NPL

≋Met Office

November 2020

Proof of Concept Study: A Framework for the Reliable Testing of CAV sensor performance and degradation in different weather conditions

First Revision: May 2021

Notifications

This report was produced by National Physical Laboratory and the Met Office and is protected by copyright. It may not be copied, reproduced, republished or transmitted in any way except for your own personal, non-commercial use. This is provided on an "as-is" basis without any warranty of any kind.

The authors would like to thank the organisations who took part in workshops or interviews and provided valuable input for this work.

Authorisation page

Written By:	• Dr Dave Jones, Principal Scientific Consultant, Observations R&D, Met Office • Dr Richard Dudley, Science Area Lead and Principal Research Scientist, Electromagnetic Technologies Group, National Physical Laboratory • Louise Wright, Head of Science for Data Science, National Physical Laboratory • Andre Burgess, MBA, Strategy Directorate, National Physical Laboratory
Contributing Specialists	• Michael Szczepanski, Senior Business Development Manager, Met Office • Dr Stephan Havemann, Senior Scientist, Met Office • Nawal Husnoo, Senior Scientist, Met Office • Robert Scovell, Senior Scientist, Met Office • Dr Henry Odbert, Senior Scientist, Met Office • Jill Dixon, Senior Scientific Consultant, Met Office • Dr Jeremy Price, Scientific Manager, Met Office • Dr. John Molloy, Senior Research Scientist, Electromagnetic Technologies Group, National Physical Laboratory (to March 2020) • Dr Imran Mohamed, Higher Research Scientist, Electromagnetic Technologies Group, National Physical Laboratory • Dr Manoj Stanley, Higher Research Scientist, Electromagnetic Technologies Group, National Physical Laboratory • Dr Fengping Li, Higher Research Scientist, Electromagnetic Technologies Group, National Physical Laboratory • Dr Carmine Clemente, Senior Lecturer, Chancellor's Fellow Centre for Signal and Image Processing, Department of Electronic and Electrical Engineering, University of Strathclyde
Reviewed by	• David Webb, Head of Innovation, CCAV

Table of Contents

1. Definitions

The following industry-acknowledged terms and definitions apply in this document.

automated driving system (ADS)
hardware and software that are collectively capable of performing the dynamic driving task on a sustained basis, regardless of whether it is limited to a specific operational design domain

automated driving system entity (ADSE)
organization or individual that puts an automated driving system forward for authorization for use and is responsible for its safety

Connected and Automated vehicle (CAV)
a vehicle designed or adapted to be capable, in at least some circumstances or situations, of safely driving itself on roads or other public places in Great Britain.

Dynamic driving task (DDT)
real-time operational and tactical functions required to operate a vehicle safely in on-road traffic

edge case
rare but plausible independent parameter value within a scenario

operational design domain (ODD)
operating conditions under which a given driving automation system or feature thereof is specifically designed to function

scenario
description of a driving situation that includes the pertinent actors, environment, objectives and sequences of events

sensor fusion
process of combining information from multiple sensor types in order to improve performance over that obtainable from a single sensor type

simulation
computer generated environments used to test components, systems or human behaviours

validation
means by which it is proven beyond reasonable doubt that an end product meets its design intent and stated performance specification

vehicle
motorised, wheeled conveyance that is mechanically propelled and intended or adapted for use on roads

verification
evaluation of a system to prove that it meets all its specified requirements at a particular stage of its development

For this document, the following terms have the specified definitions:

Controlled Environmental Test Facility (CETF)

also known as CETF refers to a planned large-scale (in terms of size) controlled environment test facility where weather conditions (e.g. rain, fog…) are emulated and are to some degree controllable

Framework
refers to the usable and reliable framework for understanding how well AV sensors perform in different weather-related conditions, including when the sensors cannot be relied upon

Real World Weather Testbed
fully instrumented test range in the outside world. Includes measurement of all relevant and available meteorological variables

Study
the Study is this report, which constitutes the findings of research undertaken by Met Office and NPL to prove the concept for the Framework

Stage 1
refers to the first stage of the project to produce the Framework, the output of which is the Study

Stage 2
refers to the second stage of the project to produce the Framework, which is based on adopting the recommendations made in the Study

2. Introduction

Note: For consistency the term Connected and Automated Vehicles (CAV) will be used to refer generally to Autonomous and Automated vehicles in this document.

With the advent of 'Automated' or 'Self-Driving' Vehicles, a new paradigm is being created requiring entirely new types of tests for the systems which are replacing the human driver.

One key challenge arising from this is the reliance on a range of sensors for safety critical applications in CAV. How these sensors perform and where they might fail must be clearly understood; failure to do so may lead to serious safety issues. [1]

Having the right tools to determine when sensors will fail is therefore vital for CAV to operate safely. The weather is a dominant aspect of the changing conditions that may affect sensors. This is an especially complex challenge and the same need has been observed across regulators and CAV developers around the world[2], indicating there is an opportunity for the UK to build and demonstrate leadership in this area.

Two national agencies, Met Office and National Physical Laboratory have undertaken this research project ('The Study') on behalf of CCAV to specify what those tools should be: a usable and reliable framework for understanding how well sensors perform in different weather-related conditions, including when the sensors cannot be relied upon ('The Framework'). When fully developed, this Framework will support validation, safety assurance and simulation testing of CAV, across the UK. NPL and Met Office, as neutral agencies and leaders in their respective fields of Metrology and Meteorology, have a vital coordinating role in developing this work.

An underlying principle of this Study and the subsequent actions to adopt its recommendations, is to ensure the Framework's usability and acceptability, early in its development; this is being achieved through engagement and collaboration across current UK stakeholders and trials, under the direction of CCAV. The Study includes a description of engagement activities and findings to date.

The study recommends creating an integrated Ecosystem of Test environments, where linked tests of different levels of complexity are conducted on sensors and full systems, to achieve the required level of confidence in sensor performance; and this being subsequently adopted as a feature of CAV Safety Assurance. This will be enabled through a combination of existing and newly developed test capabilities.

This 'Test Ecosystem' will be underpinned by a series of critical success factors, defined in the Study, as well a minimum set of measurements for each type of test and an entirely new Taxonomy for weather impact on sensors. This will also involve the development of new testbed infrastructure.

Full recommendations can be found in Section 5. Adoption of the recommendations must be undertaken in close collaboration with industry, testbeds and regulators to ensure success.

---

[1] Thatcham.org, 2020, https://www.thatcham.org/thatcham-research-and-association-of-british-insurers-urge-government-to-revise-plans-for-introduction-of-automated-lane-keeping-systems-to-ensure-road-safety/

[2] AAA, 2019, *Automatic Emergency Braking with Pedestrian detection,* , https://www.aaa.com/AAA/common/aar/files/Research-Report-Pedestrian-Detection.pdf

3. Study Overview

This section describes the background and need for the Study, as well as the relevance to the safety of CAV's. The objectives and scope of the Study are described also.

3.1. Context of the study

All CAV sensors have their performance reduced by adverse weather to some extent or another and so weather information is a key consideration in the development and operation of CAVs, as well as other intelligent mobility solutions. This reduction in performance is often the result of complex interactions between the electromagnetic spectrum used by the sensors and the weather at its finest scales e.g. individual raindrops. Managing the implications of this complexity effectively and pragmatically offers the opportunity of developing a world-leading test ecosystem and avoiding wasted investment in capability.
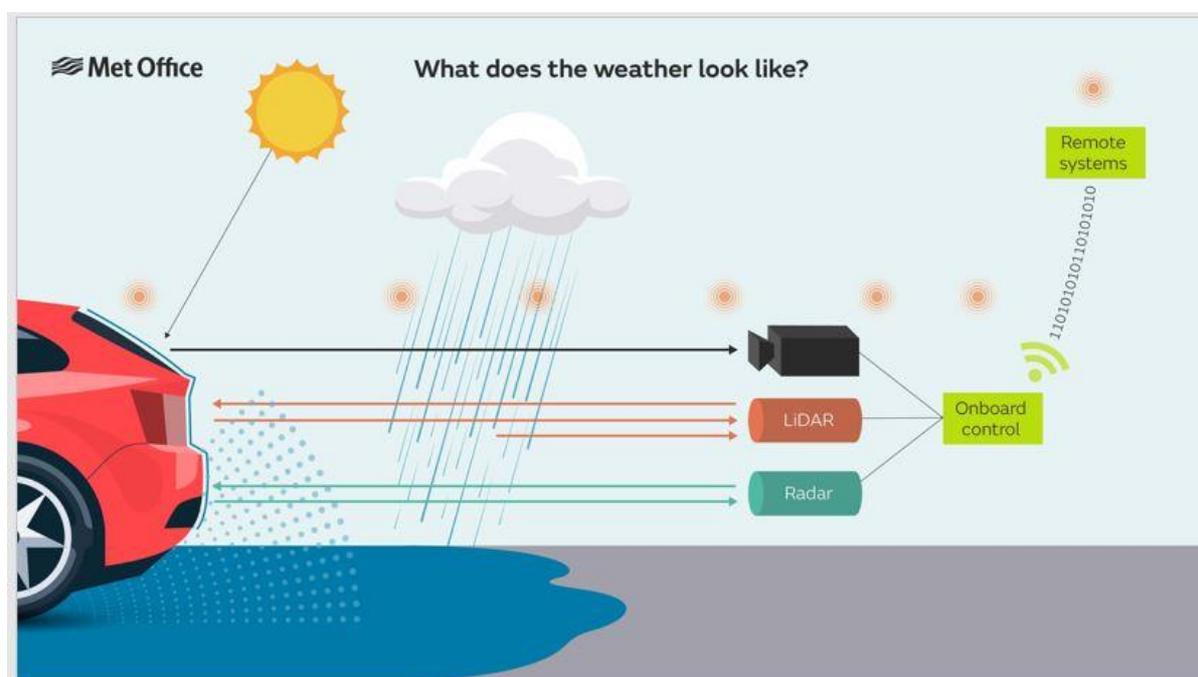
Figure 1. Impact of weather types on sensors using rain as an example. *The red dots are indicative of where the impact of rainfall might be considered. From left to right these are: water film on the target vehicle, attenuation/glare from road spray, attenuation/glare/backscatter from falling rain, water film on the sensor, net impact on processed sensor output, the impact on the ADS and finally the impact of rain on infrastructure.*

The recent Feasibility Study produced by NPL and Connected Places Catapult for CCAV on Performance Testing for CAV Sensors [3] identified the following:

- There is a need for a more structured and quantitative approach that provides a traceable link from observable weather to CAV impacts at the vehicle, fleet and network level; and in turn to understand and validate sensor performance and downstream AI-based perception systems under different conditions.
- There is a need to develop and agree common standards and taxonomies for different types of weather. Moreover, a new technical language is only useful if it is widely used in the field.

---

[3] NPL and Connected Places Catapult, 2020, *Performance Testing for Sensors in Connected and Autonomous Vehicles: Feasibility Study*

- Industry has acknowledged the importance of developing a set of common methodologies and definitions for characterising sensor performance under different conditions.

The Feasibility Study recommended the development of a usable and reliable Framework for characterising sensor performance in different weather-related conditions, including the ability to assess performance outside the design envelope ('The Framework'), along with the creation of associated physical test infrastructure. Uses of this Framework include validation, safety assurance and simulation testing of CAV. This would support a standardised approach to testing CAV in the UK and underpin future sensor testing infrastructure, whilst positioning the UK as a leader in this space internationally.

This Study is the first stage of the development of the Framework (Stage 1) and serves as a proof of concept of the Framework. To deliver the Framework in full - to the point of industry deployment - will require a subsequent project – Stage 2 (see Section 9). The focus of research for the Framework development is as follows:
- Characterisation of the relationship between observable weather phenomena and CAV system impacts at the "traffic scale", in order to adequately inform the ODD
- Development of a UK (and world) climatology of edge case weather, fully expressed in terms of CAV impacts
- Development of meteorologically based CAV sensor (and AI) performance standards
- Development of test protocols that provide a traceable link between CAV standards and real-world system performance, including the creation of a "CAV meteorological testbed" and reference virtual environments
- Make best use of 'traditional' meteorological data and CAV sensor data and insights to ensure the safe and efficient operation of individual CAV vehicles and contribute to the enhancement of the UK National Meteorological Service for the wider public good
- Development of demonstration meteorological data and consultancy services which might form the basis of a future market in CAV information services

A fundamental underlying principal of the research approach is that: The propagation of uncertainties in sensor performance through to the performance of perception algorithms and autonomous decision making must be understood and then reflected **in the setting of pragmatic industry standards.** Through direct engagement with current UK stakeholders and trials, this will ensure the framework's usability and acceptability, early in its development.

3.2. Relation to existing Safety Activities

Demonstrating safety is key, not only for legislative acceptance of CAV, but also for public confidence and uptake of the technology. Various UK bodies have generated guidance and tools to support this demonstration and in 2019, the UK Government announced the development of an assurance system CAV PASS, to ensure self-driving vehicles are safe and secure by design and minimise any defects ahead of their testing, sale and wider deployment on UK roads. The work reported here feeds into several aspects of these activities.

The CertiCAV[4] project aims to develop a framework to define and assess the safety of advanced levels of automation, along with demonstrations of the critical aspects of this framework. Their overall vision is that the framework will support scenario-based testing, including a curated database of parameterised scenarios for both real and virtual tests that will be used to seek regulatory approval.

The Study provides two key clarifications to support this approach. The *first clarification is a discussion of the challenges associated with parameterising weather* within these scenarios, and in

---

[4] Connected Places Catapult, 2020, https://cp.catapult.org.uk/events/the-certicav-safety-assurance-framework-for-cavs-consultation-workshop/

particular explains that defining exactly what (for instance) "heavy rain" means, requires a combination of spatial, temporal, and intensity statements that will potentially affect the vehicle, its sensors, and the surrounding infrastructure in complex and interlinked ways.

The **second clarification is the characterisation framework** proposed in this report, which provides a quantified and traceable approach to ensuring that the models of the vehicle sensor suite used in the virtual testing scenarios are an accurate reflection of the true performance of the sensor suite under different weather conditions. Use of the framework of measurements defined in this document will provide appropriate data for development of sensor models that do not rely on idealised assumptions about sensor performance in (for instance) rainy conditions, but instead can capture performance degradation and associated uncertainty in a way that can be reproduced in the virtual testing environment.

Another significant safety activity is taking place under the aegis of BSI, in the development of a small number of complementary Publicly Available Specifications (PASs) aimed at supporting the safe trialling and introduction of AVs.   This study is fully aligned with these:

Firstly, the production of "PAS 1880:2020[5], Guidelines for developing and assessing control systems for automated vehicles" has defined a structure for assuring the safety of the various aspects of AV systems. The work reported here is particularly relevant to what PAS 1880 terms "sensing operations", which state that safety of sensing operations requires that:
"It is demonstrated that, throughout the mission, the AV:
a) is able to determine that it is operating in compliance with its ODDs; and
b) is able to provide the data required by the AV planning operations."

Secondly, "PAS 1883:2020, Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) – Specification"[6], includes the first attempt at describing the operating environment for CAVs.   This study contributed much of the wording to the weather elements of the taxonomy and the framework defined in this document provides a structured approach to gathering evidence that the sensing operations are in compliance with those weather parameters. In addition the approach is able to identify the weather conditions under which the sensing operations cease to provide sufficiently good (i.e. accurate and low uncertainty) data to satisfy the needs of the CAV planning operations.

In general, the framework in this document is consistent with much of the existing CAV safety projects as it is scenario-driven, evidence-based, and provides quantified uncertainties to feed into a risk assessment approach. It provides a traceable methodology to assess complex systems reacting to complex conditions.

3.3. Study Objective and scope
The objective of this study is to prove the concept of, and create an evidence-based specification for, the Framework.

The Study has, therefore, focussed on rainfall as an exemplar, and its impact on a single KPI (maximum range) for two sensors (lidar and mm-wave radar) is demonstrated. By drawing on existing knowledge from metrology and meteorology, the nature of the underlying complexity is demonstrated along with how it can be managed within the test framework. Although a single weather element (rain) is considered for the Study, other weather impact pathways have been identified as part of the industry engagement process.

---

[5] BSI, 2020, https://www.bsigroup.com/en-GB/CAV/pas-1880/
[6] BSI, 2020, https://www.bsigroup.com/en-GB/CAV/pas-1883/

The study has achieved the objective as follows:

- The findings from the research have allowed the participants to understand and articulate the nature of the challenge of weather-related sensor degradation and confirmed that industry sees this as a major challenge which needs to be addressed
- The project has demonstrated how the complexity demands that a range of test approaches must be used *in concert* in order to characterise sensor performance to an acceptable level of uncertainty
- Successful and meaningful engagement across different industry and regulatory stakeholder groups has taken place to ensure the framework is relevant and will have buy in from all stakeholders. The engagement process, which has included involvement in advisory bodies to the development of BSI: PAS for CAV and MUSICC[7] and CertiCAV initiatives, has demonstrated the need for bottom up research and to inform the development of regulation and approvals
- One significant impact has been raising awareness of the role of uncertainties and how these relate to the definition and operational use of the ODD, in particular the description of rainfall in the ODD taxonomy in PAS 1883 as described earlier
- The project has enabled the definition of a generalized framework structure to address this, which has gained acceptance from industry stakeholders, and which is aligned to current programmes to develop standards for CAV assurance and testing
- Undertaking this first stage project has provided a clear definition of what is required for Stage 2 - the development of the full framework – and reduced the risk to that stage, which will in turn support successful deployment of CAV in the UK

3.4. Out of Scope

This report does not explicitly discuss testing of AI systems in any great level of detail. AI systems, usually machine learning algorithms, are used in CAVs to control the vehicle but are also used on individual sensors to identify objects. Testing of, reliability of, and evaluation of the uncertainty associated with the results of machine learning algorithms is an ongoing area of academic research. The topic is too complex to address within this project, but some aspects are discussed in brief in section 8.5.

This limitation is related to the concept of functional testing, discussed in more detail in section 6.7. The aim of much of the framework proposed here is to characterise the response of the sensing elements of the sensor system (i.e. the parts of the sensor that send and receive the signal) under different weather conditions. **This demarcation at the lower levels of testing supports root cause analysis at the higher levels of testing, so that progress to rectify problems can be made more quickly. If a sensor fails to identify a car in the rain during a test, but the sensing elements of that sensor have been shown to have adequate performance in the rain, then we can have confidence that the problems lies with the software not the hardware and can work to rectify the problem accordingly.**

It is acknowledged that access to the raw signal may be seen as undesirable by sensor manufacturers, but it will enable more rapid product development and product acceptance if this base case testing can be carried out easily.

3.5. Study Activities

The working approach has been to establish the minimum viable end-to-end thread, **sharing this with stakeholders at the earliest opportunity,** and then progressively building additional depth and capabilities incrementally.

---

[7] Connected Places catapult, 2019, https://cp.catapult.org.uk/case-studies/musicc/

3.6. Document Structure

The document structure follows the evolution of the theoretical approach during the process of engagement with industry. The resulting recommendations (section 5) link to other sections to provide more detail on how the recommendations will be fulfilled.

Section 5 and Section 6 - Recommendations and Framework Methodology - are the key sections, which describe what is required in terms of process, infrastructure and innovation and provide a roadmap to establish a useable and traceable framework for the characterisation of CAV sensors with respect to adverse weather. The subsequent sections, Section 7 and Section 8 provide the evidence to support these recommendations and the appendices contain substantial technical information related to sections 7 and 8.

Section 9 provides more insight on the range of activities required to deliver the recommendations; these would form the basis for more formal planning and resourcing for Stage 2, the development and demonstration of the application of the Framework.

4. The Industry Context

An underlying principle in the development of the Framework is to ensure its usability and acceptability, early in its development; this is being achieved through engagement and collaboration across current UK stakeholders and trials, under the guidance of CCAV. This section and the accompanying appendix explain the parameters and process of engagement during the Study.

4.1. Challenge Statements

The following reflect the main challenges which have been identified through the engagement process and which the Study focuses on addressing:

a) ***Functional testing alone is not sufficient to validate sensor performance.*** The Automotive industry is a globalized and very influential industry, where the long-established process of Vehicle **Type Approval** is used to confirm that production samples of a design will meet specified performance standards. However, in the past the perception and navigation (i.e. decision making) was undertaken by a human which is not part of the vehicle type approval until now – for example, the testing of CAV performance in bad weather will need to characterise impactful weather thresholds from the point of view of a sensor, which is not the same as that of a human. Furthermore, certification of CAV is expected to depend on the accumulation of multiple pieces of evidence that the vehicle performs safely in a set of pre-defined scenarios. The use of an evidence chain rather than a single pass/fail test makes it less likely that the vehicle will be designed specifically to pass the test rather than to drive safely in general. The scenario set for full autonomy is expected to be large, varied, and to vary geographically with common features.

b) ***Current simulation environments cannot provide a reliable link to real world testing because they have insufficiently defined sensor models.*** It is broadly acknowledged[8] that the simulation (virtual and physical) of automated driving functions is the only practical way to assess the many possible scenarios that shall comprise a sensor system design verification plan (DVP). Consequently, it is important that a) a computer model of a sensor should accurately represent its behaviour under all relevant circumstances, including poor weather; and b) physical simulation of driving scenarios should recreate equivalent environmental conditions and be able to do so on a repeatable basis. All uncertainties need to be quantified and then sampled in virtual scenarios. (see section on Uncertainty)

c) ***ODD Taxonomies which relate to sensor performance degradation are not yet sufficient for purpose.*** Interactions between sensors and the weather is very complex. Therefore, a useable assurance framework must manage this complexity on behalf of the industry, only exposing it where absolutely essential and being clear about what unavoidable uncertainties remain. The framework must relate clearly to ODD taxonomies. For example, weather, using rain as an example, is not as simple as "heavy rain"/"rain level 3" etc, because it varies spatially and temporally. So, a shower that is "on average" apparently safely within an ODD threshold may contain local and short instances of very heavy rain that are outside the ODD.

4.2. Engagement process

It was - and will continue to be – a core requirement to engage as widely as possible with relevant stakeholders, including: ADS developers; sensor manufacturers; OEMs; test centres, proving grounds; simulation operators; regulatory entities and standards bodies. To date over 50 organisations have been involved in the engagement process.

Due to the short timeframe and the disruption caused by the C-19 Pandemic, the engagement process has been truncated to two activities:

---

[8] *Nidhi Kalra, Susan M. Paddock; "*Driving to Safety , **How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?"** RAND Corporation (2016)

<u>Activity 1:</u> Phone interviews. Using the informal summary, the logic flow on page 1 (see [Appendix B](#)), which involved a series of statements, were discussed with each interviewee. The aim was to identify where there was agreement on a statement; and where there wasn't, to engage in further discussion as to alternative reasoning.

<u>Activity 2:</u> Workshop. A technically-focused workshop was held on the 5th March 2020 in central London with 30 participants, drawn from those who had contributed to interviews. The workshop objectives were to explore how a standardised solution can be developed for reliably characterising sensor performance in different weather conditions and to ensure it would be useable by all stakeholders. The workshop focused on discussing problem statements and underlying questions to the participants for general round-table discussion.

4.3. Engagement Findings

A brief summary of common themes and outputs arising from both engagement activities is provided below. The full engagement summary is available in [Appendix A](#).

Stakeholders were asked to consider the statements made in the logical flow in the discussion document ([Appendix B](#)).

The feedback was that the participants generally agreed with the following:
- The challenge over modelling weather combinations is significant and the existing models are averaged over time and region. Models of weather impacts on sensors overall are immature
- It is not just the specified weather event testing (e.g. rain/sun/frost/wind) that is important to ensure the safe operation for sensors and systems but also how they deal with (or not) the transition between weather states, which can often be a complex combination of weather types

A significant majority of stakeholders asked agreed with the following:
- Testing (including virtual testing) of CAVs needs to take weather into account. Hence vehicle test specifications, sensor characterisation, and virtual testing environments need to take weather into account
- Better confidence in sensor performance can be achieved by a series of *linked tests* & characterisation exercises at different levels (as per [Figure 3](#)) where sensors and full systems are tested both in reality and virtually. In this context "reality" may include testing in both the natural environment AND user-controlled environmental test facilities
- Because weather demonstrates significant small-scale variability and sensor response has associated uncertainties, uncertainty calculations are critical and must be reflected appropriately in the virtual simulation

The key outcome was the general agreement around the principle that linking the right tests could help increase overall confidence in test results, indicating a solution for the challenge of how to test whether perception systems are providing reliable outputs in different weather conditions.   This underpins the approach taken in this Study. Further, the feedback continues to be gathered and updated as the number and range of stakeholders engaged increases.

Further points to highlight:
a) Another point which was made by those engaged was that three important aspects of an "approval" / "verification" / "certification" method for sensors or perception systems must be borne in mind when considering the process:
   - Cost: The cost of the process to the Vehicle Manufacturer / ADSE requesting "approval" must be reasonable
   - Timing: Assuming that the system meets the requirements, it would be unreasonable for the process to delay vehicle development or launch excessively

- Unified standards: All the organisations providing certified testing services must conform to vehicle sensor safety testing standards which are traceable to the primary established vehicle functional safety standards such as ISO 26262

b) It is important to be able to detect when and how sensor is being impacted.

c) In setting performance standards, it is essential to safeguard against poor performing systems on the road. One of the challenges in this regard is that much of the type approval for automation involves assessment of technical documentation from the manufacturers, which may not contain detailed information. This could be because such information has not been mandated, or because there are IP concerns by the manufacturer.

d) Individual sensor performance testing had to be considered in the overall context of the V shaped system design and test (Equivalent of unit/component tests). An indicative approach was discussed, i.e.: vehicle requirements are x $\Rightarrow$ therefore against known vehicle dynamics define the situational awareness requirements $\Rightarrow$ This then defines mix and spec of sensors.

This V model has been promoted to be the reference model that forms the basis of ISO 26262[9]. Although ISO 26262 and its V framework generally reflect accepted practices for ensuring automotive safety, fully autonomous vehicles present unique challenges in mapping the technical aspects of the vehicle to the V approach which has been summarised by P. Koopman, et.al[10]
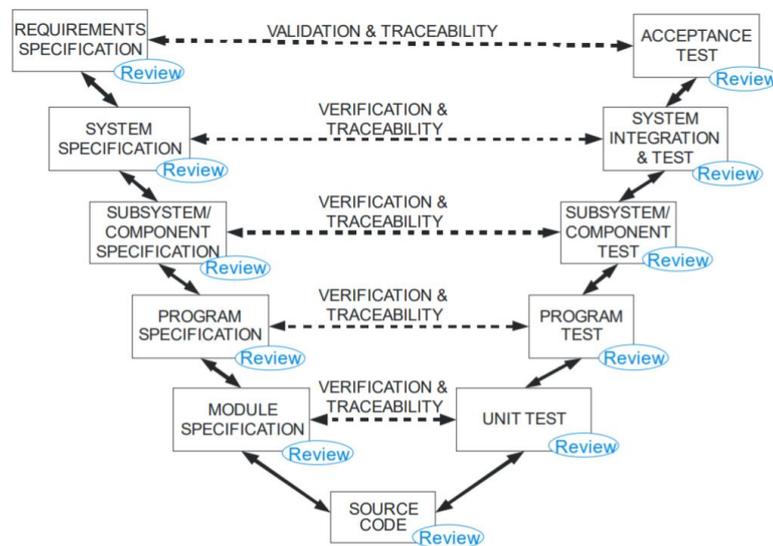
Figure 2: A generic V model (from ISO 26262)

e) Sensors are set to be increasingly relevant to numerous future standards workstreams. As it moves into the next stage, the project should ensure collaboration with other stakeholder groups at a committee level to define and agree the methodologies and tools and then propagate them through standards organisations and industry bodies. A particular focus will be given to standards bodies. Subsequent engagement with bodies exploring standards has indicated that, although approvals, regulation, testing etc use a top-down definition, to understand the challenges and constraints to frame such regulation requires a bottom-up research approach first.

---

[9] ISO 26262: 2018 Road Vehicles – functional safety, https://www.iso.org/standard/68383.html

[10] "Challenges in Autonomous Vehicle Testing and Validation", Philip Koopman & Michael Wagner

5. Recommendations

The following recommendations have been generated through the combination of learnings derived from industry engagement and the outputs from the research and investigation undertaken by NPL and Met Office. **Collectively the recommendations form a roadmap to develop and deploy the Framework in support of CAV testing and safety assurance activities in the UK.** Therefore, coordination and alignment with activities and organisations driving safety assurance and related future regulations is paramount. At the same time, the recommendations also take account of industry requirements for any new testing approaches to be accessible, budgetarily practical and not undermining their ability to protect IP. This research project has reached a point where there is sufficient confidence in following through with the recommendations.

5.1. Create the Ecosystem of Test environments

A series of linked tests and characterisation exercises at different levels of complexity, where sensors and full systems are tested both in reality and virtually, will achieve the required level of confidence in sensor performance (The 'Test Ecosystem'). This could therefore be adopted as a feature of CAV Safety Assurance. It should be noted that in this context "reality" may include testing in both the natural environment and controlled environments where weather conditions are recreated.

The principle is that confidence in sensor and sensor system performance is built up by doing a large number of simple tests under controlled conditions (lowest layer), and as the complexity of the tests increases, the number of tests required decreases because the lower level tests have given confidence in system performance.

- The information derived from the combination and linking of tests is vital for reliable virtual testing of the autonomous system, because virtual testing must reflect the likely variability of the sensor responses.
- Such tests can highlight the gaps in performance that may need to be covered by complementary sensor technologies and help to identify conditions under which the complementarity of the sensors means the system as a whole is still safe even when the performance of one sensor is degraded.

Diagrammatic examples and a description of how the ecosystem could be structured is given in Section 6.1, with a virtual walkthrough in Section 6.2

The Test Ecosystem must include the elements outlined in the table in section 6.4, where more detailed descriptions are provided. The table demonstrates some of the strengths and weaknesses of each environment, which when used in combination, will deliver assurance of the sensor performance in the full range of weather scenarios. It should be noted that whilst, where possible existing facilities should be used for testing, some of these environments currently do not exist. As part of adopting this recommendation they will need to be developed (see Section 9). In particular this includes a Real World Weather Testbed and a CETF, where the physical world is emulated.

5.2. Critical Success Factors for the Test Ecosystem

The following recommended critical success factors are necessary to ensure the Test Ecosystem in 5.1 will deliver the required confidence in sensor performance.

- *Cost and duration of tests*: For a testing regime to be useable it must be affordable to the industry and also efficient enough to enable competitive advantage through innovation
- *Traceability*: The individual elements of the test environment ecosystem must include linkages that **ensure that the characterisation information from the different tests can be traced back to national standards and combined within a well-defined uncertainty framework.** Traceability is key to demonstrating that measurements and characterisation are trustworthy, and for international acceptance of standards. (For more information see Section 6.5)

Traceability in this application is an ongoing process rather than a one-off demonstration, and requires:

- o Sensor calibration using standard ISO calibration charts, which traces back to standards such as ISO 12233:2017[11], makes the measurements traceable back to national standards for sensors such as camera and LiDAR. Calibration targets included in ISO 19206-2:2018[12] makes the vehicle radar measurements traceable. (Section 8)
- o Repetition of the same test across multiple environments, e.g. running the C ETF dry to compare to lab tests, simulation of tests in the virtual environment for validation, etc.
- o Periodic use of fully characterised (reference) sensors across environments as well as the CAV sensors themselves
- o Use of the same KPIs at all levels of testing to ensure direct comparison is straightforward
- o Sufficient number of tests repeated in the well-controlled environments to ensure baseline/ground truth performance is well understood and to obtain repeatable results, so that the effects of weather can be well isolated and quantified
- o Data sharing infrastructure and methodologies to enable linking of test data, including recommendations for standardisation of data and metadata formats

- *Quantifying uncertainty*: Characterising uncertainty of the sensor response with respect to weather is critical to ensure informed safety, but also enables decisions to be made about the balance of investment across the Test Ecosystem and potentially identifies areas of future focus for development of improved sensors. (Section 6.5)
- *Linkage to the ODD:* The characterisation framework must relate unambiguously to the definitions of weather within the ODD, such as those being developed in BSI PAS 1883. In particular, the accuracy to which the weather elements of the ODD can be measured when the CAV is in operation should directly inform the level of detail to which the characterisation with respect to the weather is performed. This is key to both enabling affordability of the sensor testing **and** also ensuring that the uncertainties in ensuring a CAV is within its specified ODD are fully understood. (Section 7.4)

5.3. Defining the Measurements and Equipment for each testbed

The requisite set of measurements and associated equipment are to be defined for each testbed. The measurements for each testbed differ and will be completed in accordance with tests defined in Section 6.4. Measurements are required for the following three areas, with indicative examples provided.

- **Meteorological**: Basic (traditional) meteorological thermodynamic variables (temperature humidity, wind, pressure); rain gauge and operational C-band rainfall radar network to measure rainfall rate; particulate/droplet measurements (including disdrometers, fog spectrometer, aerosol spectrometer) for hydrometeor size distribution; spectroradiometer for light wavelength and amplitude measurements; direct and indirect short and longwave flux measurements. Section 7 and Appendix G provide more information relating to Meteorological measurement requirements
- **Reference targets**: e.g. ISO test charts: Geometric ISO 17850:2015[13], Resolution and Spatial Frequency ISO 12233:2017, may be used for Camera; extruded versions may be suitable for Lidar, standard spherical and triangular trihedral type (corner reflector) targets would be suitable for radar. Section 8 and Appendix D provide more information relating to reference targets
- **KPI tests**: Attenuation coefficient for interrogating wavelength, noise floor, dynamic range, maximum and minimum detectable range, range resolution, angular resolution, field of view,

---

[11] ISO 12233:2017 - Photography — Electronic still picture imaging — Resolution and spatial frequency responses, https://www.iso.org/standard/71696.html

[12] ISO 19206-2:2018 - Road vehicles — Test devices for target vehicles, vulnerable road users and other objects, for assessment of active safety functions.. https://www.iso.org/standard/63992.html

[13] ISO 17850:2015 - Photography — Digital cameras — Geometric distortion (GD) measurements, https://www.iso.org/standard/60819.html

target contrast threshold, modulation transfer function probability of detection, probability of false alarm. Section 8 and Appendix E provide more information relating to KPIs

5.4. Producing a reliable, structured taxonomy of weather impact pathways

The Study includes a proposed structure for a tabular (by sensor type and degradation mechanism) **taxonomy** of weather impact pathways, in section 8.2. This is required to enable priority mechanisms to be explored early. Through investigations conducted by Met Office and NPL alongside the industry engagement process, it is clear that each weather element provides a range of pathways to sensor degradation, often in combination with other weather elements and sometimes as secondary mechanisms e.g. spray, films of water on sensor surfaces etc. Furthermore, there are effects of weather on infrastructure, such as water films forming on road signs, that may affect the vehicle's ability to detect or identify common objects.   The Taxonomy will also incorporate data on weather mitigation effects on the sensor, where such mitigation technologies are available for testing and can be properly characterised.

5.5. Collaboration and Deployment

This Study was only possible through engagement and collaboration with CAV sector stakeholders. To move this forward to the next stage will require deeper and wider collaboration, in particular as this progresses closer to implementation, including the formation of an industry steering group. CCAV will need to ensure the relevant parties can be involved in development and deployment of the Framework on an equitable basis.

6. Framework Methodology: Confidence in Sensor Performance

This section describes how the constituent parts of the Framework fit together and how the Test Ecosystem works as a whole to deliver the required outputs – i.e. confidence in sensor performance. It also provides an overview of the important roles that traceability, uncertainty and correlation play in delivering confidence, as well as the relationship between the Framework and Functional testing.

6.1. Overview - The Golden Thread

The Framework must be relevant to the operational context in which the results will be used. Specifically, it must link unambiguously to the ODD. This study demonstrates that even for a single environmental variable such as rainfall, there are many degrees of freedom in the way that sensors might respond to a headline parameter in the ODD.   This knowledge is used to inform the design of the test system and also constrain it in order to avoid unnecessary investment. An underlying principle of the approach requires that the level of detail in assessing sensor performance with respect to a given weather parameter must be proportionate with our ability to measure that parameter when the vehicle is on the road.

With the above in mind, the study has adopted an approach of using existing expertise and models in novel ways to make high level recommendations about the test ecosystem.   The research approach has been to probe some of the key sensitivities to weather *sufficiently* to make recommendations about the components of the ecosystem and next steps with a high level of confidence. At this point, these models should not be considered to be candidate sensor response models for the virtual test environment.   The test ecosystem is illustrated in Figure 3 and Figure 4 (below). Figure 3 shows an example of the testing pyramid concept.
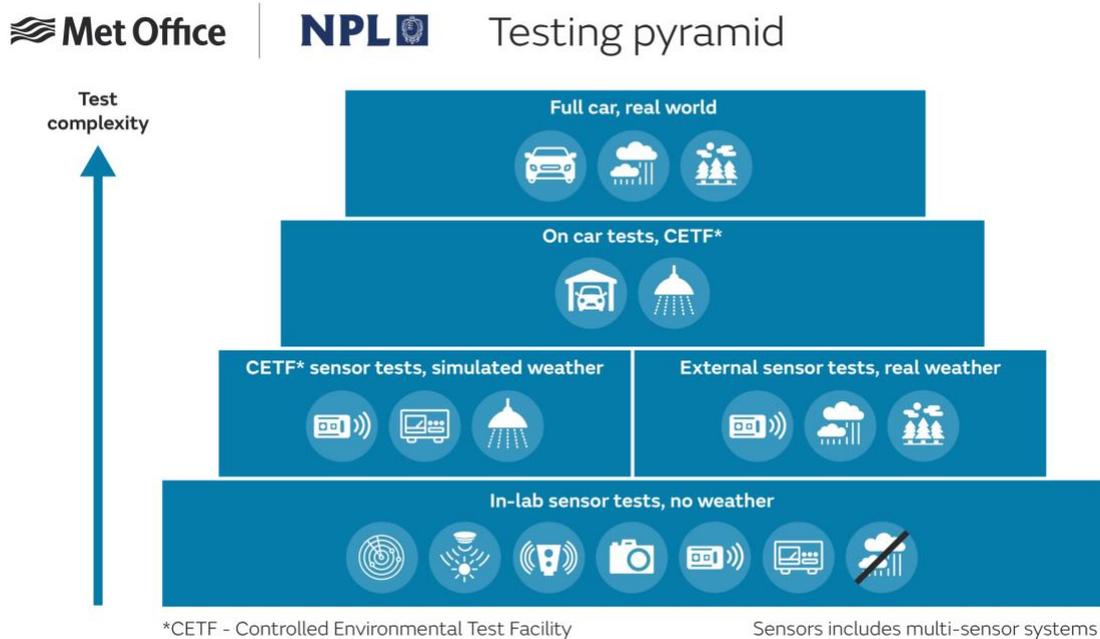
Figure 3: An illustrative (only) example of a testing pyramid. The width of the bars broadly indicates the volume of tests of each kind. (Note it is not a given that the testing regime is sequential from bottom to top, and it is expected that information from each type of test will contribute to the assurance i.e. it is not a case of pass-fail before passing to the next level)

The principle is that confidence in sensor and sensor system performance is built up by doing a large number of simple tests under controlled conditions (lowest layer), and as the complexity of the tests increases, the number of tests required decreases because the lower level tests have given confidence in system performance.   Confidence in the individual sensors under a variety of operating conditions can therefore be built up through repeated in-lab tests, including tests with simulated weather. Tests

of systems of sensors on vehicles under controlled conditions build up confidence that the effects of vehicle motion, sensor mounting (e.g. occlusion by paint layers etc) and other effects only seen on a real vehicle can be characterised; the combination of the lab test data and the on-car test data can be used to quantify these effects.

