging to quantify the effect and make different rain intensities comparable.

2. Evaluation Methodology

To answer the research questions, a methodology based on the so-called DVM introduced by Rosenberger [17] (pp. 118,119) is used. The corresponding process is visually illustrated in Figure 1. In the first step, the measurement data for the different experiments is collected. We gather multiple measurements per experiment setup. After pre-processing and position-based filtering of the measurement data, the number of resulting measurement data points is compared between the respective measurements or experiments. We name this filtered area region of interest (ROI). The comparison of the overall number is due to the condition that this deviation must be small for the DVM to be used [17] (p. 103). The DVM is based on a comparison of the EDFs of the measured data and, therefore, the underlying distribution. Rosenberger's approach uses the area validation metric of Roy et al. [18] (p. 307) and extends it by a second factor to evaluate the EDFs. In contrast to Rosenberger, we do not summarize the individual distribution functions of the measurements, but evaluate the differences of the EDFs individually. Therefore, the area between all the resulting distribution functions is determined, which is a measure for the deviation of the mean value denoted as d_{bias} of the measurements and experiments. Therefore, the mathematical description for d_{bias} results with $\boldsymbol{\mathcal{F}}$ denoted as the first EDF, $\widetilde{\boldsymbol{\mathcal{F}}}$ denoted as the second EDF, d^+ denoted as the area between $\widetilde{\boldsymbol{\mathcal{F}}}$ and $\boldsymbol{\mathcal{F}}$, where $\widetilde{\boldsymbol{\mathcal{F}}} \geq \boldsymbol{\mathcal{F}}$ and d^- denoted as the area between $\widetilde{\boldsymbol{\mathcal{F}}}$ and $\boldsymbol{\mathcal{F}}$, where $\widetilde{\boldsymbol{\mathcal{F}}} < \boldsymbol{\mathcal{F}}$, in

$$d_{\text{bias}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}) = d^- - d^+. \tag{1}$$

If subsequently the EDFs are shifted by d_{bias} and the resulting area between the corrected EDFs $\widetilde{\boldsymbol{\mathcal{F}}}_{\text{c}}$ and the first EDF $\boldsymbol{\mathcal{F}}$ is calculated, a metric d_{CAVM} for the shape deviation of the distribution functions is obtained, giving information about the scattering difference [17] (pp. 116–119).

Therefore, the following mathematical expression is obtained as

$$d_{\text{CAVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}) = d_{\text{AVM}}(\boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\mathcal{F}}}_{\text{c}}) = d_{\text{c}}^{-} + d_{\text{c}}^{+}. \tag{2}$$

These two parameters give information about the similarity of the different distribution functions in the unit of the analyzed measurand. In a final step, the distribution of d_{bias} and d_{CAVM} is visualized and analyzed via box plots.

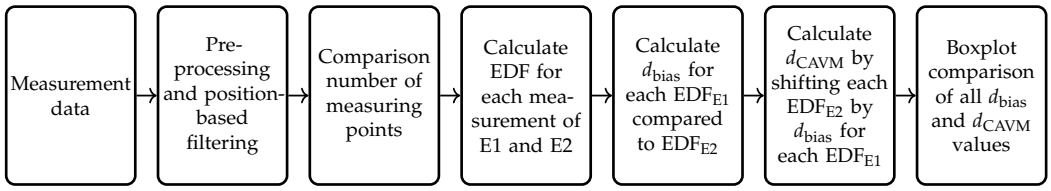


Figure 1. Used Methodology for the comparison of different experiment setups (E1 and E2), measurements and input data on radar cuboid as well as detection level.

Five different evaluation levels are considered input data for the process:

1. All range azimuth bins with the Doppler component 0 of the radar cuboid → whole radar cuboid (WRC);
2. All detections in the dimensions distance, azimuth, and Radar Cross Section (RCS) → all detections (WD);
3. A region of interest, which only includes the bins where the object is present in the measurement, on radar cuboid level with the Doppler component 0 → ROI of the radar cuboid (ROI RC);
4. A region of interest, which only includes detections where the object is present in the measurement, on detection level → ROI of the detecions (ROI D);
5. Each range azimuth bin combination separately on the radar cuboid level → each bin of the radar cuboid (EB RC);

The different evaluation levels are selected according to their relative deviation of the number of data points or detections of all measurements for the respective experiment. Furthermore, the additional information content plays a role in whether the evaluation is performed. On this basis, the evaluation scheme in Table 1 is derived, and the evaluation is performed in the following Section 4.

Table 1. Overview of the application of the different evaluation levels. In the first column, the compared experiments and objects are listed, which are explained in detail in Section 3.

Experiment	WRC	WD	ROI RC	ROI D	EB RC
CCR\ CCR	X	X	X	X	X
Day 1\ Day 2	X				X
CCR\ XC90	X				X
XC90\ XC90 rotated	X		X		X
XC90\ Vito	X	X	X		X
XC90\ XC90 rain	X				X

3. Experimental Setup

The validation experiments are performed at the August-Euler airfield in Griesheim. Except for the experiments on Day 1 and Day 2, all experiments are performed with the same sensor position on the airfield. The setup and the orientation of the sensor are very

crucial for the analysis of the data because, with effort, the epistemic uncertainties can be reduced and thus increase the comparability between the experiments. The measurement setup consists of a radar sensor mounted on a sensor rack and an object positioned in front of it. The radar sensor is a series of active cruise control (ACC) radars with a maximum range of approx. 200 m. The field of view (FoV) is approx. 12.8°. For each measurement, a measurement time of 60 s is reproduced by the measurement computer. However, due to measurement data failures in the recording, this number may fluctuate.

The frame on which the radar sensor is mounted can be leveled in all directions by means of four adjustable feet. In addition, the height, azimuth angle, and elevation angle of the sensor can be adjusted. The elevation angle is set to 0. The orientation of the azimuth angle of the sensor is set by a centered far point target, and the orientation for the positioning of the point target and the sensor is given by the driving stiff mark. The height is based on the mounting position of the sensor in the vehicle. Therefore, a height of 70 cm is chosen. For the objects, a distinction is made between a point target in the form of a CCR and an extended object in the form of two different vehicles. The vehicles are a Mercedes-Benz Vito (van, production year 2020) and a Volvo XC90 (sport utility vehicle, production year 2019). The chosen objects are exemplars, but the developed methodology is applicable to any kind of object.

The Day 1 and Day 2 experiments do not contain an object. This is to quantify the effects of setup and dismantling, as well as to study the static environment itself. The exact setup of the different experiments can be found in Table 2.

Table 2. Different parameters in the experimental setups. In the first column the compared objects are listed with the corresponding parameters.

Experiment	r in m	ϕ in °	ψ in °
CCR\CCR	29.56\29.56	0\0	0\0
Day 1\Day 2	-	-	-
CCR\XC90	29.56\29.50	0\0	0\0.5
XC90\XC90 rotated	29.50\29.68	0\0.1	0.5\13.7
XC90\Vito	29.50\29.59	0\0	0.5\−0.2
XC90 \XC90 rain	48.60\48.60	8.1\8.1	−1.1\−1.1

The experimental setup follows the scheme shown in Figure 2. The center of the sensor defines the origin of the drawn S-coordinate system. The object positions are defined relative to this coordinate system. For the description of the experimental setups, a two-dimensional experimental space is chosen, which can be defined by the polar coordinates, distance (r), and azimuth angle (ϕ) of the sensor. The reason for this is that the FoV of the sensor is also divided into polar coordinates. As the evaluation is also done at the radar cuboid level, a logical positioning with respect to the bins makes sense. The bins result from the resolution of the sensor in range and azimuth. The real positions of all structures are measured during the experiments with an RTK-based global navigation satellite system (GNSS) antenna. The position of the CCR as a point target is thus described by the two parameters r and ϕ. Due to the size of the vehicles themselves, the position of the vehicle must also be described by the rotation around the vertical axis ψ. For the vehicles, the wheel contact points are measured, and the center of the vehicle and the rotation ψ are calculated accordingly. The direction of rotation of the angles ϕ and ψ is defined in a mathematically positive sense and takes the x-axis of the sensor as a reference.

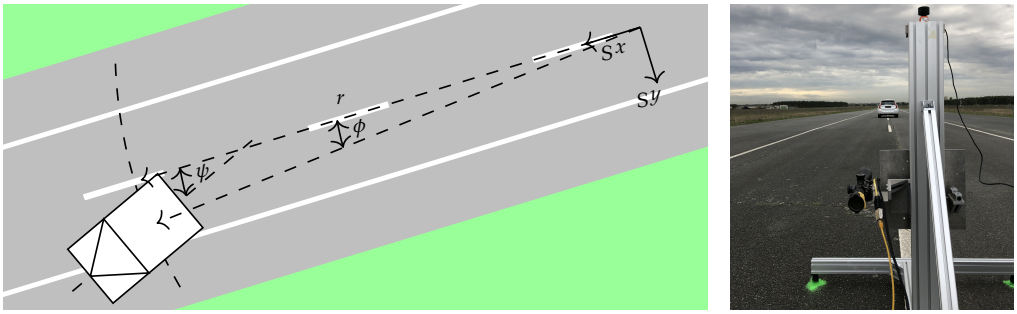


Figure 2. On the left side a sketch of the experimental setup is shown. The Cartesian sensor coordinate system origin is denoted as ${}_{S}x$ and ${}_{S}y$. The parameters for the different experiments are the polar coordinates distance r and azimuth angle ϕ. The yaw angle of the object is defined as ψ. On the right side a picture of the measurement setup of the XC90 on the August-Euler airfield is visualized.

4. Evaluation

This chapter contains the evaluation results based on the method proposed in Section 2. The chapter is structured as subsections based on the research question's order.

4.1. RQ 1: Repeatability of Measurements and Reproducibility of Measurement Setups

In the following subsections, five measurements of a CCR are compared with each other to check repeatability (Section 4.1.1). Afterwards, reproducibility is considered on the basis of two sets of eight individual measurements, where the measurements differ in that the sensor is set up and taken down on Day 1. On the following Day 2, the sensor is set up again in the same weather conditions and in as similar a position as possible using markings on the ground (Section 4.1.2).

4.1.1. Comparison of CCR Measurements

In the first step, the number distribution of the five measurements on the different evaluation levels is compared. Figure 3 shows the deviation of the data points. The number of bins from the radar cuboid is nearly identical, with a small deviation of 2 timeframes, corresponding to a percentage deviation of 2.5 ‰. Since the radar cuboid is also used for the evaluation of the ROI and each bin, and it is the same object class for both experiments, the deviation is the same. The number of detections for the whole field of view varies more significantly compared with the data points in the radar cuboid over the different measurements, with a maximum deviation of 11.9 %. Nevertheless, an evaluation of the detections is carried out in the sense of the methodology as an example. For the ROI at the detection level, however, the relative deviations are small, with a maximum of 1.2 ‰.

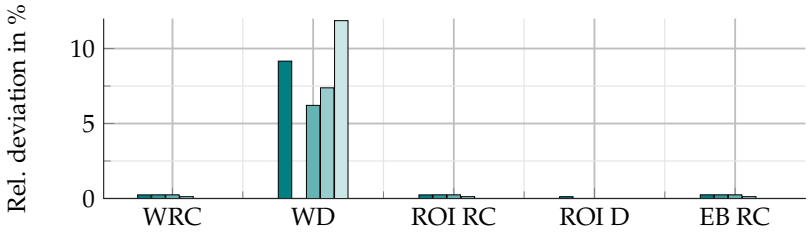


Figure 3. Plot of the relative deviation for the number of the data points or detections of all measurements of the CCR experiment. The evaluation is shown for the WRC, WD, the ROI RC and the ROI D, as well as EB RC. The gradations of the respective color delineate the different measurements.

Whole Radar Cuboid

Figure 4 shows the different EDFs, the box plots for d_{bias} and d_{CAVM}, as well as the number of data points in the different colored distributions. Since the EDFs are very close to each other, it appears in the representation as if only one EDF is shown, whereas the

opposite is the case. Due to this closeness, a visualization of the by d_{bias} shifted EDF is omitted.

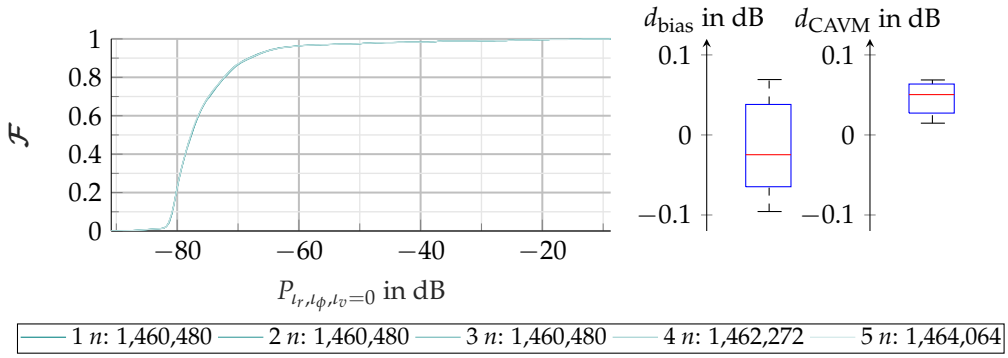


Figure 4. EDFs of all five measurements of the CCR on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} is shown.

Both the EDFs and the by d_{bias} corrected EDFs are close to each other, indicating a high degree of agreement between the measured and examined data. The statistical distribution of d_{bias} and d_{CAVM} can be seen in the box plot on the right-hand side of Figure 4. There are only small differences between the five measurements for the two evaluation criteria. Compared to d_{bias}, d_{CAVM} shows a smaller scatter. However, the median of d_{bias} is lower than that of d_{CAVM}. The values obtained show that the spread for d_{bias} is 0.16 dB, while for d_{CAVM} it is only 0.05 dB. In summary, it can be concluded from the analyses carried out that the measurements show high repeatability with respect to the whole radar cuboid. There are no outliers, which means that several successive radar measurements with this sensor can be considered almost identical based on the analysis of the whole radar cuboid.

Whole Detections

In Figure 5 the EDFs and box plots for distance, azimuth, and RCS are visualized. It can be seen that the EDFs for distance are close to each other, but a spread is visible. At a distance of 29.65 m, there is an accumulation of detections, which shows the influence of the CCR. In general, the detections are distributed over a distance of up to 147 m. Close to the sensor, detections are located on the road surface, and with higher distances, the number of detections increases at the transition between vegetation and the road. Furthermore, in the vegetation itself, multiple detections are present. A decrease in the number of detections above 110 m indicates a lower signal-to-noise ratio (SNR), making detections random. Above 147 m, the SNR is so low that there are no more detections in the environment. The median of all d_{bias} is similar to d_{CAVM} in absolute terms. However, the scatter of the d_{bias} values is larger compared with d_{CAVM}. In addition to the very different number of detections and the large scatter in the metrics, the repeatability of the measurement results for the distance of all detections is very low. The influence of the environment is too large to draw conclusions from the distance evaluation.

As with distance, the azimuth distribution is affected by the environment. Between $-5°$ and $5°$ the CCR is notable with the step at 0. The visible shoulder area next to 0 indicates a few detections due to the road. The second step close to 0 at negative azimuth angles can be associated with detections at close-range as well as at twice the distance between the CCR and the sensor. The median of d_{bias} is about $0.5°$, which is a relatively high value in relation to the sensor's FoV, indicating, in addition to the scatter, a low repeatability of the measurements. The scatter deviation of d_{CAVM} confirms this impression.

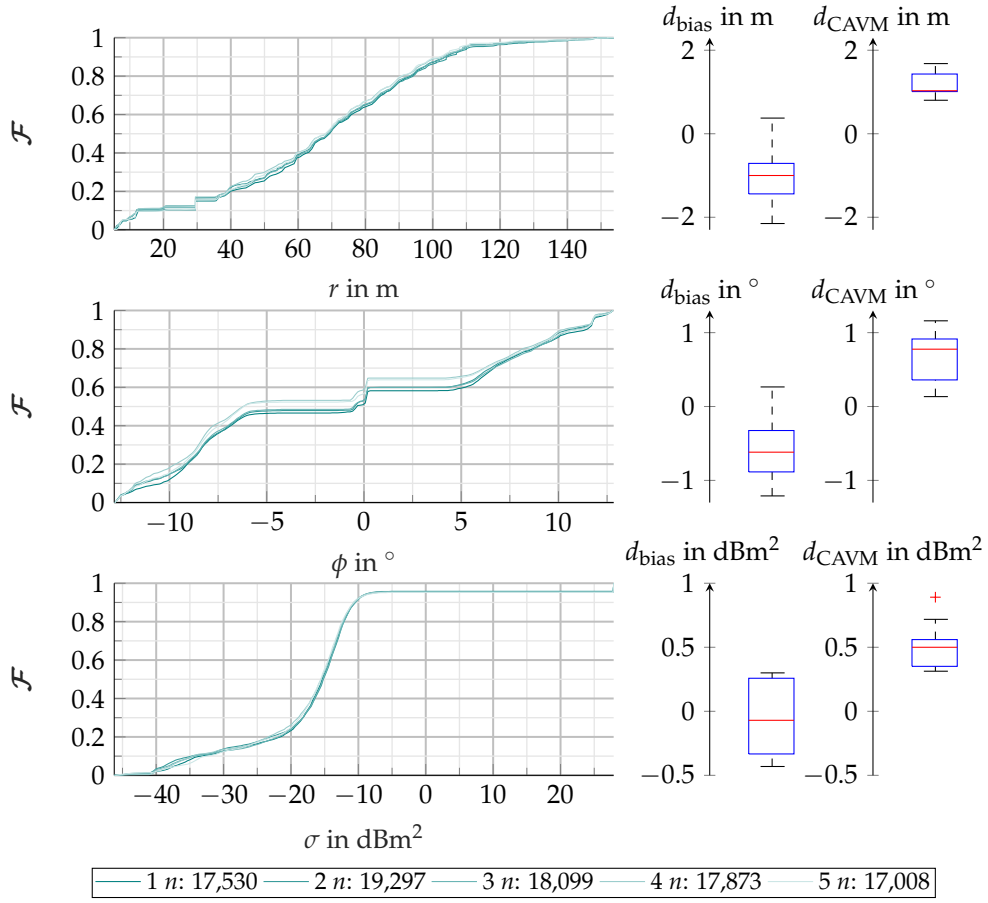


Figure 5. EDFs of all five measurements of the CCR on the detection level. In the first row the distance distribution, in the second row the azimuth distribution and in the third row the RCS distribution with the corresponding box plot for d_{bias} and d_{CAVM} are shown. The aggregated number of detections is listed below the diagrams.

The CCR is also evident in the RCS EDFs. 5 % of the detections are due to the point target with an approximate 27 dBm^2. This indicates high repeatability with respect to the reflectivity characteristics. However, especially for detections at close-range due to asphalt and vegetation, different distributions show the stochastic behavior of the majority of the detections. As before, d_{bias} and d_{CAVM} are difficult to interpret because there is too much randomness in the way the detections are generated by the environment. In particular, there is an outlier in d_{CAVM}, which is due to the comparison of measurements 2 and 5. Here, the number of detections is the most spread over all measurements.

From the evaluation of all detections, it can be concluded that a comparison of all detections is not useful for the validation of this specific radar sensor. The environment has too much influence on the detection distribution. This evaluation experiment can be used in future validation studies to assess whether an evaluation of all detections is useful in the studied environment in the context of validation measurements.

ROI Radar Cuboid

Figure 6 shows the same structure as in the previous evaluations. However, in this case, only the bin of the radar cuboid and therefore the power where the CCR is located are analyzed. The distribution functions show a slight deviation from each other. This is also reflected in the two metrics. The deviations in d_{bias} show a spread of 0.045 dB and a much smaller spread in d_{CAVM}. This shows that the repeatability of the experiments for the CCR is even higher in comparison to the whole radar cuboid. This evaluation is therefore suitable for validation measurements in the case of this sensor for point targets.

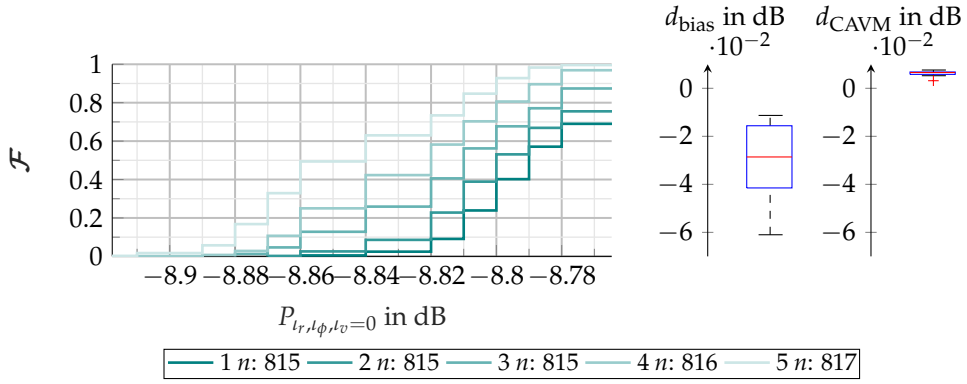


Figure 6. EDFs of all five measurements of the CCR on the radar cuboid level filterd by an ROI are shown in the diagram. The power distribution with the corresponding box plot for d_{bias} and d_{CAVM} is visualized on the right side. The aggregated number of bins is listed below the diagram.

ROI Detections

Figure 7 visualizes the RCS distribution of the CCR detections based on an ROI. Due to the point target characteristics and the small, defined ROI, an analysis of the distance and azimuth is not discussed. The deviations in d_{bias} are 0.01 dbm^2, and an even smaller scatter in d_{CAVM} appears. Finally, the ROI on detection level shows a high repeatability of the measurements. For point targets, it can therefore be concluded that a ROI can be used as a filter on the detection level for validation measurements.

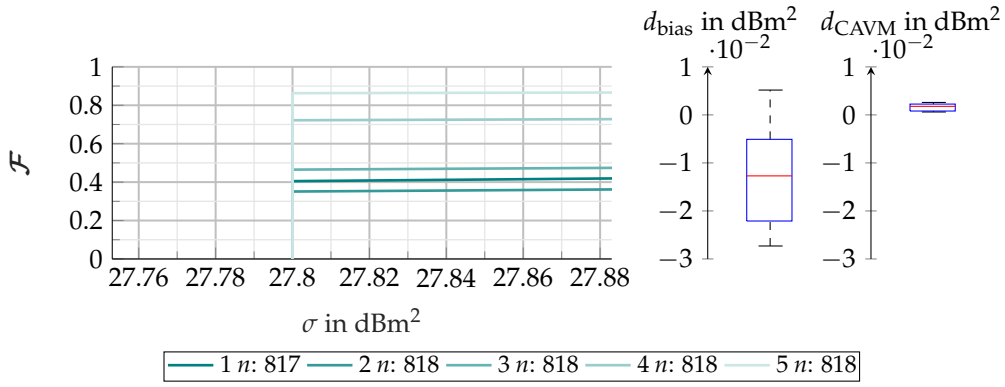


Figure 7. EDFs of all five measurements of the CCR on the detection level filterd by an ROI are shown in the diagram. The RCS distribution with the corresponding box plot for d_{bias} and d_{CAVM} is visualized on the right side. The aggregated number of detections of all timeframes is listed below the diagram.

Each Range and Azimuth Bin on Radar Cuboid Level

For a detailed analysis of the environmental influence on the repeatability of the radar cuboid, the EDFs are compared using d_{bias} and d_{CAVM} for each individual range azimuth bin. For this purpose, the respective maximum of $|d_{\text{bias}}|$ and d_{CAVM} of the corresponding range azimuth bin are visualized with a satellite image and the radar position measured by the GNSS sensor. This makes it possible to interpret the causes of anomalies and assign them to areas in the radar FoV. The result is shown in Figure 8.

It can be seen that for the CCR position, there are small deviations across all azimuths and ± 2 range bins. This can be explained by the results of the ROI of the CCR. Due to the high back scattering power, stochastic effects are less pronounced. Due to the window functions used in the radar sensor, the power is smeared into the neighboring bins of the radar cuboid, resulting in the high repeatability in the previously identified range for d_{bias} and d_{CAVM}. Only in front of and behind this area do deviations occur in both metrics in the center of the FoV. In the near range, the largest deviations occur for d_{bias} and d_{CAVM}. These

variations in the radar cuboid lead to the close-range detections identified in the whole detection analysis. In general, the transition from road to vegetation causes the largest deviations. Here, the deviations decrease with increasing azimuth space from the grass. The road surface, on the other hand, shows a very high repeatability of the measurements, with only minor deviations. Cracks and vegetation on the runway are possible explanations for this observation. Smaller deviations at the road junction confirm this impression when the opposite side with the grass is also taken into account. As the distance increases, the deviations in the two metrics decrease, and the distributions in the different cells become more similar.

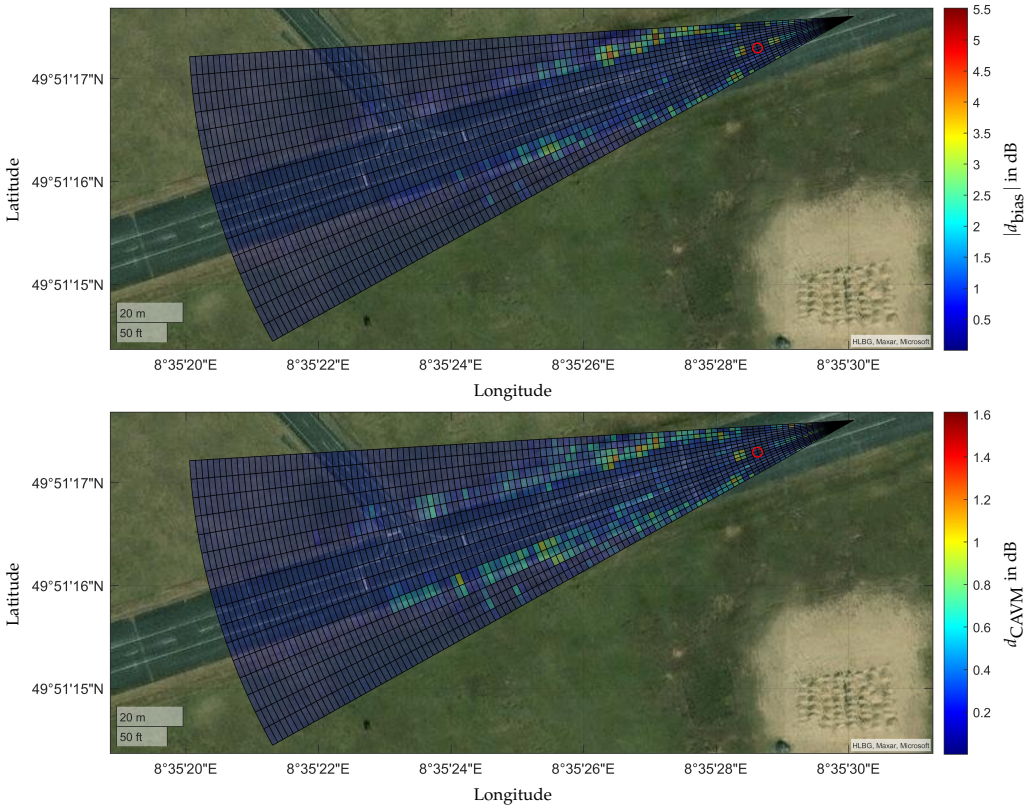


Figure 8. Satellite image fused with $|d_{\text{bias}}|$ and d_{CAVM} coded in colors for each range azimuth bin. The red circle visualizes the position of the CCR.

4.1.2. Dismantling and Setting Up on Two Different Days

In this chapter, measurement data from two consecutive days with eight measurements each is compared. No object is placed in front of the radar sensor, and only the surrounding area is measured. A dismantling and reassembly of the measurement setup between the two days is performed. In this way, the reproducibility of the radar model validation measurement setup is checked, and the second part of *RQ 1* is addressed. Compared to the CCR measurements, the sensor position on the runway is different, but this does not affect the research objective.