Both of these sets of tests also generate data that will be used within the virtual test environment, ensuring that the sensor models within the environment accurately reflect the true system behaviour. Finally, full car real world tests can be used to test safe situations and can also be used to validate the virtual testing environment by simulating the real-world test and comparing results to reality. **We refer to the linking and iteration of all of the tests as the 'Golden Thread'**

Figure 4 below shows the parts of the ecosystem and how they might link in more detail. The first step in developing the Golden Thread is to define the scope of the combined set of sensor performance KPIs and weather scenarios that must be captured. These specifications would use a vocabulary similar to that used in section 7.1 and would lead to an unambiguous definition of the required information.
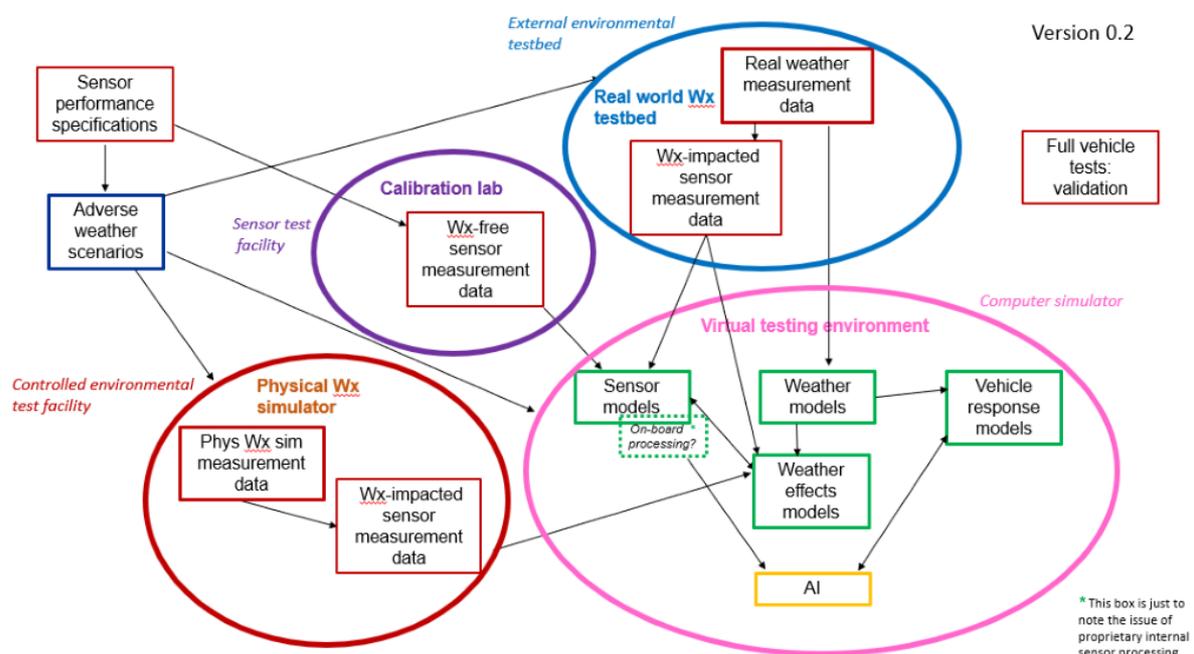
(*__Wx__ = shorthand for weather)

Figure 4: A representation of interlinked test environments for CAV sensors – the figure is explained in more detail below

The sensor specifications (section 8) will feed into the definition of the sensor tests, and the weather specifications (section 7) will feed into the definition of all the tests. The single sensor tests fall into two groups: a set of weather-free tests that are envisaged as being carried out in calibration laboratories (circled in purple in figure 4), and a set of simulated (and controllable) weather tests envisaged as taking place in a CETF (circled in red in figure 4) and uncontrollable real-world weather tests at an external testbed (circled in blue). The tests of multi-sensor systems would also fall into the same groupings and would be carried out in the same facilities.

The full vehicle tests would occur under controlled weather conditions (e.g. a C ETF, which would fall into the red circle) or in the real world where the conditions are measured but not controlled (blue circle in figure 4). It is expected that the real-world testing will take place in a safe and controlled (i.e.

not accessible to the general public) environment such as an outdoor test bed. Access control ensures that the test scenario is fully controlled (other than the weather conditions) and is risk-assessed to avoid general harm.

The pink circle denotes the virtual test environment. This environment incorporates: the sensor and weather scenarios to define its operations parameters; the results of all of the sensor tests for development of its sensor models; and the results of the tests under real and simulated weather tests to define its weather models and its models of the effects of weather.

Whilst the use of a pyramid in Figure 3 implies a hierarchy and an order of testing, it is important to note that information can flow both ways in this pyramid: for instance, if a vehicle test produces unexpected performance in the CETF, more single sensor tests may be performed to investigate and quantify this phenomenon.

An important consideration for the Golden Thread is that the testing and characterisation of individual sensors requires access to raw (or at least minimally processed) data. Some sensors include object identification algorithms as standard, but these need to be removed from the initial characterisation tests so that the tested sensor is essentially a device that sends and receives a signal. The main reasons for this are:
- Characterising the send/receive characteristics alone improves confidence in the sensor before the next stage of sensing and provides an easier route to standardisation of testing
- The amount of testing required of the full sensor (likely to be more expensive than characterisation of the send/receive aspect due to the variety of scenarios that require testing) is reduced. If it has been shown that an object identification algorithm can function well with a signal that is attenuated by a certain amount, then the root cause of that attenuation does not matter. The send/receive system can be characterised under conditions that may cause attenuation, to build confidence in the system under those conditions, but the more complex object identification system need not be tested
- Characterisation of the send/receive characteristics simplifies the construction of the sensor model for virtual testing because it enables the effects of weather to be moved into the virtual testing environment (see sensor model section)
- If a sensor that has only been tested as a black box fails an on-car test, it is not possible to know whether the failure is due to a software problem or a hardware problem, so fault root cause analysis becomes more complex and more expensive

6.2. Virtual Walkthrough
This virtual walkthrough shows how the linking of tests can increase confidence in the results. The following walk-through uses the results of a set of calculations based on assumptions that are reasonable for real world sensor systems. The details of the calculations are given in Appendix C.

We wish to assess a vehicle fitted with a 77 GHz radar and a 905 nm lidar. The lidar manufacturer's specification sheet says that it has a typical range accuracy of 2 cm and a measurement range of 120 m. The radar manufacturer's specification sheet says that it has a range of 250 m and an accuracy of 1.8 m for its far range, and a near range of 100 m with an accuracy of 0.4 m.

The first step is to test the base sensor performance by carrying out the standard tests for the key performance indicators [KPIs] as described in section 8.1. These tests will include a complete uncertainty analysis and could include use of targets of varying reflectivity to characterise the sensor performance more extensively. The base sensor tests show that the lidar has a maximum detectable range of 110 m with an associated uncertainty of 0.2 m, and the radar has a maximum detectable range of 250 m and an associated uncertainty of 1 m for a high reflective target and an uncertainty of 5 m for a low reflective target. It is not uncommon for tests to highlight aspects of sensor performance

that the specification sheets have not fully captured since they are not designed to contain that level of detail. Such tests can highlight the gaps in performance that may need to be covered by complementary sensors. The testing will also help to identify conditions under which the complementarity of the sensors means the system as a whole is still safe even when the performance of one sensor is degraded.

It is decided that this level of performance is adequate and that the next level of testing should be undertaken. The next level consists of testing the sensors, either singly or as a suite, under real and under simulated weather conditions.

The simulated weather environment has been set up to approximate a commonly assumed idealised raindrop size distribution for a rain rate of 100 mm hr^{-1}. This rate does not vary in time or space. The tests show that the water droplets attenuate the signal such that the maximum detectable range of the lidar drops to 70 m and that of the radar decreases to 140 m, so the range of both sensors has approximately halved.

The outdoor weather environment is uncontrolled, so the rain rate at a given point varies chaotically over time in a way that the Controlled Test Chamber cannot reproduce and does not exactly correspond to the idealised drop size distribution. The outdoor tests show that this variability causes a variation in performance of the sensor. During showery conditions, all with a headline rainfall rate of 100 mm hr^{-1}, the maximum detectable range for the lidar varies between 66 m and 78 m and that of the radar varies between 128 m and 158 m.

The above assumes for simplicity that the outdoor testbed experiences the rainfall rate of interest during the limited time available for testing. While the chances of this will be maximised by choosing an appropriate geographical location, it cannot be guaranteed. In practice therefore, it is the combination of selecting a range of CETF rainfall rates along with the use of well-characterised reference sensors in both the CETF and external testbed that will allow this comparison.

It should be noted that sensors, and particularly sensors for Self-Driving Vehicles, are increasingly likely to include advanced processing options such as object recognition and trajectory estimation. These tasks are typically carried out using advanced techniques such as machine learning. Testing of, reliability of, and evaluation of the uncertainty associated with the results of machine learning algorithms is an ongoing area of academic research. The topic is too complex to address within this project, so all sensors are assumed to return an electrical signal with minimal processing rather than information in a format interpretable by humans. A short discussion of object identification sensor testing is given at the end of this section.

These tests supply confidence in two aspects of the sensors. The first is the typical performance in repeatable rainfall scenario, which can be supplied by the Controlled Test Chamber tests. The second is the variability of that performance, and the associated increase in the uncertainty associated with the sensor response. **This information is vital for reliable virtual testing of the autonomous system, because virtual testing must reflect the likely variability of the sensor responses.** The testing will also help to identify conditions under which the complementarity of the sensors means the system as a whole is still safe even when the performance of one sensor is degraded.

The final tests involve testing the full vehicle in the CETF and ultimately in the real world. The CETF vehicle tests can identify whether any aspects of the wet weather adversely affect the sensor suite performance. The radar system on our vehicle is mounted behind a bumper, and the test has shown that that the radar performance, and in particular the maximum detectable range, decreases significantly due to formation of a water film on the bumper.

It is likely that this type of problem would only be picked up at a full vehicle test. The real world tests will pick up further examples of degradation, potentially due to the variability of weather conditions, to mechanical and vibration conditions that only occur under real driving conditions, and temperature-based effects whether due to solar heating or formation of ice in cold weather.

6.3. Benefits of the approach

The information from all of these tests including virtual simulation testing, collectively increases confidence in the safety of the system, and is used for validation and verification of the various models used in the virtual testing environment. The virtual testing environment, once validated and verified, can be used to test scenarios that are dangerous or near-impossible to reproduce in the real world.

The golden thread approach has different levels and complexities at its heart. The system being certified is the complete vehicle, but the required confidence in the complete system can only be built by testing individual sensors and sensor combinations to ensure that the system has been characterised correctly within the virtual testing environment (see section 6.7 on the relationship to functional testing). The majority of CAVs use multiple complementary systems specifically because different sensors are better at different sensing tasks, and so it is essential to test the entire system as well as the individual components.

6.4. Testbed definitions and purpose

The approach outlined in this section has been developed by considering the different aspects of testing that are relevant to CAVs. Testing has several "dimensions" that determine the nature of the test. These include:

- Real world vs virtual
- Single sensor vs sensor suite vs sensor suite in the entire vehicle
- No weather vs simulated weather vs real weather

Each test type has strengths and weaknesses and generates information that can be used for specific purposes. The table demonstrates some of the strengths and weaknesses of each environment, which when used in combination, will deliver assurance of the sensor performance in the full range of weather scenarios.

Core environments			
Environment	Description	Purpose	Strengths (S) / Weaknesses (W)
Calibration lab	Sensor tests in idealised "weather free" conditions	Baseline characterisation of the sensor performance with respect to reference targets	S: Low uncertainty characterisation of sensors W: Long way from real world conditions W: Real world environments are scaled in size and velocity in addition to difficulty in emulating weather conditions

Physical weather emulator (**CETF)**	This is a physically large scale controlled environment test facility (CETF). Weather conditions (e.g. rain, fog…) are created and are to some degree controllable. All key meteorological parameters are measured as if they were in the real world. Majority of the KPI tests can be performed except those KPIs which rely on vehicle motion	To span the range of meteorological conditions, albeit imperfectly in a quasi-repeatable manner	S: Quasi-repeatable and enables testing on demand across a wide dynamic range W: Is not 100% faithful to real world W: Likely limitation with respect to representing weather in motion with respect to the sensor
Real world weather test bed	Fully instrumented test range in the outside world. Includes measurement of all relevant meteorological variables	To capture real-world variability and sensor response	S: Provides physical realism to the ecosystems W: Dependent on the weather coming to site and therefore may need multiple sites (globally) W: Likely cannot test all sensors across whole dynamic range in these environments W: Ability to test impact of weather when sensor is in motion is uncertain
Virtual environment	Testing part or whole vehicle systems in a user controlled digital environment in a wide range of scenarios	To test the vehicle in situations that cannot be tested safely in the real world, and to replicate tests for validation purposes	S: Can simulate rare and hazardous events safely S: Can selectively test partial or complete systems (i.e. 'inject' virtual signal to any part of the modelled system) W: Requires proof that virtual world replicates reality

| Full vehicle on the road | Exposure of sensors mounted on the vehicle in the real-world environment, but it is likely that meteorological measurements will not be of operational quality (i.e. as used to determine the ODD) | Corroboration of the virtual models.

The only way of potentially discovering unmodelled aspects or new degradation pathways | S: Demonstrates whole vehicle integration

W: Expensive, and impossible to selectively sample full range of weather scenarios |

Desirable environments

Environment	Description	Purpose	Strengths (S) / weaknesses (W)
Mobile test bed	"Storm chasing" – a redeployable realisation of the real-world weather test bed, likely with a reduced set of meteorological sensors and reference targets	Increases the probability of exposing the sensor tests to high impact weather, especially where these have some degree of predictability of occurrence	S: Increased sampling of rarer high impact weather W: Deployability may come at the expense of calibration and measurement sensitivity (e.g. due to alignment of sensors to targets etc)
Observation fleet	Vehicles with enhanced meteorological measurement systems in addition to sensor suite for deployment for more ad hoc sampling. Reference targets not likely to be included, although some may be implemented as roadside infrastructure	To enhance sampling of sensor degradation in real world conditions	S: Experiences real-world conditions (along with well-characterised weather) S: Informs our understanding of measurability of ODD weather parameters with respect to the traditional weather observations networks W: Hard to make KPI measurements
Opportunistic	Vehicles with sensors but no bespoke weather measurement	Characterisation of sensor degradation over operational lifetime and identification of failure modes in broadly characterised weather types	S: Very large sampling population W: Little or no measurement of KPI performance W: Dependent on lower quality

			meteorological measurements W: Potential resistance to data sharing

Table 1: Description of strengths and weaknesses of each test/testbed

Further explanation of this assessment is given below:

Virtual tests are an efficient way of addressing the need for vehicles to demonstrate their safety during rare and dangerous situations. Unusual behaviour by pedestrians, cyclists and other vehicles can be simulated without needing to put any of these road users at risk. The main difficulty with virtual testing is verification. Demonstrating that the virtual world is sufficiently similar to the real word to be confident that the safety in the virtual world guarantees safety in the real world is a complex and challenging task. A related problem is ensuring that the virtual world includes all relevant safety-critical phenomena: in particular, the challenges (weather, infrastructure, driving culture, wildlife, etc.) involved in driving will vary hugely from country to country. The costs associated with virtual testing may mean that a different suite of tests is required to certify CAVs in different regions.

Confidence of safety in the real world cannot be obtained without real world tests. Virtual worlds inevitably involve simplification and assumptions and reflect the developers' best belief about reality but this best belief may be inaccurate. Only real-world testing can check these assumptions are valid. Further, many aspects of virtual testing depend on real world tests for input data. In particular, sensor models need to be based on real world tests in order to accurately reflect true performance and associated uncertainty.

Single sensor tests enable the key performance indicators described in Section 8 to be evaluated, and they provide the data to feed into sensor models. Sensor suite tests allow for exploration of the complementary nature of the sensors. For instance if the sensors are mounted in positions equivalent to their locations on a typical car, it would be possible to design a test to check for blind spots, redundancy, or conflicting detections (e.g. one sensor detects an object and the other doesn't even though single sensor tests suggest they should both detect it). Full vehicle tests allow for consideration of real-world effects that are difficult to mimic such as occlusion or reflection from the vehicle body, vibration of the chassis, and the effects of sensor coverings added for aesthetic purposes.

No-weather tests provide a baseline for the performance of a given sensor or system so that the effects of weather on the KPIs can be quantified. Simulated weather tests (i.e. tests that artificially recreate weather conditions in a controlled manner such as sprinklers for rain, aerosols for fog, fans for wind, etc.) create an understanding of how the sensors respond to weather with known and controlled characteristics, and thus enable validation of models of the effects of weather on sensors.

Furthermore, the results may suggest methods of improvement of the "weather effects" models, and they may feed information into data-driven models of weather-affected sensors. Real weather tests (i.e. tests in the open air under non-controlled conditions) provide further validation of "weather effects" models, in particular by providing some insight into how well the approximations made in the development of those models matches reality.

There are further questions that may affect the balance between the number of tests at each level. In particular, it should be noted that some aspects of the testing and characterisation should not be a once-in-a-lifetime occurrence. Sensors on vehicles require testing and calibration on a regular basis

because they are safety-critical systems[14]. In addition to ensuring ongoing safety, these procedures will generate useful information about sensor drift and permanent degradation (i.e. not weather-based) of sensors in use that may help to guide policy on sensor replacement.

Another related aspect is retesting in the event of component replacement. This aspect will not be discussed in detail here since it needs to be considered on a case by case basis, but it is possible to envisage a minimal subset of the full range of tests that could provide enough confidence that the car response is unaffected by a change of sensor and the vehicle is therefore still safe, in the same way as it may not be necessary to retest the crashworthiness of a standard vehicle if the only component altered in the new version is a wing mirror.

6.5. Traceability, Uncertainty & Risk

The purpose of traceability is to guarantee that measurements are accurate, by demonstrating that their outputs can be linked back to internationally agreed standards, via a traceability chain. For most sensor types this traceability chain is already well established for the sensor itself. At each step of the chain, it is necessary to show not only the traceability of the measurement, but also that the conditions that affect the measurement (such as temperature, vibration, and electromagnetic fields) are controlled and measured. For testing of sensor performance in rain, this effectively means that the rain must be controlled or measured, and ideally both. The entire purpose of the test is to characterise how the sensor reacts to rain of a known drop size distribution and intensity, so these parameters need to be measured, and ideally measured traceably.

Measurement of conditions becomes more challenging in "real weather" tests. Even though testing of stationary sensors in real weather eliminates the problems associated with defining and measuring the weather conditions a moving vehicle is experiencing, for the measured weather to be meaningful it needs to be measured sufficiently close to the sensor that the sensor can be said to be experiencing the same weather as is being measured. As has been mentioned above, the fractal nature of weather may make this requirement challenging. For a moving vehicle in real weather there are the challenges of spatial resolution and time synchronisation. Sensors may have integration times that are comparable to the time period over which rain intensity varies, so that the link between weather as measured and sensor response becomes more difficult to define.

Uncertainty is unavoidable in any situation involving measurement, and the random nature of weather makes this uncertainty even larger. In safety critical systems such as CAVs, it is vital to take the uncertainty into account when defining the criteria to judge a vehicle as "safe" under a given set of circumstances. Specifically, the probability that the vehicle will take a course of action leading to harm, whether to the occupants or to other road users, needs to be calculated in a way that takes the variability of weather and sensor response into account. This calculation is almost certain to be computationally expensive, but the risk to life involved in neglecting this calculation is sufficiently large that the accuracy should not be compromised without extremely good reason.

Uncertainty evaluation is key to reliable decision making, and a consideration of uncertainty is essential for safety critical applications. Broadly speaking, there are two sources of uncertainty that must be considered within this environment: uncertainty arising from measurement uncertainty, and uncertainty arising from model choices.

Uncertainty arising from measurement uncertainty is generally quantified and well understood, and can be propagated through the virtual testing environment using sampling methods such as Monte Carlo simulation that select values of the uncertain quantities in a statistically justified manner, run the testing scenario, and use the results to obtain statistical information about the scenario outcome.

---

[14] Abari et Al., Lyft.inc, 2019, Mobile Sensor Calibration Patent Application, US2019/0204425, https://www.freepatentsonline.com/20190204425.pdf

For the virtual testing environment, it is expected that the scenario outcome would be a classification of the manoeuvres made by the vehicle as either safe or not safe (perhaps also including near miss as a category), and so the statistical information could be as simple as a statement of the percentage of simulations that were placed in each category.

Uncertainty arising from model choices is more difficult to quantify. A typical approach is to validate the model against reality without considering the errors arising from model assumptions, and to consider the model choice uncertainty to be negligible if sufficiently good agreement is achieved. However, in a virtual testing environment, much of the uncertainty arises from simplifications made to reduce computational expense, and this type of error can be quantified.

Virtual testing for safety certification will require careful definition of the scenarios to be tested. As has been noted elsewhere, different scenarios are likely to be important in different geographical areas due to the variations in weather, infrastructure, driving culture, wildlife, and so on. It is also important to note that a completely different way of thinking about unsafe scenarios are necessary, because an AI has a much narrower field of experience than a human and has a completely different set of sensory information. The differences in sensory information can be taken into account by defining the key performance indicators for each sensor class, as has been done above, and considering the environmental conditions and other situations that would lead to a reduction in performance.

The narrowness of experience is considerably more difficult to limit. A human being can make deductions about dangerous situations from their knowledge of the world beyond the driving environment, whereas an AI only knows what it has been shown. As an example, consider the 2015 incident at the Shoreham air show, where an aircraft crashed onto an A road and caused multiple fatalities. A CAV would be unlikely to be monitoring airborne objects and would be unlikely to have experienced an airborne danger during its training, whereas a human driver would be more likely to identify that a potentially dangerous event was taking place. Similarly, a human driving in high winds would be more likely to be aware of the dangers of windblown objects than an AI, unless gale situations had been deliberately included in training.

These differences make identification of risk a more challenging proposition. It may be that the initial specification of driving conditions under which an AI is safe is conservative, and that the testing suite and the associated specification expand as it becomes more evident where the gaps in safe performance truly lie.

Certification of CAVs is expected to depend on the accumulation of multiple pieces of evidence that the vehicle performs safely in a set of pre-defined scenarios. The use of an evidence chain rather than a single pass/fail test makes it less likely that the vehicle will be designed specifically to pass the test rather than to drive safely in general. The scenario set for full autonomy is expected to be large, varied, and to vary geographically with common features.

Figure 5 gives examples of how the evidence chain might be delivered by the test ecosystem through the use of common measurements and references between the different environments. This is also discussed in Section 7.
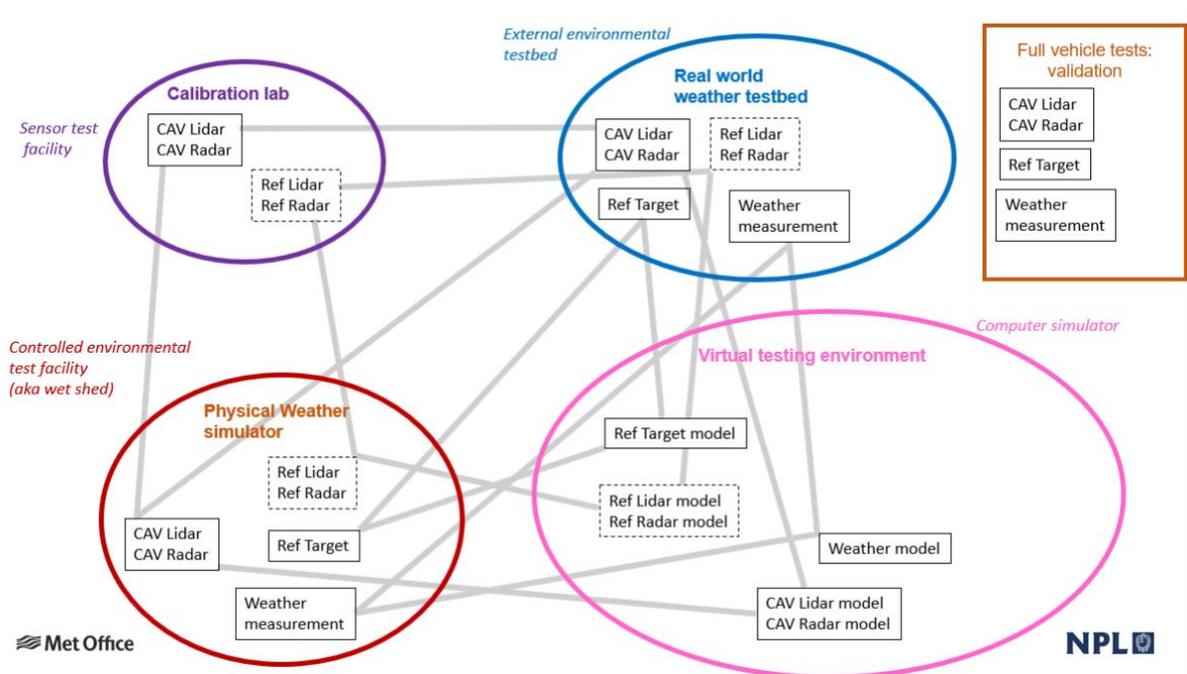
Figure 5: Indicative examples of how traceability can be strengthened in the test ecosystem through the use of common measurements and references in the different test environments

It is also possible that the definition of "safe" behaviour will be broad (e.g. avoidance of damage to the vehicle and other road users) rather than prescriptive (e.g. vehicle stops within 20 metres) as this will also make designing solely to pass the test less likely. "Safe" may also be different for different operational scenarios, for instance different vehicle configurations may require different definitions of safe, and cargo-carrying vehicles such as vans and trucks may require a different set of criteria from passenger vehicles.

6.6. Correlation

A common (and often incorrect) assumption when considering the implications of measurement and modelling uncertainty on our confidence in the outputs of a model is that all uncertainties are uncorrelated. In the case of CAV sensor response to weather, this cannot be assumed as it may result in over (or under) confidence in the abilities of a combined sensor set to cope with adverse weather.

As is mentioned before, weather is a random and fractal phenomenon. Most weather conditions will affect the performance of more than one type of sensor. When two sensors are affected by the same phenomenon their responses and associated uncertainties are correlated. In general, when two quantities are correlated due to their dependence on a third phenomenon, it is simpler to model the third phenomenon directly. This approach is the most likely option within a virtual testing environment.

Within a virtual testing system each sensor will have its own model. The virtual testing environment will return the weather-affected response of the environment appropriate to the sensor model, including any random effects directly. This approach generates the correlation: the same virtual weather affects the sensors, so their responses are linked, but the direct use of a model of the weather means that the link is produced without any need of a correlation matrix linking the sensor responses.

Correlation becomes more problematic if the weather is not modelled directly. If the sensor responses are treated as random variables, then they will need to be linked via a correlation matrix and potentially via a systematic term. This approach is unlikely to be used in practice within a virtual testing environment precisely because of these systematic terms: it is expected that the predominant effect

of rain (for instance) on a radar pulse will be increased scattering and a decrease in the intensity of the received signal, and the decreased signal intensity will cause a systematic change to the sensor response as a whole. The presence of the systematic effect makes it necessary to treat weather as a control variable.

An estimate of any correlations could be made from combining several sources of information. For rain, calculations of scattering and absorption for different signal types will show the correlations between the received signals directly, and testing of multiple sensors in the same weather conditions will allow the correlation between the sensor responses to be evaluated.

In addition to the correlation between the uncertainties associated with the sensor responses, there will also be a correlation between the sensor responses that is due to the approximations made within the weather and other environmental models within the testing environment.

As an example, consider the modelling of the effects of rainfall on lidar and radar sensors. The ideal model of the effects of rain on these sensors would calculate an exact value for the attenuation of a lidar or radar signal by a single raindrop using Maxwell's[15] equations to solve an electromagnetic scattering problem, and would combine the results of these individual scatterings via integration using a perfect description of the raindrop size distribution to define the probability of the laser or radar hitting each drop size.

In reality, the model of drop size distribution is not perfect, and so the calculated values of the attenuation for each frequency will not be perfect. Further, because the true raindrop distribution may have a systematic error (e.g. the distribution over-estimates the probability or a drop having a given radius) as well as random errors, there may be a significant systematic error in the values, and these errors in each frequency range will be correlated because they come from the effects of the rain. The likely size of the effect can be investigated using different distribution parameterisations within the integration model.

In general, it is not obvious whether the nature of the effects of these approximations is random or systematic. For the raindrop example, it is likely that the difference between true raindrop size distribution and the approximate size distribution has both types of components. The sensitivity of the results can be estimated to some extent, or at any rate bounded, by carrying out identical calculations for different drop size distributions and looking at the variation of the outputs. These calculations can include a consideration of wavelength dependence, and hence correlation between sensors. Again, these errors and correlations can be folded into the description of the weather-affected sensor response by treating the effects of the rain drop size distribution as a random quantity to be sampled during uncertainty propagation calculations.

6.7. Relationship with functional testing

The Framework offers an essentially Bayesian approach to assurance, i.e. where confidence in the assurance of the sensor (e.g. the maximum range KPI) increases with the addition of new information; in this case the different tests from the different parts of the ecosystem.

The approach is not at this stage focused on testing the functional capability of the system to **recognise** or **identify** a car. This can come later in the pyramid.

One of the challenges around post processing of raw data, especially with Machine Learning (ML), is that if you always use a blue jumper and a white car in your functional test, the ML will proceed to associate blue blobs with people and white blobs with cars and the smaller the blob the farther

---

[15] Maxwell, James Clerk (1865). "A dynamical theory of the electromagnetic field". *Philosophical Transactions of the Royal Society of London*. **155**: 459–512

away. The challenge here is that this tells you very little about whether the ML will recognise a grey jumper on a child facing away or a parked silver minibus reflecting the blue sky.

This can (and does) create real challenges in trying to distinguish the source of the problem: sensor, board processing, data fusion, ML control system etc.

The premise that the outlined approach takes is that quantitative assurance (with error bars) is built through a consistent use of reference targets. Traceability demands this. The outcome of this fully quantitative approach sets out to be that if a car such as a Ford Focus *is* put on a test range, the results are fully explainable from (i.e. predicted by) the rigorous approach. In this context the Ford Focus test is a necessary but not sufficient demonstration of performance.

When the information from each new test is added, the KPI value changes **as does the level of confidence**; when high quality information is added it has more influence on the KPI value and also tightens the error bars on the assurance.   A Ford Focus test could therefore be used in two ways;
1. Assurance: it can be added into the Bayesian mix, in which case it will influence the results, but relatively little;
2. Reassurance: we can check whether the results, with their error bars, are consistent with the results and error bars from those from the rest of the ecosystem

The aim of the targets that are used is to establish the performance envelope of the sensors in robust, reproducible, traceable fashion.
- In the first instance in compact chambers – this supports low(er) cost interventions in the development stage and troubleshooting, where there is a high degree of control which supports confidence; this is where we are essentially looking at the "silicon"
- In the second instance the tests are examining the performance envelope of the archetypal sensors in a degraded environment (i.e. in open air)

Together these will define the limiting cases (including maximum range) for **detecting** a target regardless of class (person, cyclist, bike, lamppost etc.).
- Having established this performance envelope in a traceable and high confidence fashion, it will be possible to produce trustworthy sensor models for use in synthetic testing environments
- At this point it is possible to test the functional performance of the control system (ML/AI etc..) to **recognise** and **identify** a target and take safe and appropriate action – building confidence in the technology
- Successfully completing synthetic testing, the whole vehicle will be permitted to engage in functional test e.g. Euro NCAP including against soft targets. It would probably be useful to produce a model of the soft target for transfer between synthetic and real world

In summary, the approach outlined here is built on assurance, while, where it's a practical demonstration (and not linked with other, different tests), the use of recognisable targets constitutes reassurance.

7. Weather: developing a reliable and usable description

This section summarises the key considerations when attempting to describe and measure the weather in a way that is suitable for inferring its impact on sensor performance. Using the example of signal attenuation through falling rain, it describes the sensitivity of sensor response to the small scale detail, the resultant implications for the testbed ecosystem and the need to consider the ability to measure the weather components of the ODD in an operational context.

Appendices F-H contain further supporting data.

7.1. The Golden Thread

We have adopted the term *Golden Thread* to describe the linkage between test environments that will, in combination, deliver the assurance of CAV sensors.

The interaction between the weather and the performance of any given sensor can be highly complex and the magnitude of the impact may be highly dependent on the detail of the weather at very small scales. Furthermore, each weather element (e.g. fog, rain, snow etc) causes the different sensors to respond very differently due to the strong dependence on the wavelength of the sensor in question. For this study, a single aspect of the weather – falling rainfall (as opposed to spray) – and its impact on the maximum range KPI for lidar and mm-wave radar were selected to demonstrate the key elements of the golden thread. This choice was made because rainfall carries many of the challenges around very high levels of variability over the time and spaces scales that are relevant to CAVs, it has a well-understood climatology[16] and its impact on these wavelengths can be quantitatively modelled sufficiently well using established theory.

7.2. Linking Rainfall to the maximum range KPI - overview

The maximum range KPI provides an easily understood example to illustrate the impact of rainfall on a sensor. Intuitively, in rainfall, the path of a beam of lidar or radar radiation will be obscured by the intervening rain drops. The energy may be absorbed by the raindrop or scattered out of the beam (It may also be scattered back into the beam and continue its journey, although possibly with a small delay, which might also have an impact on the signal processing.). The testing regime must capture these effects, but it must also be agnostic to the details of any given sensor i.e. the job is to test the impact of rain on the KPI. This is critical as the job of the test regime is to characterise the overall performance of the sensors, not stand in the way of the innovation that allows one sensor to be more competitive than another.

Noting the point above, it is not feasible to directly model the response of every sensor to rainfall explicitly (as indicated in the choice of approach 'C' in Section 8.5). However, it *is* possible to identify a variable that relates rainfall intensity to the *degree* of impact on the sensor, without needing to fully model every aspect of the response. The variable chosen is the 2-way attenuation, which is a measure of the total power of the lidar or radar beams that is intercepted on its journey from and back to the sensors during the interaction with the raindrops. It is a physical property of the intervening atmosphere and is sensor-independent apart from the choice of wavelength i.e. it is the same for a good or bad sensor; a more performant sensor can simply cope with more attenuation. As such it provides a good guide to the test set-up required to determine the maximum range KPI value.

Note: For completeness, it is stressed here that over short distances and short pulses involved in the CAV context, the impact on a sensor may manifest itself as signal loss, noise and/or pulse broadening. Appendix F also illustrates that, especially at lidar wavelengths in rainfall, scattering by

---

[16] Dixon, J (2019) *An investigation into short-period, extreme rainfall in the UK,* Report for Innovate UK, Met Office, pp79. Available at: https://www.metoffice.gov.uk/services/transport/cav

raindrops may be concentrated in a forward peak that is smaller than the angular resolution of the sensor. In summary, using the attenuation as we have done in the walkthrough, as a simple intensity loss, is a simplification.

7.3. Implications of sensitivity to the details of rainfall

Ultimately the questions that we are trying to address are "If we wish to determine the maximum range of the detector for a rainfall rate of R, as measured or forecast for the location of the CAV on the road,

> (a) what range of rainfall rates do we need to test in the ecosystems to achieve a certain confidence?
> (b) and what is the precise definition of the rainfall rate as observed in the real world (i.e. over what time and space scales is it an average)?"

Appendix F builds up the components of the challenge from a single raindrop upwards to the monitoring of the ODD in operations. The key messages are:

- The degradation of the signal to/from a CAV sensor is due to the cumulative effect of the interaction of the electromagnetic radiation passing (from and) to the sensor with each individual raindrop. The magnitude of this interaction is strongly and non-linearly related to the relative size of the drop and the wavelength
- The degradation can therefore be very sensitive to the detail of how the rainfall is distributed across the rain drop sizes, the *drop size distribution (DSD)*. This may have particularly significant implications when considering the *relative* performance of two or more different sensors working on the same CAV, e.g. assumptions of sensor redundancy in edge case weather, because of the correlation between the channels
- In the test environments, it is not sufficient to merely characterise the rainfall in terms of its rain rate in mm/hr. The DSD must also be measured (as described in the traceability section)
- In nature, the DSD is very strongly influenced by the type of rainfall e.g. frontal/stratiform, orographic or convective. Therefore, sensor assurance expressed only in terms of rain rate, will necessarily result in larger uncertainties regarding real world performance. It will also introduce correlated uncertainties in the assurance of sensors at other wavebands *(The study therefore ensured that BSI PAS 1883 includes rainfall type in the ODD description.).* This is demonstrated in Fig 6 below, which is also given in Appendix F
- It <u>cannot</u> be assumed from the outset that a CETF will be capable of producing the full range of naturally occurring DSDs with sufficient fidelity that it can deemed equivalent to testing in the real world
- It is therefore essential to have an external testbed within the ecosystem, the purpose of which will be either or both of:
    a. verifying the realism of the CETF DSD and corresponding (multi)sensor response; and
    b. *complementing* the test data from a CETF by combining the information from both in a manner that reflects the limitations of both
- The two environments will contribute most positively to confidence if:
    a. the "manufactured" rainfall is broadly able to control the DSD in the most "active" areas of the DSD (i.e. those drop diameters that contribute most to signal degradation) simultaneously at all wavelengths of interest
    b. the <u>shape</u> of the raindrops with size is realistic in order to capture differences in polarisation response – in practice this means that a minimum fall distance (chamber height) is required in the CETF
    c. the measurement of the key meteorological elements (especially DSD) in the CETF and the external testbed should be as close to identical as possible. Appendix G, provides an indicative list, which will also extend to the characterisation of fog
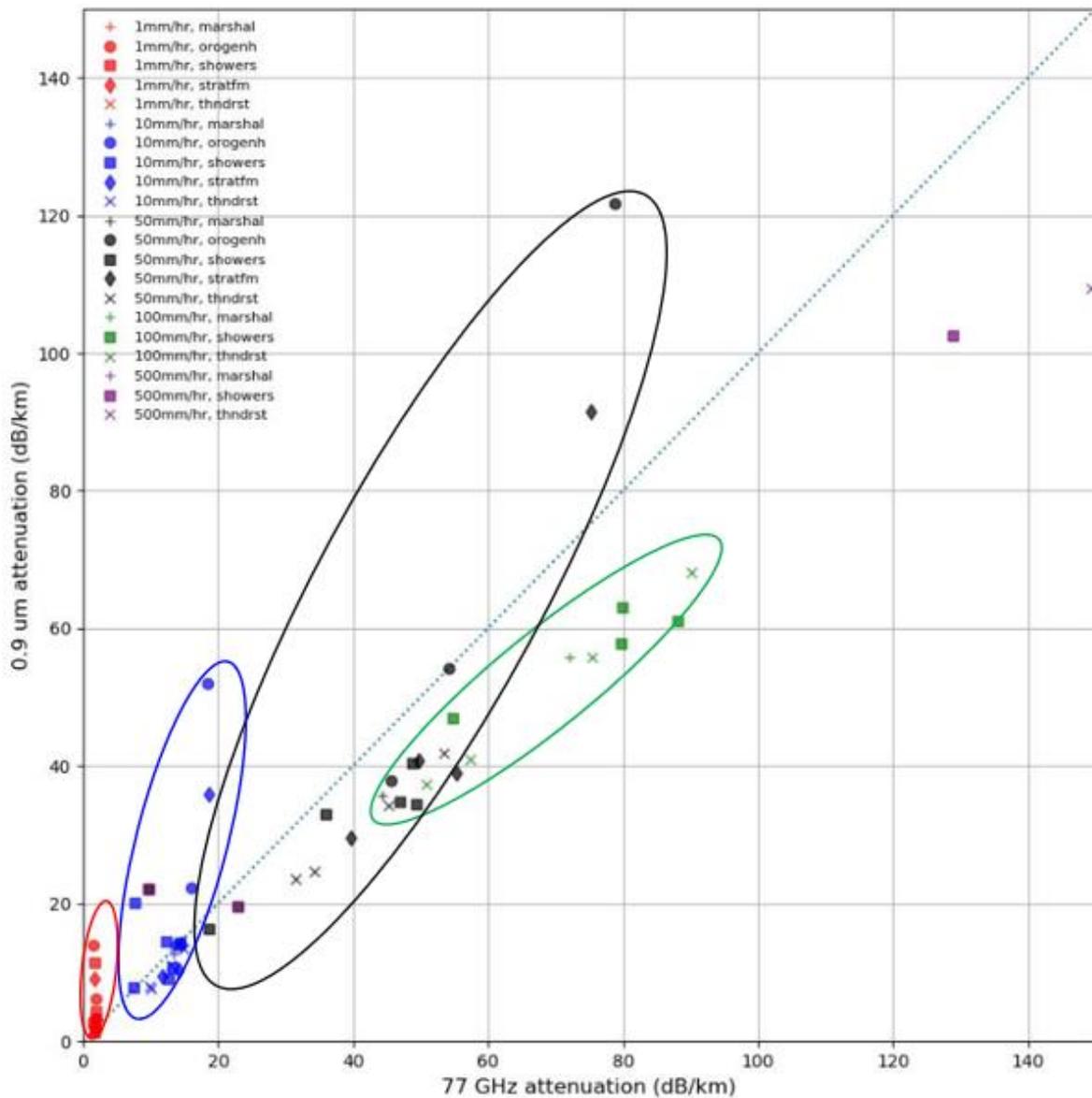
Figure 6: Scatterplot of 2-way attenuation for lidar (y-axis) and radar (x-axis). The plot shows how attenuation varies for 4 different rainfall rates (1, 10, 50, 100 mm/hr), with the scatter at each rain rate due to the assumption of DSD. The shapes refer to the type of rainfall, which will have a strong influence on the distribution of rain across smaller and larger droplet sizes. The ellipses highlight the range of values for a given rain rate and approximate to the error correlations between channels should a single value of attenuation for a given rain rate be chosen.

7.4. Linking Sensor Assurance to the ODD

The above section addresses the question "For a given steady state (unvarying in time and space) rainfall rate, what are the key drivers for the maximum range KPI?" It is already clear that a single rain rate can result in a range of attenuation values simply by distributing water across the different rain drop sizes.

If it is assumed that we are willing to accept, for now, the use of a headline rain rate, and set aside the variability due to DSD, it is still necessary to answer the question "How does this rainfall rate relate to what we actually observe in the real world?" [17] This is essential because we *must* link the performance

---

[17] In reality both the DSD shape and the rain rate may vary, however, we assume for demonstration purposes that the DSD "shape parameters" are constant

characterisation in the test ecosystem to a **practically measurable** quantity within the ODD. It is non-trivial because of the following factors:

- the rapid variations over time and space that rainfall exhibits
- the relatively small time and space scales that are relevant to CAVs
- the fact that traditional rainfall measurements were not designed with this end usage in mind

Appendix F further develops these concepts. Rain rates as used in conventional meteorological applications are typically measured by rain gauges and (C-band) radar networks that are sampling the volume of rain over *very* different time and space scales (~10's cm^2 over a minute versus ~1 km^2 quasi-instantaneously, respectively). CAV sensors effectively represent a third sampling regime (a narrow beam over 10's of metres quasi-instantaneously). Put simply, CAV sensors experience rainfall very differently to meteorological networks (and humans) and to ignore this is dangerous.

The following key points emerge:

- The spatial and temporal variability of rainfall rate is fractal in nature - it demonstrates similar levels of detail at smaller and smaller and space scales. The primary impact of this is that a CAV sensor may experience a higher rainfall than the average value provided by a meteorological observation (or meteorological data service), Figure 7
- By way of an example, a CAV sensor that operates nominally in 50 mm/hr rainfall in a CETF may fail in a real rain event where a meteorological radar estimates the rainfall to be the same value. This is explored fully in Appendix F
- It is not a given that the rainfall developed in a CETF will be able to demonstrate fully realistic time and space variability, especially in highly inhomogeneous conditions such as strong convective events. It may however be desirable, if possible, to develop such functionality to simulate the rapid *onset* of heavy rain
- This limitation further supports the need for external testbeds that have a high degree of commonality with the meteorological measurements in the CETF. In this regard, as before, the role of the external testbed will be either or both of:
  a. verifying the realism of the CETF rainfall variability and the corresponding (multi)sensor response; and
  b. *complementing* the test data from a CETF by combining the information from both in a manner that reflects the limitations of both

- It is essential to link the measurements of rain rate made in the external testbed and/or CETF to those that will be used to assess if the CAV is within its ODD when it is operational. It would be premature to attempt to define a rainfall "standard" for such a purpose at this point although it is possible to provide useful guidance on how this variability might inform safety margins. An example approach is given in Appendix F. *(It is for this reason that BSI PAS 1883 suggests that the rainfall value in the ODD taxonomy for CAV includes a description of the spatial and temporal scale of the rainfall value but does not mandate what that is.)*
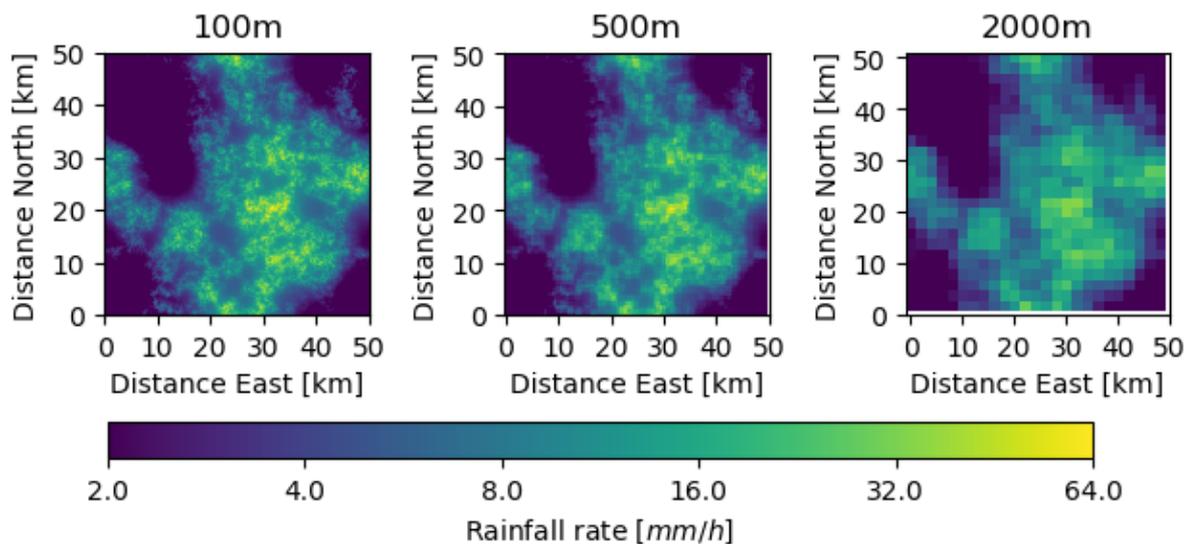
Figure 7: The effect of spatial averaging of rainfall on maximum expected rainfall rate. Instantaneous maximum intensity decrease as the pixel size increases from 100m through 200 m to 2 km. (These images use simulated data and are explained further in Appendix F)

7.5. How much detail is useful?

The implications of the fractal nature of rainfall on different observing systems is a science specialism in its own right. The ultimate application here is to be able to relate sensor performance to an **affordably measurable ODD** using a **cost and time efficient assurance** approach. The primary considerations are:

- It is not possible to capture all naturally occurring rainfall size distributions in the assurance process for every sensor under test
- Quantitative knowledge of the operating environment of the CAV (for comparison with the ODD) will carry significant uncertainty – it is not possible to know the weather everywhere. Knowledge of this may become the dominant error term in the assessment of exceedance of the ODD. Also, secondary impact pathways (e.g. spray, water films) may be significant
- A balance must therefore be struck in terms of the complexity of the test environments, however…
- It remains critical that the uncertainties and their correlations between sensors are characterised thoroughly

8. Characterising Sensor Performance and KPIs

This section describes the methodologies, key performance indicators, impact mechanisms and calibration approaches which collectively allow the reliable characterisation of sensor performance. A description of how this can enable the generation of reliable sensor models for virtual testing is also provided.

Underpinning the development of the Framework would be a standardised methodology for the characterisation of sensors, such as the one outlined below, which requires a common approach for testing, calibration, definitions, facilities and data quality.

Sensor characterisation - methodologies

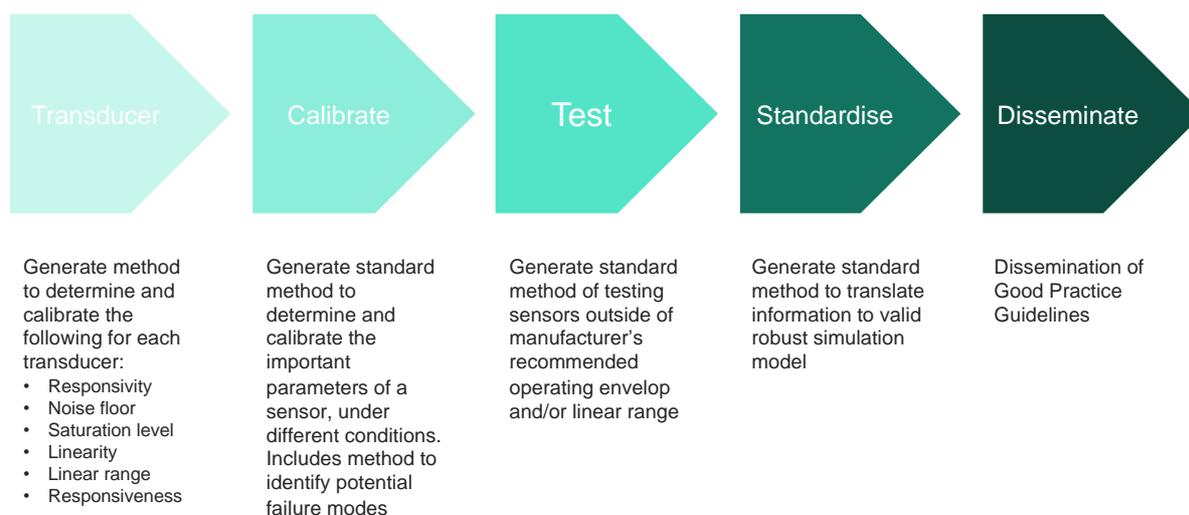| Transducer	Calibrate	Test	Standardise	Disseminate
Generate method to determine and calibrate the following for each transducer: • Responsivity • Noise floor • Saturation level • Linearity • Linear range • Responsiveness	Generate standard method to determine and calibrate the important parameters of a sensor, under different conditions. Includes method to identify potential failure modes	Generate standard method of testing sensors outside of manufacturer's recommended operating envelop and/or linear range	Generate standard method to translate information to valid robust simulation model	Dissemination of Good Practice Guidelines

Figure 8: 5 stage methodology for reliable Sensor characterisation

In this 5-stage process, the calibration stage requires an accurate and traceable (i.e. back to ground truth) definition and technical specification of these different conditions, most of the important ones being weather related. The propagation of uncertainties in sensor performance through to the performance of perception algorithms and autonomous decision making must be understood and then reflected in the setting of pragmatic industry standards.

8.1. Sensor characteristics

Sensor capabilities can be defined and quantified using a set of key performance indicators (KPIs). These indicators describe the ability of the sensors to detect objects, distinguish distinct objects, and the features of the sensor that may lead to incorrect decisions such as noise or poor contrast. The change in these KPIs under differing weather conditions will provide vital information for verification of the system as a whole. The use of KPIs enables sensors of different modalities to be compared because they relate to the capability of the sensor to detect objects rather than relating to the particular technical aspects, such as signal scattering or absorption, that affect sensor performance.

A possible set of key performance indicators appropriate for sensors on CAVs are:
- Angular/spatial resolution: capability to discriminate between two adjacent targets
- Range resolution: capability to discriminate between two targets at different distances
- Speed resolution: capability to discriminate between two targets at different speeds
- Maximum detectable range: the maximum range at which a target can be detected
- Minimum detectable range: the minimum range at which a target can be detected
- Maximum unambiguous speed: the maximum speed that can be measured unambiguously

- Minimum detectable speed: the minimum speed that can be measured
- Update rate/responsiveness: the speed with which the systems provides new outputs
- Tracking capability: the capability of the system to target/targets in the scene
- Tracking capacity: the number of targets that the system can track at the same time
- Contrast: the variance in luminance between the target and the scene background
- Sidelobe levels: Ratio of main lobe and first sidelobe peaks in radar's ambiguity function
- Integrated Sidelobe Ratio: Ratio of main lobe power and the sum of the power of all the sidelobes in radar's ambiguity function
- Antenna patterns: the radiation pattern of the systems sensor antennae
- False Alarm Probability: the probability of detecting a target when it is not present
- Detection probability: the probability of detecting a target when it is present
- Dynamic range: ratio of maximum to minimum value of received signal power that the system can measure
- Noise floor: signal created from the sum of all the noise sources and unwanted signals measured by the sensor
- Linear range: the range in which the output of systems applies a linear function to the input
- Linearity: expression of deviation of the measured response curve departs from the ideal straight line
- Field of view: the angle through which the system can detect electromagnetic radiation

Sensor Characteristic	Camera	Lidar	Radar	Ultrasound
Angular/spatial resolution	✓	✓	✓	✓
Range resolution	✓	✓	✓	✓
Speed resolution	✓	✓	✓	✓
Maximum detectable range	✓	✓	✓	✓
Minimum detectable range	✓	✓	✓	✓
Maximum unambiguous speed	✓	✓	✓	✓
Minimum detectable speed	✓	✓	✓	✓
Update rate/responsiveness	✓	✓	✓	✓
Field of view	✓	✓	✓	✓
Contrast	✓	✓		
Dynamic range	✓	✓	✓	
Noise floor	✓	✓	✓	✓
Linear range	✓	✓	✓	✓
Linearity	✓	✓	✓	✓
Antenna patterns	✗	✗	✓	✗
Sidelobe levels	✗	✗	✓	✗
Integrated Sidelobe Ratio	✗	✗	✓	✗
False Alarm Probability			✓	
Detection probability			✓	
Tracking capability	✓	✓	✓	
Tracking capacity	✓	✓	✓	

Table 2: KPIs relevant to the broad classes of CAV sensor

8.2. Degradation table: Weather Impact Taxonomy

All CAV sensor and communication systems experience performance reductions when exposed to weather. Quantifying the level and range of degradation is key to building sensor redundancy and situational awareness confidence. Table 3 presents a method for describing sensor type and

degradation mechanisms to produce a **taxonomy** of weather impact pathways. It is currently in draft form and will be updated and completed in Stage 2 to include the full quantified range of degradation levels and methods for detection. This is required to enable priority mechanisms to be explored early. Through investigations conducted by Met Office and NPL alongside the industry engagement process, it is clear that each weather element provides a range of pathways to sensor degradation, often in combination with other weather elements and sometimes as secondary mechanisms e.g. spray, films of water on sensor surfaces etc. Furthermore, there are effects of weather on infrastructure, such as water films forming on road signs, that may affect the vehicle's ability to detect or identify common objects.

The table 3 on the following page is currently only indicative of the type of information to be included in a final version. When the Taxonomy is completed, the table will be updated to describe the mechanism and KPIs against each affected sensor, accompanied by greater detail.

Degradation Item	Sensor(s) affected	Mechanism (examples)	Some Impacted KPI (examples)	Detectability (indicative)
Sun (direct sunlight)	Cameras (visible, IR, EO) Lidar Radar Ultrasound	Over exposure. Exceedance of camera dynamic range. Heating of sensors and sensor housing. UV degradation of sensor windows.	Reduced contrast and dynamic range, usable field of view, maximum detectable range, linearity, responsiveness, tracking capability and detection probability. Increased noise floor.	The use of LUX meters and signal levels detected installed camera systems (auto-contrast level for example).
Sunlight and clouds	Cameras (visible, IR, EO) Lidar	Rapid changes in brightness	Reduced contrast and dynamic range, maximum detectable range, linearity, responsiveness, tracking capability and detection probability.	As above
Rainfall	Camera (visible, IR, EO) Lidar Radar Ultrasound GPS/GNSS Comms	Absorption Refraction Scattering Reflection Coating on sensors and targets with water and dirt.	Reduced angular and range resolution, maximum detectable range, tracking capability and detection probability. Increased sidelobes.	Installed radar & lidar can provide feedback on rain density w.r.t. angular distribution.
Fog and suspended water	Camera (visible, IR, EO) Lidar Radar Ultrasound	Absorption Scattering Coating of sensors and targets resulting in absorption, refraction, scattering and reflection.	Reduced angular and range resolution, maximum detectable range, reduced tracking capability and detection probability.	As above
Water spray	Camera (visible, IR, EO) Lidar Radar Ultrasound Especially if mounted low on vehicle	Absorption Scattering Coating of sensors placed low on vehicle body, and targets causing absorption, refraction, scattering, reflection and occlusion.	Reduced angular and range resolution, maximum detectable range, reduced tracking capability and detection probability. Radar clutter.	As above
Surface water (e.g. standing water, film, damp)	Camera (visible, IR, EO) Lidar Radar	Reflection and glare Temperature Scatter Production of spray	Reduced contrast and dynamic range, usable field of view, maximum detectable range, tracking capability and detection probability	Depending on cross correlation of sensors. Polarised lenses.
Snowfall	Camera (visible, IR, EO) Lidar Radar Ultrasound GPS/GNSS Comms	Absorption & scattering Over exposure Coating of sensors and targets resulting in absorption, refraction, scattering, reflection and occlusion.	Reduced contrast and dynamic range, usable field of view, maximum detectable range, tracking capability and detection probability. Increased sidelobe level.	Medium detect Quite detectable

Weather	Sensors	Degradation mechanism	Effect	Mitigation
Surface snow (slush and dry)	Camera (visible, IR, EO) Lidar Radar	Reflection and glare Temperature Scatter	Reduced contrast and dynamic range, usable field of view, maximum detectable range, tracking capability and detection probability	
Frost and icing	Camera Radar Lidar Ultrasound GPS/GNSS Comms	Absorption Distortion Occlusion Reflection and glare Refraction Coating on sensors and targets	Reduced field of view, tracking and detection probability.	Vehicle temperature sensors can provide probability of frost/ice.
Hail	Camera (visible, IR, EO) Lidar Radar Ultrasound	Absorption Scattering Occlusion Physical damage to sensors and sensor window	Reduced angular and range resolution, maximum detectable range, Increased radar clutter.	Radar/Lidar data
Lightning strikes	Camera Radar Lidar Ultrasound GPS/GNSS Comms	Electromagnetic discharge Physical damage to sensor or sensor housing	Sensor failure.	Lux, camera intensity and audio detection.
Lightning flashes	Camera (visible, IR, EO) Lidar	Over exposure Rapid change in brightness	Reduced contrast and dynamic range, maximum detectable range, linearity, responsiveness, tracking and detection probability.	As above
Wind (mostly in combination with e.g. rain and dust)	Camera (visible, IR, EO) Radar Ultrasound	Absorption Scattering Increase in airborne objects	Reduced tracking and detection probability. Increased false alarm probability.	x
Temperature	Cameras (visible, IR, EO) Lidar radar Ultrasound	Heating of sensor and sensor housing. Deformation of sensor housing and window. Heating of surface air (mirages)	Increased noise floor and false alarm probability. Reduced linearity. Changes in angular and range resolution.	x
Dust (airborne)	Camera (visible, IR, EO) Lidar Radar	Absorption Scattering Coating on sensor and targets	Reduced angular and range resolution, maximum detectable range, tracking and detection probability, field of view.	Radar/Lidar data
Humidity	Camera (visible, IR, EO) Lidar Radar Ultrasound	Absorption Temporary coating of sensors with condensation	Reduced angular and range resolution, maximum detectable range, reduced tracking capability and detection probability.	x

Table 3: A sample of weather-dependent degradation pathways for CAV sensors

The complexity of weather-related degradation for a range of sensors utilised with a CAV are presented in Figures 9.1-9.4.   This indicates how, for example, the daylight camera can on a bright clear day experience a range of degradation from zero to 100% depending if the sun is in front or behind the camera. Once cloud cover increases the flat lighting reduces the range of degradation for

all scenarios and performance might be considered optimal. Radar and Lidar have little dependence on solar energy and therefore have an optimum performance when the weather is free from precipitation. In these charts, the levels of degradation have been estimated, but in Stage 2 further quantification will be enabled across the weather patterns and thus sensor priority with a CAV can be better defined. The chart showing combinations of chart traces highlights optimum operational windows for all weather scenario and when a 'CAV blind' event is most probable.
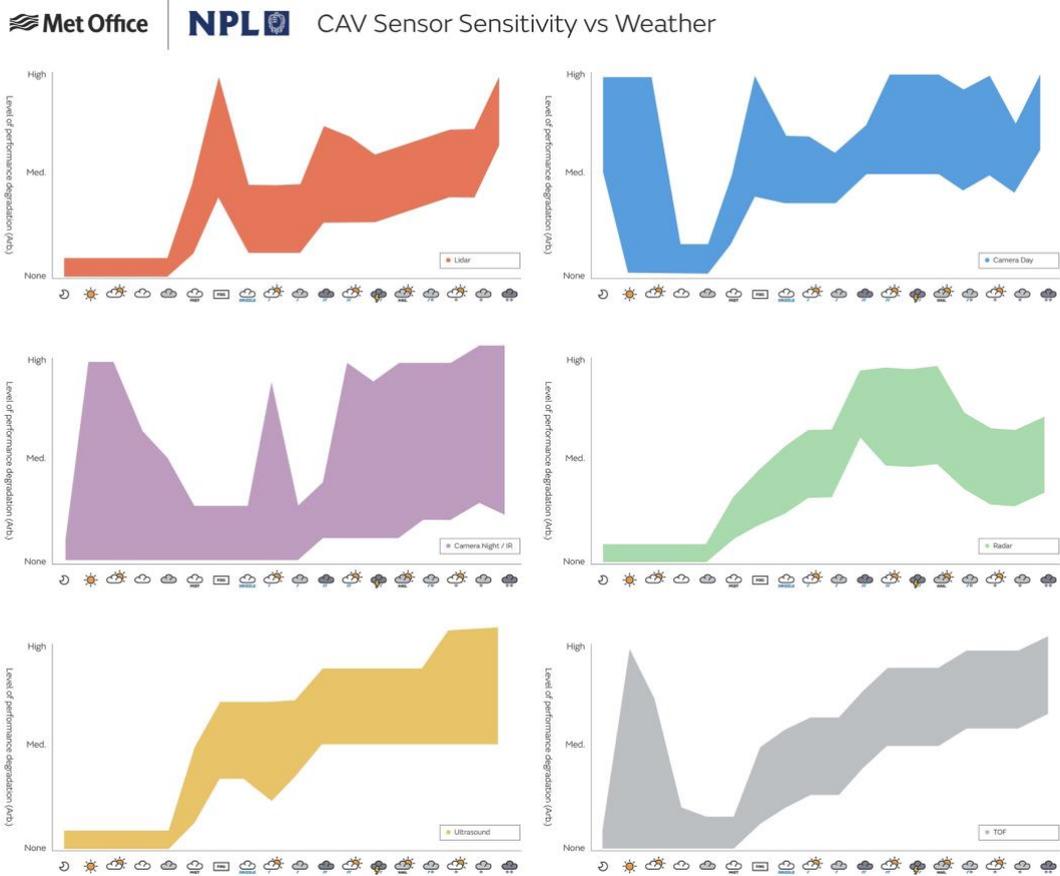
Figure 9.1: Simplified view of weather vs performance degradation for six key sensor types
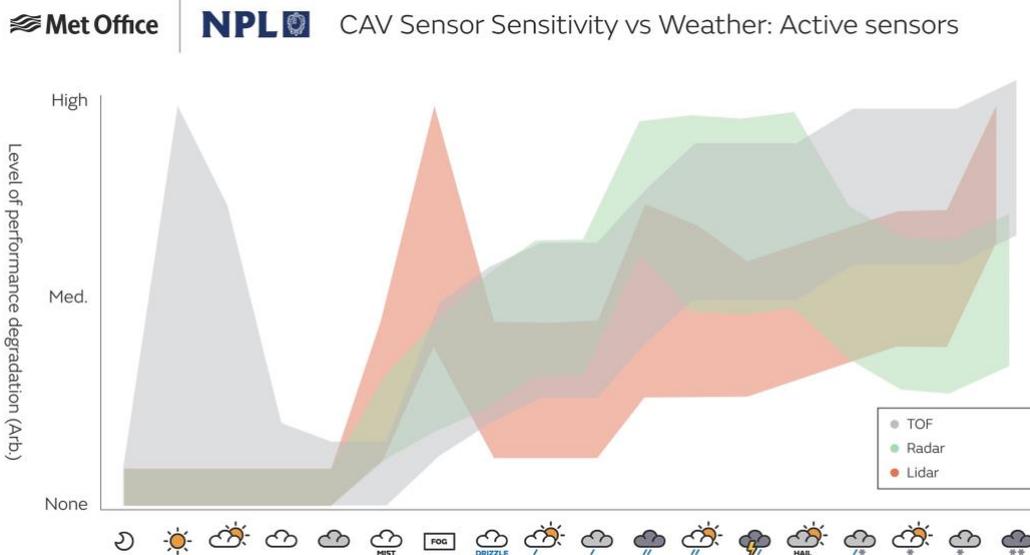
Figure 9.2: Simplified view of weather vs performance degradation for Active sensor types combined
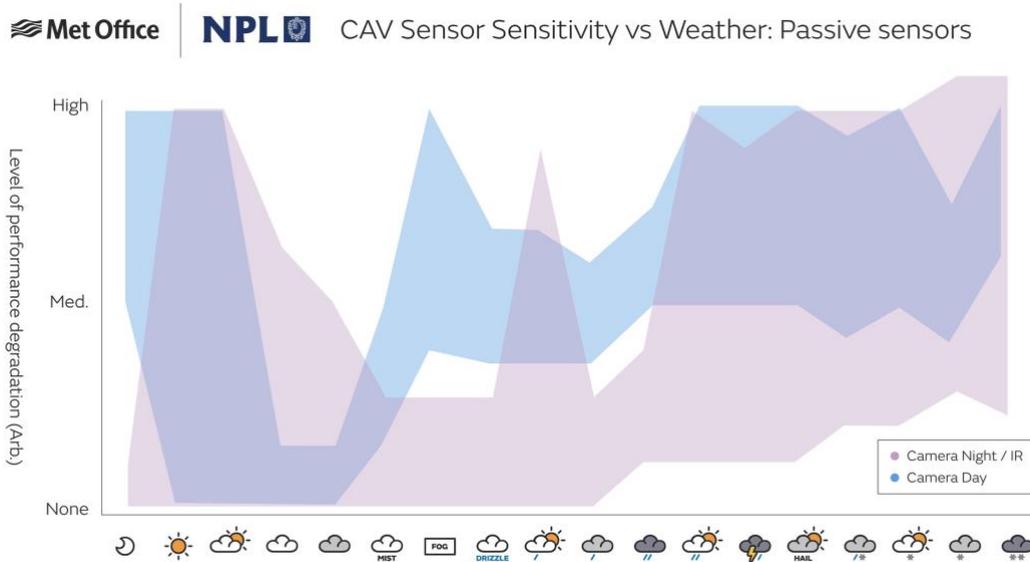
Figure 9.3: Simplified view of weather vs performance degradation for Passive sensor types combined
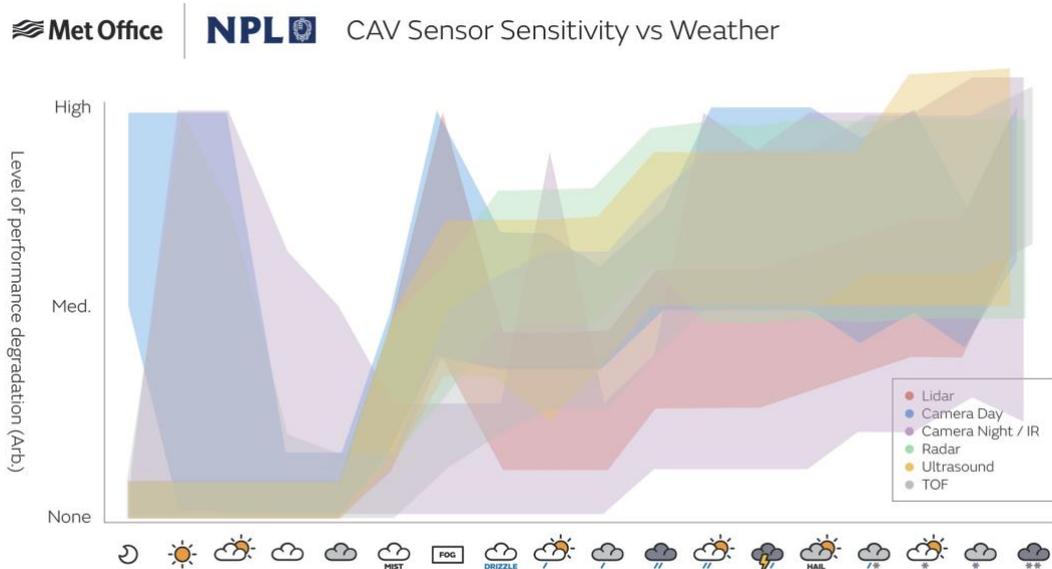

Figure 9.4: Simplified view of weather vs performance degradation for six key sensor types combined

Note the images here are indicative only.

8.3. Sensor calibration

Robust environment perception is one of the essential tasks which an CAV must accomplish [1]. To achieve this goal, various sensors such as cameras [2], radars, LiDAR's, and inertial navigation units are used, and information thereof is often fused. Essential tasks such as simultaneous localisation and mapping (SLAM), detection and tracking of moving objects, odometry, etc. are often improved by sensor fusion.

A fundamental step in the fusion process is sensor calibration, both intrinsic and extrinsic. The former provides internal parameters of each sensor (e.g. focal length of a camera, bias in LiDAR range measurements), while the latter provides relative transformation from one sensor coordinate frame to the other. The calibration can tackle both parameter groups at the same time or assume that sensors are already intrinsically calibrated and proceed with the extrinsic calibration. Additionally, temporal synchronisation of the sensors is sometimes performed within the calibration.

Intrinsic parameters are related to the working principle of the sensor. Therefore, methods for finding intrinsic parameters do not share many similarities between different types of sensors. On the other hand, parametrisation of extrinsic calibration, i.e. homogeneous transformation, can always be expressed in the same manner, regardless of the sensors involved. Despite that, solving the extrinsic calibration requires finding correspondences in the data acquired by the sensors which can be challenging since different types of sensors measure different physical quantities. After correspondence registration, optimisation steps are performed to estimate the calibration parameters. While some methods require intrinsically calibrated sensors to find the extrinsic calibration, others perform optimisation on both parameter groups simultaneously. These methods typically try to satisfy some geometric constraints through minimisation of a problem-specific reprojection error. The geometric constraints involve nonlinearities which often cannot be solved analytically. To resolve that problem, estimators use iterative techniques to find the appropriate solution. Due to the nonconvexity of the problem caused by the nonlinearities, these methods have a risk of converging to a local minimum. To avoid that risk, some methods divide optimisation in initial rough estimates that guarantee near-optimal solutions followed by nonlinear iterative refinement steps. The success of the optimisation is highly dependent on the provided data. An important step before the data acquisition is to determine minimal requirements on the dataset for which the problem becomes identifiable (or observable in case of dynamical systems).

The calibration approaches can be target-based or targetless. In the case of target-based calibration, correspondences originate from a specially designed target, while targetless methods utilise environmental features perceived by both sensors. The former has the advantage of the freedom of design which maximises the chance of both sensors perceiving the calibration target but requires the development of such a target and execution of an appropriate offline calibration procedure. The latter has the advantage of using the environment itself as the calibration target and can operate online by registering structural correspondences in the environment but requires both sensors to be able to extract the same environmental features. Registration of structural correspondences can be avoided by motion-based methods, which use the system's motion estimated by the individual sensors to calibrate them. These methods have two main advantages: (i) they rely less on the sensors' operating principles and can be applied to different sensors, if a sensor can estimate its motion, (ii) unlike other methods, they can extrinsically calibrate sensors with nonoverlapping fields of view.

In [Appendix D](#) of this report, the different calibration techniques for the most popular sensor technologies used in the automotive industry such as camera system, LiDAR and radar are discussed.

8.4. Sensor Test conclusions

Diversity of sensors is critical for making CAVs vehicles safe and ubiquitous. Being able to self-calibrate the pose of the imaging system with respect to the road or ground ahead of the vehicle always provides us with critical information, since the degree of accuracy of the measurements of the distance to the environment elements has a huge impact on the decision-making process of ADAS or CAVs. A lack of accuracy in measuring the distances to the environment elements can have serious consequences for people and vehicles in traffic scenarios. While it's still too early in the market's development to predict exactly what the future holds for self-driving cars, the rapid advancement and promise of the various 3D imaging technologies suggest that it will continue to gain ground and move toward mainstream adoption.

8.5. Sensor Models

Sensors, and particularly sensors for ADAS, are increasingly likely to include advanced processing options such as object recognition and trajectory estimation. These tasks are typically carried out using advanced techniques such as machine learning. Testing of, reliability of, and evaluation of the uncertainty associated with the results of machine learning algorithms is an ongoing area of academic research. The topic is too complex to address within this project, so all sensors are assumed to return an electrical signal with minimal processing rather than information in a format interpretable by humans. A short discussion of object identification sensor testing is given at the end of this section.

The following discussion of sensor modelling assumes that the sensor models are to be used in a virtual testing environment. The goal of the sensor model is to accurately reproduce the performance of the sensor so that the same information is supplied to the main AI by the sensor model as would be supplied to the main AI by the real sensor. The virtual environment interacts with the sensor as illustrated in Figure 10. The sensor sends a query to the environment, defining what type of sensor it is and any parameters (such as wavelength, field of view etc) associated with the sensor. The environment calculates the response, including effects of weather and any other environmental conditions, and sends the sensor model a definition of the received signal. The sensor model then processes the received signal to provide a sensor response to the main AI.

The nature of these interactions, and of the virtual testing environment itself, will not be discussed in detail. The environment will need to create simulated versions of a received signal, as is done in the "vehicle in the loop" version of testing. This approach still has many open challenges. For example, existing computational resources and approaches are not sufficient to produce a photo-realistic synthetic world: humans can generally tell whether they are watching a film of reality or a simulation. This lack may not be a fatal problem, however: the differences between an AI and a human may mean that photo-realism as judged by a human goes beyond the level that is necessary to judge whether an AI would act safely in a given situation.

Note that uncertainties are not explicitly mentioned in this figure, because it is expected that uncertainty propagation will be carried out through repeated running of a model that uses the information flow shown in Figure 10.
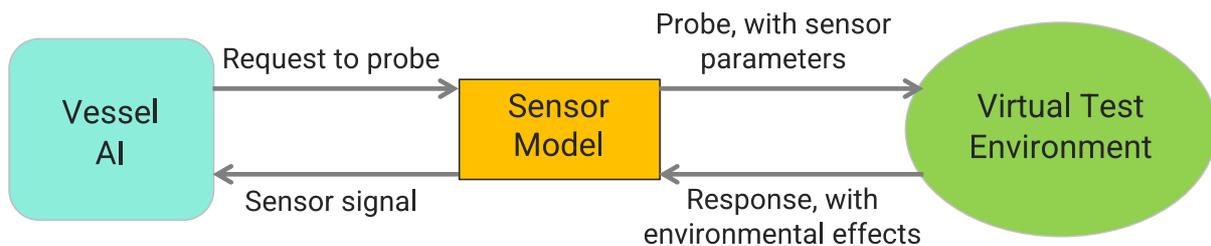

Figure 10: Sketch of interaction between sensor model, virtual testing environment, and vehicle AI

The performance of the sensor to be reproduced by the sensor model therefore has one essential aspect and two desirable ones. The essential aspect is that the sensor model must supply the AI with a signal exactly equivalent to the signal produced by the real sensor. The desirable aspects are the way the real sensor probes the real environment and the way that the sensor model probes the virtual environment should be as similar as possible, and the way the sensor model receives the response from the virtual environment and the way the real sensor receives the response from the real environment should be as similar as possible.

There are several reasons why these aspects are desirable:

- The first is that a close resemblance between the model and reality makes validation of the model simpler, because it reduces the approximation errors within the model that will affect model accuracy and so gives greater confidence in the validation
- The second reason, strongly related to validation, is that it is easier to debug a model that is similar to reality because the intermediate results are generally more likely to have an expected behaviour that can be checked for
- The third reason is that improvement of the real-world system in the case of test failure is easier if reality and the virtual system resemble one another, because the root cause of failure is often easier to identify, and alternative successful sensor specifications are easier to define. For instance, if a radar system is such that backscatter of the signal during rainfall significantly affects the signal to noise ratio, an environment that returns the scattered and reflected fields will allow the analyst to understand the cause of failure better than an environment that supplies an idealised extinction coefficient or similar single parameter
- A fourth reason is flexibility. It is easier to add new scenarios (such as new weather conditions) to an environment that is similar to reality because typically fewer approximations and extra calculations need to be made when defining the effect of the new environment on the signal to be returned to the sensor

This desire to reproduce reality leads to a conflict at the heart of virtual testing. Realistically detailed models of driving scenarios are computationally expensive. The more realistic the model, and in particular the more realistic the description of the interaction between sensor and environment, the higher the computational cost and hence the longer the model takes to run. One of the main aims of virtual testing is to reduce the time required to prove vehicle safety. Successful development and usage of virtual testing environments will therefore require finding a balance between computational cost and level of realism.

Sensor models therefore need to consider three aspects: probing, receiving, and processing. In general, the probing and receiving are likely to be physics-based because such models give the best approximation to reality and mean that a single model can be used for any situation where the physics describing the situation can be described using that model. For instance, a single physics-based model of a lidar system can be used for any form of object detection under any weather condition, provided that a) the scanning parameters of the lidar system are well understood, and b) the virtual environment has suitable physics-based descriptions of object responses to a laser under the weather condition of interest.