In itself, the measurements between Day 1 and Day 2 are consistent in terms of the number of data points (see Figure 9). However, it can be seen that switching the sensor and PC on and off for data recording has an effect on the number of data points. The recording time is always 60 s for both days. It follows that such effects must be taken into account in the quality assurance of the measured data for the validation of sensor models. Due to the fact that no objects were placed, an analysis of the ROI RC and the ROI D is not feasible.

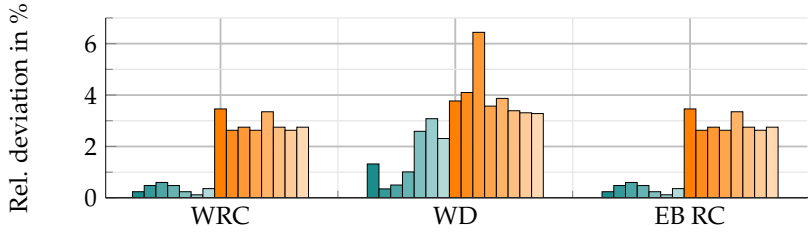


Figure 9. Plot of the relative deviation for the number of the data points or detections of all measurements of the two test days, Day 1 and Day 2. The compared data are WRC, the ROI for the ROI RC and EB RC. Experiment 1 is denoted as ■ and experiment 2 as ■. The gradations of the respective color delineate the different measurements.

Whole Radar Cuboid

In general, the EDFs of the whole radar cuboid are visually close, which is confirmed by the deviations in d_{bias} and d_{CAVM} (see Figure 10). However, a comparison of these measurements with those of CCR shows that the scatter of d_{bias} and d_{CAVM} is greater.

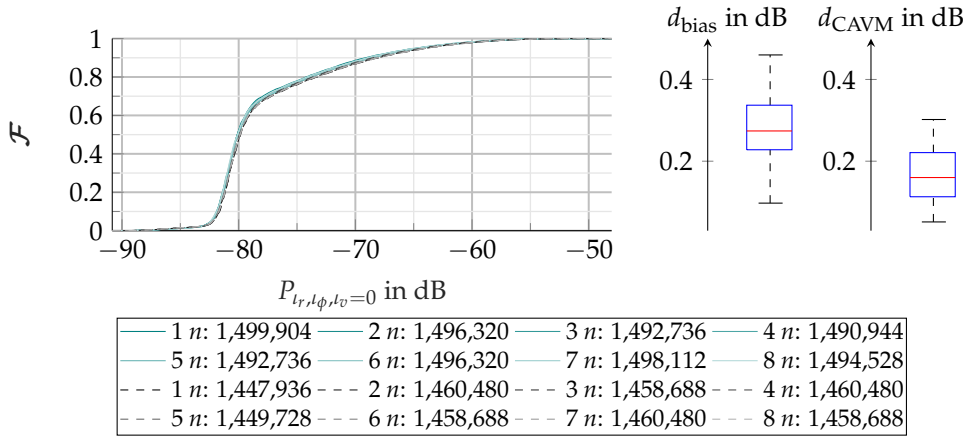


Figure 10. EDFs of all eight measurements for Day1 (——) and Day 2 (- - -) each on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.

To support this observation, Figure 11 shows the comparison of d_{bias} and d_{CAVM} for Day 1/2, Day 1/1 and Day 2/2. It can be seen that reconfiguring the measurement setup has a greater effect than taking several measurements in succession. With the method presented, this effect is quantifiable via the scatter of the two metrics, and the reproducibility is measurable.

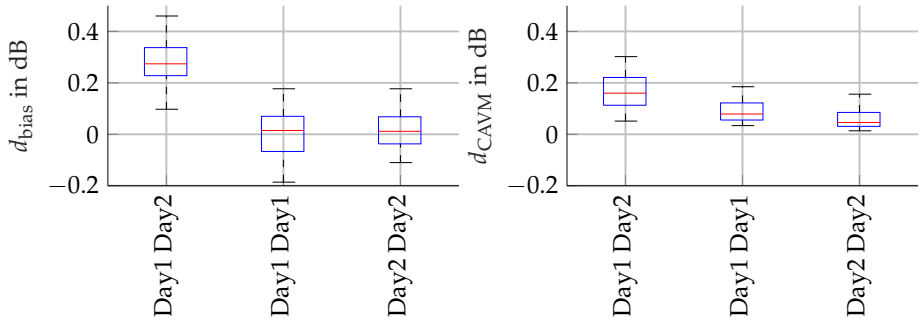


Figure 11. Comparison of d_{bias} and d_{CAVM} based on the whole radar cuboid for Day 1/2, Day 1/1, and Day 2/2. The repeatability between the measurements of each day is higher than the reproduciblity of the measurement setup.

Each Range and Azimuth Bin on Radar Cuboid Level

Figure 12 shows d_{bias} and d_{CAVM} over all range azimuth bins. The largest deviations occur in the near range of the sensor, with up to a 15 dB difference in power for d_{bias} at the edge of the FoV. Compared to the CCR measurements, where only 5 dB are present. It can be concluded that the close-range is particularly affected by the mounting and dismounting of the sensor. Especially for validation measurements with objects and environments, the reproducibility of the measurements must be checked before using them for a validity statement. For this reason, close-range measurements must be taken with particular accuracy. On the road surface, deviations of up to 10 dB in d_{bias} can be detected up to a distance of 74 m. The roadside, on the other hand, has an effect on the readings over the whole visibility range due to differences in the sensor orientation. Both d_{bias} and d_{CAVM} show deviations at identical positions in the range azimuth map.

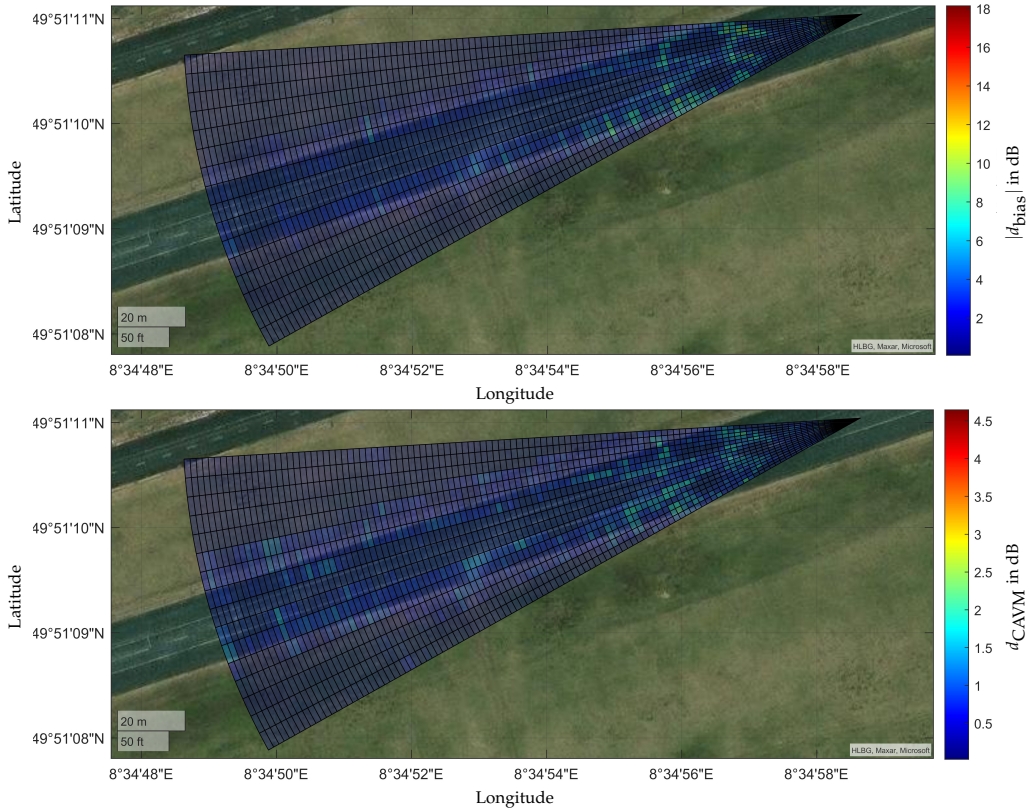


Figure 12. Satellite image fused with $|d_{\text{bias}}|$ and d_{CAVM} coded in colors for each range azimuth bin.

4.2. RQ 2: Effects of Objects on Radar Measurements

In this chapter, the influence of objects on radar validation measurements is investigated. For this purpose, the measurements of the CCR from Section 4.1.1 are first compared with those of a Volvo XC90. The center of the vehicle is almost identical to the position of the CCR. The yaw angle of the vehicle is close to 0. Next, the XC90 is observed in the identical position but rotated by a yaw angle in the second experiment. This analyzes the influence of the rotation on the measurement result. Finally, the XC90 and a Mercedes Vito are compared with each other at an almost identical position and yaw angle.

4.2.1. Comparison of the CCR and XC90

In this chapter, effects due to complex geometries compared with an ideal target are identified. Figure 13 shows the number of data points between detection and radar cuboid level. The relative deviation for the radar cuboid level is for the evaluation of WRC, the ROI RC, and the EB RC, which are similar and below 1%. At the detection level, the number

of detections differs significantly between the objects, which is why a comparison for the evaluation of WD and the ROI D is not meaningful.

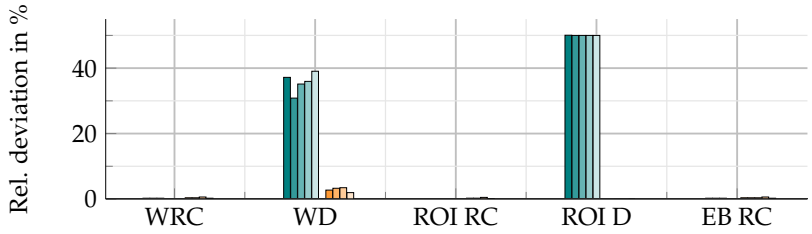


Figure 13. Plot of the relative deviation for the number of the data points or detections of all measurements of the two experiments (CCR and XC90). All evaluation levels are compared. Experiment 1 is denoted as ▬ and experiment 2 as ▬. The gradations of the respective color delineate the different measurements.

Whole Radar Cuboid

Figure 14 visualizes the result for the comparison of the vehicle and the ideal target. For the entire radar cuboid, it is evident from the EDFs that the CCR affects all bins. Already at −82 dB deviations are recognizable. At occurrences above 90%, the EDF of both objects equalizes again. At −78 dB, there is a kink in the course of the EDF of the XC90, which suggests the influence of the environment, since this kink is also evident in the measurements of Day 1/2. For d_{bias}, a median of −2 dB is shown for the whole radar cuboid. The different shape of the EDF is also reflected in d_{CAVM}. The values have significantly increased compared with the previous experiments.

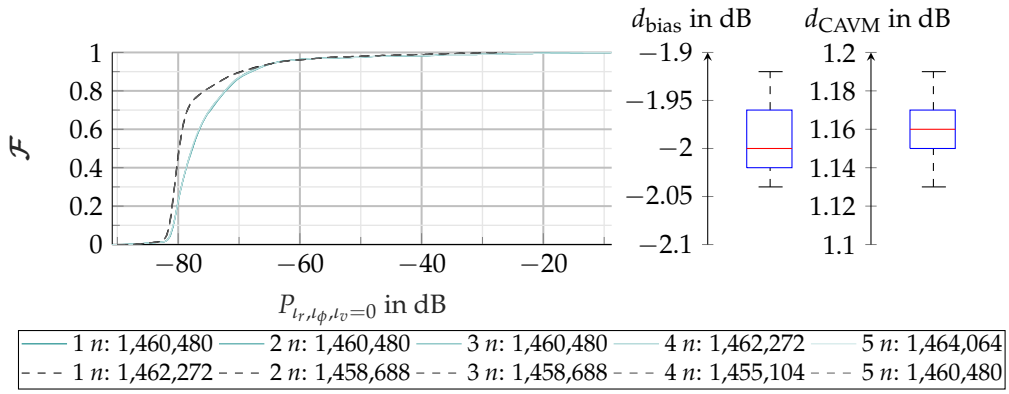


Figure 14. EDFs of all five measurements for the objects CCR (——) and the Volvo XC90 (- - -) each on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.

Each Range and Azimuth Bin on Radar Cuboid Level

The analysis of the data in Figure 15 shows significant differences in d_{bias}, especially at the object positions. These discrepancies can be attributed to the better reflective properties of the CCR. In addition, the influence of the windowing function can be seen, which smears the power across the range bins. This effect is noticeable in the central area of the FoV. The deviations at the edge of the roadway again confirm the effect already described in the evaluation of the entire radar cuboid. The CCR influences all bins. This finding is also reflected in the distribution of the d_{CAVM}. Lower deviations are seen both at the object's position and at twice the distance. This is a proof of the reproducibility of multipath propagation of radar waves due to objects. In general, it can be seen that the object is influenced by its surrounding and its characteristics.

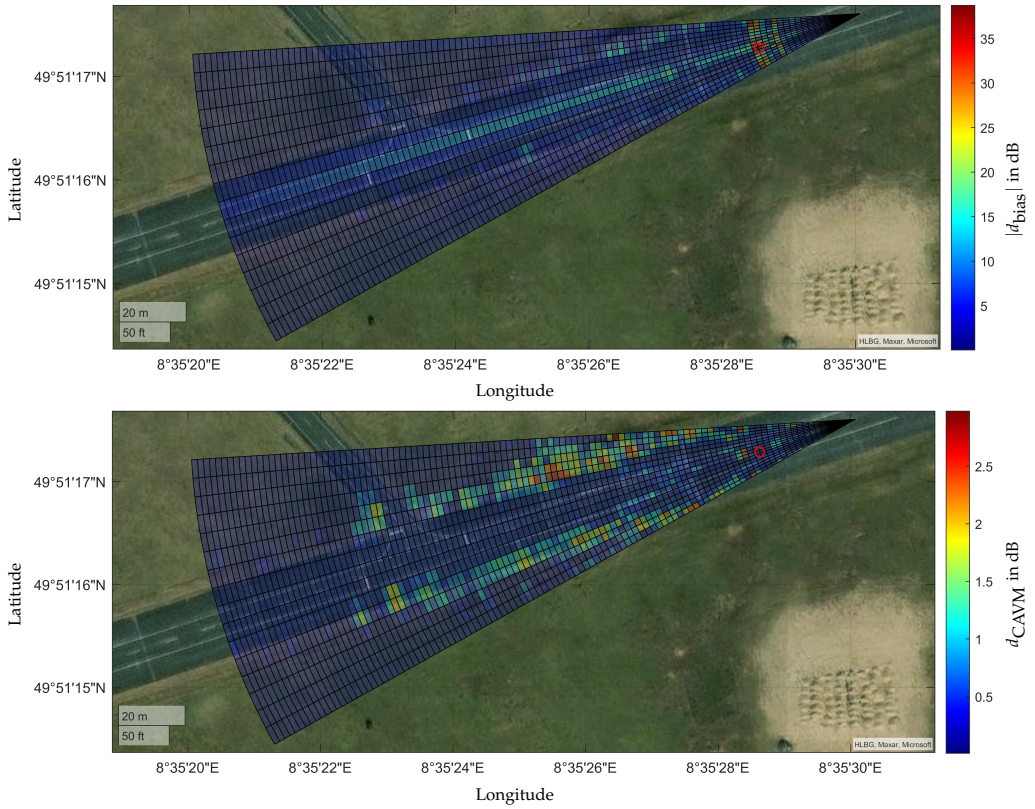


Figure 15. Satellite image fused with $|d_{bias}|$ and d_{CAVM} coded in colors for each range azimuth bin. The objects under examination are a CCR and a Volvo XC90. The red circle visualizes the position of both objects.

4.2.2. Comparison of XC90 and Rotated XC90

In this section, the rotation of an object on the wave propagation for an exemplary yaw angle is examined. However, the difference between these two experiments is the yaw angle of the same object. As in Section 4.2.1, the relative deviation on the detection level is significantly higher than the deviation o then radar cuboid level. Due to the high deviations on the detection level, see Figure 16, the comparability of the measurements and thus of the experiments is not given. Therefore, this comparison is only evaluated at the radar cuboid level. The relative deviations on the radar cuboid level for the different evaluations (WRC, ROI RC, and EB RC) are in the same order of magnitude and thereby below 1%.

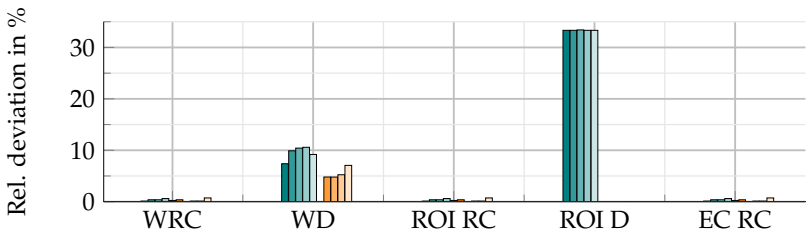


Figure 16. Plot of the relative deviation for the number of the data points of all measurements of the two experiments (XC90 and XC90 rotated).The WRC, WD, the ROI for the ROI RC and the ROI D, as well as EB RC are compared. Experiment 1 is denoted as ■ and experiment 2 as ■. The gradations of the respective color delineate the different measurements.

Whole Radar Cuboid

The rotation of the object has a small influence on the total power distribution in the radar cuboid, as visible in Figure 17. The distribution functions are again visually hardly

separable, which is reflected in the values for d_{bias} and d_{CAVM}. The scatter of d_{bias} is 0.17 dB, which puts the difference in a range also observed in the CCR measurements. This can be explained by the main influence of the environment and the disproportionate amount of road and vegetation in the measurements. An effect over the entire radar cuboid due to rotation, as in the CCR and XC90 comparisons, cannot be observed. Based on this finding, an evaluation of the whole detection is omitted, and a detailed look at the ROI is given.

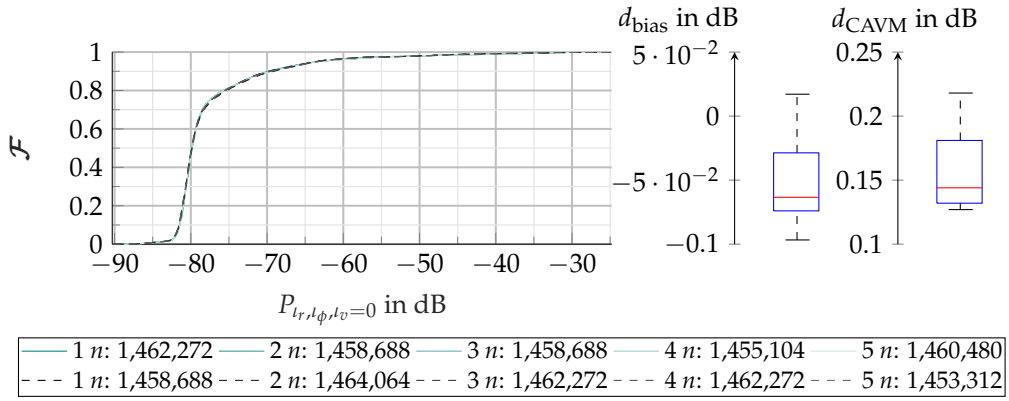


Figure 17. EDFs of all five measurements for the XC90 (——) and the XC90 rotated (- - -) each on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.

ROI Radar Cuboid

The evaluation of the ROI is shown in Figure 18. The rotation leads to a significant decrease in the power values in the filtered bins, with little fluctuation between experiments. This observation confirms d_{bias} with a significant deviation of 5.5 dB at the median. Rotation also affects the shape of the distribution functions, such that d_{CAVM} is almost equal to 2.8 dB. The results show high inter-experiment repeatability for static experiments but a significant influence of object orientation.

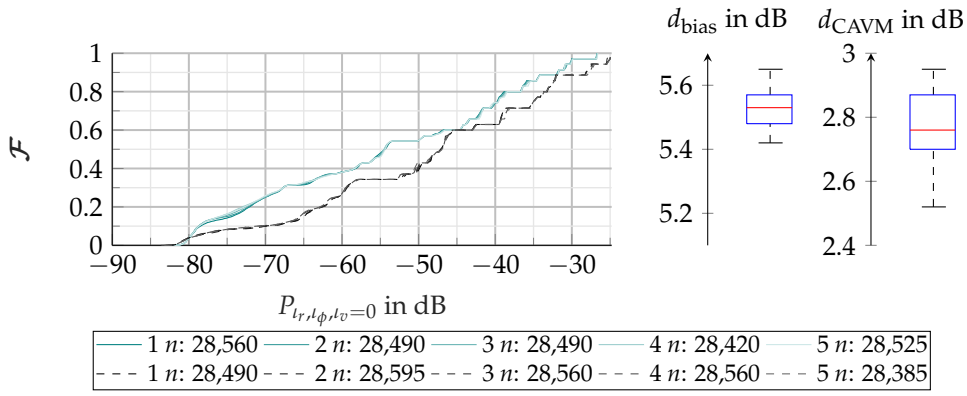


Figure 18. EDFs of all five measurements of the XC90 (——) and the rotated XC90 (- - -) on the radar cuboid level filterd by an ROI are shown in the diagram. The power distribution with the corresponding box plot for d_{bias} and d_{CAVM} is visualized on the right side. The aggregated number of bins is listed below the diagram.

Each Range and Azimuth Bin on Radar Cuboid Level

The findings from the ROI become even clearer in the satellite image (see Figure 19). Especially in the vicinity of the object, values of 25 dB are present in d_{bias}. In this area, however, d_{CAVM} shows only low values, which speaks for the previously established high repeatability of the results in this area. Interesting are the high deviations in d_{bias} at the roadside and partly in the vegetation. 15 dB are increasingly and patchily detectable. This is

further evidence of the influence of objects on the detected surroundings due to multipath propagation or effects in the signal processing of the radar sensor. The variations of d_{CAVM} are largest where d_{bias} is also high. Comparison with satellite images of CCR and XC90 shows similar values, suggesting a stochastically constant behavior of vegetation in the different experiments and measurements.

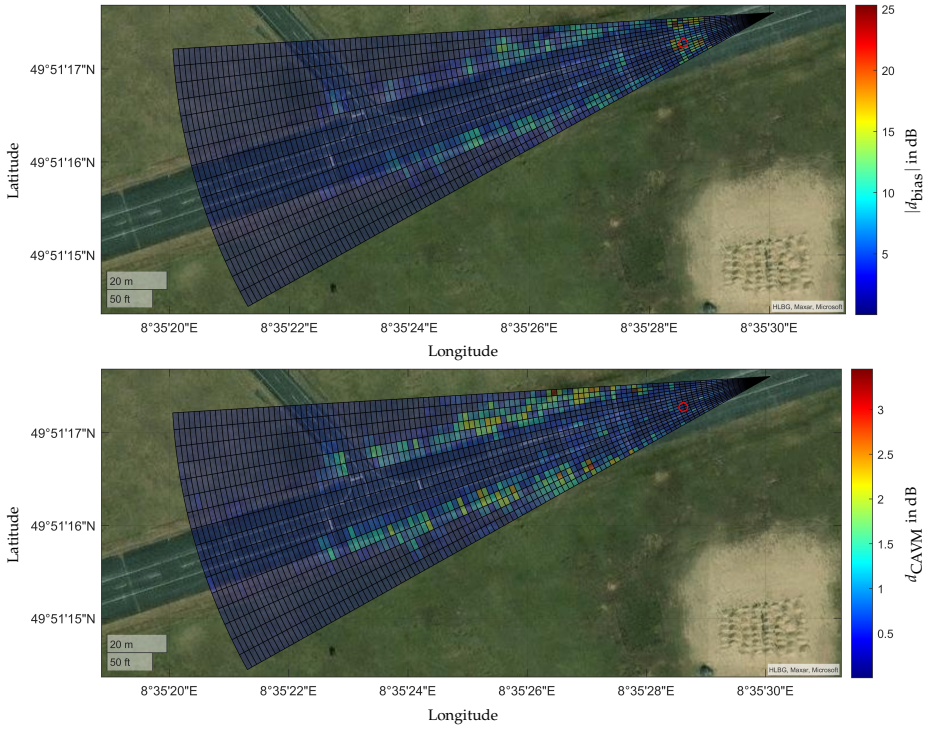


Figure 19. Satellite image fused with $|d_{\mathrm{bias}}|$ and d_{CAVM} coded in colors for each range azimuth bin. The object under examination is a Volvo XC90 with different yaw angles. The red circle visualizes the center of the vehicle's bounding box.

4.2.3. Comparison of Volvo XC90 and Mercedes Vito

Finally, the Volvo XC90 is compared with a Mercedes Vito. The effects of different object geometries and sizes are analyzed regarding their effects on validation measurement campaigns. The relative deviations are comparable to the previous findings. The deviations are higher at the detection level than those at the radar cuboid level, as seen in Figure 20. However, there is a difference in the deviations of all detections compared with the deviations for the detections in the ROI, which are significantly higher. Thus, also for this comparison, an evaluation is performed on the radar cuboid level with a relative deviation for the number of data points below 1 % and the whole detections.

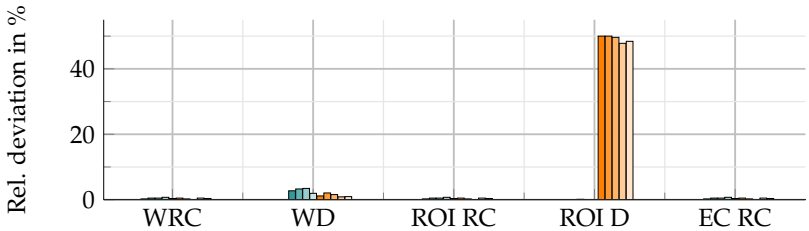


Figure 20. Plot of the relative deviation for the number of the data points of all measurements of the two experiments (XC90 and Vito). The WRC, WD, the ROI for the ROI RC and the ROI D, as well as EB RC are visualized. Experiment 1 is denoted as ▬ and experiment 2 as ▬. The gradations of the respective color delineate the different measurements.

Whole Radar Cuboid

Furthermore, the entire radar cuboid is analyzed. From Figure 21, it can be seen that the object has an influence on all bins, as in the comparison of CCR and XC90. This can be seen from the higher power distribution of the EDFs in the Mercedes Vito. The shape of the object (hatchback compared with vertical rear) therefore has a direct influence on the measured environment. In addition to the rear, the underbody can also influence the measurement result. The overall effect can be seen in d_{bias}, which shows a relatively high deviation with a low spread. For both metrics, it can be seen that the repeatability of the measurements of the different objects is very high, as the scatter is low.

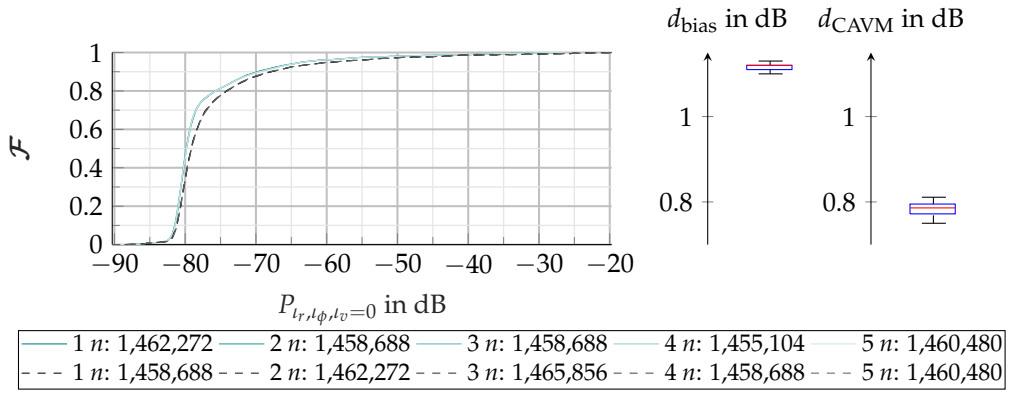


Figure 21. EDFs of all five measurements for the Volvo XC90 (——) and the Mercedes Vito (- - -) each on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.