The model of the processing is generally best described using a data-driven approach. The main reason for this is that a physics-based model is inevitably an idealisation and a simplification of the sensor, meaning that it is unlikely to capture the full details of the sensor response, particularly under challenging conditions. A further reason is that development of a physics-based model would require access to the inner workings of the sensor, which is proprietary information and unlikely to be shared.

A data-driven approach to sensor model development depends on testing to generate the relevant data. The nature of these tests is discussed in more detail elsewhere in this document (section 6), but they should be sufficient to determine the true values of the key performance indicators (see section 8.1) under a variety of weather conditions.

The figure 11 below shows how we believe weather and weather-affected sensors are best described within a virtual testing environment and has informed our approach to the work.
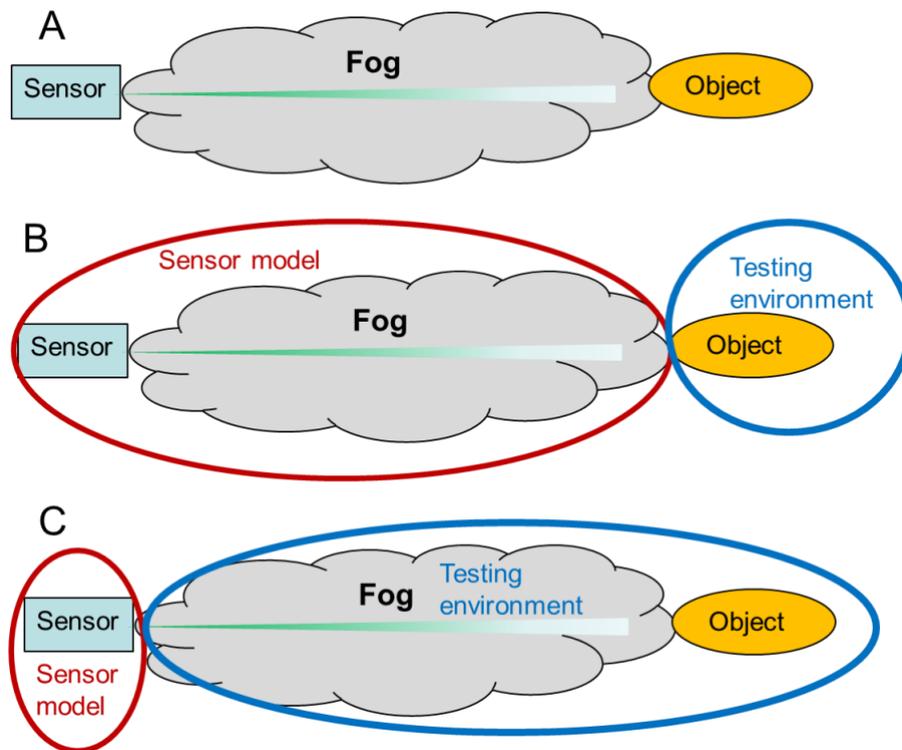
Figure 11: Perspectives on the treatment of weather sensor models

Figure 11 shows a simplified illustration of a sensor interacting with weather. This figure has been used to inform the collaboration between NPL and the Met Office throughout this work. The principle is that the sensor works by sending out a signal and then receiving back a scattered or reflected version of that signal. Assume that this signal is affected by the presence of fog, and that fog is present between the object to be detected and the sensor, as shown in Figure 11A.

In order to make a model of this situation, consisting of a sensor model and a virtual testing environment so that you can carry out virtual tests on the AI driving your CAV, there are two approaches, labelled B and C. Either:

- (figure B) the effects of the weather on the sensor signal are rolled into the sensor model, in which case you need a comprehensive set of test data to describe how the sensor behaves in fog, or
- (figure C) the effects of the weather on the sensor signal are described within the virtual testing environment.

The research approach between NPL and the Met Office takes the second approach (C), because it has a number of advantages (flexibility to weather type, accuracy, easier to swap sensors, etc.). The aim of the project is to illustrate how you might construct the models and (more importantly) how you would obtain the data/information to go into the models and to validate the models such that you could demonstrate that your model of a sensor on a vehicle in fog accurately reflected the actual behaviour of the sensor on the vehicle in fog.

(Note, the Study involves rain not fog, but the concept is the same for other weather types, hence it being a Proof of Concept project).

This definition of the weather as a feature of the virtual environment makes sensor model development, and hence generation of data for sensor model development, simpler. If the weather effects are folded into the sensor model then "weather" needs to be an input parameter to the sensor

model, and adequate data needs to be gathered to give accurate sensor model results for all values of that parameter. Given the near-infinite variety of weather (particularly in the UK), this is a challenging requirement. If the weather is shifted into the virtual testing environment, then the senor model only needs to reproduce the response of the sensor to a degraded signal, irrespective of the cause of the degradation. An additional benefit is that degradation of a signal is considerably easier to parameterise and quantify than weather.

It has been noted elsewhere in this report that many sensors have built-in object identification algorithms. Testing these algorithms, and characterising them for modelling purposes, is a challenge. If manufacturers are prepared to share a black box version of the algorithms for use in the virtual testing environments that takes in a simulated signal and returns what the real sensor would return, then integration and virtual testing are straightforward. If they are not, it will be necessary to run a sufficiently large set of real tests of the sensor with the object identification in place to characterise its accuracy and uncertainty under all weather conditions for all classes of object it can identify. Again, this is a significant amount of effort.

It is possible that CAV manufacturers would prefer to integrate the raw signal data into their AI algorithms, since this means that they maintain full control over the available data. Full control of the raw data means that sensor fusion algorithms can be more flexible, uncertainty evaluation can be handled more effectively, and full system testing becomes simpler. It also potentially reduces legal complications regarding responsibility in the event of an accident. This preference would mean that sensor manufacturers may offer sensors with and without embedded algorithms to different markets.

9. Roadmap for Framework delivery

This section explores what activities are required to successfully adopt the recommendations in section 5 and deliver a working example of the Framework – defined as **Stage 2**. It is essential that any activities are undertaken in close consultation with stakeholder groups, in order to ensure the Framework meets the criteria of reliability and usability.

In addition to establishing a practical and transparent structure for engagement, it is necessary for there to be a process of implementation and employment of the Framework. It is recommended that this is undertaken by Industry, supported by National PSRE's with the oversight and coordination of Government Agencies (CCAV, regulatory etc.).

The timeframe for the completion of Stage 2 can only be confirmed after a formal engagement and scoping exercise, but it is estimated to require 3 years. Major factors in determining the timeframe for completion are the degree and quality of external engagement and the timing of external milestones, for example funding calls for projects which have a dependency on the Framework.

There are three clusters of activities which can be identified:
- Alignment, engagement and scoping. (section 9.1 below)
- Research and Development: Build and demonstrate the Framework (section 9.2 below)
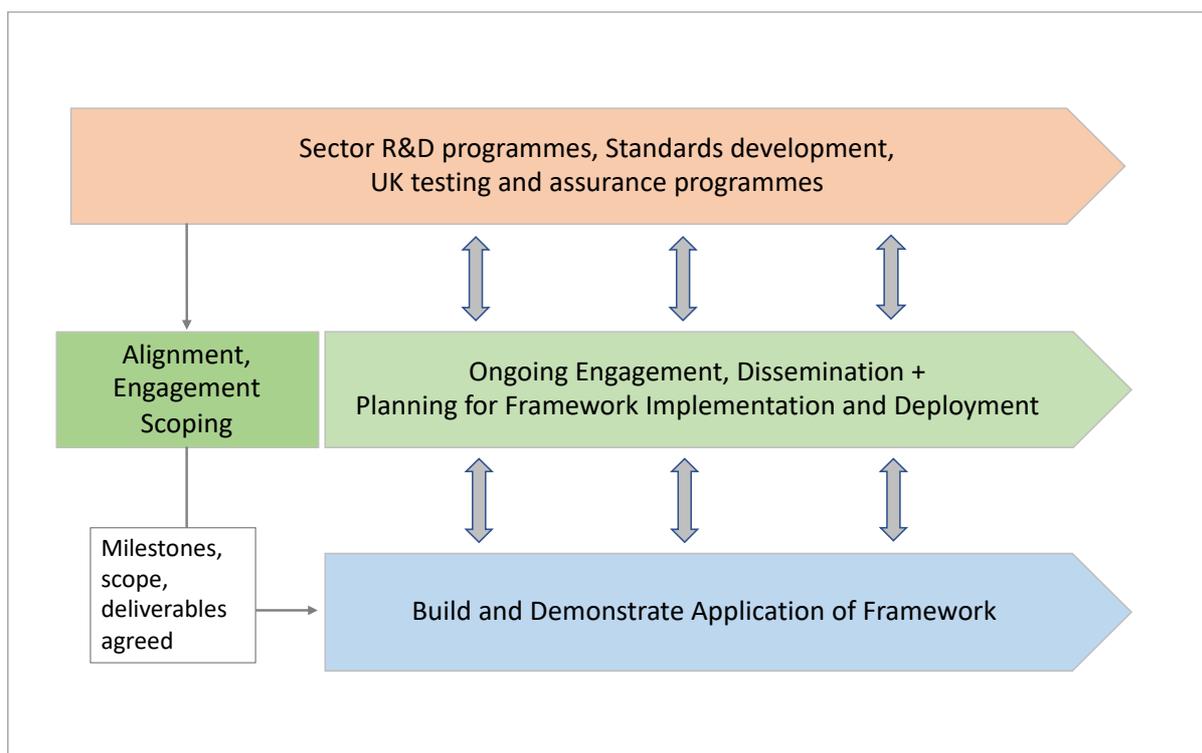- Implementation and Deployment (section 9.3 below)

Figure 12: Relationship between project, stakeholder engagement and sector roadmaps

9.1 Alignment, Engagement and Scoping

Engagement and collaboration across different stakeholder groups is a prerequisite to developing a successfully deployed Framework. At the same time, UK and International agencies are generating standards, regulations, guidance and tools to support the assurance of CAV. In 2019, the UK Government announced the development of an assurance system CAV PASS, to ensure self-driving vehicles are safe and secure by design and minimise any defects ahead of their testing, sale and wider

deployment on UK roads. The Development of the Framework supports the first two elements of the CAVPASS process – vehicle technical standards and regulation + vehicle testing, approval and in-service compliance.

CAVPASS
The Connected and Automated Vehicle Process for Assuring Safety and Security

| Vehicle technical standards & regulation | Vehicle testing, approval and in-service compliance | Safe Use | Enablers: skills & facilities | Advanced Trials | Process delivery and oversight |

Figure 13: CAVPASS process

- The outputs from Stage 1 will need to be reviewed, discussed and integrated into the planning for a successful Stage 2.   It will be necessary to build on current engagement to get input and guidance from the safety assurance bodies, regulatory, certification, and insurance sectors for example. Engagement will also increasingly focus on organisations who will be deploying / making use of the framework themselves, requiring them to understand what is required of implementation and to provide feedback on any issues with adoption as the project progresses
- The timings and scope of the Framework delivery must be aligned with and integrated into related R&D roadmaps, safety assurance and testing and evaluation activities.   It should also be developed in the context of the evolving standards landscape. (BSI CAV standards roadmap)
- A Sector Steering Group (SSG) or similar should be established, managed through CCAV, to ensure appropriate levels of governance, transparency and communication are maintained. Likewise, resource will need to be available to manage a thorough and equitable engagement process
- The above will enable completion of the scoping of activities, milestones, deliverables and timeframes for 9.2 - Development

9.2 Stage 2 Research and Development: Build and Demonstrate the Framework

At this stage the following can be identified as required to achieve the development and working demonstration of the Framework. It is noted that the engagement and alignment activities in 9.1 may result in additional requirements being identified.

A. **Completion of pathways to degradation Taxonomy** (as per recommendation 5.4). This will require lab research and data from demonstration of the test ecosystem.

B. **Specification and demonstration of the Test Ecosystem** (recommendation 5.1). This will provide the specifications for future permanent facilities and will enable the definitive set of measurements to be defined for each testbed. It is also required to enable a full demonstration walkthrough of the Framework. This will involve the following elements:
  - **Undertake Calibration Lab Tests:** The data obtained from the "dry" calibration lab would be a precursor to make comparisons with the CETF to understand the weather impact on the sensor performance.   It should be noted that an ideal calibration lab itself would be different for the different sensors

- o **Build Demonstration external environment reference testbed.** While the detailed measurement of meteorological processes at dedicated testbeds is a well-established practice, the acquisition and analysis of these measurements for this new purpose is less well-developed. It is therefore sensible to de-risk larger investment this area through a limited demonstration facility which would require being sited at a Met Office Research Unit. Basic requirements for a demonstrator include:
  - Access to meteorological and measurement expertise and equipment
  - Enables testing over a range of ~200 m
  - Proximity to UK Testbeds for ease of industry engagement

  This Testbed demonstrator represents one of the more complex deliverables of Stage 2 which is a reflection of its uniqueness and importance:
  - Informs final recommendation for the instrumentation (both meteorological and CAV sensor performance) required of a permanent facility, extending to the measurement of the other primary impact weather parameters (e.g. fog, dust, frozen hydrometeors etc)
  - Recommendations for possible geographical regions (or specific locations) of this facility – the balance of expediency over "guaranteed meteorology" may shift in favour of the latter as the Technical Readiness Level (TRL) of the test facility increases
  - Recommendations of the measurement set required in a Controlled Environmental Test Facility (CETF) in order to maximise both traceability and assurance confidence
  - Will inform the minimum measurement set for a possible mobile technical platforms
- o **Generate Specifications for Weather emulation for CET F**: this is to enable early competition/funding call for creating those elements of the CETF. This will include the weather emulation equipment and chamber specifications
- o **Explore other useful testing options**: This activity focuses on identifying what quantitative information might be provided from additional test options, including mobile data capture and fixed roadside infrastructure
- o **Creating the linkage into the virtual environment**: This should focus on the need to find a pragmatic middle ground between simplistic treatment of weather effects and requiring levels of evidence that are too computationally intensive to be practical

C. **Generate the Definitive set of measurements:** (recommendation 5.3) The requisite set of measurements must be defined for each testbed and are required for the following three areas; **Meteorological**; **Reference targets** and **KPI tests**. These will differ for each testbed as explained in Section 6. The measurement data will be used for the Physical walkthrough of the Framework. This activity addresses all relevant weather types, as well as correlating data captured from in situ representative sensor suites with weather types/conditions. It will require multiple years to establish and model a full suite of conditions and will require ongoing activities to keep up to date and improve testing methodologies. This will also require specification for data sharing infrastructure and methodologies to enable linking of test data, including recommendations for standardisation of data and metadata formats.

D. **Physical Walkthrough of Framework:** (recommendation 5.2) The physical walkthrough will define the data model and interfaces used to link the test results and to link the physical and virtual spheres. The walkthrough should be used to investigate and demonstrate the effects of uncertainty (measurement and weather-generated) and decision-making parameters on the decision. For the sake of efficiency, a direct interface to a virtual testing environment should not be required, but it is expected that the data flow between a sensor model and the virtual environment would be specified, and that a dummy sensor model would be created to illustrate this.

9.3 Framework implementation and deployment: integration into funding calls

The process of commissioning the framework into the market, will be industry/government led for use across UK stakeholders. This should include the deployment of standardised test methodologies for weather effects on sensors, as well as guidelines for supporting data quality standards, and enable the verification of the equipment specifications for a planned UK environmental testbed facility, although the approach and timing will be determined during 1.2 above and rests outside the scope of Stage 2.

10.   References

Main report references only

1. Thatcham.org, 2020, https://www.thatcham.org/thatcham-research-and-association-of-british-insurers-urge-government-to-revise-plans-for-introduction-of-automated-lane-keeping-systems-to-ensure-road-safety/
2. AAA, 2019, *Automatic Emergency Braking with Pedestrian detection,* https://www.aaa.com/AAA/common/aar/files/Research-Report-Pedestrian-Detection.pdf
3. *NPL and Connected Places Catapult, 2020, Performance Testing for Sensors in Connected and Autonomous Vehicles: Feasibility Study*
4. Connected Places Catapult, 2020, https://cp.catapult.org.uk/events/the-certicav-safety-assurance-framework-for-cavs-consultation-workshop/
5. BSI, 2020, https://www.bsigroup.com/en-GB/CAV/pas-1880/
6. BSI, 2020, https://www.bsigroup.com/en-GB/CAV/pas-1883/
7. Connected Places catapult, 2019, https://cp.catapult.org.uk/case-studies/musicc/
8. *Nidhi Kalra, Susan M. Paddock; "*Driving to Safety , How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?" RAND Corporation (2016)
9. ISO 26262: 2018 Road Vehicles – functional safety, https://www.iso.org/standard/68383.html
10. "Challenges in Autonomous Vehicle Testing and Validation", Philip Koopman & Michael Wagner
11. ISO 12233:2017 - Photography — Electronic still picture imaging — Resolution and spatial frequency responses, https://www.iso.org/standard/71696.html
12. ISO 19206-2:2018 - Road vehicles — Test devices for target vehicles, vulnerable road users and other objects, for assessment of active safety functions.. https://www.iso.org/standard/63992.html
13. ISO17850:2015 - Photography — Digital cameras — Geometric distortion (GD) measurements, https://www.iso.org/standard/60819.html
14. Abari et Al., Lyft.inc, 2019, Mobile Sensor Calibration Patent Application,   US2019/0204425 https://www.freepatentsonline.com/20190204425.pdf
15. Maxwell, James Clerk (1865). "A dynamical theory of the electromagnetic field". Philosophical Transactions of the Royal Society of London. 155: 459–512
16. Dixon, J (2019) *An investigation into short-period, extreme rainfall in the UK,* Report for Innovate UK, Met Office, pp79. Available at: https://www.metoffice.gov.uk/services/transport/cav
17. In reality both the DSD shape and the rain rate may vary, however, we assume for demonstration purposes that the DSD "shape parameters" are constant

11.   Appendices

Appendix A: Summary of engagement findings

When considering the statements made in the logical flow in the discussion document (See Appendix B), there was a general consensus that linking the right tests could help increase overall confidence in test results; thereby pointing to a solution for the challenge of how to test whether perception systems are providing reliable outputs in different weather conditions.

There was general support for the following points:
- Testing (including virtual testing) of CAVs needs to take weather into account. Hence vehicle test specifications, sensor characterisation, and virtual testing environments need to take weather into account
- Better confidence in sensor performance can be achieved by a series of *linked tests* & characterisation exercises at different levels where sensors and full systems are tested both in reality and virtually. In this context "reality" may include testing in both the natural environment AND user-controlled environmental test facilities
- Because weather demonstrates significant small-scale variability and sensor response has associated uncertainties, uncertainty calculation is critical and must be reflected appropriately in the virtual simulation

Sensors are set to be increasingly relevant to a number of future standards workstreams. As the project moves into the next stage, the project should ensure collaboration with other stakeholder groups at a committee level to define and agree the methodologies and tools and then propagate them through standards organisations and industry bodies. A particular focus will be given to standards bodies.

General agreement that in setting performance standards it is essential to safeguard against poor performing systems on the road. One of the challenges in this regard is that much of the type approval for automation involves assessment of technical documentation from the manufacturers, which may not provide relevant or detailed enough information.

It has been widely acknowledged that:
- The challenge over modelling weather combinations is significant
- It is not just the specified weather event testing (e.g. rain/sun/frost/wind) that is important to ensure the safe operation for sensors and systems but also how they deal with (or not) the transition between any two weather states
- Weather models overall are immature

Understanding the types of weather that impact sensor performance and what those impacts are / how to measure them is considered an important first step. The Tables in Appendix 5 constitute a first step in building out the relevant definitions, and also in defining separation between different aspects of weather (e.g. standing water v water in the air). Further steps and discussion will be undertaken to consider how this feeds into standards.

With regards to weather definitions and impact there has to be a clear link into ODD language. A problem with weather conditions in relation to CAV performance and the ODD is that weather elements may not be detectable by a CAV: for example there is a difference between understanding regional and national rainfall and the rainfall in front of a vehicle.

Detectability:
- Important to be able to detect when and how sensor is being impacted

- Is it possible for vehicles sensors to define/detect whether there is weather which is altering performance?
- Sensor needs to know there is something to look at
- Question of how to test that vehicle can see as far as it needs to: "needs to" is likely to be speed and weather dependent (e.g. wet roads mean longer braking distance)

Recognition that the issues defined here represent a current problem with the advent of ALKS and other 'hands free' or 'self-driving' ADAS regulations.

With regards to the linking and hierarchy of tests: Support for a framework which could generate a continuum between different types of tests, which would align with the general best practice approach for tests to be conducted across simulation, proving ground and real world. A requested next step identified was to demonstrate (theoretically) an example of how a roadmap of testing would work. A first draft hierarchy was pulled together which appears in Figure 3 in Section 6.

The presented approach on linked tests at the sensor level was acknowledged as providing several advantages over functional tests alone – at the same time, functional testing was still a valid part of testing overall. This approach also helps create robust sensor models for simulation testing and promotes an audit trail around testing.
- Bench testing: positioning is a factor. A well-positioned sensor i.e. good air flow, optimum view would be expected to behave better than exactly the same sensor array in a poor location. Needs to be reflected in bench testing
- Real world weather testbed – may need to be more mobile
- General support for a 250m environmental facility, combining wind with weather for example
- The use/purpose of virtual environments; is this for validating the perception system or the ADS?

In the context of current approaches by OEM's a review of the hierarchy of testing indicated:
- OEM focus of approach is on the whole vehicle sensors/perception system on a car. Overall perception system performance is considered important rather than a single sensor. Agreed that this presents a challenge if a single sensor fails, but in itself is not a critical component (whereas perception system is): The challenge is how is this tested?
- OEM focus on outputs: validating what the perception system is seeing and the use of data gathering (real time) and viewing tools to support this. This looks at establishing a minimum requirement for positioning/range performance, then also validating the outputs of perception system
- Multiple views expressed that individual sensor performance testing had to be considered in the overall context of the V shaped system design and test (Equivalent of unit/component tests). An indicative flow is: vehicle requirements are x $\Rightarrow$ therefore against known vehicle dynamics define the situational awareness requirements $\Rightarrow$ this then defines mix and specification of sensors
- A complete systems test should sit at the top (come last)
- This is not a single linear hierarchy of tests, but a cyclical one
- At a low level (sensors), there is probably a minimum specification that needs to be met – this give pass/fail at a "component" level
- At a higher level (systems), there needs to be an evidence-based argument for the safety of the system
- Caveat – a system comprising a set of sensors, each of which meets the minimum specification (all "Pass"), may well not constitute a sufficient system

In general, the V model represents a methodical process of creation followed by verification and validation. The left side of the V works its way from requirements through design to implementation. At each step it is typical for the system to be broken into subsystems that are treated in parallel (e.g., there is one set of system requirements, but separate designs for each subsystem). The right side of

the V iteratively verifies and validates larger chunks of the system as it climbs back up from small components to a system-level assessment. This model has been promoted to be the reference model that forms the basis of ISO 26262. Although ISO 26262 and its V framework generally reflect accepted practices for ensuring automotive safety, fully autonomous vehicles present unique challenges in mapping the technical aspects of the vehicle to the V approach below which has been summarised by P. Koopman, et.al in "Challenges in Autonomous Vehicle Testing and Validation".
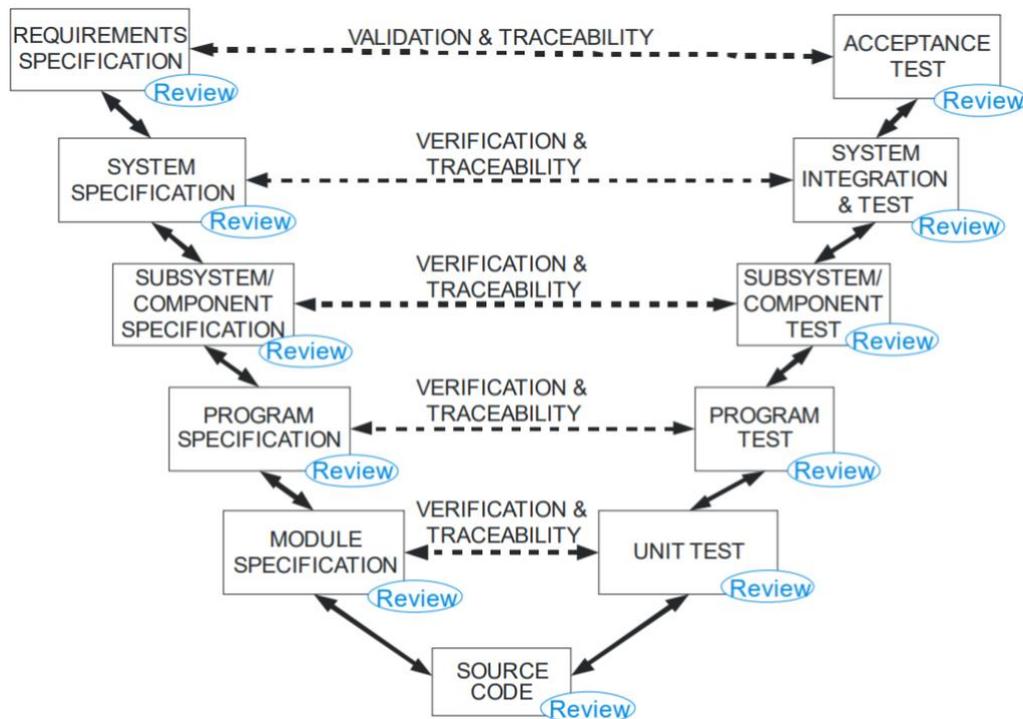
Two important aspects of an "approval" / "verification" / "certification" method for sensors or perception systems must be borne in mind when considering the process:

- Cost: The cost of the process to the Vehicle Manufacturer / ADSE requesting "approval" must be reasonable
- Timing: Assuming that the system meets the requirements, it would be unreasonable for the process to delay a vehicle launch excessively

General feedback on sensor data fusion was that the ability to test and safety assure was not yet developed. What data from sensor is most important?

With regards Uncertainty:

- Both confidence and pass/fail will be relevant: pass/fail is a function of confidence but not clear how to measure confidence. Statistical analysis and the safety argument both factors
- Uncertainty needs to capture in certification so you are not assuming perfect perception. i.e. Challenge around multiple sensors in a system and whole system validation
- Positive feedback on the idea of having 'error bars' (uncertainty) in simulation
- Recognised complexity of propagating uncertainty across a wide range of phenomena/materials
- Use component level tests, then feed this into a simulation, from which one can understand the permutations around sensor performance and how this affects the rest of the system's performance
- Concern about effects in Simulation being computationally expensive which limits number of tests (suggested use of real-world testing to create abstract models which can be used more cheaply in simulation)
- Identifying failure points through Monte Carlo simulation, which should be used sensibly

- Repeatability of virtual tests is key. Need to be able to keep retesting to identify the issue

The view is that as the number of connected vehicles and their on-board sensors increases, there will be an unprecedented ability to form a detailed and timely picture of the CAV-relevant environment.

References

[1] "Challenges in Autonomous Vehicle Testing and Validation", Philip Koopman & Michael Wagner

Summary document used in phone interviews and as preparation for the Workshop (note the version here is a longer version developed for the Workshop)

CAV Sensor testing & models / Weather degradation Stage 1 project

<u>Preamble</u>
This is an informal document to provide context to the 5th March workshop. It is an updated version of a note that was used in a number of stakeholder Skype calls to help shape the workshop, so the content will be familiar to some delegates. We ask that the following is borne in mind when reading it:

- It focusses on top level themes rather than the use of definitive/standardised language. If you feel we have inappropriately used a term, then please (a) forgive us and (b) let us know
- It is not intended to be a comprehensive presentation of the Framework, not least because that requires the information gathered at this workshop. If you feel something is missing it might be because of the level of simplification but, equally, please let us know your thoughts

The project
Our aim is to prove the concept of, and create a specification for, <u>a usable and reliable</u> Framework for characterising sensor performance in different weather-related conditions, including the ability to assess performance outside the design envelope. Uses of this framework include validation, safety assurance and simulation testing of CAV. In order to test the approach, in this project we have focussed primarily on the single weather element of **rainfall**, although we expect the workshop discussions to extend to other weather factors.

The "quick read": The logic flow
- Testing (including virtual testing) of CAVs needs to take weather into account
- Hence vehicle test specifications, sensor characterisation, and virtual testing environments need to take weather into account
- Weather, using rain as an example, is not as simple as "heavy rain"/"rain level 3" etc, because it varies spatially and temporally. So, a shower that is "on average" apparently safely within an ODD threshold may contain local and short instances of very heavy rain that are outside the ODD
- Testing of CAV performance in bad weather will need to characterise impactful weather thresholds from the point of view of a sensor, which is not necessarily the same as that of a human
- Better confidence in sensor performance can be achieved by a series of *linked tests* & characterisation exercises at different levels (as per Figure 1) where sensors and full systems are tested both in reality and virtually. In this context "reality" may include testing in both the natural environment AND user-controlled environmental test facilities
- Because weather demonstrates significant small-scale variability and sensor response has associated uncertainties, uncertainty calculation is critical and must be reflected appropriately in the virtual simulation
- We have developed a draft outline framework and are currently filling in what each of the elements within the framework, and the links between them, need to look like to achieve confidence in outputs of the testing regime
- The framework must relate clearly to ODD taxonomies

More detailed thoughts – the "longer read"
We are seeking to establish an end-to-end description of the linked tests, which we are calling our *golden thread*.

The golden thread is the minimal set of evidence that demonstrates that the vehicle as a whole (i.e. sensors, AI, response) is safe in (a given form of) adverse weather conditions. This requires a set of scenarios that define adverse weather conditions that might lead to safety problems.

The scenarios need to be designed based on actual sensor characteristics and a consideration of what affects the signal that is sent to the AI. This could be reduction of power, spreading of directionality, delay or temporal spreading, extra reflections, etc., and will include uncertainty and error terms (i.e. statistical and deterministic aspects). It should be noted that the scenarios may not be the same as those that would cause a human driver problems, so the scenario definition needs to start from the sensor characteristics. The scenarios would consist of a set of weather conditions and objects (e.g. other vehicles, pedestrians, infrastructure etc) with trajectories or other behavioural definitions if required.

Safety cannot be proved through physical testing alone, because testing can never cover the full range of conditions that might occur. Scenarios that *can* be tested physically *should* be tested physically, but this will not always be possible and will be costly when it is possible. Further, the cost is likely to be such that repeat testing (to assess variability and uncertainty) is likely to be impossible, meaning that confidence in the results is not complete. Hence there has to be virtual testing.

Virtual testing requires models of sensor response in adverse weather conditions and models of vehicle response in adverse weather conditions. This requirement will have a weather component, a sensor model component, and a vehicle component.

At the CAV scale, weather has a significant random component and needs to be simulated as a sample from some form of distribution (for instance for rain this would include droplet size distribution, intensity/velocity, duration, and possibly others). It should *not* be treated using average values, not least because the safety or otherwise of the system will be most tested at the extreme values. However, worst case scenarios are not the only useful information: testing extreme conditions can tell you about the highest risk rare incidents, but this approach doesn't answer "what can the vehicle cope with?"

The sensor model cannot be developed assuming perfect physical performance: it needs to take real performance, and hence uncertainty, into account. It should be based on measurements not just physical theory.

There also needs to be a method for defining how the weather affects the response of the sensor 'probe' (e.g. radar pulse, lidar illumination, etc) to the environment. This can to a significant degree be physics based because the probe is generic with measurable characteristics (wavelength, scan speed etc) rather than sensor-specific although it is most likely that this model would also be validated by measurement.

The sensor model will then have to take in the received weather-affected signal and process it in the way that the real sensor would, which will depend on things like integration time, field of view, etc. It is likely that some of this processing information will be commercially sensitive and hence not generally available, but it may be possible to obtain relevant information through a carefully designed test plan.

Once the weather-affected signal is processed, a set of values equivalent to the (likely) real sensor signal will be passed to the AI. The response of the AI to the sensor data determines safety. Safety may need to be determined over a prolonged time period as effects may be cumulative. It should be noted that the safety of the response of the AI also depends on the response of the vehicle in its entirety to the weather conditions and is not an unvarying yes/no criterion.

We are aware that models of vehicle response in adverse weather conditions already exist, since

driving simulators are a mature technology. There are still open questions about the nature of the weather models used in these simulators: as noted above, weather is a random/high variability phenomenon and the weather models needs to use samples at appropriate space and time scales rather than relying on average statistics.

All of these components have associated uncertainties, requiring that the virtual testing environment is essentially stochastic. In particular:
- weather is significantly random, or at best highly varying on small time and space scales
- sensor probe response to environment (e.g. for lidar & radar) may have an associated uncertainty, which may be correlated between sensors
- random weather acts on random signal to produce random response
- sensor processing adds random effects
- vehicle response will have a random effect (which will probably encapsulate unknown/unknowable characteristics of the road/environment)

In summary, the components of a testing framework would be:
- scenarios that might cause problems
- measurements of sensors
  - ideally including some under controlled weather-simulating conditions for validation
  - ideally including some full-car tests for validation too
- models of sensors based on measurements, possibly guided by physical understanding during development
- model of weather, including spatial and temporal characteristics of any defining parameters that will affect signal response
- method of generating a weather-affected signal based on weather characteristics and sensor probe characteristics
- model of processing of received signal so that weather-affected signal can be passed to AI

One possible visualisation of all the above is given below:
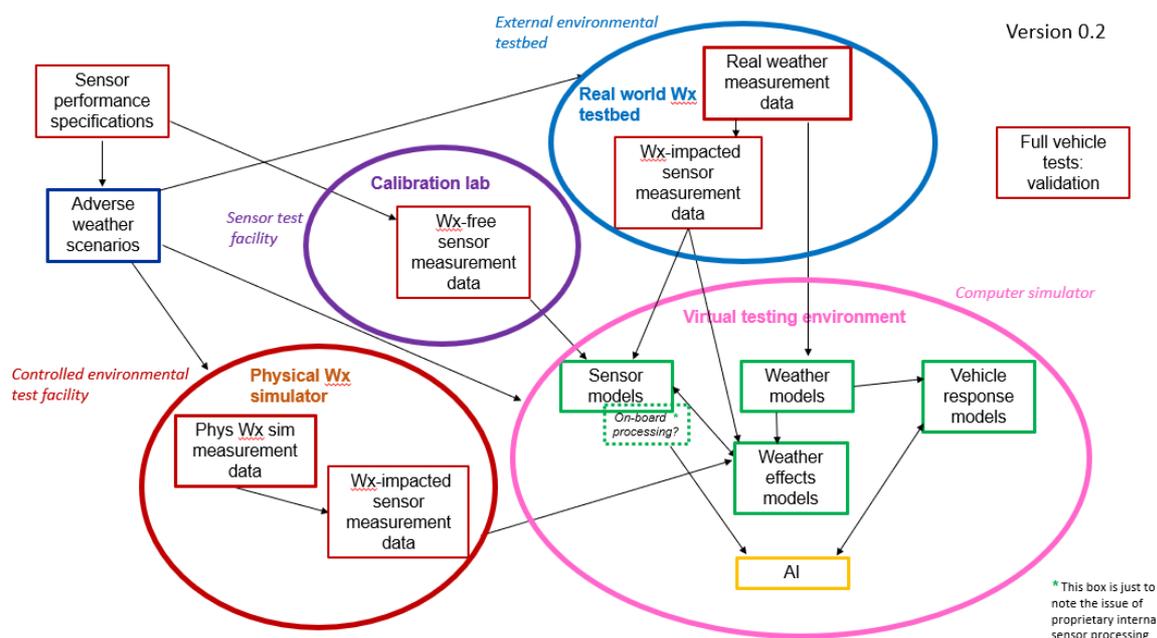
(***Wx** = shorthand for weather)

Figure 3: A representation of interlinked test environments for CAV sensors

The diagram will be discussed at the workshop, however the following may help with interpretation in the meantime:

Calibration lab: Testing of sensor performance against reference "targets" in idealised laboratory conditions i.e. in the absence of impactful weather.

External environmental testbed: Testing of sensor performance against reference targets, fully exposed to the elements. This would include ancillary measurements of all the impactful meteorological parameters (e.g. rain droplet size distributions, wind, temperature etc). Carries the advantage of sampling "real weather" but it is not controllable and requires the weather to come to the testbed site.

Controlled environmental test facility: Often referred to as a "wet shed". Testing of sensor performance against reference targets, exposed to manufactured/emulated weather (e.g. rain from a sprinkler system). Carries the advantage of "controlling weather" however may not be able to reproduce realistic rainfall in terms of droplet size (although the nature of the rainfall in such facility may be measured). This may have significant implications if considering relative performance/sensor redundancy when testing combinations of sensors operating at different wavelengths (e.g. lidar & radar) or passive and active systems (e.g. camera & lidar).

Virtual testing environment: Testing of virtual realisations of sensors, sensor fusion & ADS (and ADAS) in modelled scenarios that attempt to include the weather.

Many industries that involve construction of complex systems use a pyramid approach to testing, as illustrated in the following figure. This approach builds up confidence in the complete system by carrying out a large number of tests of the materials and individual components within the system, and progressively testing a smaller number of more complex sub-assemblies, before carrying out a small number of tests of the complete structure. The benefits of the approach include increased understanding of the sub-systems and a reduction in the number of the costliest full-scale tests.

The aim and structure of the workshop

Any Framework must be simultaneously both rigorous **and** practically usable to all stakeholders. The workshop is designed for us to gain as much insight as we can from all attendees to ensure that is the case.   It will be broken down into three self-contained workshops covering the following topics

1. Sensors and how weather conditions impact their performance
> *including causes of weather-related degradation, maturity of understanding & multi-sensor aspects*

2. The Balance of tests and how to ensure transfer and traceability
> *including what the pyramid might look like, investment priorities, strengths/weaknesses of*
different tests

3. Virtual testing and handing uncertainty
> *including pass/fail vs confidence levels, the role of simulation (including in characterising*
uncertainty)

Appendix C: Walkthrough Calculation details

Details of calculations of maximum detectable range
The attenuation coefficients have been calculated such that the attenuation along a path in one direction is

$$I = I_0 \exp -\gamma L$$

where I_0 is the signal power leaving a point, L is the distance the laser beam has travelled, and γ is the attenuation coefficient. For a LiDAR, it is assumed that the signal travels in a straight line so the only reduction of power between the signal leaving the sensor and reaching the target is this attenuation. When the signal reaches the target, it is absorbed and scattered so that the scattered signal at a distance R is given by

$$I_S = \frac{I_0 \sigma}{R^2} \exp -\gamma(L + R)$$

where σ is scattering cross-section with units m². When the scattered signal reaches the sensor, the amount of signal that enters the sensor depends on the optics, meaning that an extra factor of K (0 < K < 1) is introduced. Then

$$I_R = \frac{I_0 K \sigma}{L^2} \exp(-2\gamma L)$$

It is assumed for the purpose of this demonstration that no attenuation occurs in dry air (γ_0=0) so if the minimum detectable signal power is known then the product $K\sigma$ can be calculated from the maximum detectable distance.