Whole Detections

Figure 22 shows the EDFs for the distance, azimuth, and RCS, as well as a corresponding box plot for them. A look at the EDFs for the distance shows that these are similar to each other, but a deviation is still visible. In the distance of the vehicles at 29.65 m, a small step and thus an accumulation of detections can be seen. All detections extend over a maximum distance of 150 m. In the vicinity of the sensor, these detections result from reflections of the road surface and, with increasing distance, from the transition between road and vegetation. Whereby it can be seen from the course that at distances over 115 m, fewer detections can be recognized. Above 160 m, only isolated detections with a low RCS are present, which argues for purely stochastic causes. The median values for d_{bias} and d_{CAVM} are similar when considered as absolute values. However, the scatter of d_{bias} at 5 m is significantly larger than the scatter of d_{CAVM} at 1 m.

Looking at the EDFs for the azimuth, it is noticeable that the EDFs overlap better for angles below 0 than for angles above 0. Furthermore, the steps for the accumulation of detections by the two vehicles are different, so the step for the XC90 is at 1.5° and not as significant as the one for the Vito at 0. Basically, however, there is an accumulation of detections around 0 for both vehicles, which corresponds to the vehicle position and reflections from the road surface near the sensor. For angles below −5° and above 5°, detections can be seen by the transition from the road surface to vegetation. The plateaus between −5° and 0, and between 0 and 5°, are due to the road surface. The median values of d_{bias} and d_{CAVM} are similar for an absolute observation. However, it can again be seen that the scatter is larger for the d_{bias} than for the d_{CAVM}.

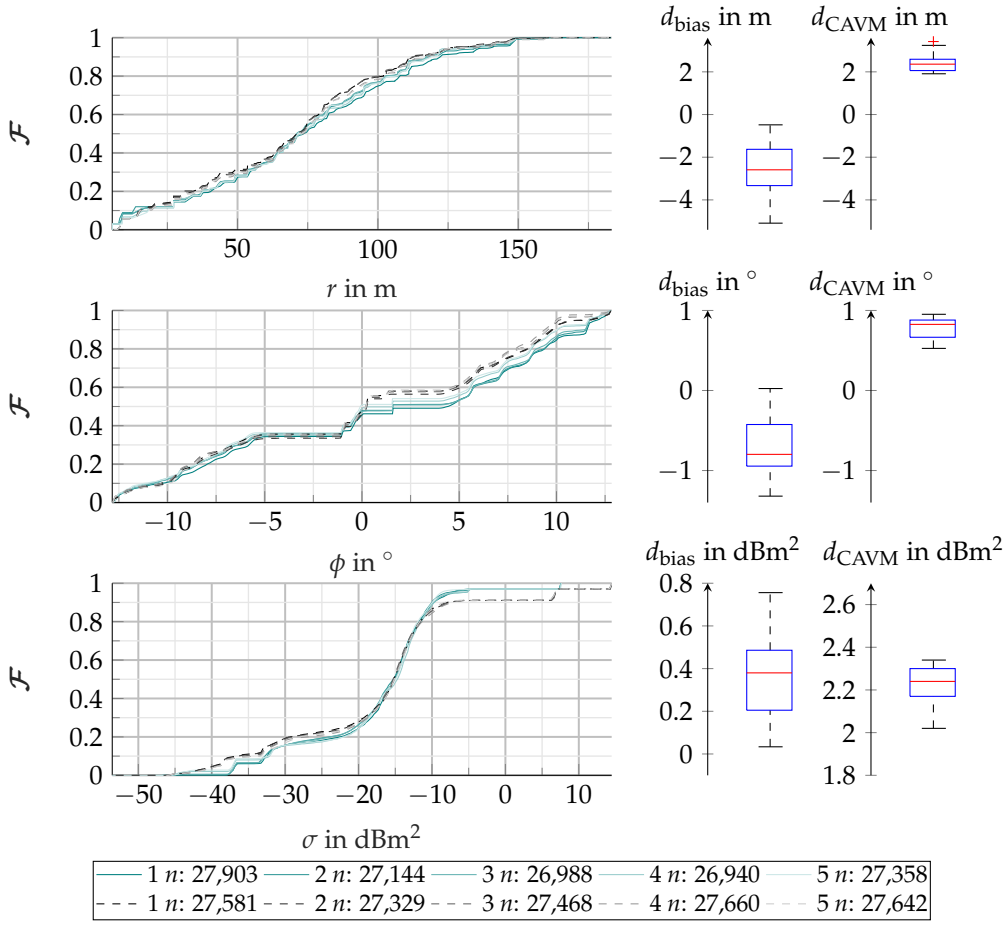


Figure 22. EDFs of all five measurements of the XC90 (——) and Vito (- - -) on the detection level. In the first row the distance distribution, in the second row the azimuth distribution and in the third row the RCS distribution with the corresponding box plot for d_{bias} and d_{CAVM} are shown. The aggregated number of detections is listed below the diagrams.

When looking at the EDFs for the RCS, these are close to the two vehicles between $-30\,\text{dBm}^2$ and $-10\,\text{dBm}^2$ and have a similar trend, but a deviation is visible. Below $-30\,\text{dBm}^2$, the EDFs of the XC90 are determined by more significant steps compared with the Vito. Above $-10\,\text{dBm}^2$, the curve is similar, but there is a clear offset between the two vehicles, indicating more detections on the Vito, proving the findings in Figure 20. The steps above $-10\,\text{dBm}^2$ indicates the RCS for both vehicles. A difference in the maximum RCS for the detections can be observed; the RCS for the XC90 is $5\,\text{dBm}^2$ less than the RCS of the Vito. In addition, a difference in the number of detections between the XC90 and the Vito can be seen in the level of the step. The medians for d_{bias} and d_{CAVM} differ by $1.75\,\text{dBm}^2$, with this being larger for d_{CAVM}. However, the dispersion for d_{bias} is $0.2\,\text{dBm}^2$ larger than for d_{CAVM}. As noted in the comparison of the two CCRs, it is difficult to interpret the d_{bias} and d_{CAVM} due to the randomness that comes from the vegetation-generated detections.

ROI Radar Cuboid

Figure 23 shows the ROI at the radar cuboid level. It confirms the impression from the analysis of all bins of the radar cuboid. In general, the Vito generates more power in the bins around its own position. Thereby, a high level of repeatability between the individual measurements is shown. The outer bins of the ROI provide the low powers, which are similar for both objects. However, d_{bias} also increases with increasing power values. d_{bias} shows values from 3.7 dB to 3.95 dB and illustrates the deviations caused by the object geometry. Similarly, d_{CAVM} is clearly influenced by the different objects. Therefore, the shape of the distribution functions differs because of the different objects.

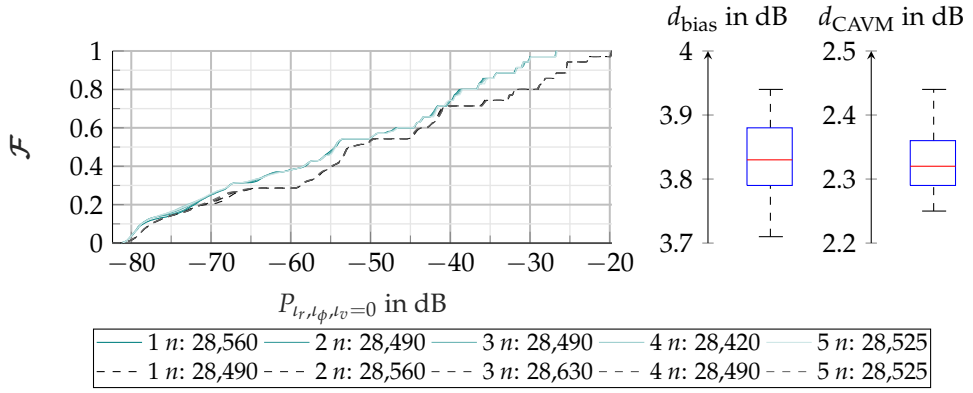


Figure 23. EDFs of all five measurements of the Volvo XC90 (——) and Mercedes Vito (- - -) on the radar cuboid level filterd by an ROI are shown in the diagram. The power distribution with the corresponding box plot for d_{bias} and d_{CAVM} is visualized on the right side. The aggregated number of bins is listed below the diagram.

Each Range and Azimuth Bin on Radar Cuboid Level

The detailed analysis of the range of azimuth bins is visualized in Figure 24. For d_{bias}, it is clearly visible that there are increased deviations in front of and behind the vehicle position. These extend over the entire azimuth range and are up to 18 dB in the center. At twice the vehicle distance, multipath propagation across the sensor is evident. These are the largest at up to 25 dB and also extend over the entire azimuth range due to the windowing function. There is also a deviation up to maximum visibility in the center of the FoV. This effect is more noticeable in the comparison between the CCR and the XC90. Here, the higher reflectivity of the Vito causes smearing in the more distant-range bins due to the sensor's windowing functions. For d_{CAVM} there are increased values in the double distance in the object position as well as for vegetation, which is comparable to the satellite images of the other object analyses.

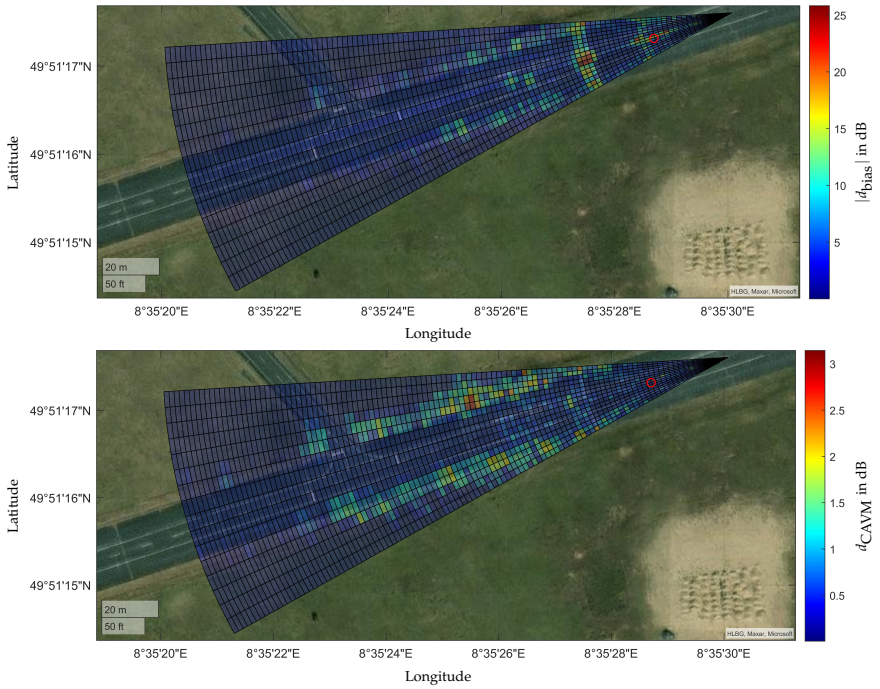


Figure 24. Satellite image fused with d_{bias} and d_{CAVM} coded in colors for each range azimuth bin. The object under examination is a Volvo XC90 and a Mercedes Vito at the same position and orientation. The red circle visualizes the center of the vehicles' bounding boxes.

4.3. RQ3: Influence of Rain on Radar Measurements

For the experiment, the XC90 is in the same position, but five measurements are recorded without rain and five measurements with rain. Subjectively, the intensity of the rain decreases over time.

The rain is of natural origin, and no rain intensity or raindrop size distribution data are available.

Also for the investigation of the environmental conditions, in this case rain, a difference of the relative deviations for the number of data points is shown (see Figure 25). On the detection level, the deviation is higher than on the radar cuboid level. However, the deviations for the evaluation of all detections are many times higher than those for the detections from the ROI. The rain increases the number of detections. Therefore, the evaluation is examined at the radar cuboid level. The relative deviation is similar for every evaluation on the radar cuboid level and below 1%.

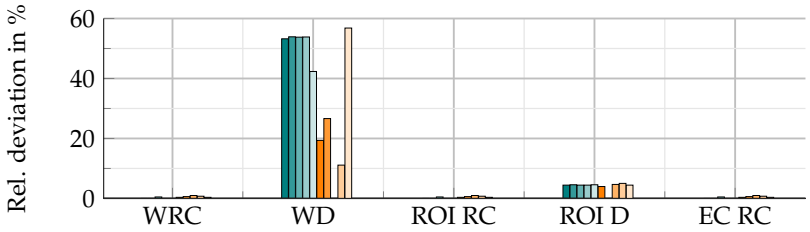


Figure 25. Plot of the relative deviation for the number of the data points of all measurements of the two experiments (XC90 and XC90 with rain). The WRC, WD, the ROI for the ROI RC and the ROI D, as well as EB RC are compared. Experiment 1 is denoted as ■ and experiment 2 as ■. The gradations of the respective color delineate the different measurements.

4.3.1. Whole Radar Cuboid

In Figure 26, the EDFs of the whole radar cuboid are shown. Rain generally increases the power of all bins. The rain measurement No. 5 lays on the measurements without rain. This may be due to the fact that the rain intensity in this measurement is already low. There is no reference data for rain intensity to make an accurate assessment. From −70 dB, the curves of all EDFs equalize, which shows the increasing influence of the XC90. This is again evident in the box plots for d_{bias} and d_{CAVM}. Rain leads to larger deviations in both metrics compared with the previous measurements in Section 4.2. For a detailed analysis, the satellite image is used.

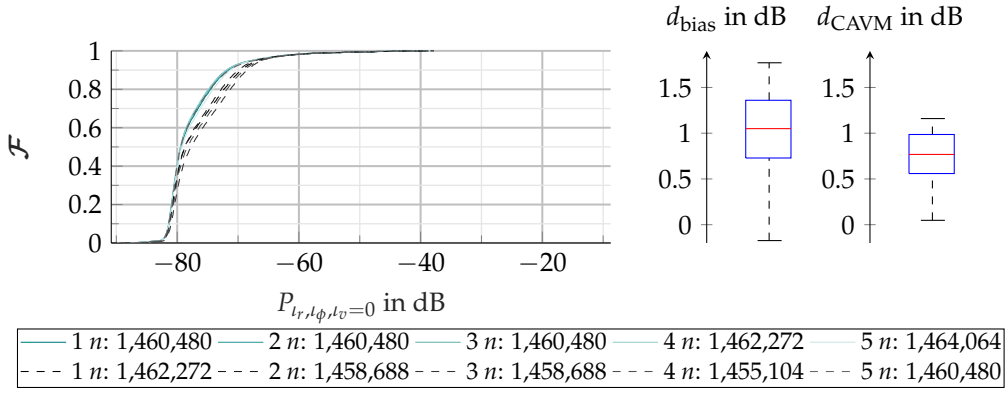


Figure 26. EDFs of five measurements of the XC90 in dry (——) and rainy (- - -) conditions on the radar cuboid level. The number of aggregated bins is listed below the diagram. On the right side the box plot for d_{bias} and d_{CAVM} in dB is shown.

4.3.2. Each Range and Azimuth Bin on Radar Cuboid Level

The satellite image in Figure 27 clearly shows the effect of the rain. The increased power is shown in both d_{bias} and d_{CAVM}. In particular, d_{bias} is increased in the vicinity of the sensor and up to about half of the FoV. The road surface is particularly affected by this phenomenon. For d_{CAVM}, the largest deviations occur in the medium range of the FoV. Furthermore, the road is more affected than the vegetation, which can be explained by the previously identified stochastic properties of the vegetation.

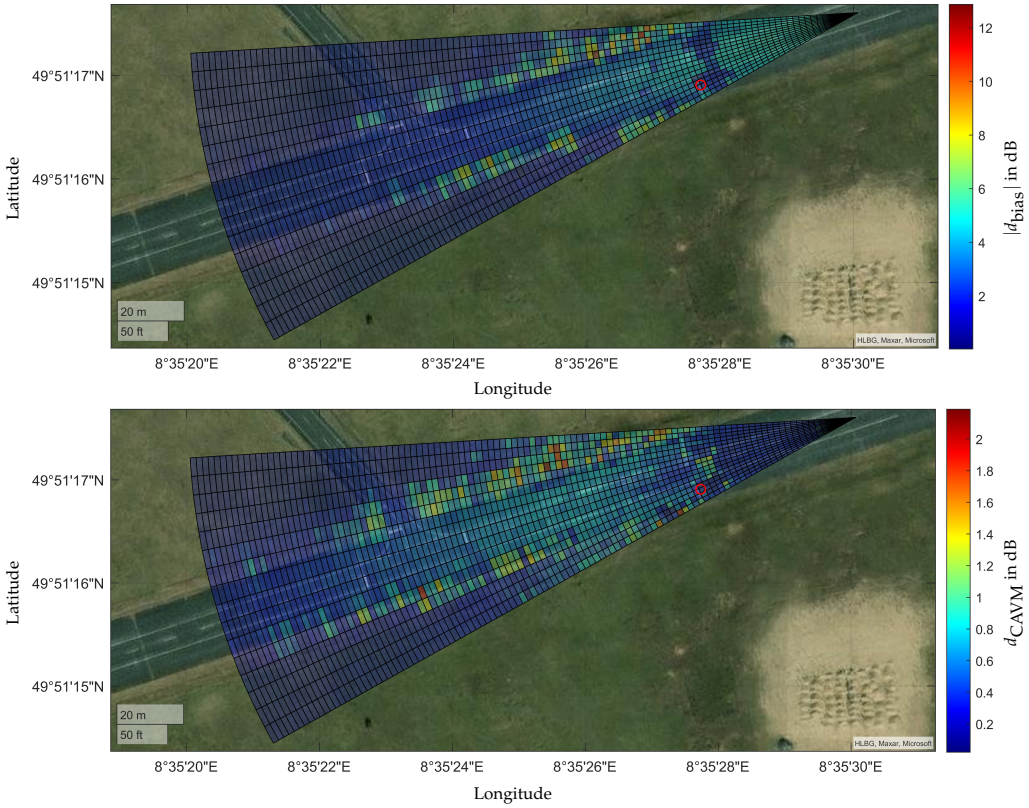


Figure 27. Satellite image fused with $|d_{\text{bias}}|$ and d_{CAVM} coded in colors for each range azimuth bin. The object under examination is a Volvo XC90 in rainy and dry conditions. The red circle visualizes the center of the vehicle bounding box.

5. Conclusions

This paper presents a methodology to make radar measurements comparable in the context of the validation of radar sensor models. The two metrics d_{bias} and d_{CAVM} are used to evaluate the similarity of the measured distribution functions. For this purpose, the number of data points is first analyzed in order to be able to make any comparisons using DVM. The analysis of the level of the radar cuboid simplifies the explanation of the different effects, especially in combination with the representation of the range azimuth bins using a satellite image. Finally, the research questions raised in Section 1.2 are answered, and the applicability of the findings to other radar sensors remains to be shown.

Research Question 1: The analysis of the measurements with a CCR shows in direct comparison that there is good repeatability of the measurements, especially near an object in radar measurements. Only vegetation and the road surface in close proximity are characterized by larger deviations in d_{bias} and d_{CAVM}. Furthermore, the dismantling and repositioning shows that smaller influences occur. Nevertheless, the measurements can be performed reproducibly with the right tools (e.g., ground markers).

Research Question 2: Extended objects can be analyzed in detail with the presented methodology, and deviations can be quantified accordingly. However, it can be seen that

objects have a direct impact on the overall FoV of the sensor, depending on the orientation and type of object.

Research Question 3: Based on our methodology, environmental influences are comparable, and their influence is quantitatively detectable. In particular, the close-range of the radar sensor shows larger deviations due to rain on average. With increasing distance, rain leads to larger scattering, whereas the effect decreases again after a certain distance and the rain merges with the background noise of the sensor.

Therefore, for validation measurements of radar sensors, the following aspects can be derived from the results:

- Measurements of static vehicles and point targets are well reproducible on radar cuboid and detection levels.
- Vegetation and the close-range of the radar sensor lead to large deviations, especially on the detection level, which is why this interface complicates the validation of radar sensor models with the methodology described here.
- The presented methodology can be used to quantify deviations between validation measurements and identify measurement outliers.
- Using ideal laboratory measurements, effects in real measurements can be identified based on the deviations in d_{bias} and d_{CAVM}, and their influence can also be quantified (e.g., multipath propagation and interference).
- With the presented methodology, a maximum achievable quality of simulation models can be determined with a quantitative value in the unit of the measurand with respect to the repeatability and reproducibility of experiments.

The previous considerations refer only to static scenarios. In particular, dynamic tests are of great importance for the radar sensor due to the Doppler effect and the associated direct velocity measurement. Therefore, in the next step, the approach and analysis will be transferred to dynamic scenarios. So far, only the usefulness of the methodology for the sensor used has been demonstrated. In the next step, the knowledge gained has to be proven with other radar sensors.

Author Contributions: Conceptualization, L.E.; methodology, L.E.; software, L.E. and J.P.S.; validation, L.E. and J.P.S.; formal analysis, L.E.; investigation, L.E. and J.P.S.; data curation, L.E.; writing—original draft preparation, L.E. and J.P.S.; writing—review and editing, L.E., J.P.S. and S.P.; supervision, S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry for Education and Research (BMBF) through the project Virtual Validation Tool Chain for Automated and Connected Driving (VIVID) under grant number 16ME0173.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACC	active cruise control
ADF	automated driving function
CCR	corner cube reflector
D	detecions
DVM	Double Validation Metric
EB RC	each bin of the radar cuboid
EDF	empirical cumulative distribution function
FFT	fast Fourier transform
FoV	field of view

GNSS	global navigation satellite system
Lidar	light detection and ranging
PDF	probability density function
Radar	Radio Detection and Ranging
RC	radar cuboid
RCS	Radar Cross Section
RTK	real time kinematic
ROI	region of interest
SNR	signal-to-noise ratio
WD	whole detections
WRC	whole radar cuboid

References

1. Winner, H. Quo vadis, FAS? In *Handbuch Fahrerassistenzsysteme*; Winner, H., Hakuli, S., Lotz, F., Singer, C., Eds.; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2015; pp. 1167–1186. [CrossRef]
2. Holder, M.F. Synthetic Generation of Radar Sensor Data for Virtual Validation of Autonomous Driving. Ph.D. Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2021. [CrossRef]
3. Eder, T. Simulation of Automotive Radar Point Clouds in Standardized Frameworks. Ph.D. Thesis, Technische Universität München, München, Germany, 2021.
4. Dietmayer, K. Predicting of Machine Perception for Automated Driving. In *Autonomous Driving*; Maurer, M., Gerdes, J.C., Lenz, B., Winner, H., Eds.; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2016; pp. 407–424. [CrossRef]
5. Holder, M.; Rosenberger, P.; Winner, H.; D'hondt, T.; Makkapati, V.P.; Maier, M.; Schreiber, H.; Magosi, Z.; Slavik, Z.; Bringmann, O.; et al. Measurements revealing Challenges in Radar Sensor Modeling for Virtual Validation of Autonomous Driving. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2616–2622. [CrossRef]
6. Schaermann, A.; Rauch, A.; Hirsenkorn, N.; Hanke, T.; Rasshofer, R.; Biebl, E. Validation of vehicle environment sensor models. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 405–411. [CrossRef]
7. Abadpour, S. Modeling Backscattering Behavior of Vulnerable Road Users Based on High-Resolution Radar Measurements. Ph.D. Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany, 2023. [CrossRef]
8. Schneider, R. Modellierung der Wellenausbreitung für ein bildgebendes Kfz-Radar. Ph.D. Thesis, Universität Fridericana Karlsruhe, Karlsruhe, Germany, 1998.
9. Magosi, Z.F.; Wellershaus, C.; Tihanyi, V.R.; Luley, P.; Eichberger, A. Evaluation Methodology for Physical Radar Perception Sensor Models Based on On-Road Measurements for the Testing and Validation of Automated Driving. *Energies* **2022**, *15*, 2545. [CrossRef]
10. Aust, P.; Hau, F.; Dickmann, J.; Hein, M.A. A Data-driven Approach for Stochastic Modeling of Automotive Radar Detections for Extended Objects. In Proceedings of the 2022 14th German Microwave Conference (GeMiC), Ulm, Germany, 16–18 May 2022; pp. 80–83.
11. Jayapal Gowdu, S.B.; Aust, P.; Schwind, A.; Hau, F.; Hein, M.A. Evaluation of scenario-based automotive radar testing in virtual environment using real driving data. In Proceedings of the 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 8–12 October 2022; pp. 2379–2384. [CrossRef]
12. Ngo, A. A methodology for validation of a radar simulation for virtual testing of autonomous driving. Available online: http://nbn-resolving.de/urn:nbn:de:bsz:93-opus-ds-127227 (accessed on 13 October 2023)[CrossRef]
13. Rosenberger, P.; Wendler, J.T.; Holder, M.F.; Linnhoff, C.; Berghöfer, M.; Winner, H.; Maurer, M. Towards a Generally Accepted Validation Methodology for Sensor Models—Challenges, Metrics, and First Results. In Proceedings of the Graz Symposium Virtual Vehicle, Graz, Austria, 7–8 May 2019.
14. Linnhoff, C.; Hofrichter, K.; Elster, L.; Rosenberger, P.; Winner, H. Measuring the Influence of Environmental Conditions on Automotive Lidar Sensors. *Sensors* **2022**, *22*, 5266. [CrossRef] [PubMed]
15. *DIN SAE SPEC 91471:2023-05*; Assessment Methodology for Automotive LiDAR Sensors. DIN Deutsches Institut für Normung e. V.: Berlin, Germany, 2023.
16. Viehof, M. Objektive Qualitätsbewertung von Fahrdynamiksimulationen durch statistische Validierung. Ph.D. Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2018.

17. Rosenberger, P. Metrics for Specification, Validation, and Uncertainty Prediction for Credibility in Simulation of Active Perception Sensor Systems. Ph.D. Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2023. [CrossRef]
18. Roy, C.J.; Balch, M.S. A holistic approach to uncertainty quantification with application to supersonic nozzle thrust. *Int. J. Uncertain. Quantif.* **2012**, *2*, 363–381. . [CrossRef]

D Paper IV:

Introducing the Double Validation Metric for Radar Sensor Models

Automotive and Engine Technology (2024) 9:6
https://doi.org/10.1007/s41104-024-00143-5

ORIGINAL PAPER

Introducing the double validation metric for radar sensor models

Lukas Elster[1] · Philipp Rosenberger[2] · Martin Holder[3] · Ken Mori[4] · Jan Staab[5] · Steven Peters[5]

Received: 20 June 2023 / Accepted: 8 May 2024

Abstract
In automated vehicles, environment perception is performed by various sensor types, such as cameras, radars, lidars, and ultrasonics. Simulation models of these sensors, as required in virtual validation methods, are available in various degrees of detail. However, proving the validity of such models is a subject of research. New metrics and methods for credibility assessment of simulation are needed to standardize the validation process in the future. The so-called double validation metric (DVM) has shown advantages and allows an intuitive interpretability of the validation results. The DVM has so far only been applied to lidar sensor models. In this paper, an extension to the DVM is introduced, which is called the DVM Map. A static measurement scenario is conducted in reality and transferred into simulation. The novel method is demonstrated on the obtained real and simulated radar sensor data. In this simple scenario special focus is put on the position accuracy of GNSS reference sensors. Therefore, their impact on the result of sensor model validation is discussed. The paper shows that the method provides a more detailed and accurate validation in comparison to the state of the art of a radar simulation, revealing previously undetected simulation errors. Errors due to the environment model, signal propagation, and signal processing are separated and satellite imagery is used for intuitive visualization of the results. This method is a complementary tool to existing validation techniques to improve the interpretability and judging the trustworthiness of radar simulations.

Keywords Radar model validation · Double area validation metric · DVM map · Radar cuboid interface · Detections interface

1 Introduction: what is so special about automotive radar modeling?

Simulation is playing an increasingly important role in proving the safety of automated vehicles. New procedures are envisaged in institutions such as UNECE [1, p. 5, 6], where simulation will be an integral part of the certification process. Automated vehicles (SAE level 3+ [2]) rely on robust environment perception with multiple sensor technologies. Radar ("radio detection and ranging") is one of these technologies, where a signal is actively transmitted to receive the echo instead of passively collecting it from other sources, like cameras. Automotive radar sensors commonly used in series production at different OEMs are based on the frequency modulated chirp-sequence principle, which is used to determine the range and the angular position of objects

✉ Lukas Elster
lukas.elster.research@web.de

Philipp Rosenberger
philipp.rosenberger@persival.de

Martin Holder
martin.holder@de.bosch.com

Ken Mori
ken.mori@tu-darmstadt.de

Jan Staab
jan.staab@stud.tu-darmstadt.de

Steven Peters
steven.peters@tu-darmstadt.de

[1] Independent Researcher, Formerly with Institute of Automotive Engineering, Technical University of Darmstadt, 64289 Darmstadt, Germany

[2] Persival GmbH, Carlo-Mierendorff-Str. 2, 64372 Ober-Ramstadt, Germany

[3] Robert Bosch GmbH, Postfach 10 60 50, 70049 Stuttgart, Germany

[4] Independent Researcher, Formerly with Institute of Automotive Engineering, Technical University of Darmstadt, 64653 Lorsch, Germany

[5] Institute of Automotive Engineering, Technical University of Darmstadt, Otto-Berndt-Str. 2, 64287 Darmstadt, Germany

Published online: 18 June 2024

via antenna patches with respect to the sensor. What makes radars special among current perception sensor technologies is their ability to measure the radial velocity via the Doppler effect and their wide use in series applications. This leads to an additional dimension of information and a radar "point cloud" therefore includes range r, azimuth angle ϕ, elevation angle θ, radar cross section (RCS) σ, and relative radial velocity v.

As shown in Fig. 1, a radar "point cloud", or more precise a radar "detection list" [4], appears more sparse and unstructured when compared to lidar. It becomes visible that modern digital beam forming radar sensors are not structuring the scans of the environment in a per layer order, as current time-of-flight lidar ("light detection and ranging") sensors do with their laser beams. The unstructured and more noisy appearance of radar detections is reasoned in the radar sensor's wave propagation characteristic and signal processing. Its front end, as shown in Fig. 2, entails the signal reception via the antennas and analog-to-digital conversion (ADC). After applying a discrete fast Fourier transform (DFFT) algorithm, the time-based signal is structured into the so-called radar cuboid. This term means a cubic multidimensional volume, often called radar cube for simplicity. It consists of multiple cells, so-called "bins", that can be divided into the dimensions range $\imath_r$, relative radial velocity $\imath_v$, azimuth $\imath_\phi$, and elevation $\imath_\theta$.

To enhance explainability, in the remainder, the content is limited to radars with range, azimuth, and radial velocity dimensions. However, the methodology can be applied to elevation without restriction, as each coordinate is anyway validated separately. The three dimensions of the radar cuboid are visualized in Fig. 3 with one single bin colored in orange and the teal color depicting the bins of the radar cuboid at a relative radial velocity of 0 ms^{-1}. Each bin contains a power ratio P in dB, which is calculated from the transmitted and the received signal. The number of bins $I_{r/v/\phi/\theta}$ of each dimension results from sensor design parameters like bandwidth, sampling rate and measurement time, as well as the configuration of the antennas. Except for digitization itself, only information due to windowing and noise, preexisting prior to the application of the DFFT, is lost at radar cuboid level. Because of this condensation of all digital available information, it is reasonable to simulate the synthetic data with chirp sequence frequency modulated continuous wave radars with uniform array antennas at this very early interface. For the required level of detail in simulation-based safety validation of automated driving, low-level interfaces must be taken into account to simulate specific perception tasks like fusion algorithms [5] due to challenging environmental conditions or object constellations. Furthermore, simulated low level interfaces enable to enhance early signal processing.

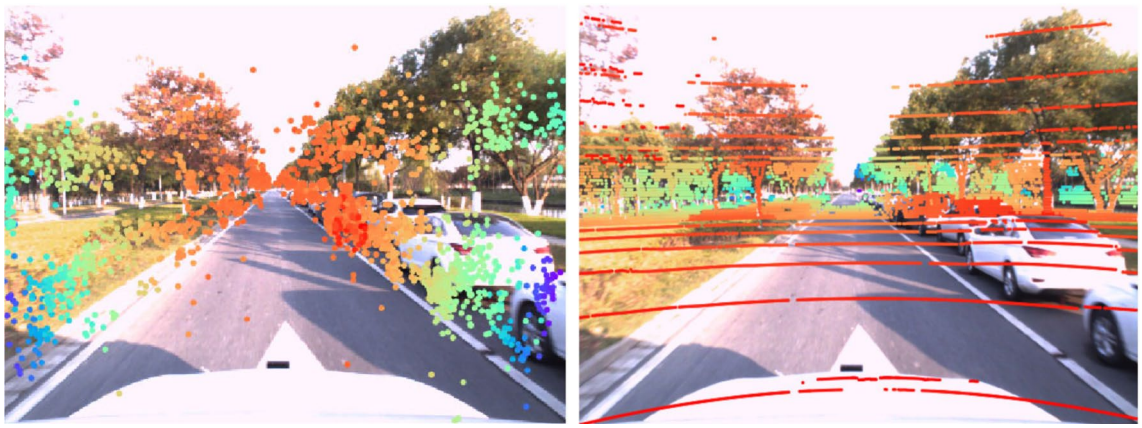


Fig. 1 Comparison of a radar (left) and lidar (right) "point clouds", projected into the captured scene with parking cars [3, p. 3]. The color represents the intensity-equivalent value per detection

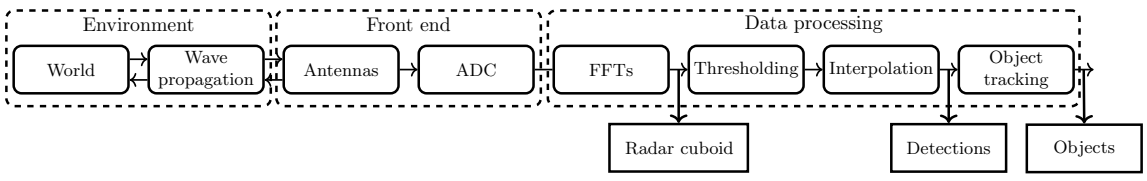


Fig. 2 An abstracted radar processing chain with elements visualized as blocks with rounded corners. The group within the processing chain is visualized as dashed rounded blocks and the sensor interfaces are marked as edged blocks

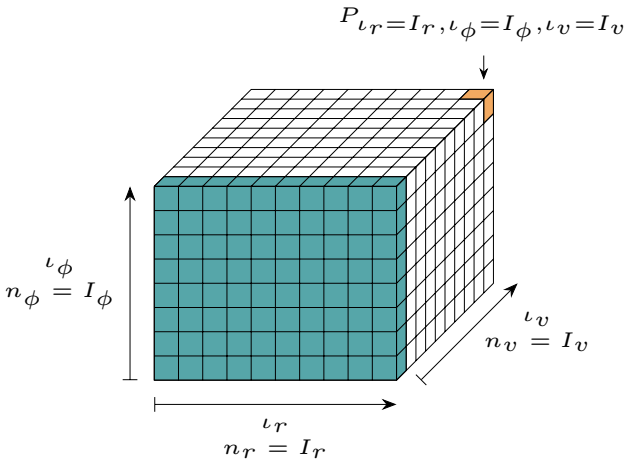


Fig. 3 Visualization of the radar cuboid with bins in range denoted as r, azimuth as ϕ and Doppler as v. The overall number of bins is I and one bin in the corresponding dimension is ι. In orange one bin with the power value P at the position I_r, I_ϕ, I_v is highlighted. The teal colored front marks the radar cuboid at the Doppler bin 0, which is of interest for static validation studies

Nevertheless, it will not be sufficient to have only a simulation model of a sensor available, but the validity of the sensor model must be proven along with its delivery. Only in this case simulation models can also be used in a trustworthy manner for safety assessments, as already envisaged by the UNECE [1, p. 5, 6]. Therefore, the complexity of radar measurements is due to noise based on multi-path propagation and RCS sensitivities. Also, the lack of public data sets and limitations in measurement repeatability is challenging [6]. Additionally, the number of detections depends on multiple causes and the noise of the sensor is complex due to e.g. the high frequency hardware components.

2 Related work on metrics for validation of active perception sensor simulation

In this chapter validation metrics from literature for radar model validation are listed and evaluated. Finally, an overview of the DVM and its application on lidar data is given.

2.1 Comprehensive review of already used validation metrics

According to Oberkampf and Trucano, validation is the "*process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model*" [7, p. 719]. Viehof [8, p. 91] introduced the so-called sample-vise validity in the context of vehicle simulation validation. Therefore, radar sensor model validation is understood as the comparison of synthetic and real sensor data with decent metrics, samplewise for a specific region of the desired parameter space of the application area. However, this means not only to be able to access the real sensor to simulate at the specific interface of interest, but also to design measurement campaigns according to the cause-effect chains that are modeled and investigated for a specific sample of the possible parameter space of the sensor model.

As described in the previous section, radar sensor modeling is a special case, as e.g. the relative velocity is directly measured and due to the complex interaction during radar wave propagation. Validation of radar sensor models is a complex task due to the complexity of radar measurements, as outlined in the previous section. Besides experiment design to optimize repeatability and reproducibility of measurements, while minimizing epistemic and aleatory uncertainties in reference data, the metrics for comparing real and synthetic data play a crucial role in validation. Aleatory uncertainty is a statistical deviation based on probability distributions in data. Epistemic uncertainty means a lack of information regarding model structure, world knowledge and measurement errors. Benke et al. [9] While epistemic uncertainties can and should be reduced by enhanced reference data collection, aleatory uncertainties describe the inherent randomness of measurements.

A first decision guidance for metric selection is provided by the seven criteria for validation metrics as refined by Rosenberger [10, p. 99] and e.g. used by Magosi [11, p. 11], which were condensed by Schaermann [12, pp. 20–21], combining the original lists of six criteria by Oberkampf and Barone [13, pp. 11–12] and the seven features from Liu et al. [14, p. 2]:

1. Metrics meet the mathematical properties of a metric as defined by Fréchet [15]. (Unbounded results)
2. Metrics are intuitive. (Plausible & output in unit of measurand)
3. Metrics are applicable to both deterministic and non-deterministic data.
4. Metrics are quantitative and objective. (No manually tuned parameters)
5. Metrics do not include acceptance criteria. (No Boolean output)
6. Metrics consider uncertainties. (Epistemic and aleatory)
7. Metrics define a confidence interval with respect to the number of measurement data.

The state of the art in validation metrics for active perception sensor simulation is extensively discussed by Rosenberger [10, pp. 60ff.]. Multiple metrics in this collection of 34 options are only indirect metrics, where detections are sorted into occupancy grids first or object detection and tracking is applied. These cannot be applied on radar data at earlier interfaces like radar cuboid or detection level. Other

metrics measure distances between points in space. These do not take into account intensity or power values. Therefore, they are not applicable to the radar cuboid, which is not a list of detections, but an equidistant distributed volume filled with power values.

Table 1 shows an excerpt of the remaining metric candidates for radar data. If a metric is capable of a category given by the column title, it is marked in a specific shade of green, otherwise the cells stay blank. The table considers the data interfaces that the metrics are or could be applied to (**D**: Detections, **O**: Objects). Then, the ability to be applied to ($\bullet$/ $\rightsquigarrow$: (Quasi) static/dynamic) scenario is provided. Additionally, in Table 1 the scale of measurement it is able to process is considered (**M**: Metric (interval or ratio), **O**: Ordinal). The uncertainties it is able to process ($\int$ /$\mapsto$: Aleatory/epistemic) are given per metric, as well. For the first four columns, it is marked if the metrics are applied without modification in literature (x), or if the metrics are applied in literature with moderate adaptions ($\star$). Additionally, the coverage of the seven criteria for validation metrics from the beginning of this section is marked in dark green.

Typical metrics that can be applied on object poses in space or detection coordinates like Manhattan distance (d_{Ma}) and overall error (OE) could be applied to the power values for a bin-wise comparison of a synthetic and a real radar cuboid. This also holds e.g. for mean error $\overline{d}$, root mean squared error (RMSE) and all other familiar error metrics from the collection that are not explicitly mentioned in this work, but listed in the original source collection [10, pp. 68–72]. Still, none of these strictly mathematical metrics accounts for aleatory or epistemic uncertainties, which disqualifies them to be applied for radar sensor model validation on the detection or radar cuboid interface, due to the stochastic and sensitive characteristics. The machine learning-based Deep Evaluation Metric (DEM) as introduced by Ngo [16] is used to measure an overall simulation-to-reality gap, but it does not cover most of the seven criteria. Additionally, due to the black-box nature of the results, there is a lack of interpretability, making it difficult to formulate safety arguments based on this metric. Also, it does not provide insights on how to calibrate and enhance a sensor model and is therefore not considered in the following. In Rosenberger's enumeration of metrics, the Mahalanobis distance d_{M} and its weighted variant are absent [17]. To achieve comprehensive coverage, Table 1 is supplemented with this metric. In the literature, this metric has already been applied to real and synthetic detection data from radar sensors [18, p. 33]. However, only the detection distribution over a binned bounding box is considered, without the inclusion of RCS or power values of the radar cuboid. Due to the metrics characteristic of not accounting for either aleatory or epistemic uncertainties, the Mahalanobis distance and its weighted form are excluded as metric candidates for radar validations.

Consequently, only the Kullback–Leibler divergence D_{KL} as applied by Schaermann [12], the Jensen–Shannon distance d_{JS} as used by Magosi et al. [11], the area validation metric (AVM) d_{AVM} introduced by Ferson et al. [19], and the Frequency of positive Kolmogorov–Smirnov tests f_{KS}, as applied by Eder [20] remain as metric candidates. This means that besides the AVM as best candidate based on 1, two families of metrics should be evaluated further, namely divergences and hypothesis testing. Clearly, both are not intuitive concepts for most people and involve some more abstract thinking compared to just computing the area between two curves, as done for cumulative distribution functions (CDFs) or empirical cumulative distribution functions (EDFs) in case of the AVM. Rosenberger presents a detailed analysis of both metric candidate families [10, pp. 105ff.], where the technique of manufactured universes [21] is used. Multiple EDFs are generated,

Table 1 Excerpt of the evaluation of metrics applied for active perception sensor simulation from Rosenberger [10, p. 99]

Metric	Interf.		Scen.		Scale		Unc.		Covered criteria						
	D	O	$\bullet$	$\rightsquigarrow$	M	O	$\int$	$\mapsto$	1	2	3	4	5	6	7
d_{Ma} (Manhattan distance)		x		x	x										
OE (Overall error)	$\star$		x	x	x										
$\overline{d}$ (Mean error)	x			x	x										
RMSE (Root mean square error)		x	x	x	x										
d_{M} (weighted Mahalanobis distance)	x														
D_{KL} (Kullback–Leibler divergence)	$\star$	x	x	x	x		x								
d_{JS} (Jensen-Shannon distance)	x			x	x		x								
d_{AVM} (Area validation metric)	$\star$	x		x	x		x								
f_{KS} (Kolmogorov-Smirnov)	x			x	x		x								
DEM (Deep evaluation metric)	x			x	x										

The Mahalanobis distance is added compared to the original table. Green color: Metric is capable of a category, Color gradient: different evaluation categories for readability, x/$\star$: Metric applied in literature without/with adaptions. The abbreviation "Interf." stands for Interfaces, "Scen." denotes Scenario, and "Unc." represents Uncertainty. The covered criteria are the seven mentioned metric criteria

where one is denoted as the real data and the others mimic simulated data to provoke edge cases for the metrics.

After showing hypothesis testing results for the different compared EDFs and a short summary on the ongoing discussions on these kinds of tests in general, this metric family is dismissed for sensor model validation due to the sometimes misleading and above all counter-intuitive results. They are not available in the unit of the measurand, which makes it less user-friendly e.g. in model specification, especially for negotiations with people with non-technical background. The same difficulties in interpretation of the results from comparing the different EDFs are present when applying Kullback–Leibler divergence or Jensen–Shannon distance, leading to not further considering them for sensor model validation, too [10, pp. 104–108].

The remaining metric candidate is the AVM that is simply the integral of the absolute difference between two CDFs $F, \widetilde{F}$ over all real and simulated sensor measurements

$$d_{\mathrm{AVM}}(F,\widetilde{F}) = \int_{-\infty}^{\infty} |F(\zeta) - \widetilde{F}(\zeta)| \, \mathrm{d}\zeta \, . \tag{1}$$

Due to the fact that the cumulated probability $F(\zeta)$ for each measurand ζ is limited to [0, 1] and unitless with m (e.g. 100) quantiles, the integral can be applied over the ordinate resulting in the mean error of all m quantiles of the CDF like

$$\begin{aligned} d_{\mathrm{AVM}}(F,\widetilde{F}) &= \int_{0}^{1} |\zeta(F) - \widetilde{\zeta}(F)| \, \mathrm{d}F \\ &= \frac{1}{m} \sum_{i=1}^{m} |\zeta(F_i) - \widetilde{\zeta}(F_i)| \, . \end{aligned} \tag{2}$$

Therefore, the AVM is very similar to the mean error of all n measurements

$$\overline{d} = \frac{1}{n} \sum_{i=1}^{n} |\zeta_i - \widetilde{\zeta}_i| \, . \tag{3}$$

As visible in Table 1, the AVM is the only metric that handles aleatory and epistemic uncertainties. This ability is reasoned by the fact that it is not only applicable on EDFs describing aleatory uncertainties reflected in the shape of the EDFs, but also on so-called probability boxes (p-boxes). A p-box is expressed by the left and right boundaries of multiple EDFs. The width of the p-box at each quantile describes the epistemic uncertainties, as in Fig. 4 for the two simulation EDFs. First introduced by Williamson and Downs [22], a p-box gives the possible interval of cumulative probabilities for a specific measurand x and for a given cumulative probability it gives a possible interval of values, as discussed in detail e.g. by Ferson et al. [23].

As epistemic and aleatory uncertainties should always be minimized during measurements, but can never be eliminated, they must be propagated through the simulation to reflect these uncertainties when the model is validated. Practically, this means that e.g. every position of a sensed object must be captured with reference sensors during the measurements to collect the real sensor data to validate the model. The uncertainty of this reference position measurement device, e.g. ±1.0 cm, is then input for multiple simulations per measurement, e.g. one with the exact reference position and two more for the edge cases of ±1.0 cm. These multiple simulations result in several EDFs and a combination of all EDFs from simulation forms the p-box. Its boundaries are composed of the maximum and minimum x-values of the set of EDFs for each y-value.

The AVM for p-boxes is simply calculated by adding the two portions where the simulated p-box $\widetilde{\boldsymbol{\mathcal{F}}}$ is higher (d^+) or lower (d^-) than the real p-box $\boldsymbol{\mathcal{F}}$ as

$$d_{\mathrm{AVM}}(\boldsymbol{\mathcal{F}},\widetilde{\boldsymbol{\mathcal{F}}}) = d^- + d^+ \, . \tag{4}$$

For simplification, in Fig. 4 the EDF F is an infinitely thin p-box. Consequently, the AVM only considers the left and the right borders of the p-box, the original course of each EDF inside is irrelevant, and the borders could actually originate from different EDFs.

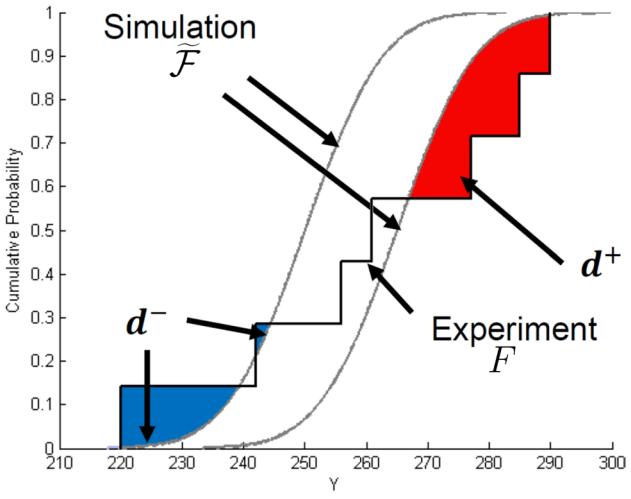


Fig. 4 Portions of the AVM, where the simulated p-box $\widetilde{\mathcal{F}}$ is higher (d^+) or lower (d^-) than the real EDF F, based on Voyles and Roy [24]

2.2 The DVM and its application in validation of lidar sensor simulation

An additional requirement on validation metrics is the ability to distinguish model bias and model scattering error to enable the structured elimination of these two different modeling errors [10, p. 72]. Model bias is an approximation of the mean deviation and model scattering error the deviation in the distribution function's shape. Figure 5 illustrates the difference between model and measurement bias and also shows the difference between the measurement standard deviation and the model scattering error. Indeed, measurement bias and scattering error are conceptually similar to the differences in mean and variance between a set of normal distributions. However, the distribution functions of the measurand can deviate from normal distributions.

Rosenberger therefore introduced the DVM that distinguishes the two components [10, p. 118ff.]. The first part is essentially the difference of d^+ and d^- in comparison to the original AVM as sum of these two portions. Voyles and Roy [24] introduced this difference that is proven to be a good estimate for the model bias as

$$d_{\text{bias}}(\mathcal{F}, \widetilde{\mathcal{F}}) = d^- - d^+ . \tag{5}$$

It eliminates symmetrically distributed area portions of the AVM, which reflect the model scattering error and therefore only keeps the model bias. Consequently, (5) can be used to estimate a "*corrected*" [24] p-box as

$$\widetilde{\mathcal{F}}_{\text{c}}(\zeta) = \widetilde{\mathcal{F}}(\zeta - d_{\text{bias}}) = \widetilde{\mathcal{F}}\Big(\zeta - (d^- - d^+)\Big). \tag{6}$$

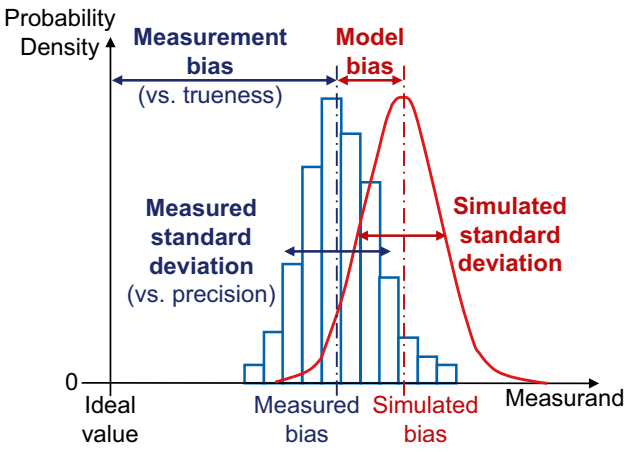


Fig. 5 Bias and scattering of measurement and model [10, p. 11]. In the case of normal distributions the two factors are the mean value and the standard deviation

Taking this idea of a corrected, bias-free simulated p-box further, a second-order AVM can be computed with $\widetilde{\mathcal{F}}_{\text{c}}$ that now only entails the remaining model scattering error. This novel metric introduced by Rosenberger [10, p. 118] is called corrected AVM (CAVM) and formulated as

$$d_{\text{CAVM}}(\mathcal{F}, \widetilde{\mathcal{F}}) = d_{\text{AVM}}(\mathcal{F}, \widetilde{\mathcal{F}}_{\text{c}}) = d_{\text{c}}^- + d_{\text{c}}^+ . \tag{7}$$

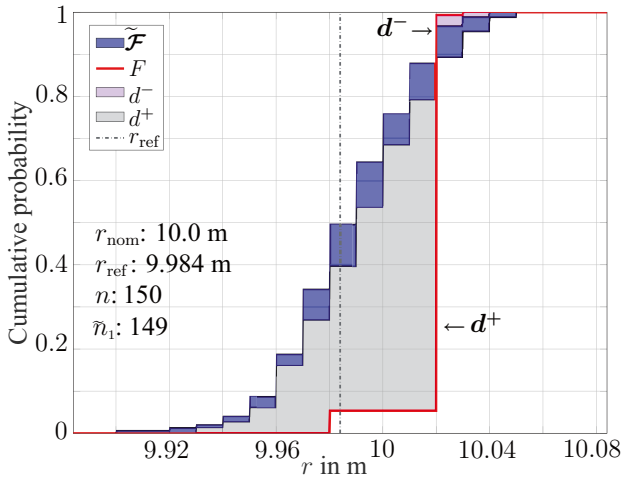
As illustrated in Fig. 6, the CAVM is a multistep process that inherently includes the calculation of the model bias on its way. It starts with the calculation of d^+ and d^- for d_{bias} (5). Then the simulated p-box is corrected by d_{bias} to get $\widetilde{\mathcal{F}}_{\text{c}}$ (6). Finally, d_{c}^+ and d_{c}^- are calculated, resulting in d_{CAVM} (7).

Consequently, the novel DVM for comparison of two p-boxes $\mathcal{F}$, or EDF F as infinitely thin p-boxes, is achieved that distinguishes model bias and model scattering error with respect to the actual sensor bias and its real scattering behavior, as

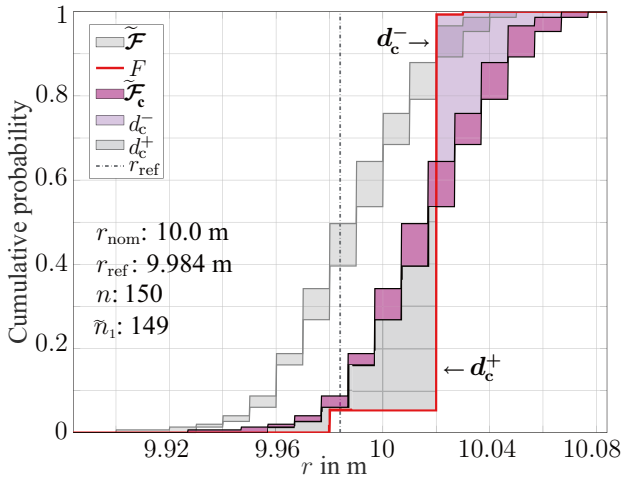
$$d_{\text{DVM}}(\mathcal{F}, \widetilde{\mathcal{F}}) = \Big(d_{\text{bias}}(\mathcal{F}, \widetilde{\mathcal{F}}),\ d_{\text{CAVM}}(\mathcal{F}, \widetilde{\mathcal{F}}) \Big) . \tag{8}$$

Aligned with distinguishing model bias and model scattering error, model validation should start simple to reach more complex scenarios later. Accordingly, Rosenberger starts with demonstrating the DVM for beam-wise model evaluation in static scenarios, like targets in different distances with no other effects taking place, to more complex and object-wise validation of synthetic lidar detections. Aside from the interpretability of the results in the unit of the measurand, the accuracies are considered by the reference tests using the p-boxes. Consequently, for radar model validation, the validation should follow this incremental approach. The experimenter should take special care to consider isolated cause-effect chains, which influence the radar signal propagation. A possible ontology to derive them is PerCollECT [25] as available on Github [26].

Besides the DVM there are other applications of the AVM described in the literature. Brune et al. compare in their application the left and right edges of the measurement and simulation p-boxes with each other [27]. Compared to the DVM, this approach offers the advantage that the p-box size is included in the metric result. This aspect is missing in the DVM according to Rosenberger. However, Brune's AVM does not explicitly consider the scattering error of the distribution function, which means that the information about the shape similarity of the p-box is lost. Figure 7 shows the AVM according to Brune et al. and illustrates how the size of the p-boxes is incorporated into the result of the AVM.



(a) Calculation of d^+ and d^- for d_{bias}.



(b) CAVM calculation with $\widetilde{\mathcal{F}}_{\mathbf{c}}$.