If we know the maximum detectable distance in dry air, L_0 say, and the attenuation coefficient for rain, and we want to calculate the maximum detectable distance when rain is present, L_R say, then we can note that the signals in both cases must be at the minimum detectable level, so that

$$\frac{I_0 K \sigma}{L_0^2} \exp(-2\gamma_0 L_0) = \frac{I_0 K \sigma}{L_R^2} \exp(-2\gamma L_R)$$

so that, as γ_0=0, L_R is the solution to

$$\frac{1}{L_R^2} \exp(-2\gamma L_R) = \frac{1}{L_0^2}$$

which is nonlinear and requires numerical solution.

Values of the attenuation coefficient have been calculated as 6.4 km⁻¹ for the Marshall-Palmer drop size distribution at 100 mm hr⁻¹, and between 4.3 km⁻¹ and 7.8 km⁻¹ for a typical shower drop size distribution at 100 mm hr⁻¹. For a maximum detectable distance of 110 m in dry air, these values equate to maximum detectable ranges of approximately 70 m, 78 m and 66 m respectively.

Radar signals are governed by a different equation because the signal they emit is less focussed than a laser, so the beam spreading needs to be taken into account in both directions. The appropriate equation is

$$I_R = \frac{I_0 C}{L^4} \exp(-2\gamma L)$$

where C is a constant (for a given sensor and target combination) with units m^4 that includes the effects of signal wavelength, the directivity (antenna gain) of the emitted signal, target reflectivity, the ability of the sensor to capture the reflected signal, and other losses. Similarly to the lidar equation, the maximum detectable range for a situation with known attenuation coefficient can be calculated from

$$\frac{1}{L_R^4} \exp(-2\gamma L_R) = \frac{1}{L_0^4} \exp(-2\gamma_0 L_0)$$

If once again the "dry" absorption is neglected (γ_0=0), this simplifies to:

$$\frac{1}{L_R^4} \exp(-2\gamma L_R) = \frac{1}{L_0^4}$$

The calculated attenuation coefficients for a 77 GHz radar are 8.3 km^{-1} for the Marshall-Palmer drop size distribution at 100 mm hr^{-1}, and between 5.9 km^{-1} and 10.4 km^{-1} for a typical shower drop size distribution at 100 mm hr^{-1}. For a maximum detectable distance of 250 m in dry air, these values equate to maximum detectable ranges of 140 m, 158 m, and 128 m respectively.

Details of calculations of beam pattern effects

The effects of the presence of a radome and a radome covered with a water film on a radar beam shape have been simulated using finite volume analysis. The simulation has used a horn antenna of transmit and receive gain of 18.2 dB with vertical polarization. These parameters are typical for automotive radar systems. The horn is made of copper and the bumper is made of a material similar to polypropelene (relative electrical permittivity of 2.3), both of which are realistic choices. The radome size has been set at 4 mm thick and 30 mm square, which is more than 20 times the size of the antenna beam footprint when the distance between antenna and the radome is 30 mm so that the edge diffraction effects are suppressed. The radome, shown as a brown layer in figure C2, is set parallel to the transmitter horn at an offset of 30 mm, which is the typical distance between the transmitter and bumper for Continental radar. A second horn antenna is placed at 6 cm offset to receive the signal, effectively in the far-field.

The simulated antenna shows excellent radiation performance and uniform radiation pattern in the 77 GHz frequency band, plotted in figure C1. The 3-dB beam width of the simulated horn antenna at 77GHz is about 22° and the antenna gain generally increases with the frequency.
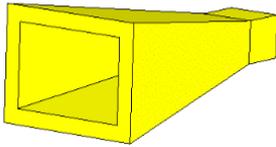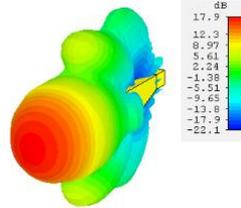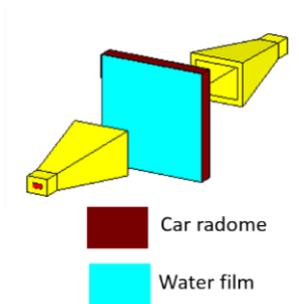
(a)   (b)

Figure C1: (a) Horn antenna designed for operation at 77 GHz; (b) Realised gain radiation pattern at 77 GHz with no radome

Antenna separation: 6 cm
Radar-radome spacing: 3 cm

Figure C2: Configuration with transmit and receive antennas (yellow), radome (brown) and water layer (blue)

The thickness of the water film, shown in blue in figure C2, has been varied between 0.1mm to 0.4mm in 0.1 mm steps. The calculated reflection and transmission coefficients are shown in figure C3 plotted against frequency. When the radome is not present (red curves in figure C3) there are minimal reflections. As the water film thickness increases, the transmission decreases (shown in the transmission coefficient plot) due to reflection and absorption of the signal by the water.
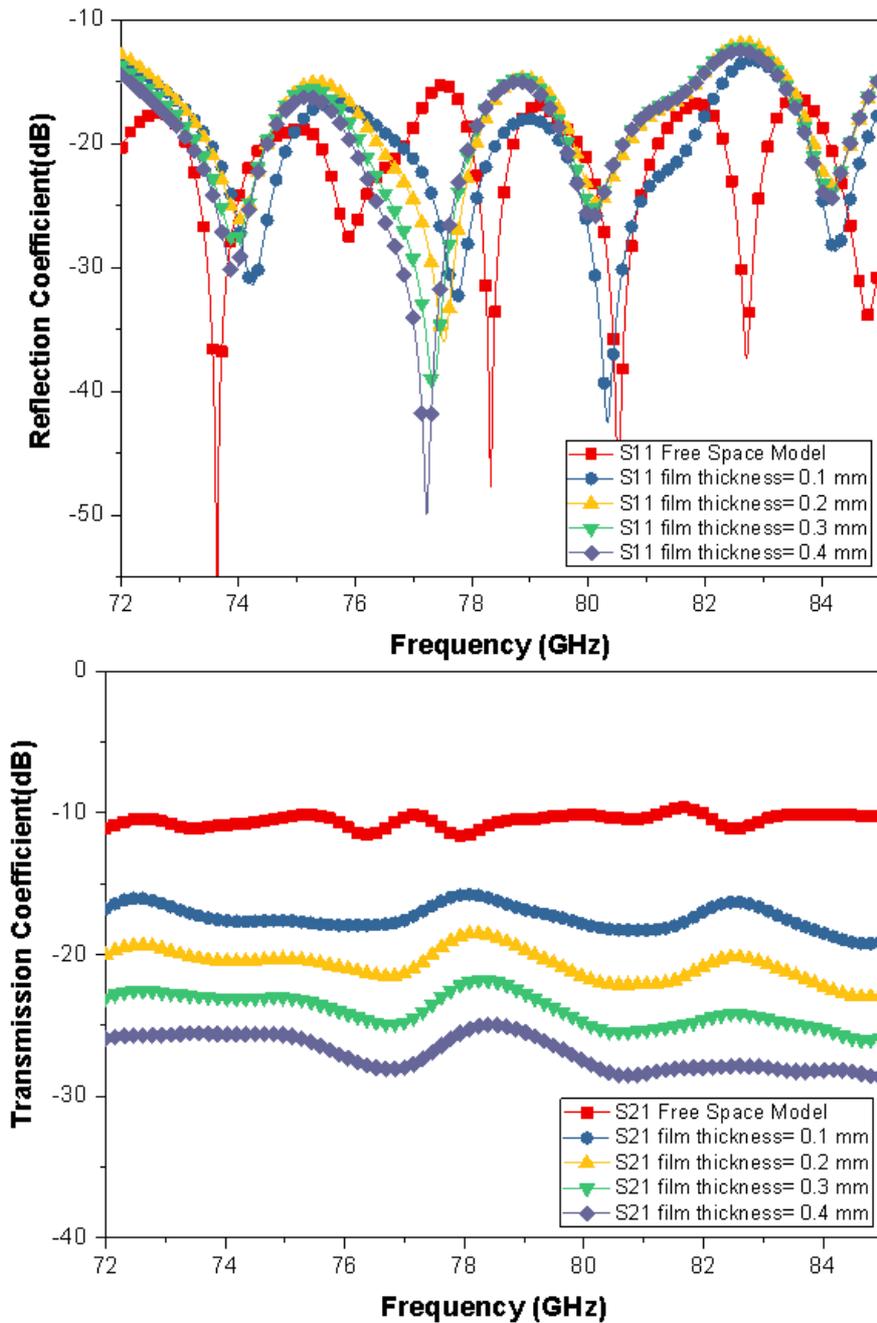
Figure C3: Reflection (upper plot) and transmission (lower plot) coefficients for various water film thicknesses plotted against frequency

Figure C4 shows the radiation pattern of the antenna at 77 GHz for no radome, a dry radome, and a radome with a 0.4 mm layer of water. For this model the radome has been placed 10 mm away from the antenna. The plots show that there is a significant back scatter component in the radiation pattern and distortion of the radiation pattern when the water film is present, which would reduce the maximum detectable range of the radar significantly.

Radar-Radome separation: **1 cm**

Free space

Car radome

Water film + car radome
Water film thickness: 0.4 mm

Reflection from the water film
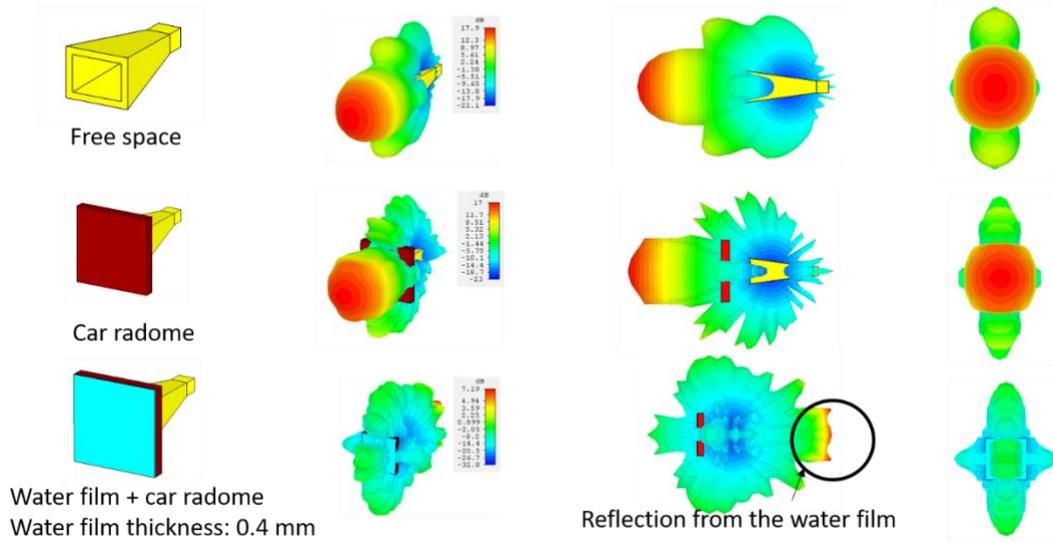
Figure C4: Radiation pattern for the radar antenna at 77 GHz for different scenarios

Appendix D: Sensor calibration

Robust environment perception is one of the essential tasks which an autonomous vehicle must accomplish [1]. To achieve this goal, various sensors such as cameras [2], radars, LiDARs, and inertial navigation units are used, and information thereof is often fused. Essential tasks such as simultaneous localisation and mapping (SLAM), detection and tracking of moving objects, odometry, etc. are often improved by sensor fusion.

A fundamental step in the fusion process is sensor calibration, both intrinsic and extrinsic. The former provides internal parameters of each sensor (e.g. focal length of a camera, bias in LiDAR range measurements), while latter provides relative transformation from one sensor coordinate frame to the other. The calibration can tackle both parameter groups at the same time or assume that sensors are already intrinsically calibrated and proceed with the extrinsic calibration. Additionally, temporal synchronisation of the sensors is sometimes performed within the calibration.

Intrinsic parameters are related to the working principle of the sensor. Therefore, methods for finding intrinsic parameters do not share many similarities between different types of sensors. On the other hand, parametrisation of extrinsic calibration, i.e. homogeneous transformation, can always be expressed in the same manner, regardless of the sensors involved. Despite that, solving the extrinsic calibration requires finding correspondences in the data acquired by the sensors which can be challenging since different types of sensors measure different physical quantities. After correspondence registration, optimisation steps are performed to estimate the calibration parameters. While some methods require intrinsically calibrated sensors to find the extrinsic calibration, others perform optimisation on both parameter groups simultaneously. These methods typically try to satisfy some geometric constraints through minimisation of a problem-specific reprojection error. The geometric constraints involve nonlinearities which often cannot be solved analytically. To resolve that problem, estimators use iterative techniques to find the appropriate solution. Due to the nonconvexity of the problem caused by the nonlinearities, these methods have a risk of converging to a local minimum. To avoid that risk, some methods divide optimisation in initial rough estimates that guarantee near-optimal solutions followed by nonlinear iterative refinement step. The success of the optimisation is highly dependent on the provided data. An important step before the data acquisition is to determine minimal requirements on the dataset for which the problem becomes identifiable (or observable in case of dynamical systems).

The calibration approaches can be target-based or targetless. In the case of target-based calibration, correspondences originate from a specially designed target, while targetless methods utilise environment features perceived by both sensors. The former has the advantage of the freedom of design which maximises the chance of both sensors perceiving the calibration target but requires the development of such a target and execution of an appropriate offline calibration procedure. The latter has the advantage of using the environment itself as the calibration target and can operate online by registering structural correspondences in the environment but requires both sensors to be able to extract the same environment features. Registration of structural correspondences can be avoided by motion-based methods, which use the system's motion estimated by the individual sensors to calibrate them. These methods have two main advantages: (i) they rely less on the sensor's operating principles and can be applied to different sensors, if a sensor can estimate its motion, (ii) unlike other methods, they can extrinsically calibrate sensors with nonoverlapping fields of view.

In Appendix D of this report, the different calibration techniques for the most popular sensor technologies used in the automotive industry such as camera system, LiDAR and radar are discussed.

1. Camera System

Cameras are passive sensors that utilise the light which goes through the lens and is detected at the optical sensor. They are a rich source of information with an affordable price that makes them commonly used in robotics and other fields. Due to their long presence and frequent usage, intrinsic camera calibration has been given a lot of research attention which resulted in camera description models and calibration techniques. While cameras with high distortion such as fisheye and omnidirectional cameras require more complex models, commonly used cameras with slight distortion are usually modelled as pinhole cameras with a previously rectified image as can be seen in the next section. This intrinsic parametrisation consists of distortion coefficients (e.g. radial distortion) and camera matrix formed by focal length, pixel scale factors, principal point and skewness between the axis. In order to retrieve depth information about the environment, two cameras are often rigidly connected to form a stereo vision system. Besides the intrinsic calibration of individual cameras, high precision of extrinsic calibration between the cameras is crucial for successful stereo reconstruction which will be described in the next section.

Principle of Stereo Vision

Assume the simplified configuration of two parallel looking 1D cameras with identical camera parameters as shown in Fig. 1.1. Furthermore, the basis, i.e. the straight line connecting the two optical centres of the two cameras, is assumed to coincide with the x-axis of the first camera.

Then, the image plane coordinates of the projections of the point $P(x^c, z^c)$ into the two images can be expressed,

$$u_1 = f \frac{x^c}{z^c} \tag{2.1}$$
$$u_2 = f \frac{x^c - b}{z^c} \tag{2.2}$$

where f is the focal length and b the length of the basis.

The pair of image points that results from the projection of one object point into the two images is often referred to as conjugate points or homologous points. The difference between the two image locations of the conjugate points is called the disparity *d*.

$$d = u_2 - u_1 = -f \frac{b}{z^c} \tag{2.3}$$

Given the camera parameters and the image coordinates of two conjugate points, the z^c coordinate of the corresponding object point *P*, i.e. its distance from the stereo camera system, can be computed by
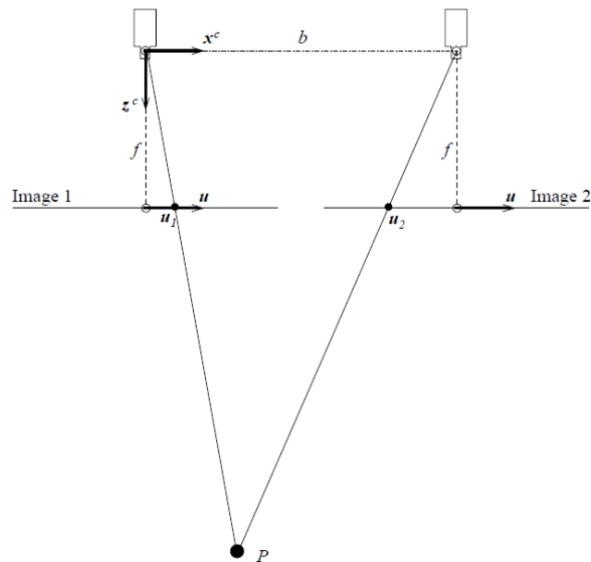
$$z_c = -f \frac{b}{d} \tag{2.4}$$

Figure 1.1 Vertical section of a binocular stereo camera system

The internal camera parameters of both cameras and the relative pose of the second camera in relation to the first camera (external camera parameter) are necessary to determine the distance of *P* from the stereo camera system. Thus, the tasks to be solved for stereo vision are:
1. To determine the internal and external camera parameters and
2. To determine conjugate points.

The first task is solved by the calibration of the stereo camera system. The second task is the so-called stereo matching process. The multi-view surface reconstruction extends the basic stereo vision principle to more than one image pair.

The way the cameras are placed influences the accuracy of the results that is achievable with the stereo camera system. The distance resolution Δz, i.e. the accuracy with which the distance z of the object surface from the stereo camera system can be determined, can be expressed by:

$$\Delta z = -\Delta d . \frac{z^2}{f . b} \qquad\qquad (2.5)$$

To achieve a high distance resolution, the setup should be chosen such that the length *b* of the basis as well as the focal length *f* are large, and that the stereo camera system is placed as close as possible to the object. In addition, the distance resolution depends directly on the accuracy Δd with which the disparities can be determined.

2. Stereo Camera System

Using the images of a stereo camera, a 3D reconstruction of the real environment can be calculated, enabling a machine to see three-dimensionally. To calculate a 3D reconstruction of a scene, precise geometric dimensions of the stereo camera, called stereo camera parameters, are required. Examples of stereo camera parameters are the focal length, the pixel size, the lens distortion, and the position and orientation of the cameras. The stereo camera parameters are nearly impossible to survey with conventional measuring tools because the cameras would have to be dismantled, the required accuracy is difficult to reach, and the effort would be immense. The solution is algorithms that can calculate the stereo camera parameters from the images of the same stereo camera. These algorithms are called stereo camera calibration algorithms.

A technique using defined calibration objects like a printed chessboard pattern has been described in [3-5]. Calibration targets are frequently used due to numerous advantages. They simplify the correspondence registration step since the number and type of correspondences is known in advance which virtually eliminates the problems associated with outliers. Additionally, target-based methods can use a prior knowledge about the target which can enhance the calibration results. Therefore, target-based methods are generally more precise than the targetless. Properties of a well-designed target are (i) ease of detection and (ii) high localisation accuracy for all the sensors in the calibration. The former ensures the success of the correspondence registration, while the latter has great influence on the quality of the results given by the optimisation step. The calibration targets supplied by some of the major car manufacturers are shown in Fig. 2.2. The calibration method using planar target is flexible and convenient, with high calibration accuracy. It has been widely applied in camera calibration, but it is unsuited to synchronously complete the calibration of multiple cameras at different positions. Use of a spherical target allows a complete outer contour of the spherical target at any angle. Such a technique is suitable for the synchronous calibration of multiple cameras at different positions, but its calibration accuracy can hardly meet high-accuracy measurement requirements. Moreover, a certain requirement must be met in the location of the spherical target: if the location is too close to the centre of the camera, the calibration will be inaccurate. However, if it is too far, the contour acquisition and elliptic fitting may be subject to camera distortion, thereby affecting the calibration accuracy. A grid spherical target (GST), as shown in Fig. 2.3(a) which combines the strengths of the spherical and planar targets allows multiple cameras to shoot simultaneously and calibrate their respective intrinsic parameters, while achieving the same calibration accuracy as that using planar target [43]. During the calibration, the linear solutions of the intrinsic and extrinsic parameters are obtained from the elliptic curves of the latitude and longitude circles in the image. Based on the coordinates of the intersecting points of the elliptic curves of the longitude and latitude lines in the image and the lens distortion, the nonlinear optimization method is adopted to obtain the optimum solutions of the intrinsic and extrinsic parameters.
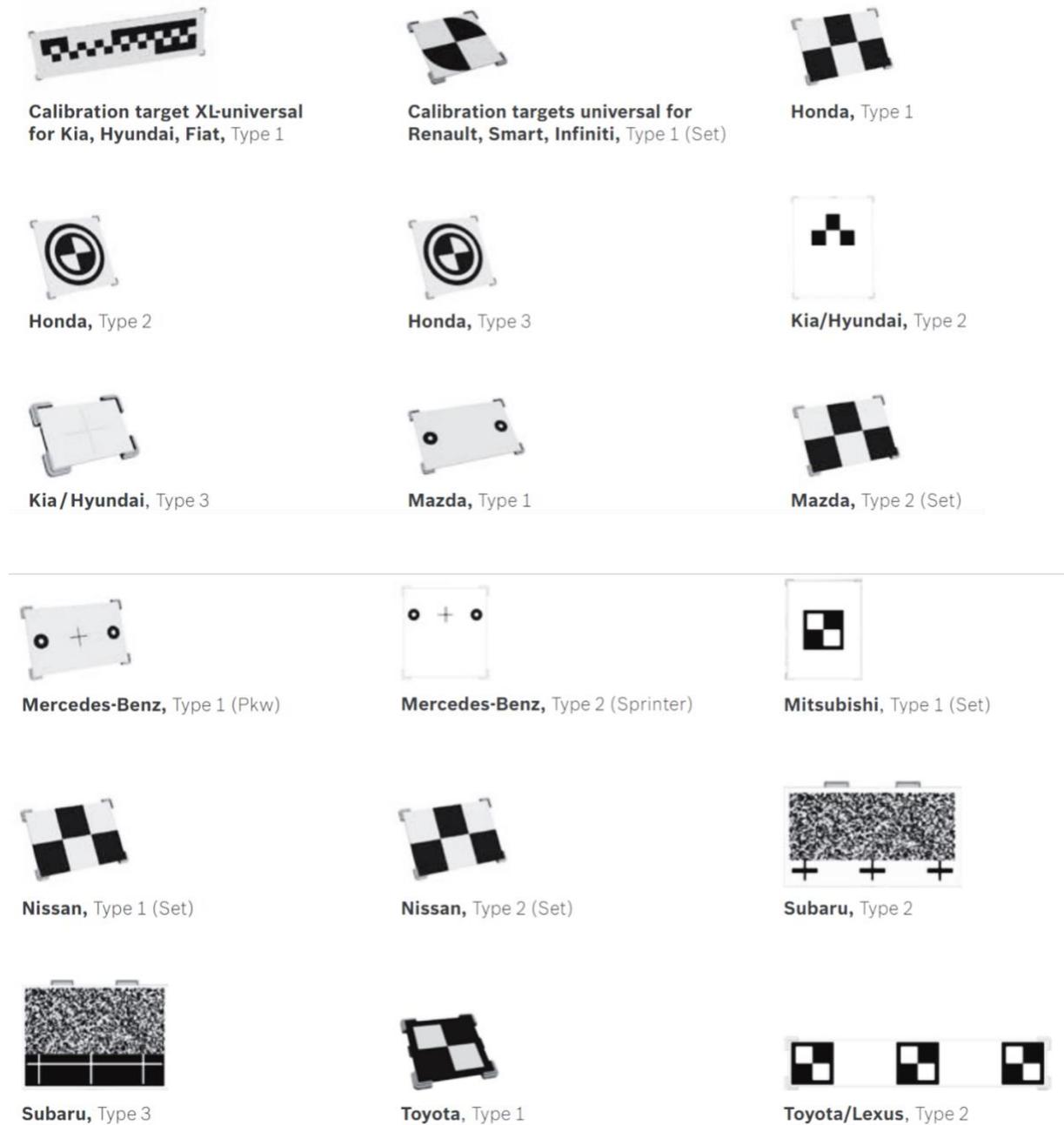
Figure 2.2 Calibration targets supplied by some car manufacturers

Another approach is the calculation of stereo camera parameters without defined calibration objects in real-world environments. [6] and [7] use epipolar geometry and nonlinear optimisation or a Kalman filter for continuous recalibration. In [8] the detection of vanishing points in the image of a mono camera is used to calibrate a roadside camera. The algorithm from [9] calculates the radial lens distortion from single images via vanishing points. In [10] bundle adjustment is applied, where the camera parameters are calculated from an image sequence containing many images of a moving stereo camera. The reprojection error and an iterated extended Kalman filter is used in [11] for continuously calibrating a stereo camera from image sequences. The algorithm in [12] uses a combination of epipolar geometry, the reprojection error, and the trilinear constraint with an iterated extended Kalman filter for continuous calibration. In [13], the three algorithms using epipolar geometry, relative world point error, and absolute world point error were proposed. These allowed the calibration of a stereo camera with only the help of a measuring tape and the data sheet of the cameras.
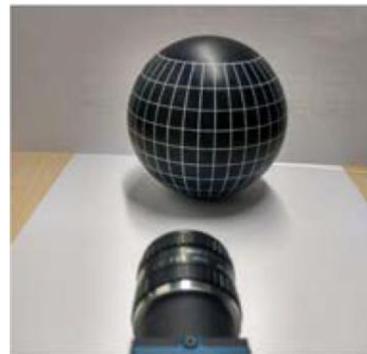
(a)   (b)

Figure 2.3 Camera calibration targets [43] (a) Planar checkerboard target; (b) Grid spherical target [43]

For autonomous cars, the calibration process can also be performed by means of road marks [28], such as lines [29-31] or dashed lines on the roadway [32], it being possible to use parking lines as the calibration pattern [33]. These methods allow the calibration process, where it is possible to recalculate the extrinsic parameters at different times and positions. The last group of methods is based on estimating the geometry of the roadway in front of the vehicle, which can be accomplished mainly in two different ways. Firstly, the three-dimensional information of the vehicle environment contained in the disparity map allows one to determine the position of the ground in front of the vehicle by means of different kinds of projections. While a second technique is based on the sampling of 3D points and subsequent adjustment to a plane [34], where both techniques can be used in outdoor applications [35,36]. Such methods allow one to find out the extrinsic parameters, avoiding the need for the calibration pattern or the road marks. Moreover, this allows recalculating the relative position of the vision system in real time while the vehicle is moving and adapting to changing parameters, as discussed above, such as vehicle load, acceleration or irregularities of the roadway.

An optimal calibration technique should produce unbiased and minimum variance estimates of the camera parameters. In practice, this is quite difficult to achieve due to different error sources affecting the imaging process. Calibration techniques using a calibration object yields the least calibration error and it is recommended to use this approach in use cases where it is possible to do so. The calibration procedure in [14] utilizes circular control points and performs mapping from world coordinates into image coordinates and backward from image coordinates to lines of sight or 3-D plane coordinates using a calibration object. This technique is used to demonstrate the test case in section 2.4. The camera model used allows least squares optimization with the distorted image coordinates.

For the calibration of the stereo camera system, each camera acquires multiple images of one or more calibration objects in different poses. It is not necessary that the calibration object is always visible in all poses for each camera. The only requirement is that the cameras can be "connected" in a chain by the calibration object poses as shown in Fig. 2.4. The camera setup between the acquisition of the calibration images and the acquisition of the stereo images of the object to be investigated should not be changed to preserve the calibration. Therefore, the cameras should be mounted on a stable platform. By taking multiple images of a calibration plate placed arbitrarily throughout the camera field of view, an explicit and precise 3D coordinate mapping of the field of view can be calculated and complex distortions can be corrected [14]. The internal and external camera parameters obtained from the calibration procedure map the image coordinates from the camera to real world coordinates.
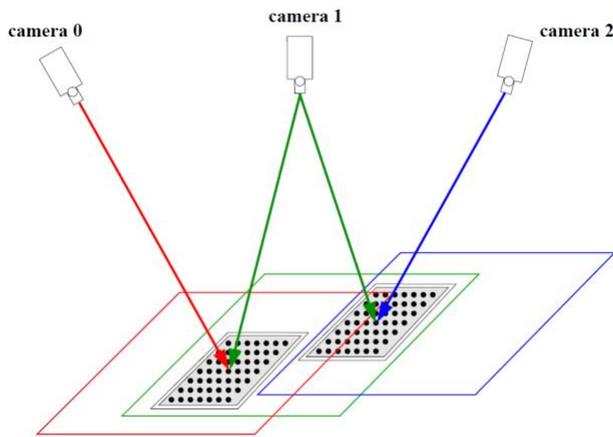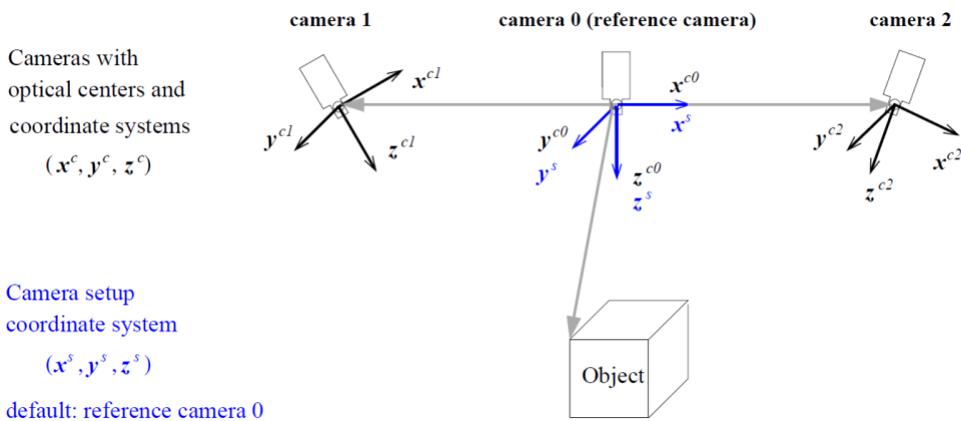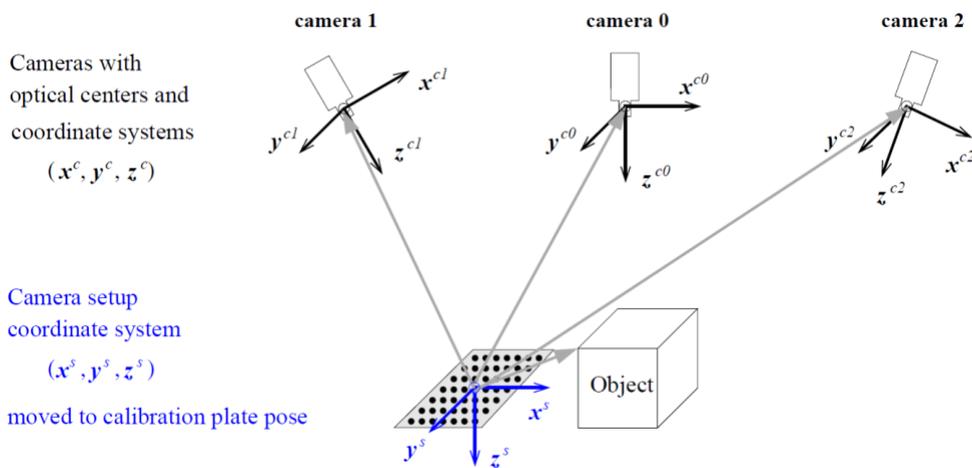
Figure 2.4 Cameras are connected in a chain

(a)

(b)

Figure 2.5 Coordinate systems of a multi-view camera setup: (a) default setup coordinate system located in the reference camera 0; (b) setup coordinate system moved to the coordinate system of the calibration plate

The camera calibration information is stored in a camera setup model, which contains the internal camera parameters as well as the relative poses between the cameras. The default coordinate system of the stereo camera setup is identical to the coordinate system of the so-called reference camera of the setup, which is typically the camera with the index 0 (as shown in Fig. 2.5(a)). The poses of the other cameras and the reconstructed coordinates are computed relative to this camera. It is desirable

to specify the pose of the desired setup coordinate system as marked by the calibration plate in Fig. 2.5(b) by moving the set-up coordinate system. This also allows a user to introduce a bounding box with respect to the calibration plate which should be tight around the volume of interest. The surface fragments lying outside the bounding box are clipped and are not returned in the final surface reconstruction. A too large bounding box results in a large difference between the minimum and maximum disparity and it would slow down the execution of the algorithm. A too small bounding box may result in clipping valid surface areas.

Reconstruction of the 3D Surface

The underlying surface reconstruction algorithm is based on fusing the information obtained from multiple binocular stereo depth maps. The objects in the scene should expose certain surface properties in order to make the scene suitable for the dense surface reconstruction. First, the surface reflectance should exhibit Lambertian properties as closely as possible (i.e. light falling on the surface is scattered such that its apparent brightness is the same regardless of the angle of view). Secondly, the surface should exhibit enough texture, but no repeating patterns.

The proper selection of image pairs has an important role for the general quality of the surface reconstruction. On the one hand, camera pairs with small base line (small distance between the camera centres) are better suited for the binocular stereo disparity algorithms. Hence, pairs with small base lines should be preferred for acquiring accurate depth information of the scene. On the other hand, the pairs should provide different points of view, such that if one pair does not see a certain area of the surface, it is covered by another pair. The number of image pairs linearly affects the runtime of the reconstruction algorithm. Therefore, use an optimum number of image pairs in order to handle the trade-off between completeness of the surface reconstruction and reconstruction runtime. Finally, the images pairs should represent a static scene, or they must be taken simultaneously to accurately reconstruct the surface.

The algorithm uses a multigrid stereo matching based disparity operator for each image pair from a predefined list of image pairs to calculate disparity and distance values for image parts that contain no texture (as long as these parts are surrounded by significant structures between which an interpolation of values is possible). The most important advantages of multigrid stereo is that it interpolates 3D information for areas without texture based on the surrounding areas and in particular for edges, the accuracy in general is higher than for correlation-based stereo, and the resolution is higher than for correlation-based stereo, i.e. smaller objects can be reconstructed. The disparity information is converted to X, Y, and Z images coordinate in the coordinate system of the respective camera. Then, for each 3D point, the normal vector is calculated. In the next step, the X, Y, and Z images as well as the normal vectors are transformed into the common coordinate system that is specified in the camera setup model. Finally, the transformed coordinate images are sub-sampled and stored in a common point cloud together with the points and normals extracted from other image pairs. The so-obtained point cloud can be additionally meshed in a post-processing step based on a Poisson solver. It creates a water-tight mesh, therefore surface regions with missing data are covered by an interpolated mesh [15]. The point cloud obtained as described above can be processed further to obtain a preferably smooth surface while keeping form fidelity. To this end, the bounding box is sampled, and each sample point is assigned a distance to a so-called iso-surface (consisting of points with distance 0). The final distance values (and thus the iso-surface) are obtained by minimizing an error function based on the points resulting from pairwise reconstruction. This leads to a fusion of the reconstructed point clouds of all camera pairs [16].

The processed point cloud is noisy as multi-view stereo methods are much more susceptible to produce noisy depth estimates due to image imperfections, triangulation inaccuracy, depth quantisation, as well as outliers due to matching ambiguities and non-diffuse surfaces. Most multi-view stereo methods refine the reconstructed depth maps, and often this is integrated into the depth

estimation stage and formulated as a (global) optimization problem [17,18]. Furukawa et al. [19] use a filter based on quality and visibility measures for merging points while handling errors and variations in reconstruction quality. The denoising step can also be implemented as a post-processing step for the 3D point cloud. Two popular common outlier removal approaches are the statistical and geometric-based filters [20]. The Statistical Outlier Removal (SOR) filter firstly computes the average distance of each point to its neighbours through k-nearest neighbours searching function. A point is considered an outlier if this distance is larger than the average distance derived from all points in the data set plus t times of the standard deviation (σ) of the average distance. Thus, the outlier removal was controlled by two threshold k and t. The geometric-based filter considers the distance from a given point to the object's surface. The algorithm locally fits a plane through each point in the data set, which was based on neighbour points of the given point extracted by either kNN search or a range search method.