Fig. 6 Illustration of the CAVM [10, p. 118] in a lidar sensor measurement campaign. A plate is placed in front of the sensor at 10 m in a static scenario. r_{nom} and r_{ref} are the nominal and measured reference range. n and $\tilde{n}_1$ are the number of detections from real data and simulation sim1, which is a control factor for the comparison of the different EDFs [10, p. 103]. These F is the EDF from real data. $\widetilde{\mathcal{F}}$ is the p-box from simulation. $\widetilde{\mathcal{F}}_{\mathbf{c}}$ is the simulated p-box corrected with the estimated model bias d_{bias}. d^+ and d^- mark areas where the simulated p-box $\widetilde{\mathcal{F}}$ is higher (+) or lower (-) than the real EDF F. d_c^+ and d_c^- mark areas where the corrected simulated p-box $\widetilde{\mathcal{F}}_{\mathbf{c}}$ is higher (+) or lower (−) than the real EDF F

2.3 Double validation metric limitations

As mentioned in Sect. 1, the measurement of the radial relative component of the velocity is possible with radar sensors. This allows the discussion of the DVM to model validation

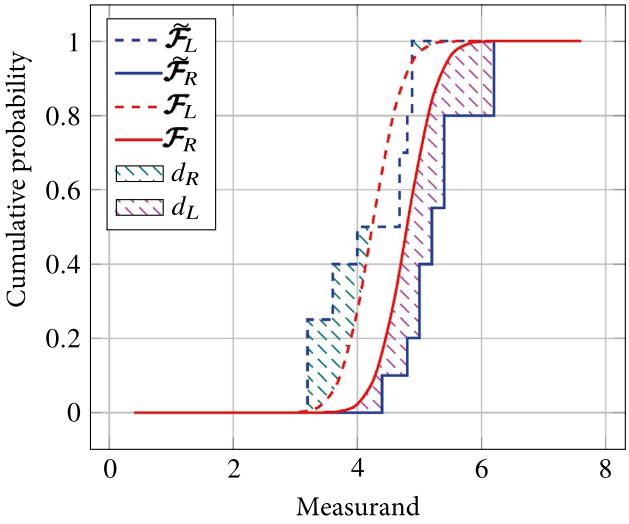


Fig. 7 Visualization of the AVM calculation d_L and d_R based on the left and right border comparison between measurement $\mathcal{F}_L$, $\mathcal{F}_R$ and simulation $\widetilde{\mathcal{F}}_L$, $\widetilde{\mathcal{F}}_R$, based on Brune et al. [27]. In the context of radar sensor data, the measurand could pertain to parameters such as the range of the detection distribution or the distribution of power values within the radar cuboid

in dynamic scenarios. As already shown by Holder, even simple measurement scenarios on a proving ground are subject to difficulties regarding repeatability of complex geometries that must be taken into account [28]. Figure 8 shows the variation of the RCS denoted as $Q_{(\sigma)}$ over the distance in the radar sensor coordinate system denoted as ${}_{\text{S}}r$ of a retroreflector, which is a corner cube reflector (CCR), and a vehicle.

Therefore, to apply the metric, it is advisable to introduce p-boxes for the measurement data to consider the limited measurement repeatability and take them into account in the metric result. Based on Fig. 8 and the limited reproducibility the size of the measurement p-box and the distribution of EDFs must be part of the metric.

Additionally, the size of the simulation p-box is a factor influencing the quality of the DVM. These properties are missing in the DVM according to Rosenberger because these characteristics are lost during the transformation of EDFs to p-boxes. Theoretically constructed ECs based on the method of manufactured universes substantiate the previous remarks.

These ECs can appear in the application of the validation methodology to radar due to the sensitivity of the model to small changes in reference sensor measurement uncertainties, but also due to the problem of reproducibility of measurements.

Figure 9 shows the first ECs with the results of the Rosenberger's DVM and the evaluation according to Brune et al. [27]. It addresses, on the one hand, the size of the p-box and the overlap of the simulation and measurement p-box. To

proof the independence of the methodology regarding the number of simulation and measurement EDFs, the combination of EDFs is varied.

Figure 9 illustrates EC 1 consisting of a large simulation p-box in comparison to the measurement. Additionally, the right side of the simulation p-box equals the left side of the measurement p-box. The AVM and the CAVM are both 0, and therefore, the model is valid based on the DVM. The extension by Brune of the AVM covers this EC by resulting in a d_{L}, which has the same value as the simulation's p-box size.

Figure 10 shows EC 2 with a large measurement p-box in comparison to the simulation. A concentration of measurements EDFs on the right side is also present. For Rosenberger's DVM the simulation results are only the deviation between the right simulation and left measurement EDF.

The AVM is small and the CAVM is almost 0, because the shape of the right simulation and left measurement EDF are nearly identical. Also, the AVM definition by Brune fails in this case. The value for d_L covers just the offset between the two p-boxes.

Figure 11 visualizes EC 3 consisting of two simulation EDFs at the left and the right side, which are very similar

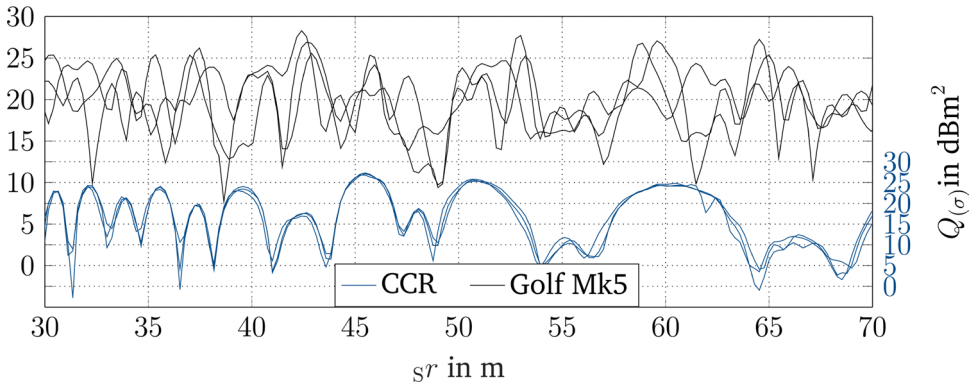


Fig. 8 RCS experiment trials of a CCR in blue and a Golf Mk5 in black from Holder [28]

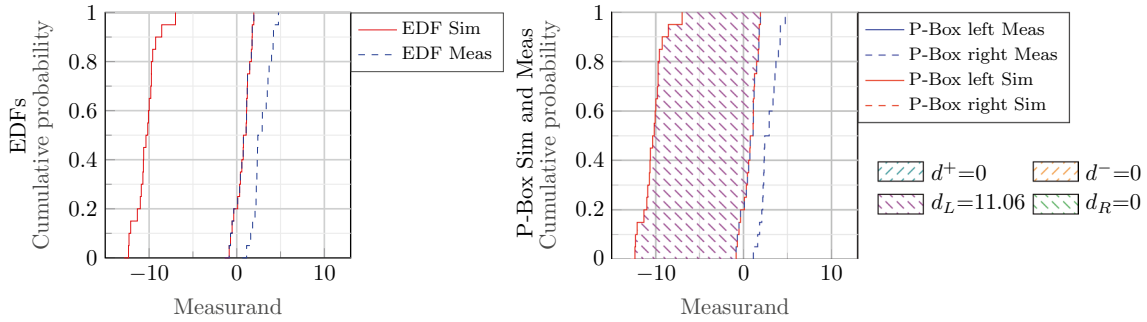


Fig. 9 Edge case (EC) 1 of the DVM definition by Rosenberger and AVM calculation by Brune. On the left side the EDFs are visualized. On the right side the p-boxes of an imagined measurement and simulation are shown with the corresponding colored area

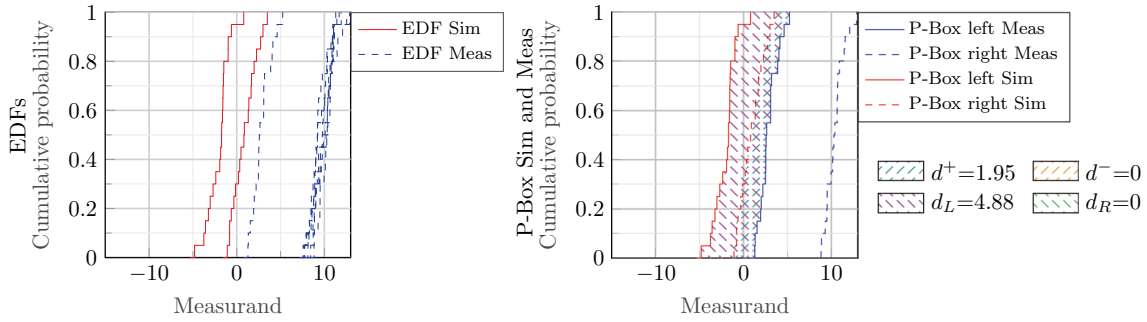


Fig. 10 EC 2 of the DVM definition by Rosenberger and AVM calculation by Brune

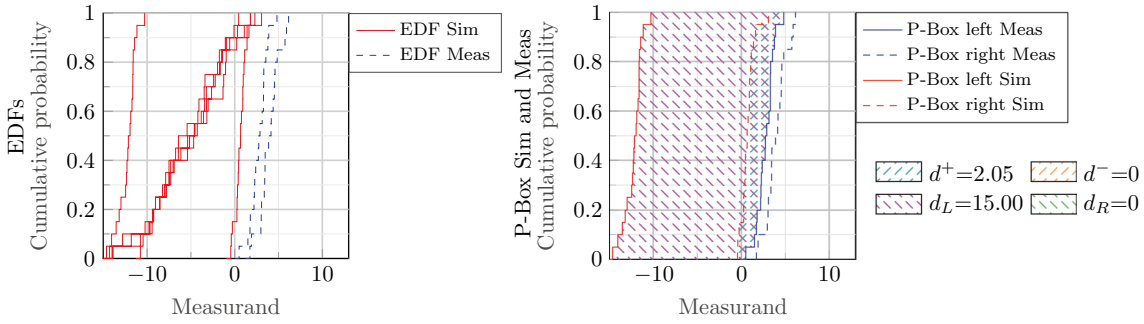


Fig. 11 EC 3 of the DVM definition by Rosenberger and AVM calculation by Brune

in shape. Between the outer distributions there are several simulation EDFs that show a completely different distribution shape. By forming the simulation p-box, the information of the inner distributions is lost.

This leads to the fact that the deviations in the shape of the distribution function are insufficiently considered in the DVM validity consideration. Brune's AVM covers the big size of the simulation p-box but fails with identifying the inner distribution functions.

Figure 12 illustrates EC 4, which is similar to number 2, where the distribution of the simulation distribution functions has no effect on the validation result itself. In the case of Brune's AVM, only the total deviation of the p-boxes is quantified.

However, the important information of the accumulation of distribution functions on the right side of the simulation p-box is lost, which could be helpful for the modeler and the experimenter. In the context of simulation, the aggregation of distribution functions suggests that the model's underlying parameterization exhibits heightened sensitivity at a specific point, resulting in a substantial deviation from other simulations. This observation is mitigated by the utilization of the p-box. Conversely, for the experimenter, such aggregation signifies an outlier within the measured data, indicating potential issues such as erroneous execution, deviation from the intended scenario, or significant alterations in environmental conditions.

The presented ECs show the shortcomings of the two validation metrics. Especially the distribution of the EDFs within the p-box borders are disregarded. Another disadvantage is that an assignment of the EDF outliers, which results from parameters varied in the simulation, to the validation result is excluded. Generally, the introduction of p-boxes obscures the validation result and leads to potential misinterpretation. Therefore, the methodology to apply the DVM has to be extended and modified.

To ensure comparability of simulated and real data based on DVM, it is essential to include the number of data points in the evaluation process. A deviation in the numbers between simulation and measurement of 10% is considered as an acceptable limit in the remainder of this paper by the authors. The deviation in data points occurs due to time effects in the radar sensor, which are not covered in the simulation model with a predefined sample frequency. At the detection level, this is particularly problematic in the case of the radar sensor, as the number of detections can vary greatly between individual measurement cycles, for example due to clutter from vegetation or rain.

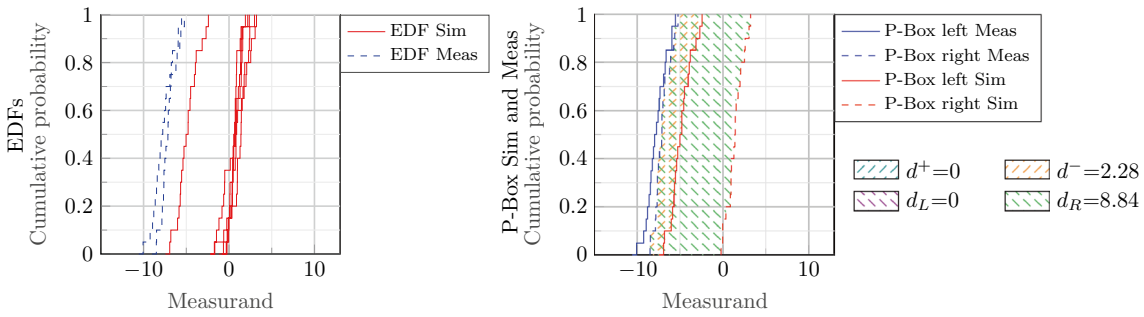


Fig. 12 EC 4 of the DVM definition by Rosenberger and AVM calculation by Brune

3 Radar validation methodology

In this chapter the methodology for determining the DVM is adapted. Furthermore, it is shown how the new methodology can be applied to radar data.

3.1 The DVM map

To overcome the DVM limitations of simulation and measurement p-box size as well as distribution and shape of EDFs an adaption of the DVM is necessary. Figure 13 shows the adapted validation methodology based on the AVM and CAVM metric. Compared to Rosenberger, the measurement and simulation p-boxes are resolved, and each simulation is compared to each measurement, deriving the new "DVM Map" as a validity tool visualized as a heat map.

In a first step, the AVM is formed for each simulation EDF in combination with each measurement and corrected by the determined d_{bias} according to (5). The absolute value of $|d_{\text{bias}}|$ is used for visualization in the DVM map so that for negative and positive model bias, the color value of the scale is unambiguous. Following the aforementioned procedure, the corresponding CAVM is formed and d_{CAVM} determined according to (7). This results in a value for the model bias and the scattering error for each simulation in comparison to each measurement. The scattering error finds an intuitive explanation in the shape deviation of the corrected simulation EDF in comparison to the measurement EDF. To calculate the overall comparison score d_{Sum} the absolute value of the model bias is added with the CAVM result as

$$d_{\text{Sum}} = |d_{\text{bias}}| + d_{\text{CAVM}}. \tag{9}$$

These deviations of the simulation can be quantified in comparison to all measurements. Therefore, the DVM Map shows the most critical measurement and simulation of the corresponding reference data uncertainty. When the results of the DVM Map need to be further processed, the maximum of d_{Sum} is formed, thus identifying the most critical combination of measured and simulation data for sample validation. To demonstrate the utility of the newly developed DVM Map, the ECs from Figs. 9, 10, 11 and 12 are re-examined.

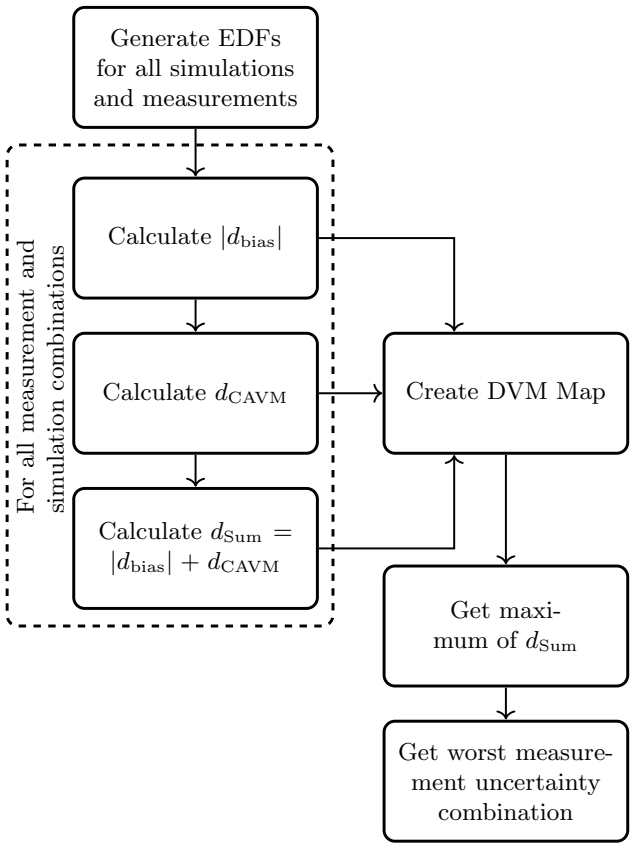


Fig. 13 This figure shows the adapted DVM methodology to address the different ECs with the new DVM Map as intermediate step

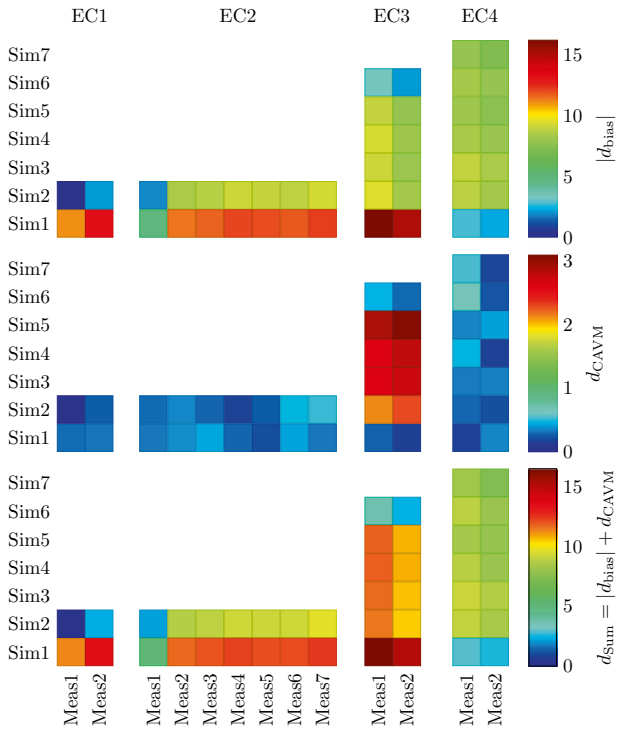


Fig. 14 DVM Map of the ECs defined in Figs. 9, 10, 11 and 12 processed with the newly introduced methodology. From left to right, the various ECs are shown. From top to bottom the DVM Map of $|d_{\text{bias}}|$, d_{CAVM} and d_{Sum} is illustrated. The coloring corresponds to the quantitative deviation of the corresponding value, where blue means low and red means high. The upper and lower limit of the scale is determined by the minimum and maximum value in the corresponding DVM Map

Figure 14 is sorted by the ECs from left to right and shows the results for $|d_{bias}|$, d_{CAVM} and d_{Sum} from top to bottom. The DVM Map of EC 1 shows that the large model bias of simulation 2 can be compensated. The CAVM shows minor deviations since the shape of all EDFs is similar to each other. It can be seen that the combination of measurement 2 and simulation 1 performs worst in terms of sample validity.

For EC 2, the focus is also on model bias examination. The accumulations of EDFs on the right-hand side are clearly evident, with simulation 1 deviating more than simulation 2. The result also shows up in d_{sum}, resolving the EC.

EC 3 shows that the distribution of the different simulation EDFs can be mapped using the d_{CAVM}. However, the influence of d_{bias} predominates for simulation 1.

EC 4 shows that the DVM Map can also reproduce high deviations of the simulation and thus reflects valuable information regarding sample validity and the influence of various aleatory uncertainties.

In Sect. 2.1, the seven criteria for metric selection are enumerated. As the DVM is based on the AVM, which is listed in Table 1, and the DVM Map is derived from the DVM, there exists an association among the criteria of the three metrics. Nevertheless, the DVM Map is evaluated based on the seven criteria outlined subsequently. Given that the DVM Map exhibits no boundary in its outcome, the first condition is met. As demonstrated by examples from the ECs, the results are plausible and consistent with the unit of the corresponding measure, facilitated by the use of the DVM. Furthermore, the metric is applicable to both deterministic and non-deterministic data. Additionally, aleatory and epistemic uncertainties are accounted for through the measurements and simulations of reference data. These two aspects are illustrated in the evaluation presented in Sect. 5. The metric itself possesses no tuning parameters and lacks any acceptance criteria. Hence, these two points from the list of criteria are also fulfilled. Confidence intervals can be defined based on the measurement data. However, this is not explicitly demonstrated in the subsequent evaluation.

3.2 DVM map application on radar sensor interfaces

Figure 15 shows the methodology application to radar cuboid and detections in a study of sample validity of a radar model. The first step is to run defined scenarios in the real world (bright red). In addition to the measurement data at detection and radar cuboid level, the measurement campaign yields the operational reference data, which is subject to epistemic and aleatory uncertainties. Operational reference data means in this context the ability to take a reference measurement with additional sensors independent of the radar sensor. The uncertainties are determined by means of reference sensors or reference sources.

The measured reference data is transferred to the simulation (light blue) in a further step. Here, in addition to the measured reference value, the epistemic uncertainties are propagated through the simulation. As a result, simulation data on detection and radar cuboid level are available, where the number of simulations depends on the number of simulated uncertainties.

Different variants of the new validation methodology can then be applied to the measurement and simulation data. First, a very rough consideration of all detections and all cells of the radar cuboid is advisable. From this, basic deviation of the sensor model from the measured data can be derived, as well as which uncertainty combination together with which measurement shows the largest measurand deviation. This allows conclusions to be drawn about gross modeling errors, as well as measurement outliers, provided the number of measurements is large enough to identify outliers. While more measurement data of longer time periods and / or repetitions (at least three repetitions should be required) is always better for the analysis, the number of measurements in practice is mostly restricted by ecological factors like available time at test tracks and personal.

Detections and the corresponding bins in the radar cuboid are of great interest of a validation study. This can be justified firstly by the fact that detections represent the input for all subsequent steps of radar processing, and secondly that here, either due to the environment or due to objects, power differences are present that a simulation model should represent.

Therefore, the time aggregated detection data from all measurements are clustered using a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and thus a region of interest is defined, which in turn can be transformed into cells of the radar cuboid. By applying the new methodology, individual areas of particular interest in the radar measurement are testable in a dedicated manner.

Finally, an application of the radar validation methodology to each individual cell of the radar cuboid is performed. Local effects and influences are the focus of the investigation and provide valuable information about the model sensitivity with respect to the reference sensor uncertainties. By matching the results to a satellite image detailed investigations of the influence of the environment are possible.

4 Application validation methodology

In this chapter the aspects considered theoretically so far are verified in the following by means of a validation study. For this purpose, a measurement campaign is carried out on the August–Euler airfield proving ground in Griesheim. The

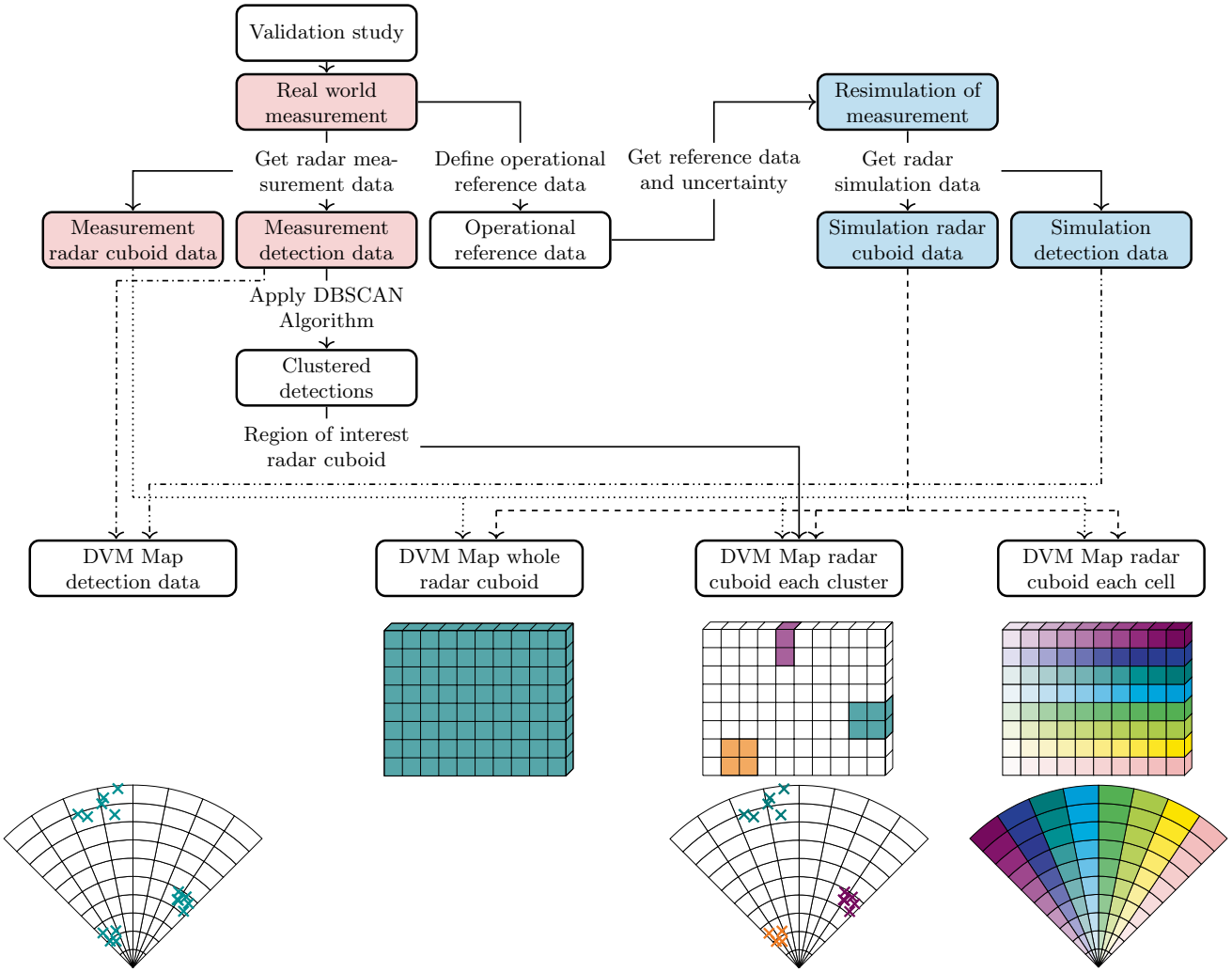


Fig. 15 Validation methodology to apply the DVM Map to different levels in the radar processing chain. From left to right the application of the DVM map to all detections, to the entire radar cuboid, the region of interests, which result from time aggregated, clustered detections, associated with the corresponding radar cuboid cells and each radar cuboid cell separately is visualized

study is intended to serve as a proof of concept of the DVM Map, therefore a static scenario is measured. Moreover, within the context of sensor model validation, it is deemed essential to commence the process by scrutinizing simple scenarios to iterative rectify any errors inherent in the model. Consequently, the selected example is designated as the initial, static experiment within a broader validation campaign. Consequently, subsequent analyses refrain from a detailed investigation of the Doppler component within the dataset, with deference given to forthcoming publications for a more comprehensive examination of this aspect.

4.1 Experimental setup

The object in the validation study is a CCR lying flat on the asphalt, which is placed on different positions in front of the radar sensor. The positions differ in the range r to the radar sensor and the azimuth angle ϕ. For r, 29.56 m and 48.33 m are chosen, respectively, due to the radial resolution, so that the CCR is located once close to the edge and once in the center of the range bins. For ϕ, positive as well as negative angles are defined so that the CCR is within the sensor's unambiguous azimuth measurement range and the width of the test site is sufficient. Thus, angles of $-8\,°$, $-4\,°$, $0\,°$, $4\,°$, $8\,°$ are obtained. Figure 16 shows a sketch of the measurement setup with the different CCR positions, which are measured one after another, on the left side and the real world measurement setup for position 1 on the right side. The position of the CCR and the sensor is obtained using a real time kinematic (RTK)-based global navigation satellite system (GNSS) antenna with a measurement uncertainty of 0.02 m. For each measuring position, 5 measurements of 60 s each are recorded with the radar sensor resulting in approx. 850 samples.