Multi-view Stereo Imaging: Precision Agriculture Scenario

Multi-view stereo imaging was utilised to capture and reconstruct an indoor and outdoor based wheat plot to demonstrate its application in wheat phenotyping for precision agriculture. The system was chosen as the trade-off technology when compared with LiDAR and Structured laser light scanner in terms of cost, resolution and capture time.
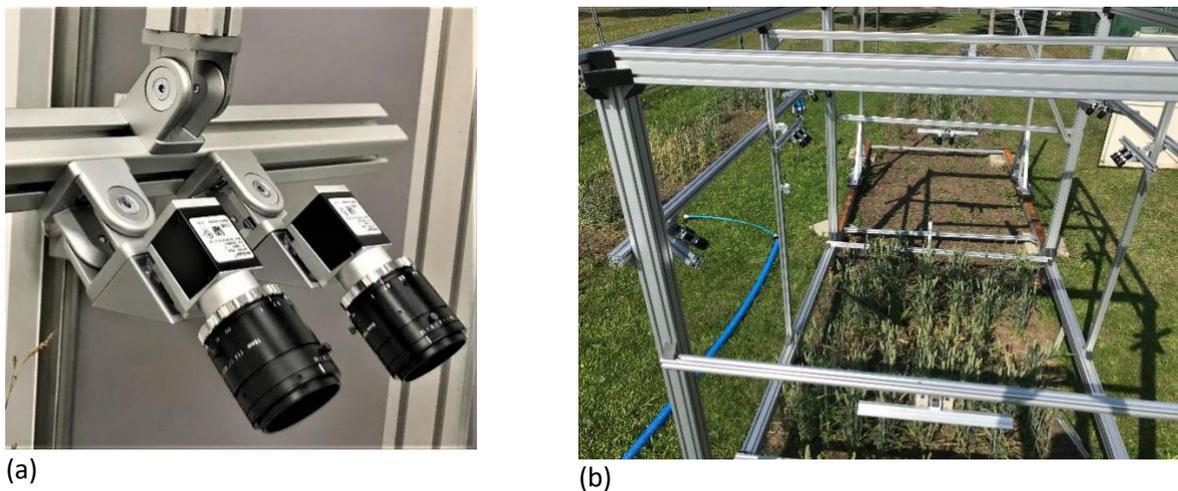

(a)   (b)

Figure 2.6 Mobile 18-camera multi-stereo imaging setup (a) One of the camera pair; (b) Entire set up for imaging the outdoor wheat plot

A NPL in-house built 18-camera multi-view stereo system utilising 9 camera pairs was used to scan the indoor wheat heads (1m x 1m) in the NPL Agritech laboratory and outdoor wheat crops (2m x 2m) in the NPL outdoor Agri-plot as shown in Fig. 2.6(a) and 2.6(b). Basler cameras (8x 2.3MP and 4x 12MP) were used to capture the images of the wheat plant as shown in Fig. 2.7(a) and 2.7(b). Prior to mounting on to the frame, the cameras' internal parameters were calculated using the calibration routine in HALCON [21]. The routine involves the use of a 32 cm × 24 cm calibration plate consisting of 5 mm white circles arranged hexagonally on a black background, with the centre to centre separation of each circle being 1 cm [22]. Images of the calibration plate placed in a variety of positions and orientations are taken. By comparing the size and position of the dots seen across in the camera's field of view from what is expected the software calculates the: focal length, radial lens distortion parameters, decentring lens distortions, the centre of the radial lens distortion, and sensor's physical dimension. A different larger calibration plate was used to calculate the external camera parameters of position and orientation of all the 12 cameras. The cameras were positioned and mounted on the mobile phenotyping platform so that all the 12 cameras were able to see the wheat plant and the calibration plate. With an exposure time of 5 ms, images from the 12 cameras were captured using a routine written in LABVIEW. The captured images were processed using the multigrid stereo matching

algorithm implemented in HALCON to produce a point cloud as shown in Fig. 2.7(c) and 2.7(d).
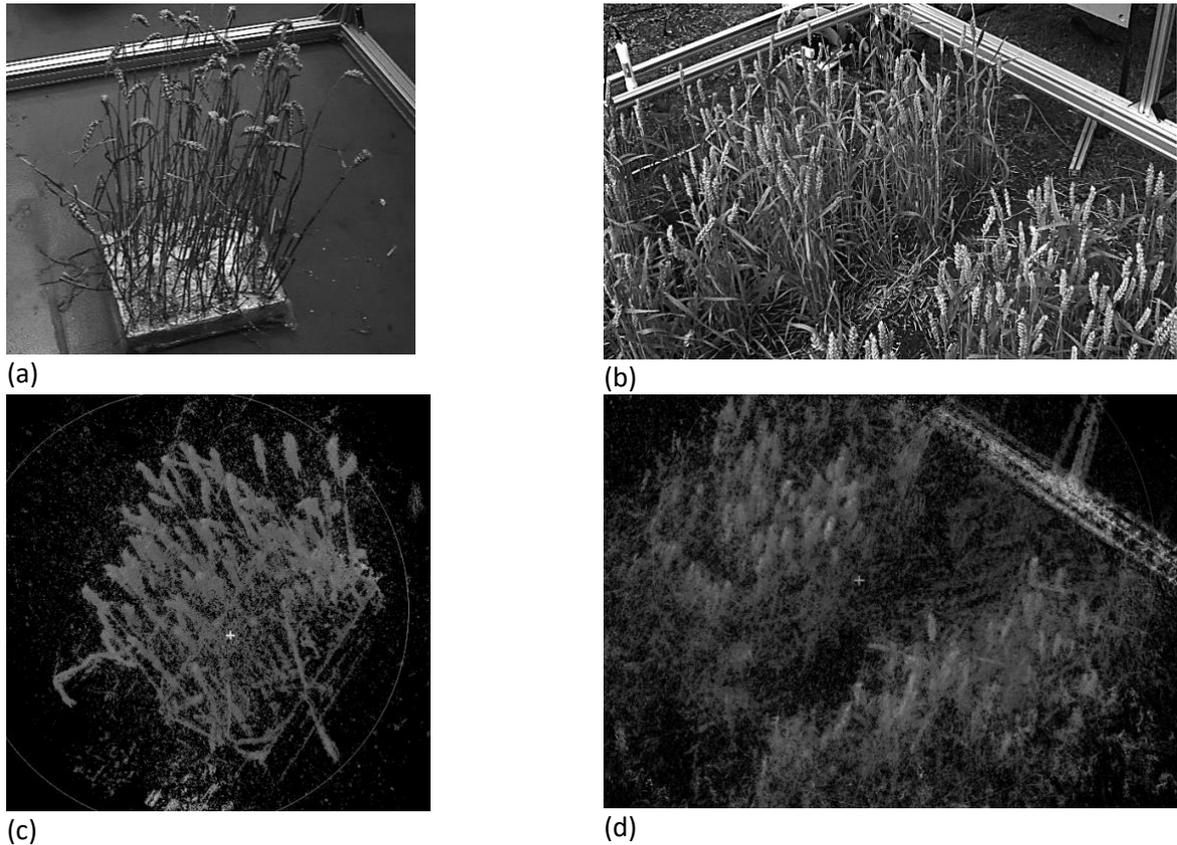

(a) (b)

(c) (d)

Figure 2.7 Images of (a) Indoor wheat heads; (b) Outdoor wheat crops and reconstructed raw 3D point cloud (c) Indoor wheat heads; (d) Outdoor wheat crops

The point clouds also contain significant noise and outliers which could be reduced by simple denoising techniques Statistical Outlier Removal filter which is implemented in CloudCompare software [23].

The calibration techniques and the 3D point cloud reconstruction methods demonstrated in the precision agriculture scenario can also be extended to an autonomous vehicle application with ease. Instead of multi-view stereo imaging, it would be a binocular stereo imaging involving a single camera pair.

3. LiDAR

Lidar (LIght Detection And Ranging) or Ladar (LAser Detection And Ranging), are optical equivalents to radar, using pulses of visible or near-infrared light to detect and measure distances to them instead of microwaves or radio waves. Lidars measure distances to objects by emitting pulses of light from a laser toward the object and measuring the time-of-flight, t, taken for the pulse to travel between the laser, the object's surface, and the sensor that detects the reflected pulse. The distance, d, is then simply calculated by $\frac{1}{2} \times t \times c$, where c is the speed of light and factor of a half is required to account for the laser pulse traveling the distance twice (Fig. 3.1).
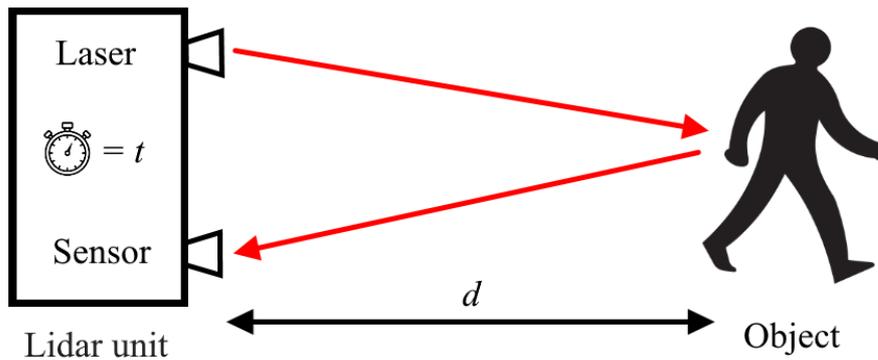
Figure 3.1: Time of flight of a laser pulse emitted from lidar, reflecting off an object, and being detected by Lidar's sensor

In contrast to cameras, lidars are an active imaging system that use their own light source rather than relying on ambient lighting, allowing them to function in low light conditions. To produce a 3D image (point cloud) reconstructing the lidar's environment, the surrounding surfaces are scanned by the laser beam either by moving the laser and sensor or by moving a mirror that reflects the laser beam. Common implementations of movement in lidars used for autonomous vehicle and robotic applications are swivelling, which allows a limited field of view to be scanned, and rotating, which allows for a full 360° scan (Oberländer et al., 2015; Sheehan et al., 2012). Building up a point cloud using a single laser beam is slow. Data acquisition and scan rates can be increased by using multibeam lidars i.e. lidars that use more than one laser. For example, the single laser Ocular Robotics RE-05 (Ocular Robotics, n.d.) can achieve a sample rate of 30 thousand points per second whilst the 64-laser Velodyne HDL-64E (Velodyne Lidar, n.d.) can reach rates up to 2.3 million points per second.

As multibeam lidars have become more widely used for applications such as autonomous vehicles, robotics, and surveying, the last decade has seen an increase in publications related to multibeam lidar calibration. The challenge is unique due to the requirement of scaling the procedure to the many lasers used in a simple and time efficient manner (Levinson and Thrun, 2014).
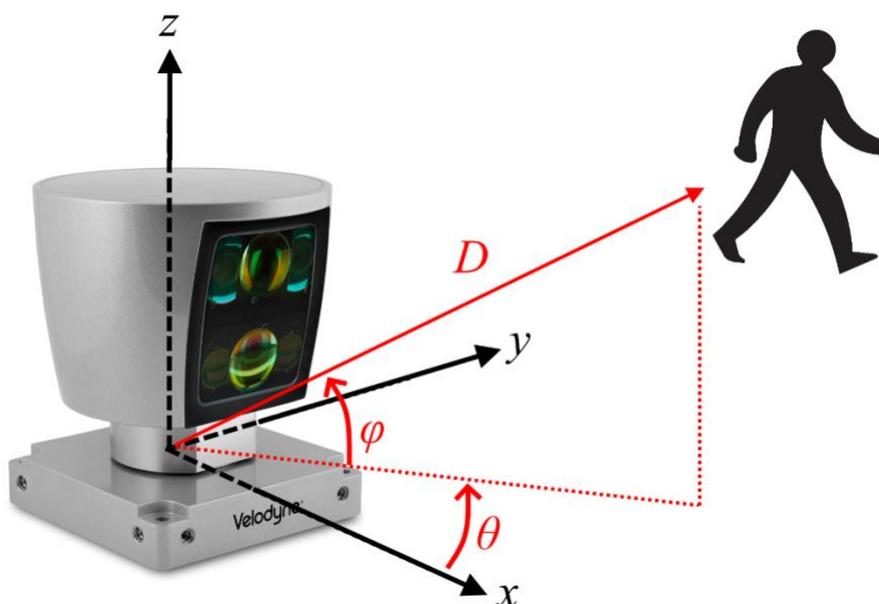
Intrinsic Calibration

Figure 3.2: Description of spherical coordinates

Various implementations for intrinsic calibration have been given for rotating multibeam lidars in the literature (Atanacio-Jiménez et al., 2011; Bergelt et al., 2017; Chen and Chien, 2012; Levinson and Thrun, 2014; Muhammad and Lacroix, 2010; Sheehan et al., 2012). At its simplest the implementations involve scanning either calibration targets or the surrounding environment itself, and then optimising the lidar's intrinsic parameters so that the resulting point cloud agrees well with the ground truth. Variations between the works arise from the choice of parameters to calibrate, method used to locate any calibration targets, the optimisation routine, and whether the lidar is stationary or moving.

Swerving and rotating lidars take measurements in spherical coordinates: D_{ToF}, the time-of-flight distance between the lidar and object, ϑ, the rotation angle of the lidar when the measurement was taken, and φ, the elevation angle of the laser from the horizon (Fig 3.2). The conversion of the spherical (D_{ToF}, ϑ, φ) coordinates into cartesian (D_x, D_y, D_z) coordinates requires the intrinsic parameters be considered. In some of the earliest attempts to calibrate a stationary rotating multibeam lidar, (Atanacio-Jiménez et al., 2011; Chen and Chien, 2012; Muhammad and Lacroix, 2010) chose five intrinsic parameters (Fig. 3.3) to characterise each of the 64 lasers in the Velodyne HDL-64 lidar:

- Distance correction, D_{corr}: Correction to the measured time-of-flight distance, D_{ToF}, due to timing delays in the lidar's electronics and data processing
- Rotational correction, ϑ_{corr}: Correction to rotation angle, ϑ, due to manufacturing differences in the lasers' positions
- Elevation correction, φ_{corr}: Correction to elevation angle, φ, due to manufacturing differences in lasers' positions
- Vertical offset, V_O: Vertical displacement of lasers from lidar's origin, typically near the base
- Horizontal offset, H_O: Horizontal displacement of lasers from lidar's origin, typically the rotational axis
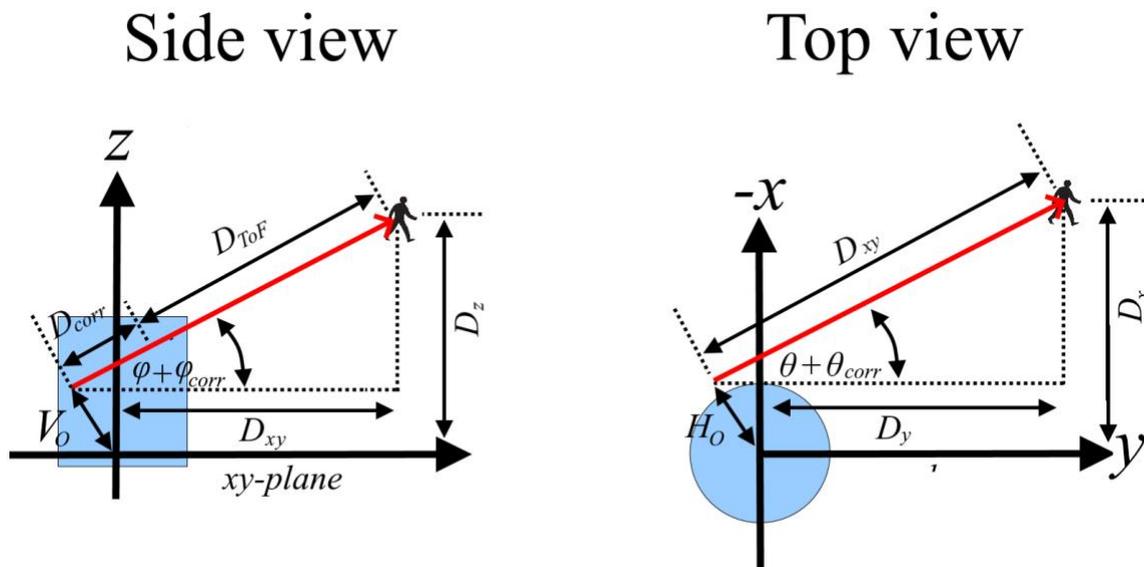

Figure 3.3 Intrinsic parameters for a single laser in a rotating multibeam lidar. Diagram modified from (Muhammad and Lacroix, 2010)

The conversion equations then take the following form:

$$D = D_{ToF} + D_{corr}$$
$$D_{xy} = D \cdot \cos(\varphi + \varphi_{corr}) - V_O \cdot \sin(\varphi + \varphi_{corr})$$
$$D_x = D_{xy} \cdot \sin(\theta + \theta_{corr}) - H_O \cdot \cos(\theta + \theta_{corr})$$

$$D_y = D_{xy} \cdot \cos(\theta + \theta_{corr}) + H_O \cdot \sin(\theta + \theta_{corr})$$
$$D_z = D \cdot \sin(\varphi + \varphi_{corr}) + V_O \cdot \cos(\varphi + \varphi_{corr})$$

According to (Muhammad and Lacroix, 2010), the choice of calibration environment and target should be based on the type of lidar system being used, the parameters being calibrated and the practicalities involved in performing the calibration. Planar surfaces such as the walls in the lidar's surrounding environment (Bergelt et al., 2017; Chen and Chien, 2012; Muhammad and Lacroix, 2010) or in a specially setup space (Atanacio-Jiménez et al., 2011) may be used as a calibration target. The existing walls of a building at a test site are advantageous due to the minimal setup required and their large surface area provide many data points in the point cloud for target detection and the calibration process. When using planar surfaces, scanning just the surfaces allows for the calibration of the laser's orientation only, but by including straight edges, the laser's position can be calibrated too. Alternatively, planar boards can be used as can poles with reflective markers (Gao and Spletzer, 2010; Xue et al., 2019), although these options appear most often for extrinsic calibration which will be covered later. Poles have the advantage of portability, but there is a sacrifice in the amount of data captured due to the pole's smaller surface area.

Several calibration targets should to be recorded (Fig. 3.4), ideally located at a range of positions and distances to ensure the calibration is independent of any position and distance related biases the lidar may have (Bergelt et al., 2017; Muhammad and Lacroix, 2010). Additionally, it is noted by (Chen and Chien, 2012) that a number of scans should be carried out to remove temporal noise.
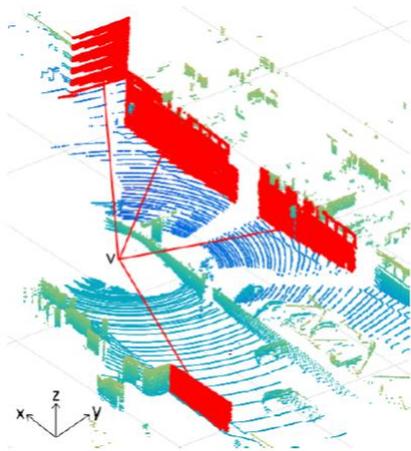

Figure 3.4 Four walls (coloured red) used as planar calibration targets in (Bergelt et al., 2017). The location of the lidar is marked by the letter "V"

Once the lidar's environment has been scanned, the targets are segmented and an optimisation on the intrinsic parameters can be performed to produce a best fit to the targets' surfaces. The cost function used to quantify the "goodness" of the fit is based on the mean distances of the points to the targets' surfaces (Bergelt et al., 2017; Chen and Chien, 2012; Muhammad and Lacroix, 2010).

Targetless procedures have also been reported (Levinson and Thrun, 2014; Oberländer et al., 2015; Sheehan et al., 2012) and are able to use any static planar or freeform surface situated in the lidar's environment. These will be explained in more detail in the Extrinsic Calibration subsection, but in lieu of fitting sections of the point cloud to specific calibration targets, these works attempt to maximise the, as (Sheehan et al., 2012) terms it, "crispness" of the whole point cloud (Fig 3.5). They do this by minimising cost functions based on the Rényi Quadratic Entropy (Oberländer et al., 2015; Sheehan et al., 2012) or an "energy function" (Levinson and Thrun, 2014) similar to one used in Iterative Closest Point (ICP) algorithms that match overlapping point clouds together. These targetless procedures are

widely used for extrinsic calibration, although some works such as (Levinson and Thrun, 2014; Sheehan et al., 2012) apply them to intrinsic calibration.
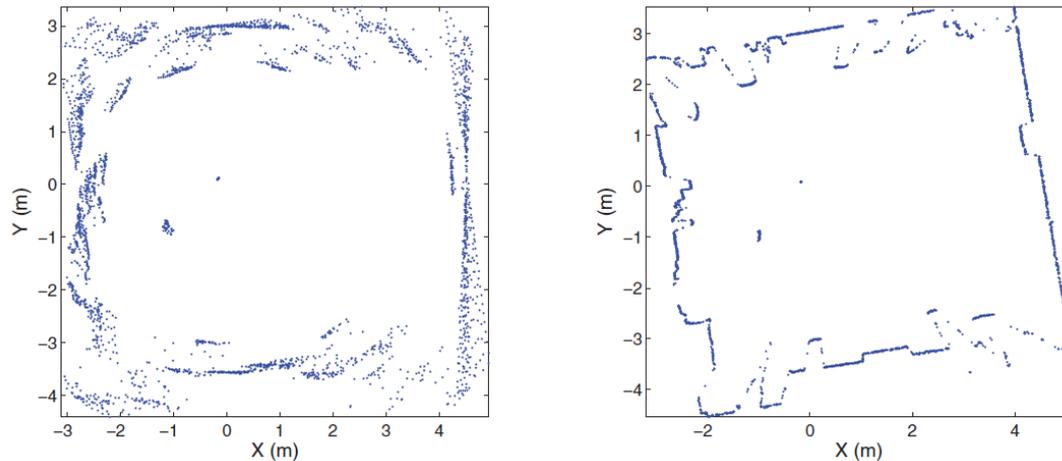

Figure 3.5: A lidar scan of a room (left) before and (right) after intrinsic parameter calibration was carried out by (Sheehan et al., 2012) to maximise the point clouds "crispness"

By carrying out their intrinsic parameter calibration, users can expect to obtain improvements in measurement accuracy over using the factory provided calibration. For example (Muhammad and Lacroix, 2010) observed a 36% reduction in the standard deviation of measurements to a target 18 m away from 2.2 cm to 1.4 cm. Similarly, (Chen and Chien, 2012) (Fig. 3.6) saw a reduction in the standard deviation of 41% from 2.37 cm to 1.39 cm. And (Bergelt et al., 2017) were able to reduce the standard deviation by 38% from 4.0 cm to 2.5 cm.
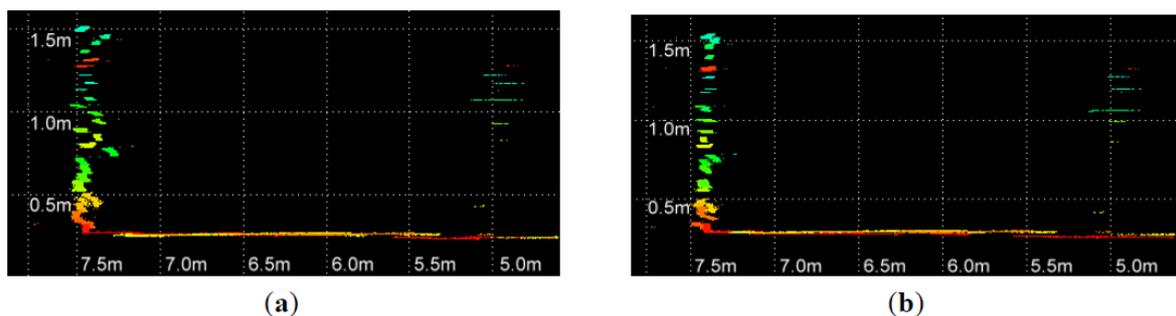

Figure 3.6: A lidar scan of a flat wall (a) using factory calibration parameters and (b) after re-calibration. A reduction in the variation of distances can be seem. Image taken from (Chen and Chien, 2012)

Extrinsic Calibration

Procedures for extrinsic calibration can be based on those previously mentioned for intrinsic calibration i.e. making use of planar surfaces (e.g. walls and ground) as targets (Atanacio-Jiménez et al., 2011; Jiao et al., 2019a; Li et al., 2019) and optimising the relevant parameters to obtain the best fit. These procedures are extended to calibrate the positions of multiple lidar systems to allow accurate merging of multiple point clouds into one, and accurate positioning of the point cloud in a vehicle's frame of reference. Differences in the procedures result from whether the lidar is required to be static or be in motion, and whether extra data is required from external sensors e.g. inertial measurement units (IMUs) or GPS.

The target-based procedure proposed by (Li et al., 2019) for a single lidar implemented a two-step approach to transforming the point cloud's frame of reference from the lidar to the vehicle (Fig. 3.7).

The only targets required are a flat horizontal ground and one vertical planar surface. In the first step, a RANSAC based fitting algorithm is used to locate the ground plane in the point cloud. The plane is used to determine the height of the lidar above the ground and the lidar's rotational orientation i.e. pitch, roll and yaw with respect to the ground. The data is used in translation and rotation matrices to relocate the point cloud's reference frame from the lidar to the ground. In the second step, the fitting algorithm is applied to the vertical surface to determine the rotation matrix to orient the point cloud from the ground to the vehicle. The translation matrix to move the point cloud from the ground to the vehicle's frame of reference is determined by manual measurement.
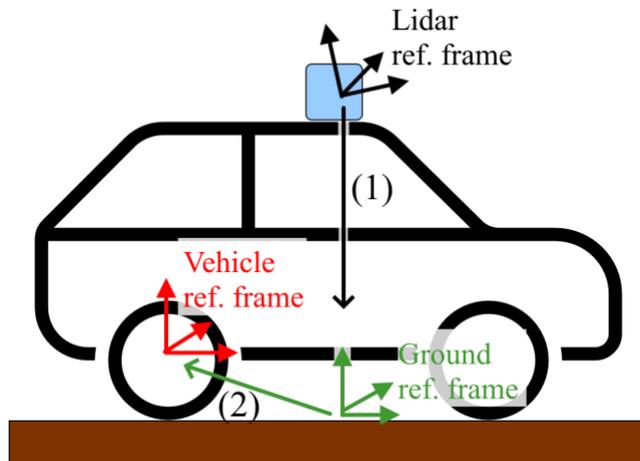

Figure 3.7: Two step procedure taken by (Li et al., 2019) to transform the point cloud's reference frame from the lidar to the vehicle

For calibrating multiple lidar systems, a selection of target-based procedures include (Gao and Spletzer, 2010; Heide et al., 2018; Jiao et al., 2019a, 2019a; Xue et al., 2019). (Jiao et al., 2019a) has a strict requirement of needing "three linearly independent planar surfaces" which in practice can be a wall corner. Demonstrating with a dual-lidar setup, the two walls and the ground are scanned and a RANSAC based algorithm is used to find the three planes. The authors then perform an optimisation to minimise the distance between corresponding planes in each lidar's point cloud to obtain the relative translational and rotational angle between the two lidars. By applying the procedure, the authors could recover the extrinsic parameters of a dual-lidar system with rotation errors less than 2.9° and translational errors less than 10 cm. In a similar manner, (Xue et al., 2019) showcase an extrinsic parameter procedure for two lidars using two poles stickered with retroreflective tape. The retroreflective tape allows the poles to be easily segmented in the point cloud, from which straight lines are fitted to the poles. The relative transformation data between the two lidars is then found by minimising the difference between the lines in both lidars.

The approach taken by (Heide et al., 2018) allows more than two lidars to be calibrated. Using initial estimates of each lidars' extrinsic parameters, allows all the point clouds to be referenced to the same origin. Each point cloud is cleaned by removing outlier points, a pair of point clouds is chosen, and a Generalized Iterative Closest Point (GICP) algorithm (Segal et al., 2009) is applied to register matching features before the point clouds are merged. The merged point cloud is then paired with another point cloud and the process repeated until all lidars' point clouds have been merged together. The order of pairing and merging is chosen to produce the largest overlapping field of view in the final point cloud. (Heide et al., 2018) state that at least two or three planar surfaces should be within each overlapping field of view to obtain a good calibration. When evaluating the procedure with a setup involving three lidars and a commercial digger, the authors found the procedure capable of recovering the distance between two of the lidars to within 3 cm of the ground truth. The final step, relocating the reference frame to that of the vehicle, is achieved by applying GICP to merge the merged point cloud with a point cloud of the vehicle.

Targetless methods for single and multilidar setups have been reported by (Jiao et al., 2019b; Levinson and Thrun, 2014; Oberländer et al., 2015). All three works require the lidar(s) to be moved whilst the lidars are scanning as part of the process for determining the extrinsic parameters. For a single lidar, (Levinson and Thrun, 2014)'s procedure utilises the knowledge that the points in a moving lidar's static surroundings will be scanned more than once and the readings of those points will not be randomly distributed in space. The lidar's relative position and motion data is tracked with either an IMU, GPS or wheel encoder, and the data used to provide a local reference frame for each lidar scan. Without a calibration target to fit to, the procedure defines an "energy function" that is calculated for the whole point cloud. The energy function sums up the positional differences of each point measured by a laser beam with respect to the neighbouring points measured by the same beam and neighbouring beams. The extrinsic parameters are then obtained by minimising the energy function. Using data from 15 seconds of movement, (Levinson and Thrun, 2014) were able to determine a lidar's translational and rotational orientation with respect to the vehicle to within 1 cm and 0.03° of the measured values. A downside to relying on external devices to measure motion is the possibility of measurement drift as seen by (Levinson and Thrun, 2014) which effects the accuracy of the calibration.

In contrast, (Jiao et al., 2019b) was able to calibrate mutlilidar setups using point cloud data alone, without additional positional data from an IMU etc. In their work, motion is initially estimated by comparing line and edge features in each successive point cloud using the LeGO-LOAM algorithm (Shan and Englot, 2018). Each lidars' extrinsic parameters are then calculated from their estimated motions. As an example (Fig 3.8), at time steps k-1 and k, two lidars, a and b, have a relative transformation between them denoted by a matrix T_b^a consisting of the relative translation and rotation. There also exists a transformation matrix relating the lidars' positions and orientations between the time steps, $T_{a_k}^{a_{k-1}}$, for lidar a and $T_{b_k}^{b_{k-1}}$, for lidar b. The extrinsic parameters can be determined by solving the equation:

$$T_{a_k}^{a_{k-1}} \mathrm{T}_{\mathrm{b}}^{\mathrm{a}} = \mathrm{T}_{\mathrm{b}}^{\mathrm{a}} T_{b_k}^{b_{k-1}}.$$
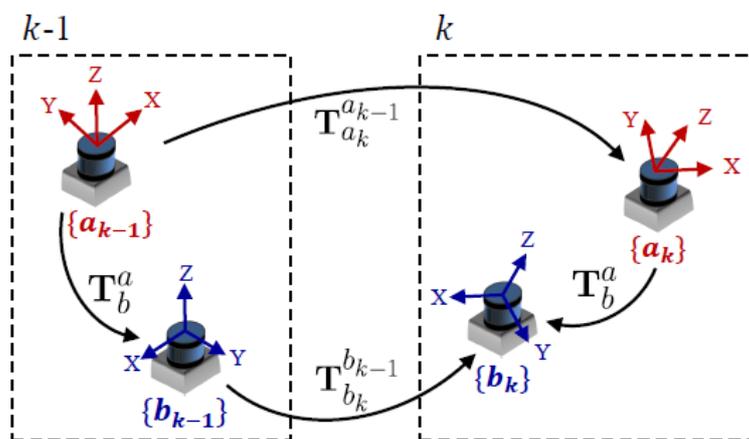

Figure 3.8: Diagram from (Jiao et al., 2019b) explaining the transformation between lidars *a* and *b*, and also the transformations they experience between time steps *k*-1 and *k*

The calculated parameters are refined further by applying constraints based on the features within the point cloud. RANSAC fitting of the ground plane allows the lidar's height to be determined, and ICP based registration of features between point clouds taken at different time steps is used to constrain the extrinsic parameters. Evaluating their procedure with a three-lidar setup, the (Jiao et al., 2019b) found the determined translational and rotational errors to be less than 1 cm and 2.3° respectively from the ground truth.

4. RADAR

Radars are active sensors which, similarly to the LiDAR, emit an electromagnetic signal and determine the range of objects in the vicinity based on the returned echo. Although being frequently used in automotive applications due to their low price and robustness, extrinsic radar calibration has not gained much research attention. The existing methods are all target-based since, for all practical means and purposes, the targetless methods are hardly feasible due to limited resolution of current automotive radar systems, as the radar is virtually unable to infer the structure of the detected object and extract features such as lines or corners. Current radars have no elevation resolution while the information about the detected objects they provide contains range, azimuth angle, radar cross section (RCS) and range-rate based on the Doppler effect. Although having no elevation resolution, radars have substantial elevation FOV which makes the extrinsic calibration challenging due to the uncertainty of the measurements.

Concerning automotive radars, common operating frequencies (24 GHz and 77 GHz) result with reliable detections of conductive objects, such as plates, cylinders and corner reflectors, which are then used in intrinsic and extrinsic calibration methods [37]. Wang et al. [38] used a metal panel as the target for radar-camera calibration. They assume that all radar measurements originate from a single ground plane, thereby neglecting the 3D nature of the problem. The calibration is found by optimising homography transformation between the ground and image plane. In contrast, Sugimoto et al. [39] considered the 3D nature of the problem. Therein, they manually search for detection intensity maximums by moving a corner reflector within the FOV. They assume that detections lie on the radar plane (zero elevation plane in the radar coordinate frame). Using these points, a homography transformation is optimised between the radar and the camera. The drawback of this method is that the maximum intensity search is prone to errors, since the return intensity depends on several factors, e.g., target orientation and radar antenna radiation pattern which is usually designed to be as constant as possible in the FOV.

RADAR KPIs and test bed definition
The following KPIs have been identified in order to characterise the Radar sensor for automotive applications:
A. **Range resolution:** Is the capability to discriminate two different targets in different ranges, it is inversely proportional to the transmitted bandwidth and provides a fundamental capability for the radar sensor.
    Test bed definition: In order to test and calibrate the range resolution, two targets spaced in range should be used. The range spacing of the two targets should be smaller than the nominal range resolution, in this case the test should demonstrate (by inspection of the range profile) that the radar is not able to discriminate the two targets in range. The test should be then performed with targets spaced in range of the nominal range resolution. In this case the test should demonstrate the capability to discriminate the two targets in range (by inspection of the range profile). In the case that it would not be possible to discriminate the targets in range, then additional tests should be made increasing the spacing in range between the targets every time by 0.5 times the nominal range resolution until it becomes possible to discriminate the two targets, the minimum distance between the two targets when these can be discriminated will define the actual range resolution of the sensor.
    Required equipment: Radar Echo Generation system (examples provided in reference Radar [1]). The operating principles of such a device are illustrated in [Figure 4.1](). The radar sensor is transmitting a signal and listens for echoes from possible targets in the area. The Radar Echo Generator receives the radar signal and generates an echo with shifted delay, frequency and amplitude compared to the original signal to simulate a target in different distance, speed and size (RCS) respectively.
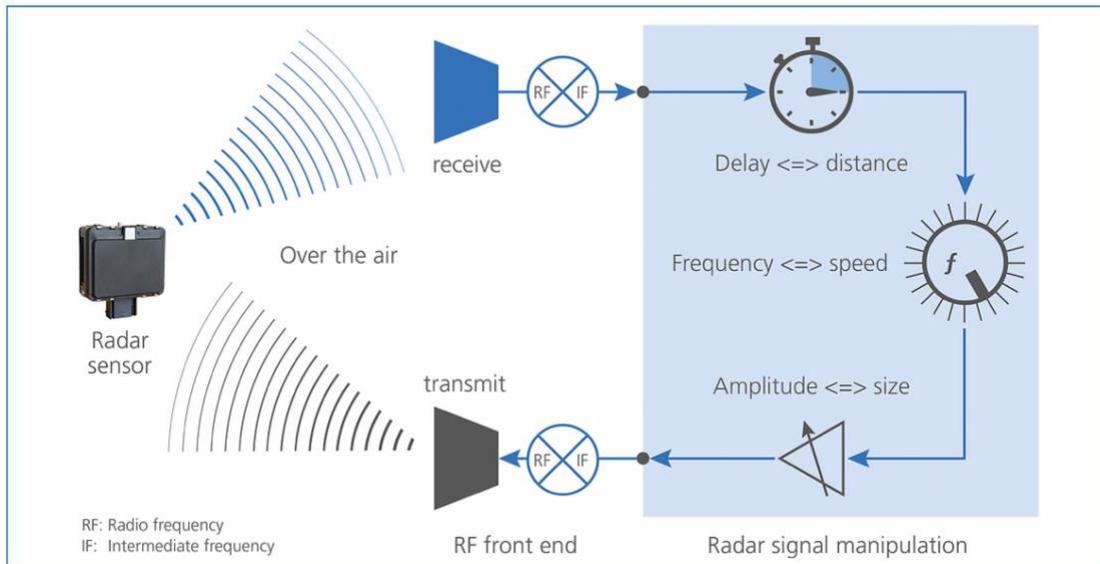
Figure 4.1 Operating principles of a Radar Echo Generator device (Courtesy dSpace [1])

B. **Angular resolution:** Is the capability to discriminate in angle the position of the target, in conjunction with the range position provides the localization of the target. It can be in both Azimuth and Elevation angles, even if most of the sensors currently available provide angular discrimination in Azimuth only, and is achievable through the use of multiple receiving channels.

Test bed definition: In order to test and calibrate the angular resolution, two targets closely spaced in angle from the sensor should be positioned in front of the sensor. The angular spacing of the two targets should be smaller than the nominal angular resolution of the sensor. In this case the test should demonstrate (by inspection of the range/angle map) that the radar is not able to discriminate the two targets in angle. The test should be then performed with the targets spaced in angle of the nominal angular resolution of the sensor. In this case the test should demonstrate the capability to discriminate the two targets in angle. If the sensor is not able to discriminate the targets in angle in this case, then the test should be re-iterated increasing the spacing of 0.5 times the nominal angle resolution until it becomes possible to discriminate the two targets. The minimum angular separation between the two targets when these can be discriminated will define the actual angular resolution of the sensor. An example of such measurement is show in Figure 4.2

Required equipment: Two Radar Echo generation systems would be required, as these have a single output channel and spatial/angle separation of the artificial echoes can be only generated by physically separating the echo generators' antennae.
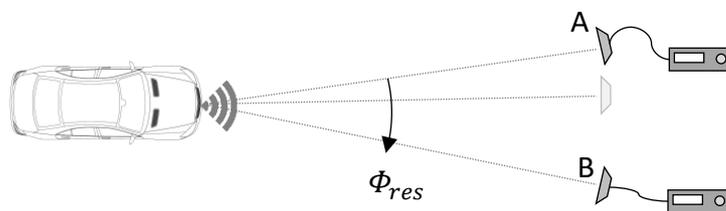

Figure 4.2 Example of angular resolution being measured by increasing the angle difference between targets A and B with the radar and finding the minimum that the sensor can separate the two targets

C. **Speed resolution:** Is the capability to discriminate targets with different speeds, corresponds to the minimum speed that targets need to differ in order be separated in speed. The relevant component of the speed is the component of the target velocity vector lying on the Line at 90 Azimuth and 0 angle degrees (perpendicular to the Line of Sight between the radar and the scene). The speed resolution is directly proportional to the Doppler resolution of the sensor that depends on the Coherent Processing Interval (CPI) that the sensor uses to perform the Doppler processing.
Test bed definition: In this case the two targets with different radial velocities be used in order to observe the sensor's output in terms of estimated target velocities. The speed difference between the two platforms should be equal to the nominal speed resolution of the sensor. The estimated two speeds of the targets should be different and equal to the ground truth speed. If it is not possible to obtain different speeds the procedure should be re-iterated increasing the relative speed separation between the two targets of 0.5 times the nominal sensor speed resolution until it is possible to discriminate the targets in speed. The speed separation between the targets when they can be discriminate will define the actual sensor speed resolution.
Required equipment: Radar Echo Generation system.

D. **Maximum detectable range:** Is the maximum distance at which the radar sensor can detect a target while respecting the desired false alarm probability. Depends on a range of design parameters, including emitted power, operating frequency, processing gain, antenna gains as well as processing requirements and environmental conditions. It is generally defined assuming a nominal target reflectivity (Radar Cross Section-RCS), a human has an RCS of 1 square meter.
Test bed definition: In order to test the maximum detectable range a target of known reflectivity (RCS) should be used to perform the testing. The target should be positioned at a range 1m closer than the nominal maximum detectable range of the sensor. If the target can be the detected, then the target should be moved further away than 1 m until it is no longer detected. Similarly, if the target cannot be detected at the first measurement then it should be moved 1 m closer until it gets detected. The furthest range at which the target can be detected defines the sensor's maximum detectable range. The same procedure can be performed with multiple targets of different RCSs in order to provide a full characterization depending on the target's size and material.
Required equipment: Radar Echo Generation system, however, as these systems have limited minimum range and maximum ranges possible to setup in the artificial target's generation in some cases it will be required to use 1 corner reflector as target with a known RCS. Corner reflectors are passive devices used to directly reflect radio waves back toward the emitting source making them very useful for calibrations. Examples of corner reflectors with different RCS are shown in Figure 4.3
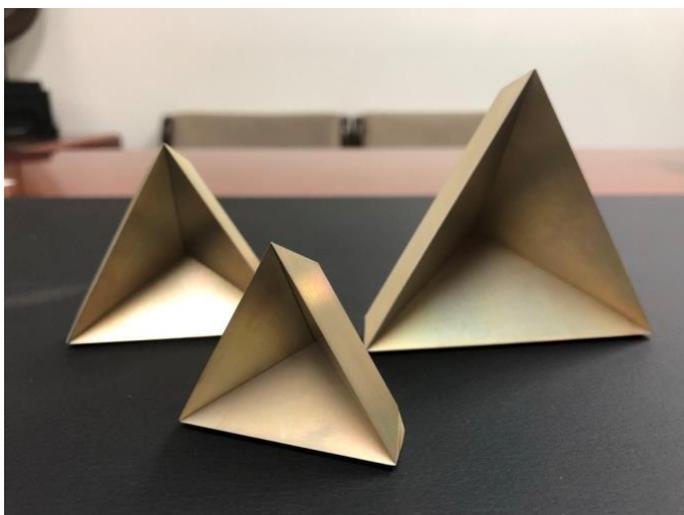

Figure 4.3 Examples of corner reflectors

E. **Minimum detectable range:** Is the minimum distance at which the radar can detect a target. Typical radar sensors for autonomous systems exploiting FMCW modulations do not have a nominal minimum detectable range as in principle it can be 0 m.
Test bed definition: The minimum detectable range can be tested using a target of known RCS (i.e.1 square meter) placed in proximity of the sensor (1 m further away than the nominal minimum detectable range) and moved towards the sensor until it is no longer detected (0.1 meters for each measurement). The closest range at which the target can be detected successfully will define the actual minimum detectable range of the sensor.
Required equipment: Radar Echo Generation system. These systems can have limited minimum range that can be emulated by the artificial targets, and it is generally higher than typical nominal minimum ranges of automotive radar sensors, then a corner reflector should be used.