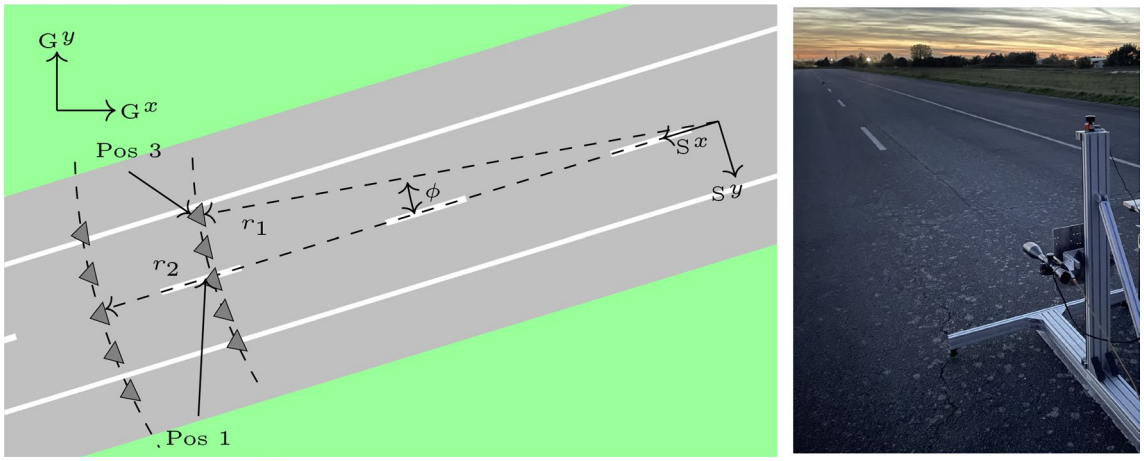


Fig. 16 Experimental setup of the validation study. CCR position 1 ($r_{\text{CCR, Pos 1}} = 29.56$ m, $\phi_{\text{CCR, Pos 1}} = 0$) is used for simulation calibration purposes and position 3 ($r_{\text{CCR, Pos 3}} = 29.56$ m, $\phi_{\text{CCR, Pos 3}} = -8$ °) is analyzed based on the presented methodology

Table 2 Measured reference data uncertainties for CCR position 3 defined in Fig. 16. The local Cartesian coordinate system G is defined in East-North-Up direction with the origin located on the August–Euler airfield in Griesheim

Reference data	Variable	Measurement	Uncertainty
Sensor azimuth orientation in G	${}_{\text{G}}\phi_{\text{S}}$	197.91 °	± 0.07 °
Sensor *x* position in G	${}_{\text{G}}x_{\text{S}}$	977.43 m	± 0.02 m
Sensor *y* position in G	${}_{\text{G}}y_{\text{S}}$	241.56 m	± 0.02 m
Sensor height in G	${}_{\text{G}}h_{\text{S}}$	12.89 m	± 0.02 m
CCR *x* position in G	${}_{\text{G}}x_{\text{C}}$	948.33 m	± 0.02 m
CCR *y* position in G	${}_{\text{G}}y_{\text{C}}$	216.46 m	± 0.02 m
Edge length CCR	l_{C}	0.240 m	± 0.005 m

The measurement setup and the existing measurement technology result in the uncertain reference data listed in Table 2. Here, some parameters are determined directly and others result from the propagation of error of several measurement uncertainties. In case of the sensor height and the edge length a measurement tape is used with a given measurement uncertainty of 0.005 m. The sensor orientation is measured via a reference target and the RTK-based GNSS antenna resulting in 0.07 °.

4.2 Simulation model

The reference data uncertainties listed in the table are transferred to the simulation in a next step. Reference data uncertainty propagation through simulation is realized by defining separate scenarios with the upper $^{+}$ and lower limit $^{-}$ of this uncertainty. Additionally, one simulation with all uncertainty-free reference measurement data denoted as *N* is integrated. For the radar simulation, the output of a black box radar ray tracing algorithm by IPG CarMaker version 9.1.1 and an adapted open source radar signal processing model [29] is used.

The input from the ray tracing algorithm is interpreted in the radar model as a delta peak in frequency space and is called "Fourier tracing" [28]. The range and angle information of the ray is used to calculate the radar cuboid bin in the different dimensions in which the delta peak is located. A windowing function is used to smear the power of the delta peak into the neighboring bins. This allows effects such as ambiguities, separation capabilities and interferences to be present in the radar model. In addition, a non-deterministic noise simulation is implemented to demonstrate the metric capabilities even for such modeling approaches. Based on measurements, the mean and standard deviation are determined for each range-azimuth cell combination. A Gaussian distribution with the determined parameters is then imposed on cells whose minimum power is below the noise floor. Figure 17 shows the determined mean and standard deviation as a range azimuth map. This result is conducted by placing the sensor on the asphalt with its front side pointing into the sky.

The radar model is parameterized using the data sheet and calibrated with position 1 of the CCR on the detection interface. Additionally, a simulation model of the August–Euler airfield in Griesheim is used for the environment simulation.

5 Results and discussion

In this chapter, the different results of the DVM Map are shown and discussed. Therefore, the methodology defined in Sect. 3 is applied to the static validation scenario. The DVM Map with the new validation methodology is applied to the different interfaces as described in Figs. 13 and 15. In Sect. 5.1 all detections are analyzed. Afterwards in Sect. 5.2 the results of the whole radar cuboid is discussed. The

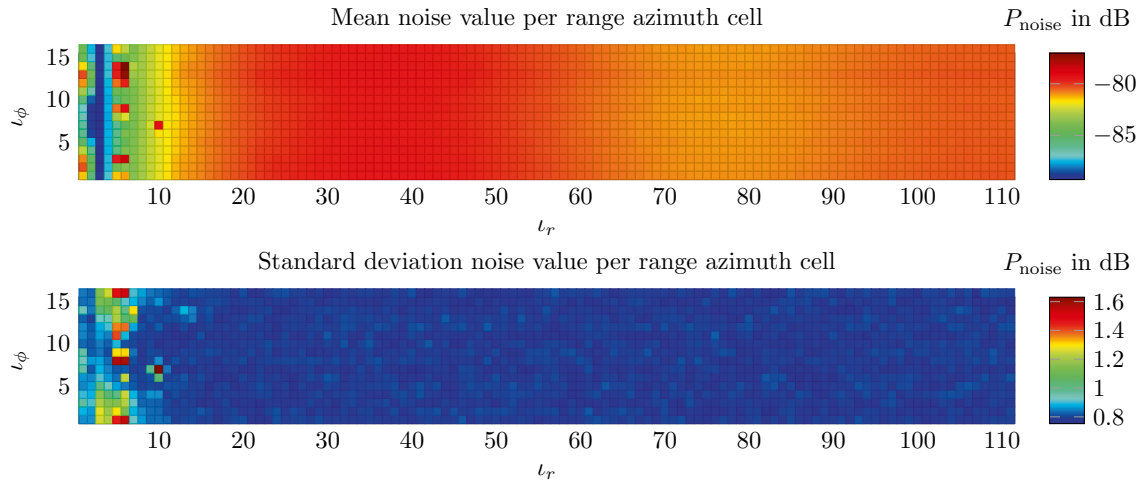


Fig. 17 Mean value and standard deviation illustrated as range azimuth map for a Gaussian distribution noise model

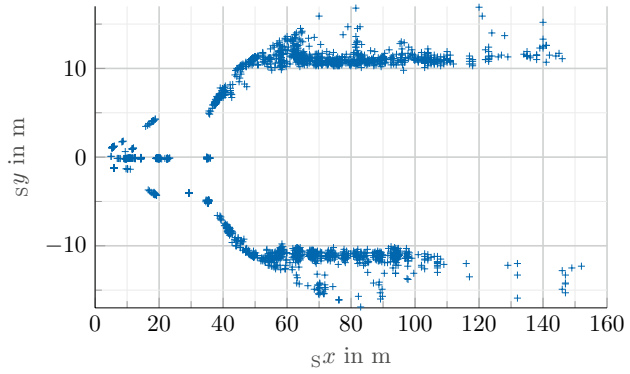


Fig. 18 Plot of all detections in the Cartesian sensor coordinate system. Especially in near range of the sensor as well as at the transition of the asphalt and vegetation clutter is present. The CCR is located at ${}_Sx = 29.4$ m and ${}_Sy = -4.2$ m

region of interest in the radar cuboid by means of clustered detections is described in Sect. 5.3. Finally, the results of each cell of the radar cuboid combined with a visualization on top of a satellite image are shown. As an example, the previously mentioned scenario of CCR position 3 at $r_{\text{CCR, Pos 3}} = 29.56$ m and $\phi_{\text{CCR, Pos 3}} = -8\,^\circ$ is evaluated.

5.1 Whole detections

This section contains the evaluation for all detections of the measurements and simulations. Figure 18 shows all detections of all measurements combined in one diagram in Cartesian coordinates, where the origin is the sensor position. Therefore, the EDFs of the different distributions of the quantities r, ϕ and σ are calculated.

The representation of the detections as distribution functions is visible the top row of Fig. 19. The EDFs of the different measurements are close to each other and the deviation of numbers of detections is less than 1%. This indicates a high reproducibility of the measurement results. However, there is a significant difference in the number of detections between simulation and measurement, which prevents a valid evaluation based on the DVM.

Nevertheless, further analysis of the data will be conducted to describe and analyze the general methodology. The distributions of the detections from the simulation are almost at the same position in every cycle for the range as well as the azimuth angle. However, deviations of the simulations in RCS show up. The analysis of the RCS shows that the model reacts very sensitively to small changes of the scenario, which again emphasizes the necessity to consider the reference sensor uncertainties.

The basically different distribution functions between measured and simulated data result from the environment model of the August–Euler airfield as well as the ray tracing simulation. The ray tracer used in the CarMaker version has no reflections from the road surface in the given setup as well as no reflections from the vegetation next to the asphalt surface due to the lack of simulated vegetation. Therefore, only rays from the CCR are processed as radar signal processing model input. Thus, only detections at the object result, whereby besides the small number of detections also the distribution in range and azimuth are limited to the position of the CCR. However, the detections of the CCR are in the same distance in measurements and simulations. This can be seen particularly well in the second row at $r = 29.56$ m as well as $\phi = -8\,^\circ$ by the step in the measurement EDFs highlighted by the red ellipse.

The second row of the Fig. 19 shows the simulation EDFs $\widetilde{F}_c$ corrected by the model bias d_{bias} based on the AVM calculation to measurement 1. This represents the second step of the methodology from Sect. 3. It is already evident, without

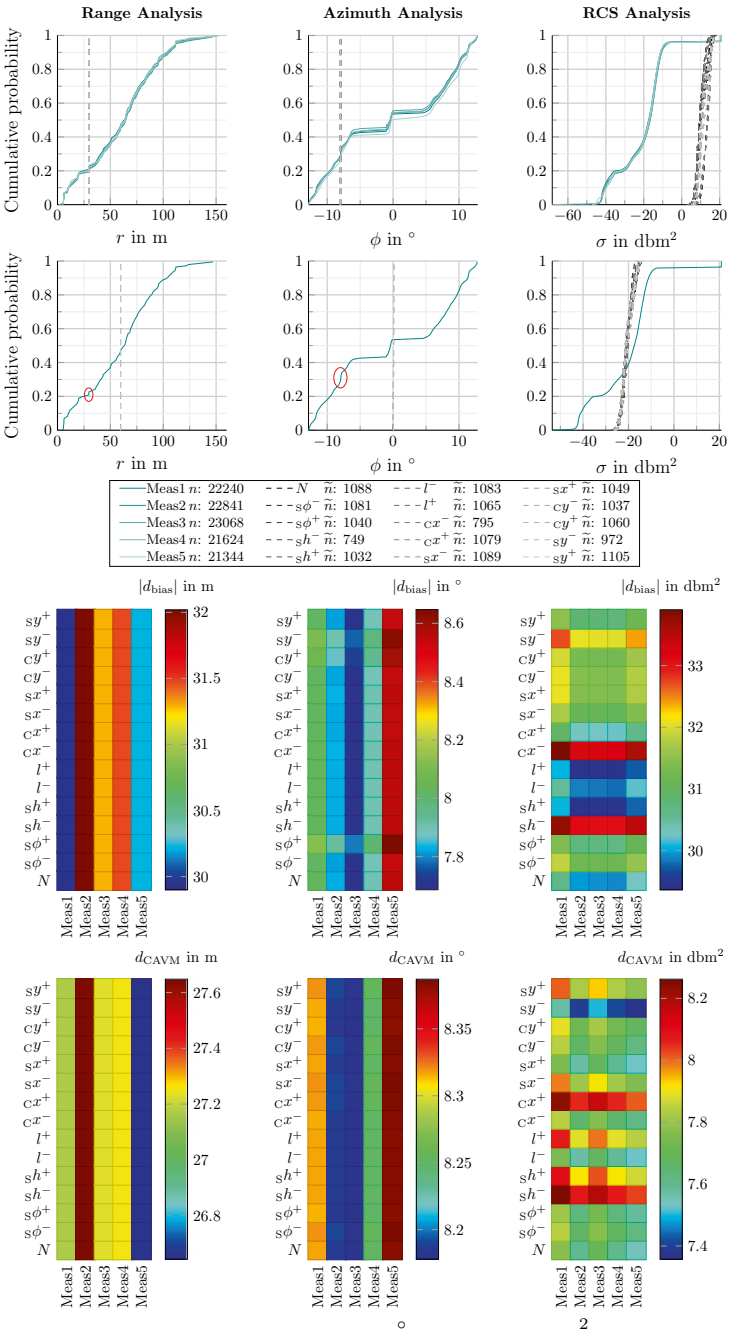


Fig. 19 From left to right are the range in m, azimuth in °, and RCS in dBm^2 results of all detections. The first row visualizes the EDF of all measurements F and of all simulations $\widetilde{F}$. In the second row the EDF of measurement 1 F_{Meas1} and the corrected simulation EDFs $\widetilde{F}_{\text{c}}$ are shown. The red circles illustrate the data points in the measurement EDFs where the CCR is located. The last two rows show the DVM Map of the above-mentioned quantities. The number of detections of the measurement n and simulation $\widetilde{n}$ is in the legend

a quantitative determination of the deviation by means of d_{CAVM}, the fundamental difference of the distribution functions.

The DVM Map in Fig. 19 verifies the findings already made from the first visual impression. In the range domain, the constancy of the simulation of d_{bias} and d_{CAVM} stands out. In the range, the simulation model is insensitive to the simulated uncertainties. This is due to the fact, that the measurement uncertainties are very small compared to the radar's range resolution of 1.8 m. Measurement 1 and 5 show the smallest deviations in the model bias and measurement 2 the largest deviation. The scattering error of measurement 2 is largest and lowest for measurement 5 for the range dimension. Thus, the differences in the results of the DVM Map are due only to the differences in the measurements.

In the case of the azimuth dimension, the model is slightly more sensitive, as shown by the minor color changes in the perpendicular components of the DVM Map. These variations are small compared to the measurement influence shown in the horizontal of the various uncertainties. Here, the negative variations of the y-position and the rotation ϕ

of the sensor are the uncertainties with the most notable influence on the result.

For the RCS, in turn, the negative x-position of the CCR variation and the negative sensor height variation have the largest influence. All parameters have a clear but different influence on the RCS distribution of detections. In addition, it is shown that the DVM Map is able to represent the different positions of the EDFs with respect to each other in a very intuitive and simple way.

The influence of the environment in combination with the ray tracing algorithm as stated above is simply too large to make a validity statement by means of the analysis of all detections. However, the analysis of all radar detections can be used to compare stochastic effects from the environment.

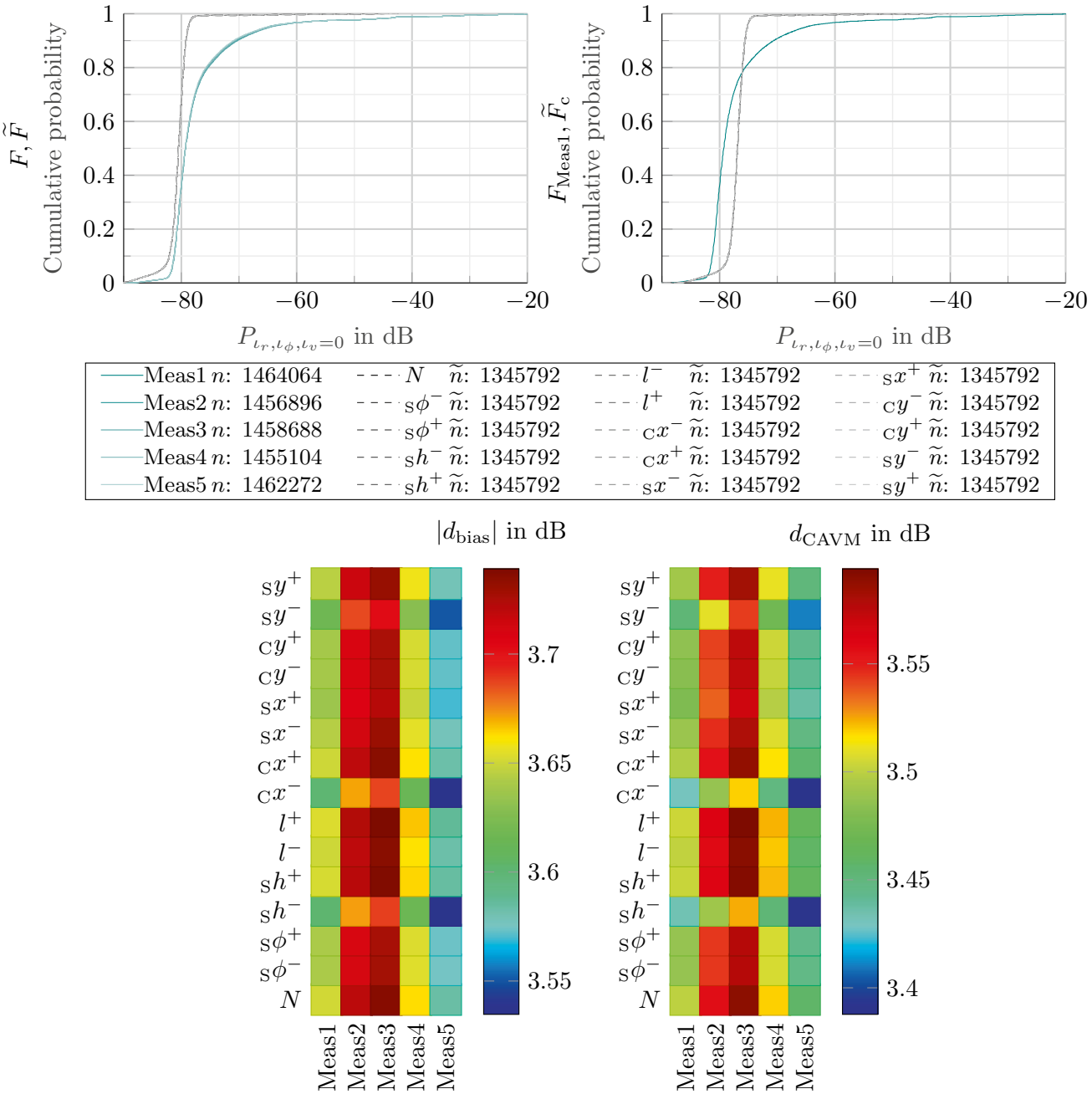


Fig. 20 The DVM Map of the whole radar cuboid in dB is shown. The first diagram visualizes the EDF of all measurements F and of all simulations $\widetilde{F}$. In the second one the EDF of measurement 1 F_{Meas1} and the corrected simulation EDFs $\widetilde{F}_c$ are shown. The second row shows the DVM Map of the above-mentioned quantities. The number of analyzed radar cuboid cells of the measurement n and simulation $\widetilde{n}$ is in the legend

This includes not only influences like vegetation but also weather influences like rain and snowfall.

5.2 Whole radar cuboid

In this section, the DVM Map is applied to the whole radar cuboid. The only dimension considered in this case is the power distribution P of all cells of the radar cuboid, with the velocity dimension reduced to one cell $\iota_v = 0$. This reduces the radar cuboid to a range-azimuth map where each cell holds a spectral power value. Figure 20 shows the EDFs of simulation and measurement as well as the corresponding DVM Map.

The course of the simulation EDFs is characterized by the noise simulation at the beginning. Only above − 82 dB the effect of the radar cuboid cells filled by the CCR with a higher power is visible. In relation to the number, however, these are represented much less frequently, so that the gray curves just above − 80 dB are just close to 100% cumulative probability. All simulation runs are very close to each other and show only minor differences. This observation is again justified based on the sensor noise simulation introduced into the model. Nevertheless, the measurement EDFs deviate from the simulation course by a few dB up to − 80 dB. Above − 80 dB the model of the environment and the ray tracing algorithm become noticeable again. Due to reflections of the environment, cells of the radar cuboid in the measurement are filled with power up to − 60 dB. Subsequently, the effect of the CCR is visible in the form of steps at − 43 dB up to − 20 dB in the EDFs.

The corrected EDFs are also close to each other, which means that only small deviations in the d_{CAVM} are to be expected. The number of cells still indicates that the cycle time of the simulation model does not yet match the real sensor. Nevertheless, the deviation in the number of data points allows a comparison because the deviation is less than 10%. The problem of comparability, as evident in Sect. 5.1 when analyzing all detections, is less present in the radar cuboid.

The values of the d_{bias} are close to the real measurement due to the noise simulation and a deviation of 3.5 dB is tolerated given the dynamic range of a radar sensor that spans over 80 dB. Measurement 2 and 3 show the largest deviations to all simulation parameters, with an increased sensitivity of the model to the uncertainties ${}_{\mathrm{C}}x^{-}$, ${}_{\mathrm{S}}y^{-}$ as well as ${}_{\mathrm{S}}h^{-}$.

d_{CAVM} is around 3.5 dB and can be justified by the aleatory uncertainties of the environment on the measurement result. Different areas on the test track produce higher powers in the measurement, which the environment simulation does not cover. In conclusion, the noise simulation distorts the influence of the environment model and the ray tracing algorithm. Thus, before integrating stochastic effects, it is recommended to analyze and optimize the whole simulation chain with ideal test objects and small region of interests.

5.3 Clustered detections on radar cuboid level

In this section, the results of clustered detections at radar cuboid level are presented and analyzed. Figure 21 shows the position of the clustered detections in the Cartesian sensor coordinate system. The cluster of CCR is number 5 and highlighted in the figure with a red circle at ${}_{\mathrm{S}}x = 29.4$ m and ${}_{\mathrm{S}}y = -4.2$ m. At close range of the sensor, some detections are visible due to the reflection of the road surface. Especially on the x-axis and in the edge region of the sensor more detections due to this effect show up. All detections from ${}_{\mathrm{S}}x = 40$ m are located at the road border where the asphalt ends and the vegetation starts. This clutter is present in all measurements and differs only slightly between measurements.

In detail, the CCR and the corresponding bins of the radar cuboid are considered below. The upper part of Fig. 22 again represents the uncorrected and corrected EDFs of simulations and measurements and measurement 1, respectively.

The step shape visible in the measurements results from the 4 different range azimuth cells analyzed in the evaluation based on the CCR's position. The variations of a cell are within a few dB over time, which can be seen in the slope of the EDFs. Furthermore, the reproducibility of the measurements is exceptionally high, which is reflected in the overlap of the courses of the measurements.

In general, the simulations have a clear model bias. Here, the first modeling errors of the signal processing are evident. The model is calibrated to the RCS of the CCR at position 1 and therefore a difference in the calculation of the RCS from the radar cuboid power to the detections exists.

Furthermore, there is a clear influence of the uncertainties propagated by the simulation. In three simulations, the step shape is similar to the measurements, but the slope itself is more substantially smeared and not as steep. All other simulations have a much lower slope after the initial step. The beginning of the simulation slopes can be explained by the noise simulation. Two radar cuboid cells are considered here, which are not yet affected by the power increase due to the CCR. From this, a modeling error can again be identified. The window functions of the real sensor differ from those of the simulation model because the power increase of the CCR does not smear as far into neighboring bins as in the measurement. The discrepancy in the number of cell values considered indicates a sampling difference between the model and the real sensor. This is reasoned by the fact that the real sensor does not exhibit fixed cycle times, and the co-simulation restricts this parameter through a specific sampling frequency.

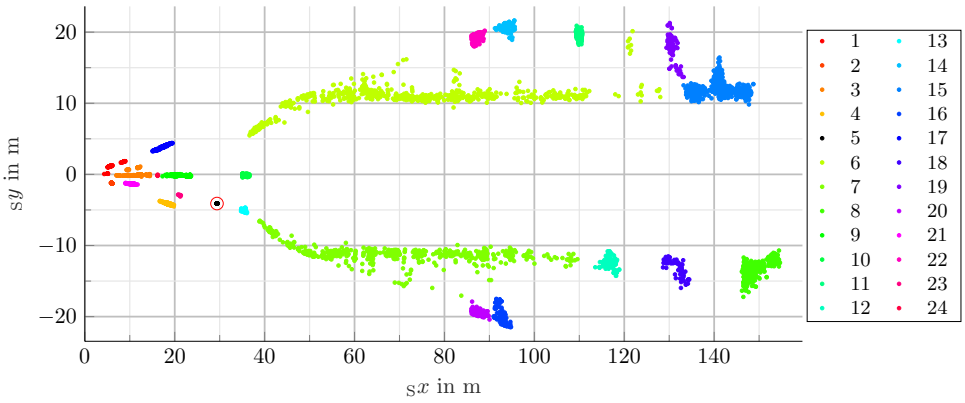


Fig. 21 Clustered radar detections of all measurements in the sensor coordinate system, where the color represents the belonging to a cluster. The red circle at ${}_Sx = 29.4$ m and ${}_Sy = -4.2$ m shows cluster 5, where the CCR is located

The previously described findings from the EDFs are also reflected in the DVM Map. The simulations of ${}_Sy^-$, ${}_Cx^-$ as well as ${}_Sh^-$ show the lowest deviations in model bias as well as scattering error. This is in agreement with the findings from Sect. 5.2, where also the mentioned uncertainties represent the smallest deviation (see also Fig. 20).

In comparison to the previous figures, the heat map of d_{Sum} is additionally shown, since the further considerations in Fig. 21 are based on these results. To compare the clusters with each other, the maximum value of d_{Sum} and its corresponding cell in the heat map is used. The values for $|d_{\text{bias}}|$ as well as d_{CAVM} of the uncertainty measurement combination are transferred to a separate bar diagram in Fig. 23.

Clusters 1 to 4, 10, 13, 17, 21 and 24 in Fig. 23 show the clear difference between simulation and measurement in the close range of the sensor. As already described, ground reflections are not further considered in the simulation model, which results in the visible difference between simulation and measurement. Clusters 6 and 7 represent the largest clusters with the main clutter due to vegetation. The influence of vegetation is not as large as the deviations in the near sensor range, since the distance is larger and thus the power in the radar cuboid approach the noise level. Nevertheless, a clear difference between simulation and measurement can be identified.

Among the clusters farthest from the sensor in the range, number 18 stands out. At this location, there is an intersection of runway and taxiway on the August–Euler airfield. The effects of the change in ground properties are thus detectable in the methodology using the DVM Map. Across all clusters, no trend in the measurements and uncertainty parameters can be detected, which on the one hand speaks for the good reproducibility of the measurements and on the other hand for the high sensitivity of the radar model. If such a trend is observed in the presented analysis, either for an uncertain parameter or a measurement, it suggests the presence of either a parameterization with high sensitivity in the simulation or an outlier in the measurements.