F. **Maximum unambiguous speed:** Is the maximum speed that the sensor is able to measure unambiguously, this parameter depends on the transmitted waveform characteristics and in particular on the operating frequency and the Pulse Repetition Frequency that defines the sampling frequency in which the Doppler returns from the targets would be sampled.
Test bed definition: In order to test the maximum unambiguous speed, the test should be performed in a similar environment as described for the Speed Resolution. In this case a single moving target should be used. The target speed should be firstly set as plus or minus the nominal absolute value of the maximum unambiguous speed of the sensor, if it cannot be measured properly then the speed should be decreased in modulus by 0.5 m/s until the speed is measured correctly. Similarly, if the nominal maximum unambiguous speed can be measured then the test should be re-iterated with a speed of 0.5 m/s larger in modulus until the speed is correctly measured. The maximum speed that can be measured unambiguously will define the actual maximum unambiguous speed.
Required equipment: Radar Echo Generation system. However in case the nominal maximum detectable speed is higher than the maximum speed that can be emulated by the Radar Echo Generation system then the latter should be assumed as the maximum "certified" detectable speed as creating real targets at speeds in order of 100s of km/h is not practical and safe.

G. **Minimum detectable speed:** Is the minimum speed that the sensor can measure, it is proportional to the minimum detectable Doppler sensor that depends on the Coherent Processing Interval (CPI) that the sensor uses to perform the Doppler processing.
Test bed definition: The minimum detectable speed will be defined by the speed resolution of the sensor. The test for speed resolution will provide also this value.
Required equipment: Radar Echo Generation system.

H. **Update rate/Responsiveness:** Is the rate at which the sensor provides its measurement outputs in terms of detections, targets' parameters estimation and tracking.
Test bed definition: In order to test the sensor's Update rate and Responsiveness, the rate of the outputs of the estimated target's parameters should be measured in a controlled environment. The sensor's output in terms of detection maps (Range-Doppler, Range-Angle) should be stored for an acquisition of a fixed duration (i.e. 10s). The acquisition duration divided by the number of recorded maps will provide the update interval that indicates the sensor's capability to respond to a change of stimulus.
Required equipment: PC to record and extract the update rate.

I. **Tracking capability and capacity:** Is the capability of the sensor to track one or multiple targets in the scene, depends on the processing algorithms used and on the accuracy of the estimated target location and speed. Capacity refers to the maximum number that the sensor is able to track at the same time.

Test bed definition: In order to test the target tracking capabilities of a sensor, a similar setup as the one used for the speed resolution test can be used. The true target's positions and speeds should be used as ground truth when compared with those measured by the radar. Error in position and speed estimation should be quantified and characterised. The tracking capacity should be tested by increasing the number of targets in the scene starting from the nominal number maximum targets that the sensor is claimed to be able to track. If the sensor is able to successfully track the nominal value then the number of targets should be increased by 1 target each time, until the sensor fails to track all targets. If the sensor does not track the nominal number of targets then the number of targets should be decreased by 1 each time, until the sensor tracks correctly all the targets. This test will define the tracking capacity.

Required equipment: Multiple Radar Echo Generation system, based on the number of targets to be tracked. If the number of targets to be tracked is high and multiple targets cannot be simulated by the Radar Echo Generator, then a hard setup using corner reflectors can be used. In order to make the test easier and more cost-effective corner reflector targets can be stationary while the sensor can be placed on a moving platform to simulate moving targets.

J. **Sidelobe levels:** Is the ratio between the peak of the main lobe and the level of the first sidelobes of the radar waveform ambiguity function, these are evaluated along both the delay (range) and Doppler (velocity) domains.

Test bed definition: In order to measure the sidelobe levels the measurement of a target should be made. The output of the sensor in the range-Doppler domain should be analysed and the side-lobe levels estimated from the range-Doppler map in both range and Doppler direction. The sidelobe levels will be measured as the ratio between the amplitude of the peak of the main lobe of the response at the position of the target and the value of the amplitude of the peak of the highest sidelobe (these are generally expressed in dBs). These values will provide the actual sidelobe levels of the radar sensor.

Required equipment: Radar Echo Generation system.

K. **Integrated Sidelobe Ratio:** Is the ratio between the power of the main lobe and the sum of the power of all the sidelobes of the radar waveform ambiguity function, these are evaluated along both the delay (range) and Doppler (velocity) domains.

Test bed definition: In order to measure the sidelobe levels the measurement of a single target should be used. The output of the sensor in the range-Doppler domain should be analysed and the side-lobe levels estimated from the range-Doppler map in both range and Doppler direction. The sidelobe levels will be measured as the ratio between the amplitude of the peak of the main lobe of the response at the position of the target and the value of the amplitude of the peak of the highest sidelobe (these are generally expressed in dBs). These values will provide the actual sidelobe levels of the radar sensor.

Required equipment: Radar Echo Generation system.

L. **False alarm probability:** Is the probability of the sensor detecting a target when it is not present. This should be designed in order to minimise false alarms while preserving good detection performance.

Test bed definition: In order to test the false alarm probability a setup with an empty scene (no targets) should be used. Existing stationary objects in the scene can be removed from the detection maps using background estimation and subtraction. The nominal false alarm probability of the sensor (NPFA) should be used as reference and 100/NPFA measurements (decisions) should be taken. The actual false alarm probability should be then estimated as the number of false detections divided by the total number of measurements/decisions. This value will provide the actual sensor false alarm probability.

Required equipment: PC to record and analyse detection maps to extract false alarm probability.

M. **Detection probability:** Is the probability of the sensor detecting a target when it is present. High probability of detection can also lead to high false alarm probability and usually the detector's design involves a trade-off between those two characteristics.
Test bed definition: In order to test the probability of detection, the setup described in False alarm probability can be used with the addition of one target in the scene. In the case of existing stationary targets, background extraction can be used from measurements without the target of in the case of stationary clutter filtering, the target should have a Doppler shift. The nominal probability of detection (NPD) should be used as reference and 100/NPD measurements should be taken.
Required equipment: Radar Echo Generation system.

N. **Dynamic Range:** Is the ratio between the maximum and the minimum value of the received signal power that the sensor can handle. A poor dynamic range means that the sensor would easily saturate in presence of close/large targets leading to blinding (lack of ability to detect weaker targets' returns).
Test bed definition: In order to test and calibrate the Dynamic Range of the receiver two targets, one with the maximum RCS (σ_{max}) and one the minimum RCS (σ_{min}) that the sensor must be able to detect, should be used. The target with the maximum RCS should be placed at the minimum detectable range (R_{min}), while the target with the minimum RCS should be placed at the maximum detectable range (R_{max}). The dynamic range (D) will be then equal to:

$$D = 10 \log_{10} \frac{\sigma_{max} R_{max}^4}{\sigma_{min} R_{min}^4}$$

If the smaller target cannot be detected then it should be placed closer (0.5 times the range resolution cell every time) until it is detected and the target's distance will represent the value of R_{min} to be used to compute the dynamic range.
Required equipment: Radar Echo Generation system.

O. **Noise floor:** Is the measure of the signal created from the sum of all the noise sources and unwanted signals measured by the sensor. Its value affects the performance of the sensor in terms of detection/tracking/recognition capabilities and it should be kept as low as possible.
Test bed definition: In order to measure the noise floor, the sensor should be tested in open space, illuminating an area without targets (i.e. pointing towards the sky). The received signal power in this case will represent the noise floor.
Required equipment: PC to record the radar data and analyse noise power levels.

P. **Linear range and linearity:** Is the range in which the output of the sensor transmitter and receiver amplifier applies a linear function to the input. Non-linearities if present would need to be taken into account and compensated at the processing stages. The linearity of the sensor is an expression of the extent to which the actual measured curve of a sensor departs from the ideal curve.
Test bed definition: In order to test the linear range (on the receiver amplifier only) the test should be performed using a target of known RCS. Measurements should be made at different ranges, starting from the maximum detectable range and moving 2 meters every time towards the minimum detectable range. The ratio of the measured received power and the expected power at the receiver (referring to the radar equation [2]) should be then considered for each measurement. The linear portion of the derived curve from the measurements will define the dynamic range. In order to test the linearity, the results of the linear range can be used in order to measure the extent that the curve deviates from the linear behavior.
Required equipment: Radar Echo Generation system.

Q. **Antenna patterns:** Are the radiation patterns of the transmitter and receiver antenna of the sensor, their 3D shape is important to identify sidelobes and back lobe insulation.

Test bed definition: In order to test the antenna pattern, standard antenna measurements procedure should be used [3]. For a complete characterisation of the radiation field of the antenna, its relative amplitude, relative phase, polarisation, and the power gain shall be measured along different azimuth and elevation points with the antenna being on the centre of the scene and the range of the measurement equipment being constant. Figure 4.4 illustrates the standard spherical coordinate system used in antenna measurements. Additionally, if the transmitted wave covers a large or different frequency bands, the radio frequency shall also be treated as a variable in the measurements. A direct method of measuring the radiation pattern of a test antenna is to employ a suitable source antenna, which can be positioned in such a manner that it moves relative to the test antenna along lines of constant elevation θ and constant azimuth ϕ. Two main configuration can be used for such measurements: *fixed-line-of-sight* configuration where the test antenna and its associated coordinate system are rotated about a suitable axis, and *movable-line-of-sight* configuration where the source antenna is moved incrementally or continuously along the circumference of a circle cantered approximately at the phase centre of the antenna under test. If it is moved incrementally, then for each position of the source antenna the test antenna is rotated and the received signal is recorded. Alternately the test antenna can be rotated incrementally, and for each of its positions the source antenna is moved continuously along its circumferential path. An example of two *fixed-line-of-sight* measurements is illustrated in Figure 4.5. All test should be performed in an anechoic environment. These measurements can be automated by employing antenna test chambers solutions.

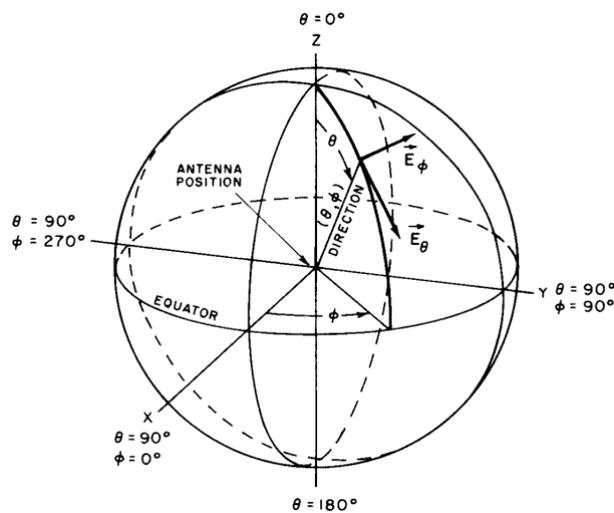**Required equipment**: Antenna test chamber (example in Radar reference [4])

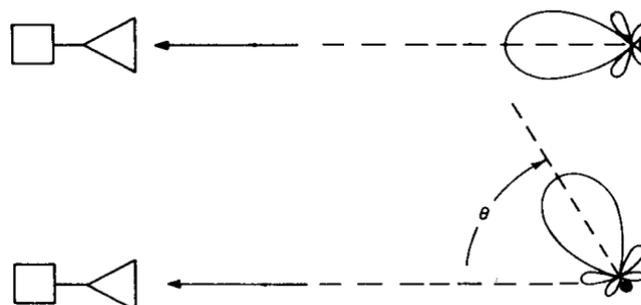**Figure 3.4 Standard Spherical Coordinate System Used in Antenna Measurements (ANSI/IEEE [3])**

Figure 4.4 Example of two fixed-line-of-sight measurements with the test antenna rotating in azimuth

Radar Standards

It is assumed that the DUT is compliant with the ISO 26262 standard. Furthermore, standards defining the test scenarios for the sensors once these are fitted on the vehicle are defined by ISO 19237 and ISO 19206. However, these only refer to dry weather conditions, so further scenarios should be defined for different weather conditions. Regarding the antenna radiation pattern, IEEE Std 149-1979 comprises test procedures for the measurement of antenna properties.

5. Combined RADAR-Camera-Lidar Calibration

While the above described radar calibration methods provide sufficiently good results for the targeted applications, they lack the possibility to fully assess the placement of the radar with respect to other sensors. Research on 3D LiDAR-radar calibration was conducted in [40] where they proposed a method which estimates a 6 degrees of freedom (DOF) extrinsic calibration of a 3D LiDAR-radar pair. The method includes a target design suitable both for the LiDAR and the radar shown in Fig. 5.1(a). It is inspired by a target constructed by Stanislas and Peynot [41] where radar performance is evaluated using a 2D LiDAR as a ground truth with a target composed of radar tube reflector and a square cardboard. The target for 3D LiDAR–radar calibration consists of a Styrofoam triangle which is invisible to the radar while it has good properties for detection and localization in the LiDAR point cloud. Radar receives the echo from the trihedral corner reflector shown in Fig. 5.1(b) which has high RCS and low orientation sensitivity. In the end, extrinsic calibration parameters are found by two-step optimisation. The first step is based on the reprojection error minimisation while the second uses space distribution of RCS, measure of the detection intensity, to estimate variables which are not observable from the reprojection error due to the lack of radar's vertical resolution.
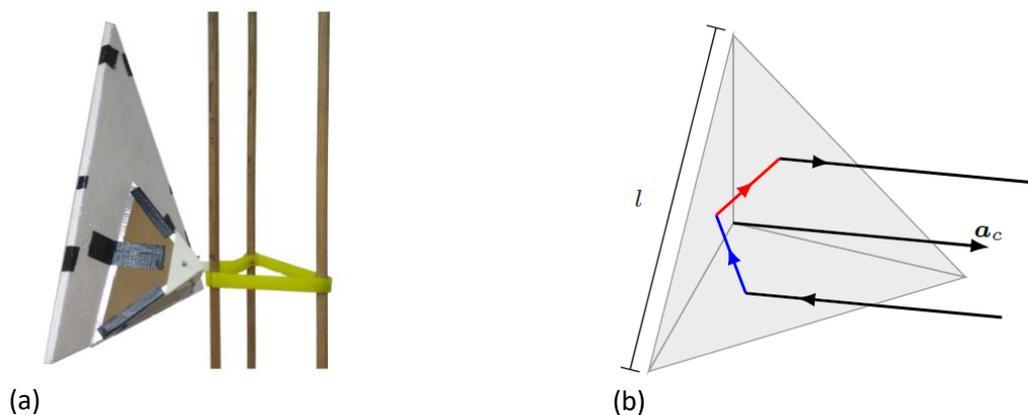

(a)   (b)

Figure 5.1 (a) Constructed calibration target from [23]; (b) Illustration of the working principle of the triangular trihedral corner reflector

Certain radar calibration targets such as the Precision Expandable Radar Calibration Sphere (PERCS) [44] which uses a geodesic polyhedron shape can be modified to accommodate unified calibration of cameras as well as LiDAR systems. The PERCS was designed as an orbit deployable HF radar target with an isotropic radar cross section for all ground-based HF radars. The PERCS target made use of Hoberman sphere technology for deployment of large objects in space. In this example, the starting point of calibration target design is an icosahedron shape consists of 20 equal-sized equilateral triangles and 12 nodes, all of which lie on the surface of a sphere. This mesh can be improved by breaking each of the original triangles into smaller triangles and projecting them onto the spherical surface. The number of the subdivisions of each of the original triangles defines the "level" of the mesh. For the second-level mesh each of the original triangles is subdivided into four triangles. The additional nodes in the second-level mesh are located above the midpoint of the original triangles. For the nth-level mesh each of the original triangles of the icosahedron is replaced with n^2 smaller

triangles. With each additional level of refinements the coarseness of the original icosahedron shape can be gradually improved to approximate a spherical shape more closely and the variation of the radar cross section with viewing direction is reduced which is useful for calibrating multiple radars at different directions simultaneously or calibrate a ground based radar when the calibration target is rotating in its axis which is the case when it's deployed in an orbit. The triangulation of levels 1 to 8 are shown in Figure 5.2. The advantage of these shapes with respect to spherical target is the additional flexibility it provides to add small, optical corner-cube reflectors to each vertex of the Hoberman sphere for LiDAR calibration and the ease at which the different faces can be painted with checkerboard like patterns for camera calibration. Optical corner reflectors provide a reflection cross section for visible light that depends on the incident angle and works similarly in principle of the microwave counterpart trihedral corner reflector as shown in Fig. 5.3(a). With one corner cube on each vertex as shown in Fig. 5.3(b), those that are facing toward a ground LiDAR system will reflect visible light back to the source as shown in Fig. 5.3(c). This geodesic polyhedron calibration target is more suitable for calibration of HF and VHF space weather radars and HF heaters. They can also be applied for laser tracking and imaging of satellites. Designing the target for calibration of automotive radar at 77 GHz is challenging due to the smaller wavelength which results in the requirement of smaller edge length of the faces as the edge length is a function of frequency of operation but is not impossible with higher order icosahedron geometries and improvements in manufacturing technologies.
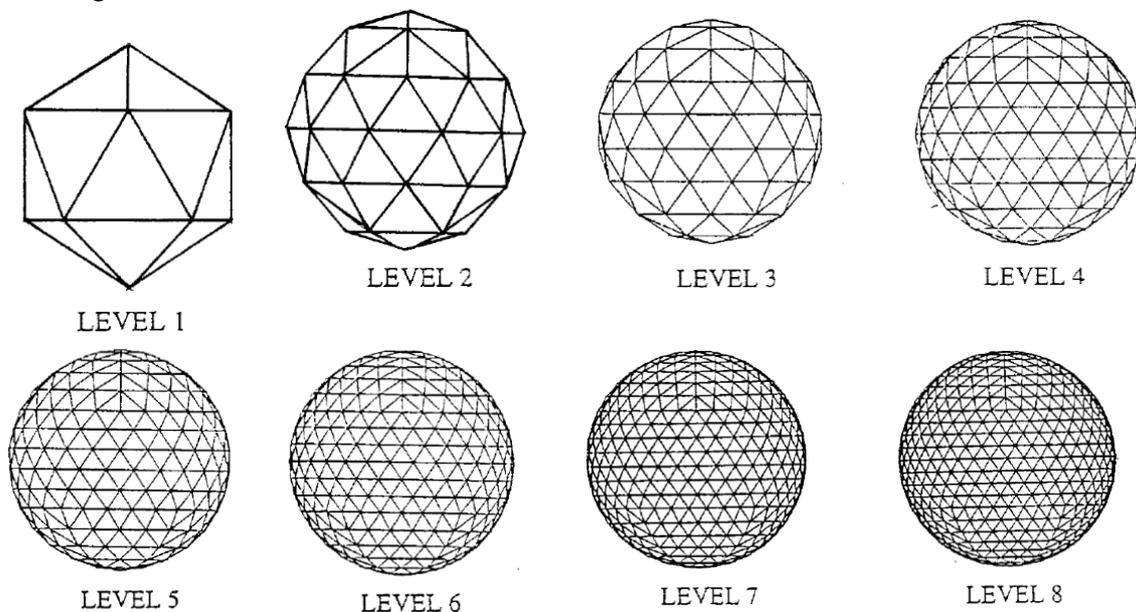

Figure5.2 First 8 levels of the equilateral triangular mesh of a sphere

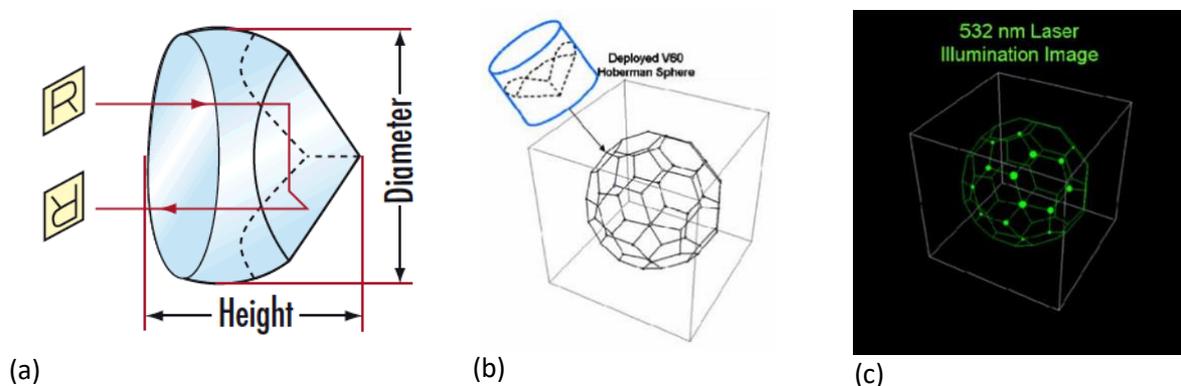(a)   (b)   (c)

Figure 5.3 (a) Principle of optical corner cube retroreflectors; (b) corner cube reflectors at each vertex; (c) Laser illumination on the calibration target

6. Measurement Confidence in Sensor Systems

While 3D imaging systems offer the potential of great positives, such as massively increased measurement speeds, equipment portability and relative ease when measuring freeform surfaces, their limitations are still being understood and international standards that describe suitable tests and procedures for their acceptance and use are yet to be developed, for example, VDI/VDE 2634 [24], the German guideline for optical 3D measuring systems addresses neither freeform surfaces nor surface finish. To support industry's increased use of and improve measurement confidence in these technologies, NPL launched a 3D optical scanner characterisation facility [25]. It comprises a purpose-built environmentally controlled laboratory (approximate dimensions: 3 m × 5 m × 2.5 m), test artefacts, test procedures and equipment to examine the performance of 3D optical scanners. The facility's capabilities included:

(a) Studying temperature effects of imaging systems and object to be imaged by measuring the separation between two spheres attached to a Zerodur bar as shown in Fig. 6.1(a) at different temperature.

(b) Developed multi-faceted test artefacts as shown in Fig. 6.1(b) to quantify 3D optical scanners' ability to measure surfaces with different reflectances, roughness and colours. In addition, surfaces of specific interest to industries that include aerospace, automotive, manufacturing, medical and heritage are addressed.

(c) To quantify the effects on data quality and data quantity of 3D imaging systems in environments with different and changing illumination using artefacts such as the NPL 150 mm Freeform artefact as shown in Fig. 6.1(c)[27].

(d) The resolution of the 3D imaging system is assessed by comparing scanner measurements of the NPL Bessel plate (Fig. 6.1(d)) with its CAD model.

(e) Developed a tetrahedral artefact to test for imaging system sensitivities to artefact position within the measurement volume.

(f) To find the optimum balance between scan velocity and measurement quality, NPL's National FreeForm Centre have developed a system where the articulating arm's scanning head is placed within an adjustable frame and moved along by a precision carriage. This is very useful for automotive industry to characterise the performance of their imaging systems.
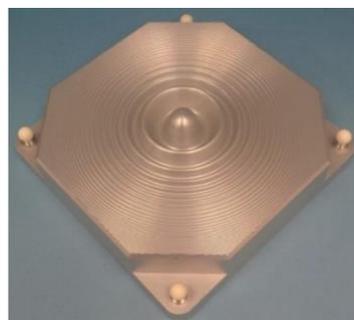

(a)

(b)

(c)
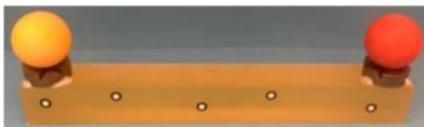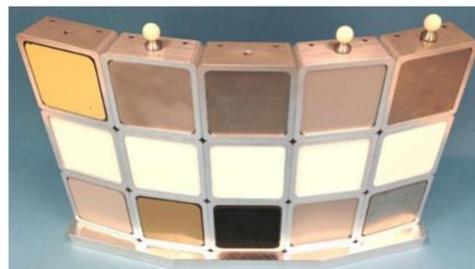
(d)

Figure 6.1 (a) Zerodur bar artefact for temperature effects measurement; (b) The NPL 3D material coupon plate; (c) A fringe pattern projected onto the NPL 150 mm Freeform artefact; (d) The NPL Bessel plate artefact

References:

[1] D. Scaramuzza, "Omnidirectional vision: from calibration to robot motion estimation," ETH Zurich, PhD Thesis, no. 17635, p. 189, 2008.

[2] Y. Wang, et.al, "Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving", arXiv preprint arXiv:1812.07179, 2018.

[3] Heikkila, J. (2000). Geometric camera calibration using circular control points. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1066–1077.

[4] Prokos, A., Kalisperakis, I., Petsa, E., & Karras, G. (2012). Automatic calibration of stereocameras using ordinary chess-board patterns. ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 45–49.

[5] Zhang, Z., "A flexible new technique for camera calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330–1334, 2000.

[6] Hansen, P., Alismail, H., Rander, P., & Browning, B. (2012). Online continuous stereo extrinsic parameter estimation. 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1059–1066.

[7] Ling, Y. & Shen, S. (2016). High-precision online markerless stereo extrinsic calibration. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1771–1778.

[8] Zheng, Y., He, Z., Yang, W. G., & Zhang, X. F. (2014). An efficient and practical calibration method for roadside camera using two vanishing points. Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), pp. 65–69.

[9] Yang, S., Rong, J., Huang, S., Shang, Z., Shi, Y., Ying, X., & Zha, H. (2016). Simultaneously vanishing point detection and radial lens distortion correction from single wide-angle images. 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 363–368.

[10] Rehder, E., Kinzig, C., Bender, P., & Lauer, M. (2017). Online stereo camera calibration from scratch. 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 1694–1699.

[11] Dang, T. & Hoffmann, C. (2004). Stereo calibration in vehicles. IEEE Intelligent Vehicles Symposium, pp. 268–273.

[12] Dang, T., Hoffmann, C., & Stiller, C. (2009). Continuous stereo self-calibration by camera parameter tracking. IEEE Transactions on Image Processing, vol. 18, no. 7, pp. 1536–1550.

[13] Marko, Thomas and Kubinger, Wilfried, Automatic Stereo Camera Calibration in Real-World Environments Without Defined Calibration Objects, Proceedings of the 29th DAAAM International Symposium, pp.0102-0108, Austria, 2018.

[14] J. Heikillä: "Geometric Camera Calibration Using Circular Control Points"; PAMI-22, no. 6; pp. 1066-1077; 2000.

[15] M. Kazhdan, M. Bolitho, and H. Hoppe: "Poisson Surface Reconstruction." Symposium on Geometry Processing (June 2006).

[16] C. Zach, T. Pock, and H. Bischof: "A globally optimal algorithm for robust TV-L1 range image integration." Proceedings of IEEE International Conference on Computer Vision (ICCV 2007).

[17] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In Proc. ICCV, pages 1–8, 2007.

[18] G. Zhang, J. Jia, T. Wong, and H. Bao. Consistent depth maps recovery from a video sequence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(6):974–988, 2009.

[19] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In Proc. CVPR, pages 1434–1441. IEEE, 2010

[20] S. Chen, et.al, "Outlier detection of point clouds generating from low-cost UAVs for bridge inspection", The Sixth International Symposium on Life-Cycle Civil, Belgium, October 2018.

[21] HALCON. MVTec Software GmbH.

[22] I. Mohamed, et.al, "Comparison of 3D Imaging Technologies for Wheat Phenotyping", 1st Workshop on Metrology for Agriculture and Forestry (METROAGRIFOR), 2019.

[23] CloudCompare. http://www.cloudcompare.org/.

[24] VDI 2634 Optical 3D-measuring systems Berlin Beuth 2008.

[25] Dury M, Brown S, McCarthy M, Woodward S 3D Optical Scanner Dimensional Verification Facility at the NPL's "National FreeForm Centre" Laser Metrology and Machine Performance XI (Huddersfield) pp. 191-200.

[26] Adamczyk M, Kaminski M, Sitnik R, Bogdan A, Karaszewski M. 2014 Appl Optics.53 5154-62.

[27] McCarthy M, Brown S, Evenden A, Robinson A 2011 NPL freeform artefact for verification of non-contact measuring systems Three-Dimensional Imaging, Interaction, and Measurement (San Francisco, USA, 24 - 27 January 2011) (Proceedings of SPIE vol 7864) ed Beraldin J, et al.

[28] Hold, S.; Gormer, S.; Kummert, A.; Meuter, M.; Muller-Schneiders, S. A novel approach for the online initial calibration of extrinsic parameters for a car-mounted camera. In Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, 4–7 October 2009; pp. 1–6.

[29] Coulombeau, P.; Laurgeau, C. Vehicle yaw, pitch, roll and 3D lane shape recovery by vision. In Proceedings of the IEEE Intelligent Vehicle Symposium, Versailles, France, 17–21 June 2002; Volume 2, pp. 619–625.

[30] Collado, J.; Hilario, C.; de la Escalera, A.; Armingol, J. Self-calibration of an on-board stereo-vision system for driver assistance systems. In Proceedings of the 2006 IEEE Intelligent Vehicles Symposium, Tokyo, Japan, 13–15 June 2006; pp. 156–162.

[31] Nedevschi, S.; Vancea, C.; Marita, T.; Graf, T. Online extrinsic parameters calibration for stereovision systems used in far-range detection vehicle applications. IEEE Trans. Intell.Transp. Syst. 2007, 8, 651–660.

[32] De Paula, M.; Jung, C.; da Silveira, L.G., Jr. Automatic on-the-fly extrinsic camera calibration of onboard vehicular cameras. Expert Syst. Appl. 2014, 41, 1997–2007.

[33] Li, S.; Hai, Y. Easy calibration of a blind-spot-free fisheye camera system using a scene of a parking space. IEEE Trans. Intell. Transp. Syst. 2011, 12, 232–242.

[34] Cech, M.; Niem, W.; Abraham, S.; Stiller, C. Dynamic ego-pose estimation for driver assistance in urban environments. In Proceedings of the 2004 IEEE Intelligent Vehicles Symposium, Parma, Italy, 14–17 June 2004, pp. 43–48.

[35] Teoh, C.; Tan, C.; Tan, Y.C. Ground plane detection for autonomous vehicle in rainforest terrain. In Proceedings of the 2010 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT), Kuala Lumpur, Malaysia, 20–21 November 2010; pp. 7–12.

[36] Wang, Q.; Zhang, Q.; Rovira-Mas, F. Auto-calibration method to determine camera pose for stereovision-based off-road vehicle navigation. Environ. Control Biol. 2010, 48, 59–72.

[37] E. F. Knott, Radar Cross Section Measurements. ITP Van Nostrand Reinhold, 1993.

[38] T. Wang, N. Zheng, J. Xin, and Z. Ma, "Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications," Sensors, vol. 11, no. 9, pp. 8992–9008, 2011.

[39] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave radar and its visualization on image sequence," in International Conference on Pattern Recognition (ICPR), 2004, pp. 342–345.

[40] J. Perˇsi´c, I. Markovi´c, and I. Petrovi´c, "Extrinsic 6Dof Calibration of 3D LiDAR and Radar," in European Conference on Mobile Robotics (ECMR), 2017, pp. 165–170.

[41] L. Stanislas and T. Peynot, "Characterisation of the Delphi Electronically Scanning Radar for Robotics Applications," in Australasian Conference on Robotics and Automation (ARAA), 2015.

[42] Bosch, "Calibration targets", https://www.boschaftermarket.com/gb/en/diagnostics/test-equipment/advanced-driver-assistance-systems/calibration-targets/.

[43] Z. Liu, et.al, "Flexible and accurate camera calibration using grid spherical images", Optics Express, vol. 25, No. 13, Jun 2017.

[44] P. Bernhardt, et.al, "Design and applications of a versatile HF radar calibration target in low Earth orbit", Radio Science, vol. 43, 2008.

Lidar Section References

G. Atanacio-Jiménez *et al.*, 'LIDAR Velodyne HDL-64E Calibration Using Pattern Planes':, *Int. J. Adv. Robot. Syst.*, Jan. 2011, doi: 10.5772/50900.

R. Bergelt, O. Khan, and W. Hardt, 'Improving the intrinsic calibration of a Velodyne LiDAR sensor', in *2017 IEEE SENSORS*, Oct. 2017, pp. 1–3, doi: 10.1109/ICSENS.2017.8234357.

C.-Y. Chen and H.-J. Chien, 'On-Site Sensor Recalibration of a Spinning Multi-Beam LiDAR System Using Automatically-Detected Planar Targets', *Sensors*, vol. 12, no. 10, Art. no. 10, Oct. 2012, doi: 10.3390/s121013736.

C. Gao and J. R. Spletzer, 'On-line calibration of multiple LIDARs on a mobile vehicle platform', in *2010 IEEE International Conference on Robotics and Automation*, May 2010, pp. 279–284, doi: 10.1109/ROBOT.2010.5509880.

N. Heide, T. Emter, and J. Petereit, 'Calibration of multiple 3D LiDAR sensors to a common vehicle frame', presented at the ISR 2018 - 50th International Symposium on Robotics, Messe München, Entrance East, Munich, Germany, Jun. 2018, [Online]. Available: https://www.iosb.fraunhofer.de/servlet/is/88371/.

J. Jiao *et al.*, 'A Novel Dual-Lidar Calibration Algorithm Using Planar Surfaces', in *2019 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2019a, pp. 1499–1504, doi: 10.1109/IVS.2019.8814136.

J. Jiao, Y. Yu, Q. Liao, H. Ye, R. Fan, and M. Liu, 'Automatic Calibration of Multiple 3D LiDARs in Urban Environments', in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019b, pp. 15–20, doi: 10.1109/IROS40897.2019.8967797.

J. Levinson and S. Thrun, 'Unsupervised Calibration for Multi-beam Lasers', in *Experimental Robotics: The 12th International Symposium on Experimental Robotics*, O. Khatib, V. Kumar, and G. Sukhatme, Eds. Berlin, Heidelberg: Springer, 2014, pp. 179–193.

N. Li, T. Luo, and B. Su, 'Fast Extrinsic Calibration for 3D LIDAR', in *Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering*, Shenzhen, China, Jul. 2019, pp. 1–10, doi: 10.1145/3351917.3351986.

N. Muhammad and S. Lacroix, 'Calibration of a rotating multi-beam lidar', in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct. 2010, pp. 5648–5653, doi: 10.1109/IROS.2010.5651382.

J. Oberländer, L. Pfotzer, A. Roennau, and R. Dillmann, 'Fast calibration of rotating and swivelling 3-D laser scanners exploiting measurement redundancies', in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2015, pp. 3038–3044, doi: 10.1109/IROS.2015.7353796.

Ocular Robotics, 'RobotEye RE05 3D LIDAR | Ocular Robotics'. http://www.ocularrobotics.com/products/lidar/re05/ (accessed Jul. 03, 2020).

A. V. Segal, D. Haehnel, and S. Thrun, 'Generalized-ICP', presented at the Robotics: Science and Systems V, University of Washington Seattle, USA, 2009, Accessed: Jun. 29, 2020. [Online]. Available: http://www.roboticsproceedings.org/rss05/p21.html.

T. Shan and B. Englot, 'LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain', in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 4758–4765, doi: 10.1109/IROS.2018.8594299.

M. Sheehan, A. Harrison, and P. Newman, 'Self-calibration for a 3D laser', *Int. J. Robot. Res.*, vol. 31, no. 5, pp. 675–687, Apr. 2012, doi: 10.1177/0278364911429475.

Velodyne Lidar, 'HDL-64E Durable Surround Lidar Sensor', *Velodyne Lidar*. https://velodynelidar.com/products/hdl-64e/ (accessed Jul. 03, 2020).

B. Xue *et al.*, 'Automatic Calibration of Dual-LiDARs Using Two Poles Stickered with Retro-Reflective Tape', in *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*, Dec. 2019, pp. 1–6, doi: 10.1109/IST48021.2019.9010134.

Radar Section References
[1]   https://www.rohde-schwarz.com/ph/products/test-and-measurement/radar-echo-generation/radar-echo-generation-theme_232540.html
        https://www.dspace.com/en/ltd/home/products/systems/darts.cfm#144_45161
[2]   Richards, M. A., Scheer, J. A., Holm, W. A., Beckley, B., Mark, P., Richards, A., et al. (2010). *Principles of Modern Radar Volume I-Basic Principles*. Citeseer.
[3]   ANSI/IEEE Standard Test Procedures for Antennas, ANSI/IEEE Std. 149-1979, IEEE, New York; John Wiley Distributors.
[4]   https://www.rohde-schwarz.com/ph/product/ats1500c-productstartpage_63493-713188.html

Appendix E: Quantification of the Sensor key performance indicators

A method is suggested for testing the each of the key capabilities of perception systems, differentiation for individual sensors types is provided where appropriate. If no text accompanies a sensor type, no differentiation in method is identified.

Sensor Capabilities:

I.   Angular/Spatial Resolution
      Radar

      In order to test and calibrate the range resolution, two corner reflectors closely spaced in range from the sensor should be positioned in front of the sensor. The range spacing of the two reflectors should be smaller than the nominal range resolution, in this case the test should demonstrate (by inspection of the range profile) that the radar is not able to discriminate the two reflectors in range. The test should be then performed with the corner reflectors spaced in range of the nominal range resolution. In this case the test should demonstrate the capability to discriminate the two targets in range (by inspection of the range profile). In the case that it would not be possible to discriminate the targets in range, then additional tests should be made increasing the spacing in range between the targets every time by 0.5 times the nominal range resolution until it becomes possible to discriminate the two targets, the minimum distance between the two targets when these can be discriminate will define the actual range resolution of the sensor.

      Camera

      The ability of a camera to spatial resolve an object is defined by the resolution limit of the optics, the detector and the electronics in combination. The diffraction limit or theoretical maximum resolution may be calculated directly from knowledge of the individual system components i.e. lens and detector. However real systems are non-ideal, and functional resolution should be determined by imaging appropriately illuminated ISO test charts: Geometric ISO 17850:2015, Resolution and Spatial Frequency ISO 12233:2017, and calculating the Modulation Transfer Function. Most resolution metrics are designed for grey-scale systems and colour have not been well integrated. Moreover, resolution does not account for other factors such as responsivity and atmospheric transmittance, which need to be considered in when trying to determine the accurate detection of targets.