5.4 Each range and azimuth radar cuboid cell

To increase the interpretability of the results, the outcome of the DVM Map are plotted on a satellite image. For this purpose, the measured positions of the sensor and the determined orientation are taken as origin and the range as well as azimuth resolution of the radar cuboid is used to distribute its cells over the satellite image. From top to bottom, $|d_{\text{bias}}|$, d_{CAVM} and d_{Sum} are visualized in Fig. 24. The coloring of the cells corresponds to the results of the DVM Map per radar cuboid cell. As an example, the DVM Map of cell $\iota_r = 18, \iota_\phi = 16$ is shown. As in Fig. 23 for further analysis, the maximum of d_{Sum} for the measurement and uncertainty combination is used. Therefore, this combination is used to color the cell in the value of $|d_{\text{bias}}|$, d_{CAVM} and d_{Sum}.

In the top plot it can be seen that there is a deviation of about 30 dB in the area where the CCR is located. The smearing of the power in neighbor bins due to the windowing function is included in the simulation model, but an assignment to the causal effect of the deviation is difficult. On the one hand, the power of the CCR is too low, which can be corrected by calibrating the model at radar cuboid level instead of the detection interface with its RCS value. Nevertheless, there is a model error in signal processing in the calculation of RCS from radar cuboid data, as the simulation is calibrated to a centrally positioned CCR on the runway. On the other hand, the window function in the model is iteratively determined, which means that measurement and modeling errors may also be present here.

Directly next to the highlighted cell $\iota_r = 18, \iota_\phi = 16$ there is an area with the maximum deviation between simulation model and measurement, which is 55 dB. In the simulation model there is no input data from the ray tracing algorithm and only the noise simulation fills these cells of the radar cuboid with power values. During the measurements, no objects or asphalt peculiarities were noticed that justify this increase in power. For these reasons, there has to be an effect in the signal processing of the radar sensor, which is not considered in the radar model and is triggered by the CCR.

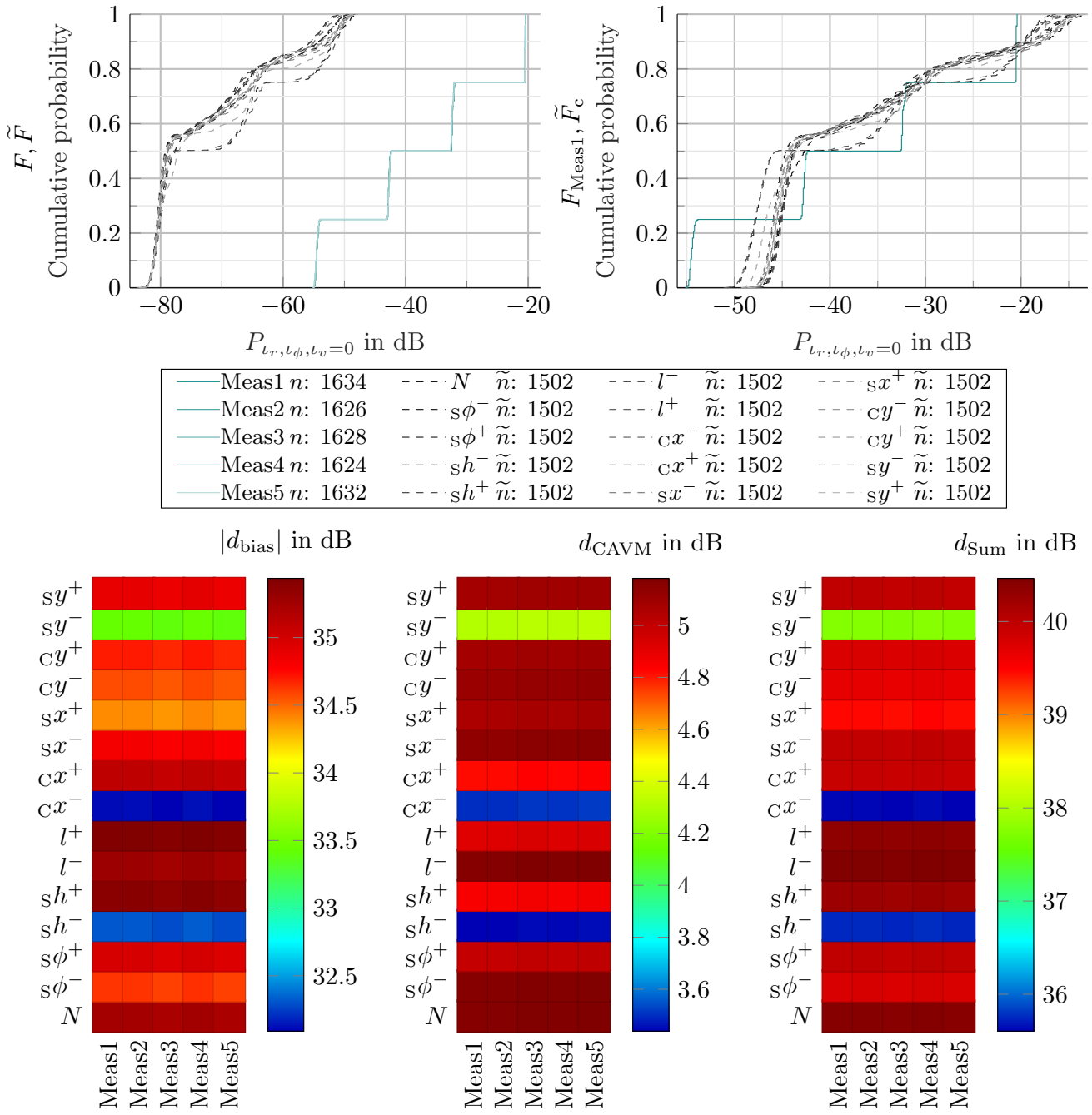


Fig. 22 The uncorrected and corrected EDFs as well as the DVM Map with d_{Sum} of the CCR cluster 5 of the radar cuboid interface in dB is shown

Therefore, the method is able to identify a systematic model error at this point. Along the runway there are still increases in the model bias, ranging from 15 to 25 dB especially in the transition area of the asphalt and vegetation.

In the second satellite plot the scattering error represented by d_{CAVM} is visualized. The highest deviation is present at the CCR. The distribution shape in the measurement looks like a step function. The simulation EDFs are not so steep, and the propagated uncertainties have a big impact on the shape especially at lower powers. This proves the high sensitivity of the radar model chain with respect to the measurement data uncertainty.

Along the runway and the transition to vegetation, notable deviations are evident, as already in $|d_{\text{bias}}|$. Of particular interest is the intersection of the taxiway with the runway highlighted by the red circle. Due to the transition between

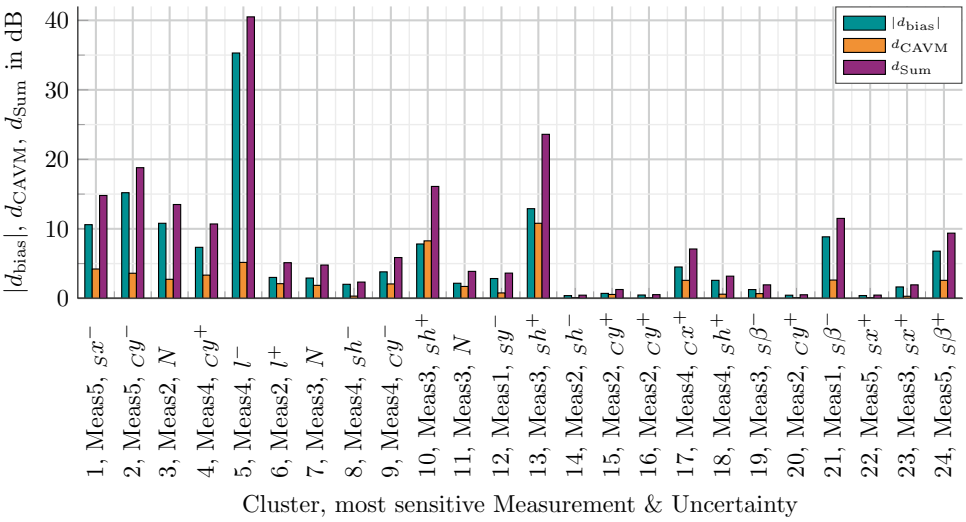


Fig. 23 Bar plot of all clusters of $|d_{\text{bias}}|$, d_{CAVM} and d_{Sum} with the corresponding cluster number as well as the biggest influence based on the measurement uncertainty parameter combination on the x-axis. The evaluated interface is the radar cuboid of the different clusters in dB

vegetation and asphalt, higher values in comparison to the surrounding cells for d_{CAVM} are shown.

In the d_{Sum} satellite plot, the differences now become even more apparent. In addition to the features of the runway, vegetation, and intersection already mentioned, the sensor's close range has significant discrepancies. This underlines the findings from the analysis of the clusters in the previous chapter.

6 Conclusion

This paper introduces the concept of the DVM Map and its application to radar data using a static scenario. Based on four ECs, the need for an extension of the existing DVM definition is presented. A methodology which allows to apply the DVM Map on radar cuboid and detection level is described. A validation study is exemplified using the described experimental setup and an adapted radar simulation model. It is evident that looking at all detections only makes sense if the environment simulation is matched with the ray tracing algorithm. It can be seen that the different number of detections is a fundamental problem of the simulation model with all its components. The evaluation of the entire radar cuboid has the advantage that the number of data depends only on the correct model parameterization. The comparison reveals whether aleatory uncertainties such as noise are modeled correctly. The clustered detections are used to analyze the areas that are particularly affected by power differences. In the case of the validation study, the analysis of the CCR shows high deviations between measurement and simulation. Therefore, objects can be identified and evaluated particularly well using this approach. Finally, all cells of the radar cuboid are analyzed. Local aleatory and epistemic uncertainties of the environment model, e.g. vegetation and asphalt, are visible. Additionally, effects of the radar signal processing model can be separated from environment model and ray tracing algorithm. In general, it can be seen that validation of a sensor model and its signal processing is only possible if the environment simulation is qualified regarding physical effects and aleatory uncertainties.

Overall, it can be seen that the DVM Map with its application to the different levels significantly increases the interpretability of scenarios in the following manner:

- The DVM Map gives the output score in units of the analyzed size.
- The DVM Map gives information about the sensitivity of each modeled reference data inaccuracy in the simulation with each measurement.
- Clustering gives dedicated information about object model and modeling errors.
- Using the DVM Map in combination with the satellite plot, errors can be spatially localized, and thus the environment model can be examined.

The DVM Map can also be used to compare measurements with each other and thus investigate stochastic effects such as rain and compare the similarity of rain conditions between measurements. Furthermore, measurement setups that have to be dismantled and reassembled can be examined and compared by a reference measurement. Particular attention should be paid to the periphery of the radar sensors' field of view, where azimuth ambiguities are present and radar detection accuracy is lower compared to the boresight. Therefore, future experiments should focus specifically on this area and evaluate the performance of the DVM Map. Additionally, an analysis of the signal-to-noise ratio offers further potential to improve the understanding of the underlying effects in the future. So far, the consideration of uncertainties is limited to the upper and lower bounds, which does not take into account mutual influences of the uncertainties. Therefore, it is recommended in a next step to combine the uncertainties with each other and

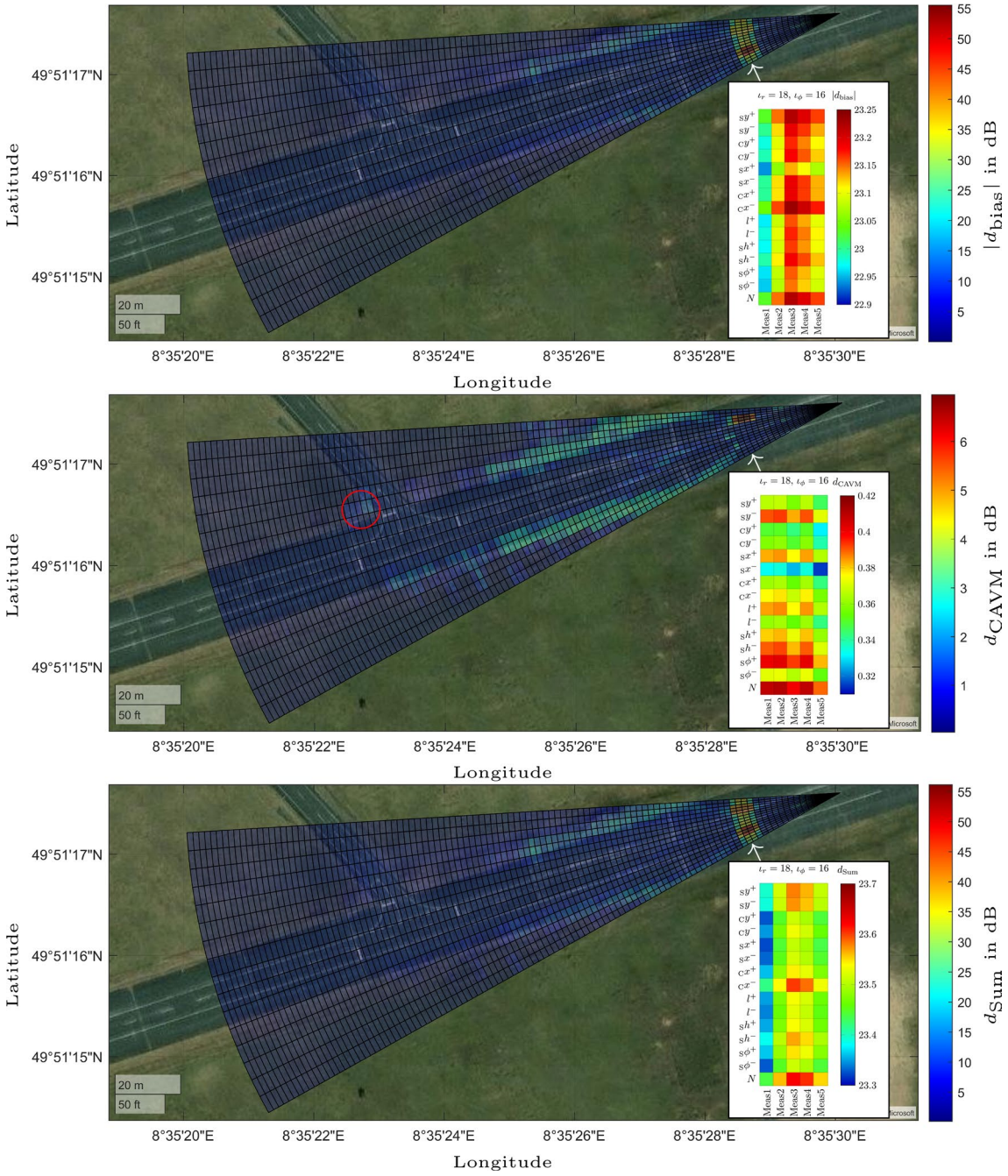


Fig. 24 Satellite image in which the results of the DVM Map of the radar cuboid interface in dB are shown. From top to bottom, the validation results for $|d_{\text{bias}}|$, d_{CAVM} and d_{Sum} are illustrated. In the lower right corner of each plot, the DVM Map of cell $\iota_r = 18, \iota_\phi = 16$ is shown as an example on which the coloring in the satellite plot is based

to limit the parameter space in the process. As soon as further uncertainties are added and not only the upper and lower bounds of the uncertainties are varied, this inevitably leads to an explosion of the parameter space. Assuming a parameter number in the present scope with five instead of three variations is examined full factorial, the number of necessary simulations for just one scenario with a CCR is

$$n_{\text{sim}} = n_{\text{var}}^{n_{\text{param}}} = 5^7 = 78{,}125. \tag{10}$$

This estimate does not include material properties or complex geometries and each simulation has to be repeated for a change in the model itself. Therefore, the parameter sensitivity of the model must be determined in advance and thus reduce the parameter space. In the future, we will extend the methodology developed here for static scenarios to dynamic

scenarios. However, this poses challenges specifically with respect to temporal aggregation of data. It is imperative that these challenges be resolved to qualify the methodology for model validation.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported in part by the German Federal Ministry for Education and Research (BMBF) through Virtual Validation Tool Chain for Automated and Connected Driving (VIVID) under Grant 16ME0173. This work also received funding from VVM of the PEGASUS project family, promoted by the German Federal Ministry for Economic Affairs and Energy (BMWK) under Grant 19A19002S.

Declarations

Conflict of interest The corresponding authors declare no conflict of interest.

References

1. Devision, U.T.: New Assessment/Test Method for Automated Driving (NATM) Guidelines for Validating Automated Driving System (ADS)—amendments to ECE/TRANS/WP.29/2022/58. UNECE. https://wiki.unece.org/download/attachments/184615115/VMAD-SG2-39-01%20SG2%20solved%20UK%20comments%20on%20NATM%20Guidelines.docx?api=v2. Accessed 17 Apr 2023 (2022)
2. Society of Automotive Engineers.: SAE-J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_202104/. Accessed 06 Jan 2022 (2021)
3. Zheng, L., Ma, Z., Zhu, X., Tan, B., Li, S., Long, K., Sun, W., Chen, S., Zhang, L., Wan, M., Huang, L., Bai, J.: TJ4DRadSet: a 4D radar dataset for autonomous driving. In: 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), pp. 493–498. IEEE, Macau, (2022). https://doi.org/10.1109/ITSC55140.2022.9922539. https://ieeexplore.ieee.org/document/9922539/. Accessed 03 Apr 2023
4. International Organization for Standardization.: ISO 23150:2021(E): road vehicles—Data Communication Between Sensors and Data Fusion Unit for Automated Driving Functions—Logical interface (2021)
5. Steinbaeck, J., Steger, C., Holweg, G., Druml, N.: Design of a low-level radar and time-of-flight sensor fusion framework. (2018). https://doi.org/10.1109/DSD.2018.00056
6. Holder, M.F., Rosenberger, P., Winner, H., D'hondt, T., Makkapati, V.P., Maier, M., Schreiber, H., Magosi, Z., Slavik, Z., Bringmann, O., Rosenstiel, W.: Measurements revealing challenges in radar sensor modeling for virtual validation of autonomous driving. In: 2018 IEEE 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2616–2622. IEEE, Maui (2018). https://doi.org/10.1109/ITSC.2018.8569423. https://ieeexplore.ieee.org/document/8569423/. Accessed 10 Nov 2020
7. Oberkampf, W.L., Trucano, T.G.: Verification and validation benchmarks. Nucl. Eng. Des. **238**(3), 716–743 (2008). https://doi.org/10.1016/j.nucengdes.2007.02.032
8. Viehof, M.: Objektive Qualitätsbewertung von Fahrdynamiksimulationen durch statistische Validierung. PhD thesis, Technische Universität Darmstadt, Darmstadt (2018). http://tuprints.ulb.tu-darmstadt.de/7457/
9. Benke, K.K., Norng, S., Robinson, N.J., Benke, L.R., Peterson, T.J.: Error propagation in computer models: analytic approaches, advantages, disadvantages and constraints. Stoch. Environ. Res. Risk Assess. **32**(10), 2971–2985 (2018). https://doi.org/10.1007/s00477-018-1555-8.
10. Rosenberger, P.: Metrics for specification, validation, and uncertainty prediction for credibility in simulation of active perception sensor systems. Ph.D. thesis, Technische Universität Darmstadt, Darmstadt. https://doi.org/10.26083/tuprints-00023034. http://tuprints.ulb.tu-darmstadt.de/23034/ (2023)
11. Magosi, Z.F., Wellershaus, C., Tihanyi, V.R., Luley, P., Eichberger, A.: Evaluation methodology for physical radar perception sensor models based on on-road measurements for the testing and validation of automated driving. MDPI Energies **15**(7), 20 (2022). https://doi.org/10.3390/en15072545
12. Schaermann, A.: Systematische Bedatung und Bewertung umfelderfassender Sensormodelle. Ph.D. thesis, Technische Universität München, München (2020). http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20200526-1518611-1-9
13. Oberkampf, W.L., Barone, M.F.: Measures of agreement between computation and experiment: validation metrics. J. Comput. Phys. **217**(1), 5–36 (2006). https://doi.org/10.1016/j.jcp.2006.03.037
14. Liu, Y., Chen, W., Arendt, P., Huang, H.-Z.: Toward a better understanding of model validation metrics. J. Mech. Des. (2011). https://doi.org/10.1115/1.4004223
15. Fréchet, M.R.: Sur quelques points du calcul fonctionnel. Rend. Circ. Mat. Palermo **1884–1940**(22), 1–72 (1906). https://doi.org/10.1007/BF03018603
16. Ngo, A.: A methodology for validation of a radar simulation for virtual testing of autonomous driving. Ph.D. thesis (2023). https://doi.org/10.18419/opus-12703. Accepted: 2023-02-07T13:58:14Z ISBN: 9781833425246. http://elib.uni-stuttgart.de/handle/11682/12722. Accessed 11 Feb 2023
17. Wölfel, M., Ekenel, H.K.: Feature weighted mahalanobis distance: improved robustness for gaussian classifiers. In: 13th European signal processing conference (EUSIPCO 2005), 4–8 September 2005, Antalya, Turkey. ISBN: 9781604238211, pp. 1–4 (2005)
18. SafeTRANS Closing the Gap Initiative: controlling risk for highly automated transportation systems operating in complex open environments (2024). https://safetrans-de.org/de/Uploads/Aktuelle_Meldungen/White_Paper_Version_1.3.pdf?m=1711536874. Accessed 09 Apr 2024
19. Ferson, S., Oberkampf, W.L., Ginzburg, L.: Model validation and predictive capability for the thermal challenge problem. Comput. Methods Appl. Mech. Eng. **197**(29), 2408–2430 (2008). https://doi.org/10.1016/j.cma.2007.07.030
20. Eder, T.: Simulation of automotive radar point clouds in standardized frameworks. Ph.D. thesis, Technische Universität München, München. Google-Books-ID: QYJSEAAAQBAJ (2021)
21. Stripling, H.F., Adams, M.L., McClarren, R.G., Mallick, B.K.: The method of manufactured universes for validating uncertainty quantification methods. Reliab. Eng. Syst. Saf. **96**(9), 1242–1256 (2011). https://doi.org/10.1016/j.ress.2010.11.012

22. Williamson, R.C., Downs, T.: Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. Int. J. Approx. Reason. **4**(2), 89–158 (1990). https://doi.org/10.1016/0888-613X(90)90022-T
23. Ferson, S., Kreinovick, V., Ginzburg, L., Myers, D.S., Sentz, K.: Constructing probability boxes and dempster-shafer structures. Technical Report SAND2002-401 (2003). https://www.semanticscholar.org/paper/Constructing-Probability-Boxes-and-Dempster-Shafer-Ferson-Kreinovick/8eff743341521cca30f6d2a48df50bf6977c96b2. Accessed 24 Mar 2023
24. Voyles, I.T., Roy, C.J.: Evaluation of model validation techniques in the presence of aleatory and epistemic input uncertainties. In: 17th AIAA Non-deterministic Approaches Conference. American Institute of Aeronautics and Astronautics, Kissimmee, Florida (2015). https://doi.org/10.2514/6.2015-1374. Accessed 18 Aug 2021
25. Linnhoff, C., Rosenberger, P., Schmidt, S., Elster, L., Stark, R., Winner, H.: Towards serious perception sensor simulation for safety validation of automated driving—a collaborative method to specify sensor models. In: 2021 IEEE 24th International Conference on Intelligent Transportation Systems (ITSC), Indianapolis, pp. 2688–2695 (2021). https://doi.org/10.1109/ITSC48978.2021.9564661
26. Linnhoff, C., Hinsemann, T., Rosenberger, P., Elster, L.: PerCollECT-LidarLimbs. PerCollECT. original-date: 2021-03-19T07:34:25Z (2022). https://github.com/PerCollECT/LidarLimbs. Accessed 03 Apr 2023
27. Brune, A.J., West, T.K., White, L.M.: Calibration probe uncertainty and validation for the hypersonic material environmental test system. NTRS Author Affiliations: NASA Langley Research Center NTRS Report/Patent Number: NF1676L-33286 NTRS Document ID: 20200002798 NTRS Research Center: Langley Research Center (LaRC) (2020). https://doi.org/10.2514/1.T5839
28. Holder, M.F.: Synthetic generation of radar sensor data for virtual validation of autonomous driving. PhD thesis, Technische Universität Darmstadt, Darmstadt (2021). https://doi.org/10.26083/tuprints-00017545. http://tuprints.ulb.tu-darmstadt.de/17545/
29. Github Open MSL: Reflection Based Radar Model. https://github.com/openMSL/sl-1-1-reflection-based-radar-object-model. Accessed 23 Mar 2023

Bibliography

Abadpour, S. et al.: Dielectric Material Characterization of Traffic Objects (2023)
Abadpour, Sevda; Pauli, Mario; Kretschmann, Marius; Iqbal, Hasan; Aust, Philip; Zwick, Thomas: Dielectric Material Characterization of Traffic Objects in Automotive Radar Applications, in: 2023 17th European Conference on Antennas and Propagation (EuCAP), pp. 1–5, 2023

Ahmann, M. et al.: Towards Continuous Simulation Credibility Assessment (2022)
Ahmann, Maurizio; Le, Van Thanh; Eichenseer, Frank; Steimann, Frederik; Benedikt, Martin: Towards Continuous Simulation Credibility Assessment, in: Asian Modelica Conference 2022, Tokyo, Japan, November 24-25, 2022, pp. 171–182, 2022

Andres, M. et al.: 3D-scattering Center Detection of Automotive Targets Using Radar Sensors (2012)
Andres, Markus; Feil, Peter; Menzel, Wolfgang: 3D-scattering Center Detection of Automotive Targets Using 77 GHz UWB Radar Sensors, in: 2012 6th European Conference on Antennas and Propagation (EUCAP), pp. 3690–3693, 2012

Arage Hassen, A.: Signal Degradation and Optimization Radar Sensors (2006)
Arage Hassen, Alebel: Indicators for the Signal Degradation and Optimization of Automotive Radar Sensors under Adverse Weather Conditions, PhD Thesis, Technische Universität Darmstadt, 2006

ASAM e.V.: ASAM OpenSCENARIO® - User Guide (2021)
ASAM e.V.: ASAM OpenSCENARIO® - User Guide, URL: https://www.asam.net/standards/detail/openscenario/, 2021, visited on 12/20/2023

ASAM e.V.: ASAM OSI® (Open Simulation Interface) - Official Documentation (2023)
ASAM e.V.: ASAM OSI® (Open Simulation Interface) - Official Documentation, URL: https://opensimulationinterface.github.io/osi-documentation/, 2023, visited on 02/28/2024

Aust, P. et al.: Fingerprints of the Automotive Radar Scattering of Passenger Cars and Vans (2023)
Aust, Philip; Hau, Florian; Dickmann, Juergen; Hein, Matthias: Fingerprints of the Automotive Radar Scattering of Passenger Cars and Vans, in: Applied Sciences, Vol. 13, 2023

Aust, P. et al.: Numerical Synthesis of Radar Target Detections (2023)
Aust, Philip; Hau, Florian; Dickmann, Jürgen; Hein, Matthias A.: Numerical Synthesis of Radar Target Detections Based on Measured Reference Data, in: 2023 20th European Radar Conference (EuRAD), pp. 26–29, 2023

BMW Group Pressclub Deutschland: Level 3 im neuen BMW 7er (2023)
BMW Group Pressclub Deutschland: Ab Frühjahr: Hochautomatisiertes Fahren auf Level 3 im neuen BMW 7er. URL: https://www.press.bmwgroup.com/deutschland/article/detail/T0438214DE/ab-fruehjahr:-hochautomatisiertes-fahren-auf-level-3-im-neuen-bmw-7er?language=de, 2023, visited on 12/18/2023

Brahmi, M.: Bewertung der Umfeldwahrnehmung für Fahrerassistenzsysteme (2020)
Brahmi, Mohamed: Bewertung der objektbasierten Umfeldwahrnehmung für Fahrerassistenzsysteme mithilfe von Referenzsystemen, PhD Thesis, Universitätsbibliothek Braunschweig, 2020

Buddappagari, S. J. G. et al.: RCS Measurements of Road Traffic Objects (2020)
Buddappagari, Sreehari Jayapal Gowdu; Schwind, A.; Stephan, R.; Hein, Matthias A.: Monostatic RCS Measurements of Representative Road Traffic Objects in the 76 … 81 GHz Frequency Band, in: 2020 IEEE Radar Conference (RadarConf20), pp. 1–6, 2020

Buddappagari, S. J. G. et al.: Monostatic RCS Measurements of a Passenger Car (2019)
Buddappagari, Sreehari Jayapal Gowdu; Schwind, Andreas; Stephan, Ralf; Hein, Matthias A.: Monostatic RCS Measurements of a Passenger Car Mock-up at 77 GHz Frequency in Virtual Environment, in: 2019 49th European Microwave Conference (EuMC), pp. 996–999, 2019

Clavenger, S.: Torc Autonomous Truck Launch in 2027 (2023)
Clavenger, Seth: Torc Lays Out Road Map to Autonomous Truck Launch in 2027, URL: https://www.ttnews.com/articles/torc-autonomous-launch-27, 2023, visited on 12/18/2023

Continental Engineering Services: Datasheet Continental ARS408 (2024)
Continental Engineering Services: Datasheet Continental ARS408, URL: https://conti-engineering.com/wp-content/uploads/2020/02/ARS-408-21_EN_HS-1.pdf, 2024, visited on 01/22/2024

Continental Press Release: Continental and Aurora Partner Autonomous Trucking Systems (2023)
Continental Press Release: Continental and Aurora Partner to Realize Commercially Scalable Autonomous Trucking Systems, URL: https://www.continental.com/en/press/press-releases/20230427-continental-aurora/, 2023, visited on 02/12/2024

Diewald, F.: Objektklassifikation und Freiraumdetektion auf Basis bildgebender Radarsensori (2013)
Diewald, Fabian: Objektklassifikation und Freiraumdetektion auf Basis bildgebender Radarsensorik für die Fahrzeugumfelderfassung, PhD Thesis, Universität Ulm, 2013

DIN: DIN ISO 3534-2 Statistics - Vocabulary and Symbols (2013)
DIN: DIN ISO 3534-2:2013-12, Statistik- Begriffe Und Formelzeichen- Teil 2: Angewandte Statistik (ISO 3534-2:2006); Text Deutsch Und Englisch, URL: https://www.beuth.de/de/-/-/192284289, 2013, visited on 11/06/2023

DIN: DIN SAE SPEC 91471 Assessment Methodology Automotive LiDAR (2023)
DIN: DIN SAE SPEC 91471:2023-05 Assessment Methodology for Automotive LiDAR Sensors, 2023

DIVP Project: Driving Intelligence Validation Platform Homepage (2024)
DIVP Project: Driving Intelligence Validation Platform Homepage, URL: https://divp.net/, 2024, visited on 02/16/2024

Eder, T.: Simulation of Automotive Radar (2021)
Eder, Thomas: Simulation of Automotive Radar Point Clouds in Standardized Frameworks, PhD Thesis, Technische Universität München, 2021

Elster, L.: RCS Measurement (2022)
Elster, Lukas: RCS Measurement, URL: https : / / gitlab . com / tuda - fzd / fzd - datasets / rcs - measurement, 2022, visited on 12/05/2023

Elster, L. et al.: Dataset Radar Scattering Characteristics (2023)
Elster, Lukas; Holder, Martin F.; Rapp, Manuel: A Dataset for Radar Scattering Characteristics of Vehicles Under Real-World Driving Conditions: Major Findings for Sensor Simulation, in: IEEE Sensors Journal, Vol. 23, pp. 4873–4882, 2023

Elster, L. et al.: RCS Measurement Dataset (2022)
Elster, Lukas; Holder, Martin Friedrich; Rapp, Manuel: RCS Measurement Dataset, URL: https: //tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/3606, 2022, visited on 12/05/2023

Elster, L. et al.: Introducing the DVM for Radar Sensor Models (2024)
Elster, Lukas; Rosenberger, Philipp; Holder, Martin; Mori, Ken; Staab, Jan; Peters, Steven: Introducing the Double Validation Metric for Radar Sensor Models, in: Automotive and Engine Technology, Vol. 9, p. 6, 2024

Elster, L. et al.: Making Automotive Radar Sensor Validation Measurements Comparable (2023)
Elster, Lukas; Staab, Jan Philipp; Peters, Steven: Making Automotive Radar Sensor Validation Measurements Comparable, in: Applied Sciences, Vol. 13, p. 11405, 2023

ENVITED OpenMSL Github: SL 1-1 Reflection Based Radar Object Model (2023)
ENVITED OpenMSL Github: SL 1-1 Reflection Based Radar Object Model, URL: https://github.com/openMSL/sl-1-1-reflection-based-radar-object-model, 2023, visited on 12/04/2023

Europäische Kommission: Durchführungsverordnung (EU) 2022/1426 (2022)
Europäische Kommission: Durchführungsverordnung (EU) 2022/1426 Der Kommission, 2022

Ferson, S. et al.: Validation of Imprecise Probability Models (2009)
Ferson, Scott; Oberkampf, William: Validation of Imprecise Probability Models, in: International Journal of Reliability and Safety, Vol. 3, 2009

GeneSys: Datasheet ADMA-G-Pro+ V35 (2023)
GeneSys: Datasheet ADMA-G-Pro+ V35, URL: https://genesys-offenburg.de/wp-content/uploads/2023/07/GeneSys-Datasheet-ADMA-G-Pro-v35.pdf, 2023, visited on 12/07/2023

Gottschalg, G.: Data Fusion Architecture for State Estimation (2022)
Gottschalg, Grischa: Data Fusion Architecture with Integrity Monitoring for State Estimation in Automated Driving, PhD Thesis, Technische Universität Darmstadt, 2022

Grant, N. et al.: Waymo, Chasing Cruise, Plans Fully Driverless Rides in San Francisco (2022)
Grant, Nico; Ludlow, Edward: Waymo, Chasing Cruise, Plans Fully Driverless Rides in San Francisco, in: Bloomberg.com, 2022

Hawkins, A.: California DMV Suspends Cruise's Robotaxi Permit 'Effective Immediately' (2023)
Hawkins, Andrew: California DMV Suspends Cruise's Robotaxi Permit 'Effective Immediately', URL: https://www.theverge.com/2023/10/24/23930629/california-dmv-suspends-cruise-robotaxi-permit-safety, 2023, visited on 11/30/2023

Heinkel, H.-M. et al.: Credible Simulation Process Framework (2023)
Heinkel, Hans-Martin; Steinkirchner, Kim: Credible Simulation Process Framework, URL: https://gitlab.setlevel.de/open/processes_and_traceability/credible_simulation_process_framework, 2023, visited on 12/06/2023

Holder, M. F.: Synthetic Generation of Radar Sensor Data (2021)
Holder, Martin Friedrich: Synthetic Generation of Radar Sensor Data for Virtual Validation of Autonomous Driving, PhD Thesis, Technische Universität Darmstadt, 2021

Holder, M. F. et al.: Digitalize the Twin (2022)
Holder, Martin Friedrich; Elster, Lukas; Winner, Hermann: Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation, in: MDPI Energies, Vol. 15, p. 989, 2022

Holder, M. F. et al.: Source Code Xosc-Converter (2021)
Holder, Martin Friedrich; Elster, Lukas; Winner, Hermann: Source Code Xosc-Converter: "Digitalize the Twin: A Method for Calibration of Reference Data for Transfer Real-World Test Drives into Simulation", URL: https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/2993, 2021, visited on 12/20/2023

Holder, M. F. et al.: Modeling and Simulation of Radar Sensor Artifacts (2019)
Holder, Martin Friedrich; Linnhoff, Clemens; Rosenberger, Philipp; Popp, Christoph; Winner, Hermann: Modeling and Simulation of Radar Sensor Artifacts for Virtual Testing of Autonomous Driving, in: 9. Tagung Automatisiertes Fahren, 2019

Honda Motor Europe Ltd: Honda Receives Type Designation for Level 3 (2020)
Honda Motor Europe Ltd: Honda Receives Type Designation for Level 3 Automated Driving, URL: https://hondanews.eu/eu/en/cars/media/pressreleases/318975/honda-receives-type-designation-for-level-3-automated-driving, 2020, visited on 12/18/2023

ISO: ISO 23150:2021(E) Data between Sensors and Fusion (2021)
ISO: ISO 23150:2021(E): Road Vehicles — Data Communication between Sensors and Data Fusion Unit for Automated Driving Functions - Logical Interface, 2021

Karlsson, K. et al.: Reducing Ground Reflection during RCS Characterization (2019)
Karlsson, Kristian; Toss, Henrik; Costagliola, Francesco: Reducing Influence from Ground Reflection during RCS Characterization of Automotive Targets, in: 2019 13th European Conference on Antennas and Propagation (EuCAP), pp. 1–5, 2019

Karlsson, K. et al.: HiFi Radar Target (2018)
Karlsson, Kristian; Toss, Henrik; Lang, John; Costagliola, Francesco; Zheng, Tian; Marel, Elias: HiFi Radar Target : High Fidelity Soft Targets and Radar Simulation for More Efficient Testing (Real and Virtual), 2018

Landron, O. et al.: A Comparison of Theoretical and Empirical Reflection Coefficients (1996)
Landron, O.; Feuerstein, M.J.; Rappaport, T.S.: A Comparison of Theoretical and Empirical Reflection Coefficients for Typical Exterior Wall Surfaces in a Mobile Radio Environment, in: IEEE Transactions on Antennas and Propagation, Vol. 44, pp. 341–351, 1996

Langen, B. et al.: Reflection and Transmission Behaviour of Building Materials at 60 GHz (1994)
Langen, B.; Lober, G.; Herzig, W.: Reflection and Transmission Behaviour of Building Materials at 60 GHz, in: 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Wireless Networks - Catching the Mobile Future. Vol. 2, 505–509 vol.2, 1994

Lehnen, M. et al.: Validating NVIDIA DRIVE Sim Radar Models (2023)
Lehnen, Matthias; Montebaur, Marius; Elsaeid, Ayman: Validating NVIDIA DRIVE Sim Radar Models, URL: https://developer.nvidia.com/blog/validating-nvidia-drive-sim-radar-models/, 2023, visited on 02/16/2024

Li, H. et al.: The Effect of Rainfall and Illumination on Automotive Sensors Detection Performance (2023)
Li, Hexuan; Bamminger, Nadine; Magosi, Zoltan Ferenc; Feichtinger, Christoph; Zhao, Yongqi; Mihalj, Tomislav; Orucevic, Faris; Eichberger, Arno: The Effect of Rainfall and Illumination on Automotive Sensors Detection Performance, in: Sustainability, Vol. 15, p. 7260, 2023

Linnhoff, C. et al.: PerCollECT - LidarLimbs (2022)
Linnhoff, Clemens; Hinsemann, Timo; Rosenberger, Philipp; Elster, Lukas: PerCollECT - LidarLimbs, URL: https://github.com/PerCollECT/LidarLimbs, 2022, visited on 12/06/2023

Linnhoff, C. et al.: Towards Sensor Simulation for Safety Validation (2021)
Linnhoff, Clemens; Rosenberger, Philipp; Schmidt, Simon; Elster, Lukas; Stark, Rainer; Winner, Hermann: Towards Serious Perception Sensor Simulation for Safety Validation of Automated Driving - A Collaborative Method to Specify Sensor Models, in: 2021 IEEE 24th International Conference on Intelligent Transportation Systems (ITSC), pp. 2688–2695, 2021

Linnhoff, C. et al.: Simulating Road Spray Effects in Automotive Lidar Sensor Models (2022)
Linnhoff, Clemens; Scheuble, Dominik; Bijelic, Mario; Elster, Lukas; Rosenberger, Philipp; Ritter, Werner; Dai, Dengxin; Winner, Hermann: Simulating Road Spray Effects in Automotive Lidar Sensor Models, URL: http://arxiv.org/abs/2212.08558, 2022, visited on 01/19/2023

Liu, Y. et al.: Toward a Better Understanding of Model Validation Metrics (2011)
Liu, Yu; Chen, Wei; Arendt, Paul; Huang, Hong-Zhong: Toward a Better Understanding of Model Validation Metrics, in: Journal of Mechanical Design, Vol. 133, 2011

Magosi, Z. F. et al.: A Survey on Modelling of Automotive Radar Sensors (2022)
Magosi, Zoltan Ferenc; Li, Hexuan; Rosenberger, Philipp; Wan, Li; Eichberger, Arno: A Survey on Modelling of Automotive Radar Sensors for Virtual Test and Validation of Automated Driving, in: Sensors, Vol. 22, p. 5693, 2022

Magosi, Z. F. et al.: Evaluation of Physical Radar Perception Sensor Models (2022)
Magosi, Zoltan Ferenc; Wellershaus, Christoph; Tihanyi, Viktor Roland; Luley, Patrick; Eichberger, Arno: Evaluation Methodology for Physical Radar Perception Sensor Models Based on On-Road Measurements for the Testing and Validation of Automated Driving, in: MDPI Energies, Vol. 15, p. 20, 2022

Mercedes-Benz: Mercedes-Benz DRIVE PILOT (2023)
Mercedes-Benz: Mercedes-Benz DRIVE PILOT, URL: https://www.mercedes-benz.de/passengercars/technology-innovation/mercedes-benz-drive-pilot.html, 2023, visited on 12/18/2023

Meteer, O. et al.: Low-Power Sign-Magnitude FFT Design for Radar (2021)
Meteer, Oğuz; Bekooij, Marco J. G.: Low-Power Sign-Magnitude FFT Design for FMCW Radar Signal Processing, in: Workshop on Design and Architectures for Signal and Image Processing (14th Edition), pp. 52–59, 2021

Mobileye: A Brief History of Autonomous Vehicles (2023)
Mobileye: A Brief History of Autonomous Vehicles – from Renaissance to Reality, URL: https://www.mobileye.com/blog/history-autonomous-vehicles-renaissance-to-reality/, 2023, visited on 11/30/2023

Ngo, A.: Methodology for Validation of Radar Simulation (2023)
Ngo, Anthony: A Methodology for Validation of a Radar Simulation for Virtual Testing of Autonomous Driving, PhD Thesis, Universität Stuttgart, 2023

Norouzian, F. et al.: Phenomenology of Automotive Radar Interference (2021)
Norouzian, Fatemeh; Pirkani, Anum; Hoare, Edward; Cherniakov, Mikhail; Gashinova, Marina: Phenomenology of Automotive Radar Interference, in: IET Radar, Sonar & Navigation, Vol. 15, pp. 1045–1060, 2021

Oberkampf, W. L. et al.: Validation Under Aleatory and Epistemic Uncertainty (2007)
Oberkampf, William L.; Ferson, Scott: Model Validation Under Both Aleatory and Epistemic Uncertainty, in: NATO/RTO Symposium on Computational Uncertainty in Military Vehicle Design, vol. Paper No. AVT-147/RSY-022. P. 26, 2007

Oberkampf, W. L. et al.: Measure Agreement between Computation and Experiment (2006)
Oberkampf, William Louis; Barone, Matthew F.: Measures of Agreement between Computation and Experiment: Validation Metrics, in: Journal of Computational Physics, Vol. 217, pp. 5–36, 2006

Peters, S. et al.: Opportunities of Automated Driving (2023)
Peters, Steven; Abendroth, Bettina; Adamy, Jürgen; Ahrend, Klaus-Michael; Findeisen, Rolf; Kassens-Noor, Eva; Peters, Jan; Rinderknecht, Stephan; Janine, Wendt: Opportunities of Automated Driving - The Darmstadt View on AD, URL: https://www.fzd.tu-darmstadt.de/media/fachgebiet_fzd/responsive_design/content/forschung_12/Darmstadt_View_on_AD_updated_1.pdf, 2023, visited on 12/18/2023

Rapp, M. L.: Messkampagne für winkelabhängige RCS-Profile (2021)
Rapp, Manuel Lukas: Messkampagne Zur Bestimmung Der Winkelabhängigen RCS-Profile von Verkehrsteilnehmern, bathesis, Technische Universität Darmstadt, 2021

Roache, P. J.: Validation in Fluid Dynamics and Related Fields (2019)
Roache, Patrick J.: Validation in Fluid Dynamics and Related Fields, in: Beisbart, Claus; Saam, Nicole J. (Hrsg.): Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives, Springer International Publishing, 2019

Robson, K.: Fully Self-Driving Cars Unlikely before 2035, Experts Predict (2023)
Robson, Kurt: Fully Self-Driving Cars Unlikely before 2035, Experts Predict, URL: https://www.verdict.co.uk/fully-self-driving-cars-unlikely-before-2035-experts-predict/, 2023, visited on 11/30/2023

Rosenberger, P.: Metrics for Simulating Sensors (2022)
Rosenberger, Philipp: Metrics for Specification, Validation, and Uncertainty Prediction for Credibility in Simulation of Active Perception Sensor Systems, PhD Thesis, Technische Universität Darmstadt, 2022

Rosenberger, P. et al.: Towards Generally Accepted Validation Methodology (2019)
Rosenberger, Philipp; Wendler, Jan Timo; Holder, Martin Friedrich; Linnhoff, Clemens; Berghöfer, Moritz; Winner, Hermann; Maurer, Markus: Towards a Generally Accepted Validation Methodology for Sensor Models - Challenges, Metrics, and First Results, in: 12th Grazer Symposium Virtuelles Fahrzeug (GSVF), 2019

Roth, E. et al.: Analysis and Validation of Perception Sensor Models (2011)
Roth, Erwin; Dirndorfer, Tobias; Knoll, A.; von Neumann-Cosel, Kilian; Ganslmeier, T.; Kern, Andreas; Fischer, Marc: Analysis and Validation of Perception Sensor Models in an Integrated Vehicle and Environment Simulation, in: 22nd Enhanced Safety of Vehicle Conference (ESV), 2011

Roy, C. J. et al.: Framework for Verification, Validation, and Uncertainty (2011)
Roy, Christopher J.; Oberkampf, William L.: A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing, in: Computer Methods in Applied Mechanics and Engineering, Vol. 200, pp. 2131–2144, 2011

Ryde, J. et al.: Laser and Radar Ranging in Adverse Conditions (2009)
Ryde, Julian; Hillier, Nick: Performance of Laser and Radar Ranging Devices in Adverse Environmental Conditions: Ryde & Hillier: Comparison of Laser and Radar in Adverse Environmental Conditions, in: Journal of Field Robotics, Vol. 26, pp. 712–727, 2009

Schaermann, A.: Systematische Bewertung umfelderfassender Sensormodelle (2020)
Schaermann, Alexander: Systematische Bedatung und Bewertung umfelderfassender Sensormodelle, PhD Thesis, Technische Universität München, 2020

Scharf, L. L. et al.: Statistical Signal Processing (1991)
Scharf, Louis L.; Demeure, Cédric: Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, Addison-Wesley Series in Electrical and Computer Engineering. Digital Signal Processing, Addison-Wesley Pub. Co, 1991

Schipper, T. et al.: RCS Measurement for Automotive Objects (2011)
Schipper, Tom; Fortuny, Joaquim; Tarchi, D.; Reichardt, Lars; Zwick, Thomas: RCS Measurement Results for Automotive Related Objects at 23-27 GHz, in: 5th European Conference on Antennas and Propagation (EUCAP), p. 686, 2011

Schöner, H.-P.: Automotive Needs and Expectations towards Driving Simulation (2018)
Schöner, Hans-Peter: Automotive Needs and Expectations towards Next Generation Driving Simulation, in: Driving Simulator Conference 2018, 2018

Society of Automotive Engineers: SAE-J3016 (2021)
Society of Automotive Engineers: SAE-J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, URL: https://www.sae.org/standards/content/j3016_202104/, 2021, visited on 01/06/2022

Stripling, H. F. et al.: The Method of Manufactured Universes (2011)
Stripling, H. F.; Adams, M. L.; McClarren, R. G.; Mallick, B. K.: The Method of Manufactured Universes for Validating Uncertainty Quantification Methods, in: Reliability Engineering & System Safety, Vol. 96, pp. 1242–1256, 2011

Toss, H. et al.: Radar Reflectivity Spatial Profile of 3D Surrogate Targets and Real Vehicles (2019)
Toss, Henrik; Karlsson, Kristian: Radar Reflectivity Spatial Profile of 3D Surrogate Targets and Real Vehicles, in: 2019 13th European Conference on Antennas and Propagation (EuCAP), pp. 1–5, 2019

UNECE: Validation Method for Automated Driving (2022)
UNECE: Validation Method for Automated Driving -39th SG2 Session (Virtual Testing), URL: https://wiki.unece.org/display/trans/VMAD-39th+SG2+session, 2022, visited on 12/18/2023

VDI/VDE: VDI/VDE 2206: Development of Mechatronic and Cyber-Physical Systems (2021)
VDI/VDE: VDI/VDE 2206: Development of Mechatronic and Cyber-Physical Systems, 2021

Viehof, M.: Objektive Qualitätsbewertung von Fahrdynamiksimulationen (2018)
Viehof, Michael: Objektive Qualitätsbewertung von Fahrdynamiksimulationen Durch Statistische Validierung, PhD Thesis, Technische Universität Darmstadt, 2018

Voyles, I. T. et al.: Model Validation in the Presence of Uncertainty (2014)
Voyles, Ian T.; Roy, Christopher J.: Evaluation of Model Validation Techniques in the Presence of Uncertainty, in: 16th AIAA Non-Deterministic Approaches Conference, 2014

Wachenfeld, W. et al.: The Release of Autonomous Vehicles (2016)
Wachenfeld, Walther; Winner, Hermann: The Release of Autonomous Vehicles, in: Maurer, Markus; Gerdes, J. Christian; Lenz, Barbara; Winner, Hermann (Hrsg.): Autonomous Driving, Springer Berlin Heidelberg, 2016

Waldschmidt, C. et al.: Automotive Radar — From First Efforts to Future Systems (2021)
Waldschmidt, Christian; Hasch, Juergen; Menzel, Wolfgang: Automotive Radar — From First Efforts to Future Systems, in: IEEE Journal of Microwaves, Vol. 1, pp. 135–148, 2021

Wei, Y. et al.: Extended Target Recognition in Cognitive Radar Networks (2010)
Wei, Yimin; Meng, Huadong; Liu, Yimin; Wang, Xiqin: Extended Target Recognition in Cognitive Radar Networks, in: Sensors (Basel, Switzerland), Vol. 10, pp. 10181–97, 2010

Williamson, R. C. et al.: Probabilistic Arithmetic (1990)
Williamson, Robert C.; Downs, Tom: Probabilistic Arithmetic. I. Numerical Methods for Calculating Convolutions and Dependency Bounds, in: International Journal of Approximate Reasoning, Vol. 4, pp. 89–158, 1990

Winner, H.: Automotive RADAR (2016)
Winner, Hermann: Automotive RADAR, in: Winner, Hermann; Hakuli, Stephan; Lotz, Felix; Singer, Christina (Hrsg.): Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort, Springer International Publishing, 2016

Zhou, Y. et al.: Towards Deep Radar Perception for Autonomous Driving (2022)
Zhou, Yi; Liu, Lulu; Zhao, Haocheng; López-Benítez, Miguel; Yu, Limin; Yue, Yutao: Towards Deep Radar Perception for Autonomous Driving: Datasets, Methods, and Challenges, in: Sensors, Vol. 22, p. 4208, 2022

Own Open Source Content

Holder, Martin; **Elster, Lukas**; Winner, Hermann: Source Code xosc-Converter: ‘‘Digitalize the twin: A method for calibration of reference data for transfer real-world test drives into simulation‘‘,
https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/2993,
2021, Accessed 03.05.2024.

Rosenberger, Philipp; Linnhoff, Clemens; **Elster, Lukas**: Modular OSMP Framework,
https://gitlab.com/tuda-fzd/perception-sensor-modeling/modular-osmp-framework,
2022, Accessed 03.05.2024.

Elster, Lukas, Holder, Martin; Rapp, Manuel: RCS Measurement Dataset,
https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/3606,
2022, Accessed 03.05.2024.

Elster, Lukas, Holder, Martin; Rapp, Manuel: RCS Measurement Evaluation Code,
https://gitlab.com/tuda-fzd/fzd-datasets/rcs-measurement,
2022, Accessed 03.05.2024.

Rosenberger, Philipp; Linnhoff, Clemens; **Elster, Lukas**: Reflection Based Lidar Object Model,
https://github.com/openMSL/sl-1-2-reflection-based-lidar-object-model,
2023, Accessed 03.05.2024.

Linnhoff, Clemens; Rosenberger, Philipp; **Elster, Lukas**: Object Based Generic Perception Object Model,
https://github.com/openMSL/sl-1-3-object-based-generic-perception-object-model, 2023, Accessed 03.05.2024.

Elster, Lukas; Linnhoff, Clemens; Rosenberger, Philipp: Reflection Based Radar Object Model,
https://github.com/openMSL/sl-1-1-reflection-based-radar-object-model,
2023, Accessed 03.05.2024.

Rosenberger, Philipp; Hinsemann, Timo; Linnhoff, Clemens; **Elster, Lukas**: PerCollECT — Lidar Limbs,
https://github.com/PerCollECT/LidarLimbs,
2022, Accessed 03.05.2024.

Linnhoff, Clemens; Hinsemann, Timo; **Elster, Lukas**; Rosenberger, Philipp: PerCollECT — RadarRami,
https://github.com/PerCollECT/RadarRami,
2022, Accessed 03.05.2024.

Elster, Lukas; Hinsemann, Timo; Linnhoff, Clemens; Rosenberger, Philipp: PerCollECT — CameraCopse,
https://github.com/PerCollECT/CameraCorpse,
2022, Accessed 03.05.2024.

Elster, Lukas; Hinsemann, Timo; Linnhoff, Clemens; Rosenberger, Philipp: PerCollECT — UltrasonicUnderwood,
https://github.com/PerCollECT/UltrasonicUnderwood,
2022, Accessed 03.05.2024.

Supervised Theses

Bai, Juntao: Umfelderkennung für Automatisiertes Fahren basierend auf Punktdetektionen von Radar.
Master-Thesis Nr. 823/21

Gröger, Simon: Analyse und Entwicklung einer Objektklassifizierung von Zusatzzeichen der StVO basierend auf Kameradaten eines automatisierten Fahrzeugs.
Bachelor-Thesis Nr. 1380/21

Hinsemann, Timo: Analyse von Effekten in Lidardaten für die virtuelle Absicherung automatisierter Fahrfunktionen.
Bachelor-Thesis Nr. 1387/21

Hofrichter, Kristof: Entwicklung eines FMCW-Lidar Signalverarbeitungsmodells.
Master-Thesis Nr. 871/23

Lüsebrink, Marcel: Entwicklung, Konstruktion und Inbetriebnahme eines modularen, teilautomatisierten Validierungsprüfstands für FMCW-Lidar Sensoren.
Master-Thesis Nr. 885/23

Park, Sochoong: Entwicklung und Analyse eines Algorithmus zur gemeinsamen Schätzung der Eigenbewegung und Ausrichtung mehrerer Eck-Radarsensoren.
Master-Thesis Nr. 855/22

Paul, Frederic: Umfelderkennung für Automatisiertes Fahren basierend auf Punktdetektionen von Radar.
Master-Thesis Nr. 832/21

Rapp, Manuel: Messkampagne zur Bestimmung der winkelabhängigen RCS-Profile von Verkehrsteilnehmern.
Bachelor-Thesis Nr. 1384/21

Schweiwe, Gunnar: Identifikation und Analyse von Beugungseffekten bei Radar-Sensoren.
Master-Thesis Nr. 810/21

Staab, Jan: Entwicklung und Anwendung einer Methode zur Validierung von Radarmodellen.
Master-Thesis Nr. 856/23

Tsai, Cheng-Ting: Sim2Real Transfer für Lidar Perzeption.
Master-Thesis Nr. 848/22

Walter, Angelo: Entwicklung eines Lidarraytracing Algorithmus mit Hilfe von OptiX.
Master-Thesis Nr. 843/22

Yörük, Oguz: Untersuchung der Einflüsse von Gischt auf Lidar- und Radarsensoren automatisierter Fahrzeuge.
Bachelor-Thesis Nr. 1389/21