      Lidar

      To test and calibrate the angular resolution, an appropriate 3-dimensional range chart may be used, with spatial frequency of the reflecting bars spanning the nominal spatial resolution, and the depth of valleys between bars being somewhat greater than the nominal range resolution. Inspection of the range/angle map should show the resolution below which the lidar is unable to discriminate between the bars.

      Ultrasound

      Ultrasonic sensors are unidirectional and do not provide detection capability off-axis. There are therefore no tests for angular/spatial resolution for individual sensors. ISO 17386 [1] describes tests for multi-sensor implementations for covering key zones at different azimuths around the vehicle. (See Range Resolution)

I.   Range Resolution

Radar

In order to test and calibrate the angular resolution, two corner reflectors closely spaced in angle from the sensor should be positioned in front of the sensor. The angular spacing of the two reflectors should be smaller than the nominal angular resolution of the sensor. In this case the test should demonstrate (by inspection of the range/angle map) that the radar is not able to discriminate the two reflectors in angle. The test should be then performed with the corner reflectors spaced in angle of the nominal angular resolution of the sensor. In this case the test should demonstrate the capability to discriminate the two targets in angle. If the sensor is not able to discriminate the targets in angle in this case, then the test should be re-iterated increasing the spacing of 0.5 times the nominal angle resolution until it becomes possible to discriminate the two targets. The minimum angular separation between the two targets when these can be discriminated will define the actual angular resolution of the sensor.

Camera

In order to test and calibrate the range resolution of stereo or multi-stereo cameras the procedure described in the radar chapter can be used, substituting using two closely spaced 18% Grey spherical targets for retro reflectors.

Lidar

To test and calibrate the range resolution, closely spaced range target reflectors should be positioned in front of the sensor. The range spacing of the two reflectors should be smaller than the nominal range resolution, in this case the test should demonstrate (by inspection of the range profile) that the lidar is unable to discriminate the two reflectors in range. The test should be then performed with the target reflectors spaced in range of the nominal range resolution. In this case the test should demonstrate the capability to discriminate the two targets in range (by inspection of the range profile). In the case that it would not be possible to discriminate the targets in range, then additional tests should be made increasing the spacing in range between the targets every time by 0.5 times the nominal range resolution until it becomes possible to discriminate the two targets, the minimum distance between the two targets when these can be discriminate will define the actual range resolution of the sensor. The tests may be combined by constructing an apparatus with targets separated by increasing range gaps spanning the nominal range resolution.

Ultrasound

While ultrasound sensors are generally not able to discriminate different sources simultaneously, ISO 17386 describes test methods for the effectiveness of target detection across predefined zones to be monitored, at the rear, front and corners of the vehicle. It also defines test grids in both horizontal and vertical planes.
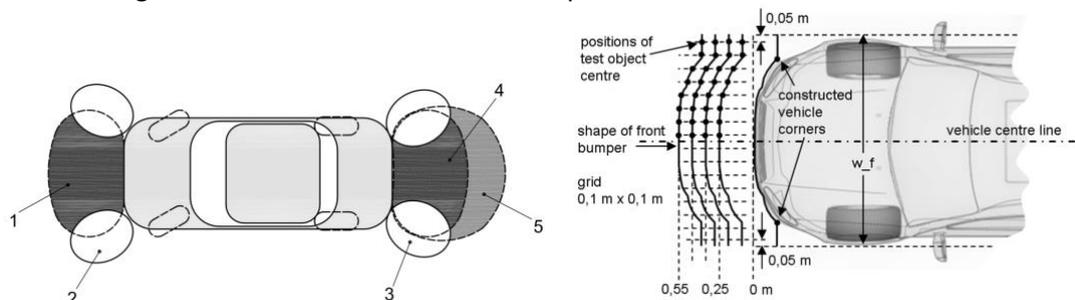

Figure 1 ISO 17386 Monitoring zones and example grid pattern identifying test locations at the front of the vehicle

The test involves the use of cylindrical test objects (diameter 75 mm, length 100 cm or the width of the vehicle under test, as appropriate), which are mounted either perpendicular

(100 cm lengths) or parallel (100 cm lengths for corners zones and vehicle width for front and back zones) to the level floor surface, at the each grid location in succession. The purpose of the multiple test positions is to map the areas where detection does and does not occur. Criteria are specified for the percentage coverage in each of the given zones.

II.   Speed Resolution
       Radar

In order to test and calibrate the speed resolution of the sensor two potential approaches can be used depending on the capability to know the sensor emitted waveform characteristics and received signal:

a.   Knowledge of the emitted waveform characteristics and access to raw data: In the case in which full knowledge of emitted pulse train a start-stop approach can be used to perform the testing. Two corner reflectors should be placed at different ranges and range profiles should be generated, while changing the position of the reflectors in range mimicking a radial motion of the targets with difference of displacements proportional to the nominal speed resolution of the sensor. The total number of range profiles acquired should be compliant with the sensor characteristics and be equal to the number of frames used by the sensor to compose a CPI. The nominal sensor CPI should then be used as reference to create a range-Doppler map. In the range-Doppler map the two targets should be then visible at different velocities and equal to the ground truth velocity emulated. If it is not possible to see the targets at different speeds the procedure should be re-iterated increasing the relative speed separation between the two targets of 0.5 times the nominal sensor speed resolution until it is possible to discriminate the targets in speed. The minimum speed separation between the targets when they can be discriminate will define the actual sensor speed resolution.

b.   Lack of knowledge of the emitted waveform characteristics and access to raw data: In this case the two corner reflectors should be mounted on moving platforms with accurate speed control in order to observe the sensor's output in terms of estimated target velocities. The speed difference between the two platforms should be equal to the nominal speed resolution of the sensor. The estimated two speeds of the targets should be different and equal to the ground truth speed. If it is not possible to obtain different speeds the procedure should be re-iterated increasing the relative speed separation between the two targets of 0.5 times the nominal sensor speed resolution until it is possible to discriminate the targets in speed. The speed separation between the targets when they can be discriminate will define the actual sensor speed resolution.

Cameras

In order to test and calibrate the range speed of stereo or multi-stereo cameras the procedure described in the radar chapter can be used, substituting 18% Grey spherical targets for retro reflectors.

III.   Maximum Detectable Range
       Radar

In order to test the maximum detectable range a target of known reflectivity (RCS) should be used to perform the testing. The target should be positioned at a range 1 m closer than the nominal maximum detectable range of the sensor. If the target can be detected then the target should be moved further away than 1 m until it is no longer detected. Similarly, if the target cannot be detected at the first measurement then it should be moved 1 m closer until it gets detected. The furthest range at which the target can be detected defines the sensor's maximum detectable range. The same procedure can be performed with multiple targets of

different RCSs in order to provide a full characterization depending on the target's size and material.

Camera

In order to test the maximum detectable range an extended black target should be positioned against a bright horizon, at a range 1 m closer than the nominal maximum detectable range of the sensor, i.e. where the contrast C\ =\ 0.05. If the target can be the detected, then the target should be moved further away than 1 m until it is no longer detected. Similarly, if the target cannot be detected at the first measurement then it should be moved 1 m closer until it gets detected. The furthest range at which the target can be detected defines the sensor's maximum detectable range.

Lidar

To determine the maximum detectable range for a target object of given size (cross sectional area) and reflectivity, the target should be positioned at a distance equal to the nominal maximum detectable range of the sensor. If the target can be detected at that range, then the target should be moved further away in 1 m increments until it is no longer detected. Similarly, if the target cannot be detected at the first measurement then it should be moved closer in 1 m increments until it is detected. The same procedure should be performed with multiple targets of different reflectivity and size to fully the dependency of maximum detectable range on size and reflectivity of an object.

Ultrasound

While there is no predefined test, the methodology of ISO 17386 can readily be adapted to evaluate the maximum detectable range. This has the advantage of utilising standardised test objects and offers compatibility with the method of assessing effectiveness of detections.

IV.   Minimum Detectable Range

Radar

The minimum detectable range can be tested using a corner reflector of know RCS (i.e.1 square meter) placed in proximity of the sensor (1m further away than the nominal minimum detectable range) and moved towards the sensor until it is no longer detected (0.1 m for each measurement). The closest range at which the target can be detected successfully will define the actual minimum detectable range of the sensor.

Camera

The minimum detectable range can be tested by moving an extended black target closer to the camera. The closest range at which the target can be detected successfully will define the actual minimum detectable range of the sensor. For a mono-camera this can be 0 m but for a stereo-setup the required overlap in fields for view will generate a real positive limit.

Lidar

The minimum detectable range can be tested using a target of known reflectivity and size placed close to the sensor (at nominal minimum detectable range) and moved towards the sensor in 10 cm increments until it is no longer detected, or away from the sensor until it is detected.

Ultrasound

While there is no predefined test, the methodology of ISO 17386 can readily be adapted to evaluate the minimum detectable range. This has the advantage of utilising standardised test objects and offers compatibility with the method of assessing effectiveness of detections

V.   Maximum Unambiguous Speed

Radar

In order to test the maximum unambiguous speed, the test should be performed in a similar environment as described for the Speed Resolution. In this case a single moving target should be used. The target speed should be firstly set as the nominal maximum unambiguous speed of the sensor, if it cannot be measured properly then a speed of 0.5 m/s slower should be used until the speed is measured correctly. Similarly, if the nominal maximum unambiguous speed can be measured then the test should be re-iterated with a speed of 0.5 m/s faster until the speed is correctly measured. The maximum speed that can be measured unambiguously will define the actual maximum unambiguous speed.

Camera

In order to test the maximum unambiguous speed for of stereo or multi-stereo cameras the procedure described in the radar chapter can be used, substituting a 18% Grey spherical target for a retro reflector.

VI. Minimum Detectable Speed
Radar

The minimum detectable speed will be defined by the speed resolution of the sensor. The test for speed resolution will provide also this value.

VII. Update rate/Responsiveness
Radar

In order to test the sensor's Update rate and Responsiveness, the rate of the outputs of the estimated target's parameters should be measured in a controlled environment. If the sensor is able to provide an output through an interface (i.e. a display) then its refresh rate will provide the measurement of the update rate of the sensor.

Camera

In camera systems frame rate is the limiting case.

Ultrasound

The same approach to that used for *Radar* is appropriate. The limiting factor is the pulse repetition rate.

VIII. Tracking Capability
Radar

The tracking capacity should be tested in the same setup as the tracking capability test, increasing the number of targets in the scene starting from the nominal number maximum targets that the sensor is claimed to be able to track. If the sensor is able to track successfully the nominal value then the number of targets should be increased by 1 target each time, until the sensor fails to track all targets. If the sensor does not track the nominal number of targets then the number of targets should be decreased by 1 each time, until the sensor tracks correctly all the targets. This test will define the tracking capacity. In order to make the test easier and more cost-effective the targets can be stationaries while the sensor can be placed on a moving platform (in this case all targets will exhibit similar speeds due to their stationary nature).

Camera

In order to test and calibrate target tracking capabilities of stereo or multi-stereo cameras the procedure described in the radar chapter can be used, substituting a 18% Grey spherical target for a retro reflector.

Lidar

Testing of tracking capabilities will require a method to establish ground truth as to a target's position and speed as the target passes through the field of view of the lidar. The ground truth can be compared against the positioning output from lidar tracking to establish tracking precision.

Ultrasound

Ultrasound sensors are primarily used to identify the distance from stationary targets or the momentary presence of a moving target (e.g. another vehicle). However, ultrasound systems are not designed to track targets.

IX.   Tracking Capacity
Radar

The tracking capacity should be tested in the same setup as the tracking capability test, increasing the number of targets in the scene starting from the nominal number maximum targets that the sensor is claimed to be able to track. If the sensor is able to track successfully the nominal value then the number of targets should be increased by 1 target each time, until the sensor fails to track all targets. If the sensor does not track the nominal number of targets then the number of targets should be decreased by 1 each time, until the sensor tracks correctly all the targets. This test will define the tracking capacity. In order to make the test easier and more cost-effective the targets can be stationaries while the sensor can be placed on a moving platform (in this case all targets will exhibit similar speeds due to their stationary nature).

Camera

The tracking capacity of stereo or multi-stereo cameras can be tested using the procedures described in the radar chapter, substituting 18% Grey spherical targets for retro reflectors.

Lidar

The tracking capacity should be tested in the same setup as the tracking capability test, increasing the number of targets in the scene starting from the nominal number maximum targets that the sensor is claimed to be able to track. If the sensor is able to track successfully the nominal value then the number of targets should be increased by 1 target each time, until the sensor fails to track all targets. If the sensor does not track the nominal number of targets then the number of targets should be decreased by 1 each time, until the sensor tracks correctly all the targets. This test will define the tracking capacity.

X.   Contrast
Camera

In order to test the contrast threshold for detection, an extended black target should be placed on the ground against a bright horizon such that the contrast $C = 0.05 = C_{Th}$ the threshold contrast where nominally the probability of detection $P_D = 50\%$. If $P_D > 50\%$ the target should be moved away from the camera, the brightness of the horizon or target size reduced until the probability is 50% or if detection probability $< 50\%$ the target should be moved towards the target or the brightness of the horizon or the target size increased and the actual contrast threshold recalculated.

XI.   Sidelobe Levels
Radar

In order to measure the sidelobe levels the measurement of a single corner reflector should be made. The output of the sensor in the range-Doppler domain should be analysed and the

side-lobe levels estimated from the range-Doppler map in both range and Doppler direction. The sidelobe levels will be measured as the ratio between the amplitude of the peak of the main lobe of the response at the position of the target and the value of the amplitude of the peak of the highest sidelobe (these are generally expressed in dBs). These values will provide the actual sidelobe levels of the radar sensor.

XII. Integrated Sidelobe Ratio

Radar

The integrated sidelobe level can be tested using the same setup used to measure the sidelobe level, by computing the ratio the main lobe over that of all the power outside the main lobe in either the range or the Doppler direction.

XIII. False Alarm Probability

In order to test the false alarm probability measurements in a number of varying empty scenes (no targets) should be used. Existing stationary objects in the scene can be removed from the detection maps using background estimation and subtraction. The nominal false alarm probability of the sensor (NPFA) should be used as reference and 100/NPFA measurements (decisions) should be taken. The actual false alarm probability should be then estimated as the number of false detections divided by the total number of measurements/decisions. This value will provide the actual sensor false alarm probability.

XIV. Detection Probability

In order to test the probability of detection of a specified class of objects, the setup described in False alarm probability can be used with the addition of one target in the scene. The location and orientation of targets should vary between test scenes. In the case of existing stationary targets, background extraction can be used from measurements without the target of in the case of stationary clutter filtering, the target can be placed on a moving platform. The nominal probability of detection (NPD) should be used as reference and 100/NPD measurements should be taken. Separate determinations of the detection probability may be made for different classes of target, or environmental conditions etc.

XV. Dynamic Range

Radar

In order to test and calibrate the Dynamic Range of the receiver two reflectors with the maximum RCS (σ_{max}) and the minimum RCS (σ_{min}) that the sensor must be able to detect should be used. The target with the maximum RCS should be placed at the minimum detectable range (R_{min}), while the target with the minimum RCS should be placed at the maximum detectable range (R_{max}). The dynamic range (D) will be then equal to:

$$D = 10 \log_{10} \frac{\sigma_{max} R_{max}^4}{\sigma_{min} R_{min}^4}$$

If the smaller target cannot be detected then it should be placed closer (0.5 times the range resolution cell every time) until it is detected and the target's distance will represent the value of R_{min} to be used to compute the dynamic range.

Camera

In order to test and calibrate the Dynamic Range, one must be obtain the Saturation Equivalent Exposure SEE, and the Noise Equivalent Exposure NEE. SEE is the exposure that is just equal to the saturation level of the detector. NEE is the exposure that is just equal to the saturation level of the detector.

Dynamic Range is defined as:

$$DR = \frac{SEE}{NEE} = \frac{V_{Max}}{V_{Noise}}$$

$$SEE = \frac{V_{Max}}{R_{Ave}} \quad \left(\frac{J}{cm^2}\right)$$

$$NEE = \frac{V_{Noise}}{R_{Ave}} \quad \left(\frac{J}{cm^2}\right)$$

Where V_{Max} is the maximum detector signal, R_{Ave} is the average responsivity, and V_{Noise} is the noise signal.

Care should be taken while testing and calibrating Dynamic Range as the device quantum efficiency is a function of wavelength, and light-source characteristics can vary considerably; consider the variation between the blackbody emission spectra of the Sun vs the spectra of Halogen lamps or Phosphor converted LEDs. Guidance is provided in ISO 15739:2017, ISO 7589:2002 and ISO/CIE 11664:2007.

Lidar

Experiments to determine the maximum and minimum detectable range, using a maximally reflective object and the minimum detectable range, and a minimally reflective object at its maximum detectable range allows calculation of the dynamic range of the sensor.

Ultrasound

Although not generally applicable to the use-case for ultrasound sensors, the dynamic range can be assessed in a laboratory environment, using the methods for linearity range and noise floor described below. The dynamic range is simply the ratio (or difference if specified in decibels) between these two parameters.

Assessment of the dynamic range is not feasible (or necessary) for Ultrasound sensors fitted to vehicles.

XVI.   Noise Floor

Radar

In order to measure the noise floor, the sensor should be tested in open space, illuminating an area without targets (i.e. pointing towards the sky). The received signal power in this case will represent the noise floor.

Camera

In order to measure the noise floor, a full analysis of the system noise (n_{Sys}) sources must be undertaken where:

$$\langle n_{Sys} \rangle = \sqrt{\langle n_{Shot}^2 \rangle + \langle n_{Pattern}^2 \rangle + \langle n_{Reset}^2 \rangle + \langle n_{Amp}^2 \rangle + \langle n_{ADC}^2 \rangle}$$

including determination of the components of pattern noise: Fixed Pattern Noise/dark current (FPN) and Photoresponse Nonuniformity (PRNU) by dark and flat-field exposures respectively, where:

$$\langle n_{Pattern} \rangle = \sqrt{\langle n_{FRN}^2 \rangle + \langle n_{PRNU}^2 \rangle}$$

Ultrasound

Although not generally applicable to the use-case for ultrasound sensors, the noise floor can be assessed in a laboratory environment. The noise floor of acoustic transducers is assessed by isolating the transducer from all sources of external acoustic stimuli in the frequency band of interest. In practice a thick-walled enclosure suitably sized to accommodate the sensor can be used. It is necessary to pass the sensor cable out of the enclosure taking precaution to not

compromise the acoustic isolation. It is then simply a matter of measuring the residual signal. The power spectral density is normally measured which can be readily integrated over the pre-requisite frequency bandwidth to obtain an equivalent sound pressure level in decibels. An audio analyser will normally compute all parameters automatically.

Assessment of the noise floor is not feasible (or necessary) for Ultrasound sensors fitted to vehicles.

XVII. Linear Range
Radar

In order to test the linear range (on the receiver amplifier only) the test should be performed using a corner reflector of known RCS. Measurements should be made at different ranges, starting from the maximum detectable range and moving 2 m every time towards the minimum detectable range. The ratio of the measured received power and the expected power at the receiver (referring to the radar equation [2] should be then considered for each measurement. The linear portion of the derived curve from the measurements will define the dynamic range.

Camera

In order to test the linear range, the test should for Dynamic Range should be performed while reducing the incident power from the Illuminant in a controlled fashion. Measurements should be made at powers from saturation to the noise floor. The ratio of the measured received power and the expected power should be then considered for each measurement. The linear portion of the derived curve from the measurements will define the dynamic range.

Ultrasound

Although not generally applicable to the use-case for ultrasound sensors, the linear range can be assessed in a laboratory environment. A method whereby the sensor under test is compared with a laboratory-grade reference ultrasound sensor (typically a hydrophone, but operated in air) can be used. The reference sensor should be known to operate linearly over the range of sound pressure levels to be tested. An ultrasound source, also known to operate linear is required to provide the test stimulus. The test then proceeds by measuring the sensor output as a function of the applied sound pressure level, as determined by the reference sensor. Alternatively, the sensor output can be analysed directly by a distortion meter.

Assessment of the linear range is not feasible (or necessary) for Ultrasound sensors fitted to vehicles.

XVIII. Linearity
In order to test the linearity, the results of the linear range can be used in order to measure the extent that the curve deviates from the linear behaviour.

XIX. Antenna Patterns
Radar

In order to test the antenna pattern, standard antenna measurements procedure should be used [3].

XX. Field of View
The angle through which the sensor can detect electromagnetic radiation may be determined by moving a target in azimuth and elevation with respect to the sensor and determining the position at which it is no longer visible.

References

[1] ISO 17386:2010 Transport information and control systems — Manoeuvring Aids for Low Speed Operation (MALSO) — Performance requirements and test procedures
[2] Richards, M. A., Scheer, J. A., Holm, W. A., Beckley, B., Mark, P., Richards, A., et al. "Principles of Modern Radar Volume I-Basic Principles." (Citeseer, 2010).

[3] ANSI/IEEE Standard Test Procedures for Antennas, ANSI/IEEE Std. 149-1979, IEEE, New York; John Wiley Distributors

Appendix F: The physical basis for rainfall testing requirements

Aim of this Appendix: The section describes the underpinning science behind CAV sensor degradation due to falling rain at lidar and radar wavelengths. Its purpose is to demonstrate the complexity of the rainfall-sensor interaction that drives the recommendations regarding the essential components of the test ecosystem, along with the need to appropriately handle the inescapable uncertainties that arise from this complexity. It is not intended to be a comprehensive summary of the topic. Some of this material also appears in the main body of the document.

Orientation

The description follows an incremental approach thus:

1. The impact of a single spherical raindrop on the beam from an active CAV sensor, its sensitivity to the size of the drop and the introduction of terms such as attenuation
2. The cumulative effect of the passage of the beam through a collection of raindrops, whose sizes are described in terms of a drop size distribution (DSD) that can be linked to the rainfall rate
3. The variability of these DSDs, by rain fall type, and its consequent impact on attenuation at different frequencies
4. The impact of non-spherical raindrops on the attenuation
5. The linkage between attenuation and the maximum range KPI
6. The variability of rainfall rate (and by implication DSD) at different scales
7. The resultant need for extreme care when considering meteorological definitions of rainfall rate for CAV sensor applications
8. Commentary on the linkage to the monitoring of the ODD
9. Implications for the test ecosystem

Single raindrop

Rainfall attenuates electromagnetic (EM) signals at all wavelengths as the EM wave travelling between the target and the sensor interacts with each raindrop in its path. The electromagnetic wave will either pass *between* the raindrops or, if it encounters one, it will be either absorbed (effectively heating the raindrop) or scattered. The combined effect of absorption and scattering is called *extinction*, and the summation of all the extinction events as the EM travels between the sensor and the target is the *attenuation*.

Both the strength of absorption and scattering caused by a <u>single</u> raindrop, and the directions in which the EM radiation is scattered, can be modelled. It is dependent on the following factors:

- The *size parameter* of the droplet, which is a measure of the size of the droplet compared to the wavelength of the radiation
- The *refractive index* of the water in the raindrop, which itself is a function of the operating wavelength of the sensor (and the temperature)
- The shape of the droplet, noting that as the droplets increase in size, they tend to squash in the vertical direction due to aerodynamic drag. This results in the attenuation also being dependent on the polarisation of the beam. It is common to make an initial assumption that the raindrop is spherical, especially in the presence of other dominant uncertainties such as the precise drop size distribution (see later)

For spherical drops, the effectiveness of a given drop in removing energy from the beam can be determined by Mie (1908) theory[18]. Fig 1, shows this for water droplets at 77 GHz (i.e. a mm-wave radar) and a 950 nm lidar.

---

[18] Mie, G. von, 1908: Beitrage zur Optic triiber Medien, speziell kolloidaler Metallosungen, *Ann. der Physi1c*, 25,377-445.
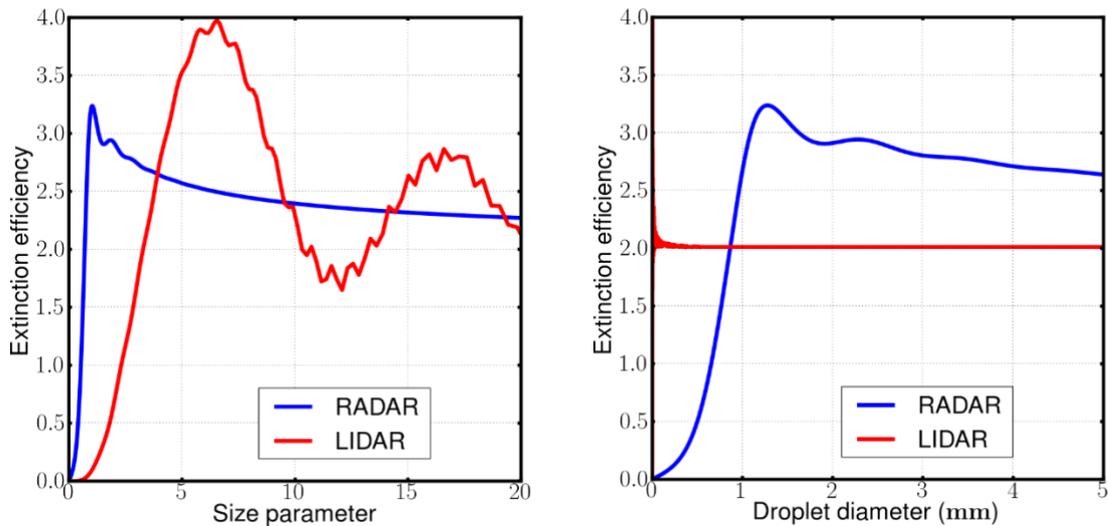
Figure 1: The variation of extinction efficiency with (a) size parameter and (b) droplet diameter. Qe is a measure of the extinction coefficient divided by the cross-sectional area of the droplet. Size parameter is π*D/λ i.e. the ratio of the circumference of the droplet divided by the wavelength

Two things are particularly noteworthy. Firstly, droplets have the potential to become very efficient at removing energy from the beam when they are comparable in size to the wavelength. Secondly, the strength of this interaction is very sensitive to the droplet size. This has significant implications to the assurance challenge because it means that the impact of rain on the attenuation, and therefore the performance of sensors, can be very sensitive to how the total rainfall is distributed across the different sizes of raindrop.

Merely considering the extinction value is not enough. As mentioned previously, extinction may comprise both absorption and scattering. It matters both how much of the lost energy is simply absorbed and, for the scattered part how directional that scattering is. Fig 2 demonstrates how the *directionality* of the scattering varies with size parameter. At small size parameters (the droplet is much smaller than the wavelength) the scattering in the forward and backward directions is equal and much of it is away from the direction of the beam. This is referred to as Rayleigh scattering and it is this that is responsible for blue skies and red sunsets. It also applies to the scattering by very small droplets at the mm-wave (radar) band.

As the droplet (and therefore size parameter) increases, the scattering is predominantly forward and as the size increases it is concentrated in a narrow forward peak. This is the case of lidar light interacting with larger raindrops. The behaviour of this narrow peak is significant for a number reasons:

- Energy that is scattered in the strongly forward direction contributes to the *calculated* extinction in the same way as energy scattered out of the beam. The resultant calculation of attenuation may therefore represent an overestimate of the true energy lost, however the wave will be slightly time delayed as a result of passing through the droplet. This has the effect of spreading the pulse of energy in time, which may affect sensor performance
- Energy that is scattered close to the forward direction may subsequently scattered back towards the main beam direction by other droplets. This also leads to spreading the pulse of energy. This is well described in Guo[19] et al 2015

---

[19] Guo J, H Zhang, X Zhang (2015) Propagating Characteristics of Pulsed Laser in Rain, *International Journal of Antennas and Propagation*, Volume 2015, Article ID 292905, 7 pages,

- The magnitude of the effect of such scattering on sensor performance will also be related to the relative angular widths of the beam and the forward scattering peak

Overleaf:

Figure 2: Phase functions (for unpolarised light) for progressively larger particles for increasing size parameters. The plots themselves are for lidar however for the purposes of demonstration they are indicative of the behaviour at other wavelengths
(a) Size parameter < 0.1 e.g cloud droplets at radar frequencies / air molecules at lidar i.e. Rayleigh scattering
(c) Size parameter ~1 e.g. rain drops at radar / cloud droplets at lidar
(d) Size parameter ~10 e.g. large rain droplets at lidar
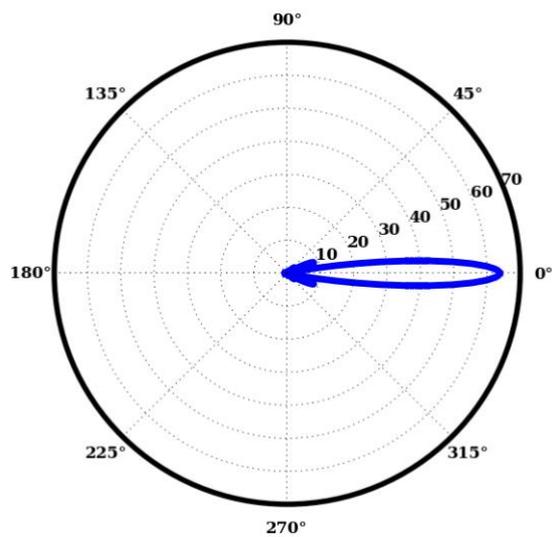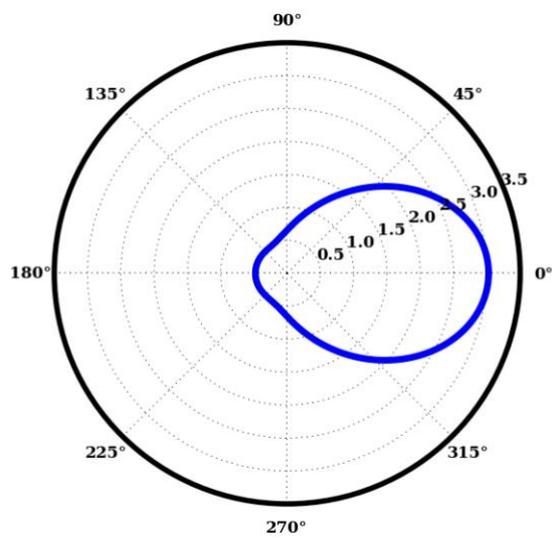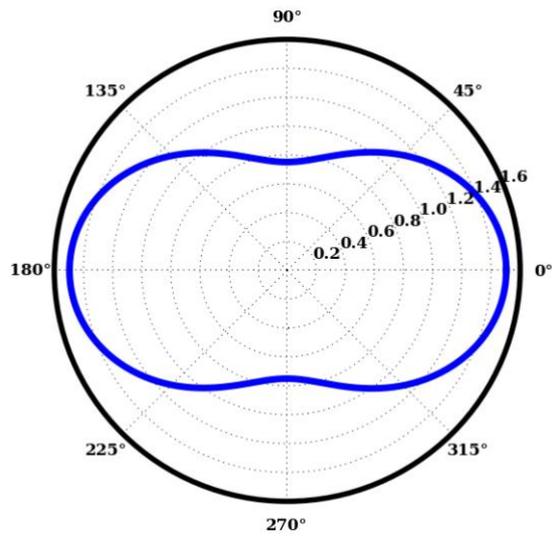
Clearly there is much complexity involved even at the single rain drop level and the sensor assurance framework must be aware of the *implications* of this in its design i.e. it must be capable of measuring

the <u>net</u> impact of all of these processes for all sensors, which will vary considerably with the precise characteristics (e.g. angular resolution, pulse width) of the sensor.

The next section looks at how this strong dependence on droplet size manifests itself when a sensor is looking through "real rain".

The sensitivity to the drop size distribution
Note: In this section, full detail has been omitted for readability. For example, in some of the equations, mathematical constants have been combined and their values not explicitly given.

There are numerous studies in the meteorological literature that attempt to parameterise the number of raindrops with droplet size (e.g. Ulbrich[20] 1983). This is referred to as the *drop size distribution* (DSD) and is usually parameterised by a mathematical function, a common form of which is the *gamma* function:

$$N(D) = N_0 D^\mu \exp\left[-(3.67 + \mu)\frac{D}{D_0}\right]$$

Where $N(D)$ is the number density (in m^{-3} cm^{-1}) for a droplet diameter D (cm), N_0 is a scaling factor related to the total number concentration, D_0 is a median volume drop diameter, and μ is a shape factor that influences the relative prevalence of large and small droplets.

The above form makes use of the relationship between the terminal fall velocity of a rain drop and its size proposed by Atlas and Ulbrich (1977)[21]:

$$v(D) = 17.67 D^{0.67}$$

It is noted here that this parameterisation is one of a number in the literature and would have been particularly suited for linking weather radar returns to measurable rainfall rate. Other studies such as van Boxel (1998)[22] specifically consider the acquisition of terminal velocity as an important consideration for desired height of a physical rainfall simulators.

The parameterisation $N(D)$ is powerful because it enables the linking of one measurable bulk property of rainfall to another. Many of these 'measurables' can be expressed as in the following general integral:

$$P = a_p \int_0^\infty D^p N(D) dD$$

where D^p is the raindrop diameter raised to the pth power and a_p is a constant that ensures the correct unit. For example:
- when $p=6$ the integral is the radar reflectivity Z (assuming Rayleigh scattering)
- when $p=3$, the integral is summation of the volumes of all the spherical raindrops and with the appropriate a_p, this yields the liquid water content

---

[20] Ulbrich, C. W. (1983) :Natural variations in the analytical form of the raindrop size distribution," *J. Climate Appl. Meteor.*, **22** (10), 1764–1775

[21] Atlas D. and Ulbrich C.W 1977: Path- and area-integrated rainfall measurement by microwave attenuation in the 1-3 cm band. *J. Appl. Meteor.*, **16,** 1322-1331

[22] Height of shed paper: van Boxel, J. (1998). Numerical model for the fall speed of raindrops in a rainfall simulator. *I.C.E. Special Report*, *1998/1*, 77-85

- when the volume of each droplet is multiplied by its fall velocity, giving $p=3.67$, the integral will produce the measured rainfall rate
- when $p=0$ the integral yields the total number of raindrops in a measured volume

Thanks to the well-behaved nature of exponential functions during integration, it is possible to derive very simple power law relationships between e.g. the radar reflectivity and the equivalent rain rate. This is the basis of using long wavelength radars for rain measurement in meteorology.

In order to calculate the attenuation or extinction coefficient values as an EM wave passes through the volume of rainfall, it is necessary to replace the simple powers of D with the Mie extinction efficiency $Q_e(D)$ multiplied by the cross-sectional area of the droplet, thus:

$$k_{ext} = a_p \int_0^\infty Q_e(D) D^2 N(D) dD$$

As can be seen in Fig 1, $Q_e(D)$ is highly variable with D. In this study, Q_e was calculated at a range of discrete droplet sizes using an extension of the Met Office's in-house Havemann-Taylor Fast Radiative Transfer Code (HT-FRTC)[23] to create a look-up table of values, which allowed for rapid computation of total extinction when combined with a range of DSDs, alongside their equivalent rainfall rates. This was performed for both lidar and mm-wave wavelengths, and also the C-band wavelength (5 cm) employed by the Met Office's operational radar system as an additional quality check (which is not reported on further here) and also compared to those from the literature[24].

A note about units and notation

This collaborative study brought together experts from the disciplines of Meteorology, Physics and Engineering, and with them their respective preferences for units and notation. In the treatment of attenuation both linear and logarithmic units are used in this report. The following provides traceability between these.

The walkthrough of Appendix C and Section 6.2 of the main report, employs a one-way attenuation coefficient γ in km^{-1}. The notation γ is interchangeable with the extinction coefficient, k_{ext}, which is more commonly used in the field of meteorological (passive) remote sensing and radiative transfer.

When considering the cumulative effect of transmission along a path through a medium of varying attenuation coefficient, it is more straightforward to work in a logarithmic unit because it enables a more simple additive approach. Also, as the sensors considered in this limited pilot are both active, therefore following an emitted *and* returned path, a 2-way attenuation value is useful. Much of the analysis in this Appendix therefore employs both a 2-way attenuation coefficient A in dB/km and a total 2-way attenuation in dB.

The conversion between A, k_{ext} and γ is:

$$A = \frac{20}{ln10} k_{ext} = \frac{20}{ln\,10} \gamma$$

[23] Havemann S, J-C Thelan, J.P. Taylor & C. Harlow (2018), The Havemann-Taylor Fast Radiative Transfer Code (HT-FRTC): A multipurpose code based on principal components, *J Quant. Spect and Radiative Transfer*, **220**, 180-192.

[24] Van de Hulst (2003), *Light Scattering by Small Particles*, Dover Publications, pp470, ISBN 978-0486642284

The impact of the DSD to rainfall relationship

Careful thought must be taken when applying DSDs from the literature to the CAV sensor domain. As described earlier, the non-linear Mie response means that the total attenuation may be very sensitive to the detail of the DSD, especially where the droplet size is close to the wavelength. Many of the DSDs in the literature were developed to give good agreement with empirically derived reflectivity-to-rain rate relationships and are heavily influenced by larger droplets due to the D^6 relationship. Also, it is important to be aware that there may be considerable variability over the short time and space scales relevant to the CAVs (more on this later). However, there is still much to be learned from the implications of the documented sets of N_0, D_0 and μ.

The μ term is the most instructive to discuss here. Fig 3 demonstrates the how the shape of the DSD, in particular the relative population sizes of the raindrops impacts the measured rainfall rate and the extinction parameters.

Figure 3 (overleaf): How μ affects key parameters. (a) The droplet number density as a function of diameter (in log-linear form) for 3 DSDs with a different μ values. μ=2 is more representative of convective rain (showers) and μ=-2 is considered to be more representative of orographic rain, where more of the rain volume is in the form of smaller drops. The widely used Marshall Palmer (1948)[25] distribution corresponds to μ=0 (b) The relative contribution of the different droplet sizes to the measured rainfall rate. (c) The relative contribution of the different droplet sizes to the extinction at lidar wavelengths. (d) As (c) but for a 77 GHz mm-wave radar. The liquid water content (i.e. the total mass of water contained in 1m3 of air is the same for all of these and, as can be seen by plot (b), the rainfall rates are also effectively identical
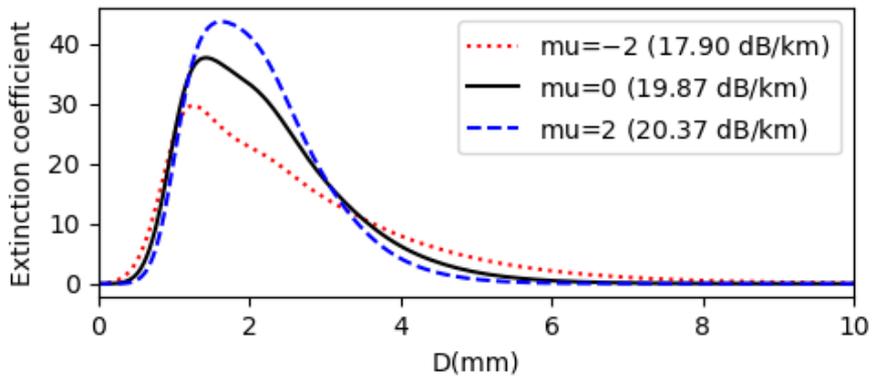
[25] Marshall, J.S., and W.M. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165-166
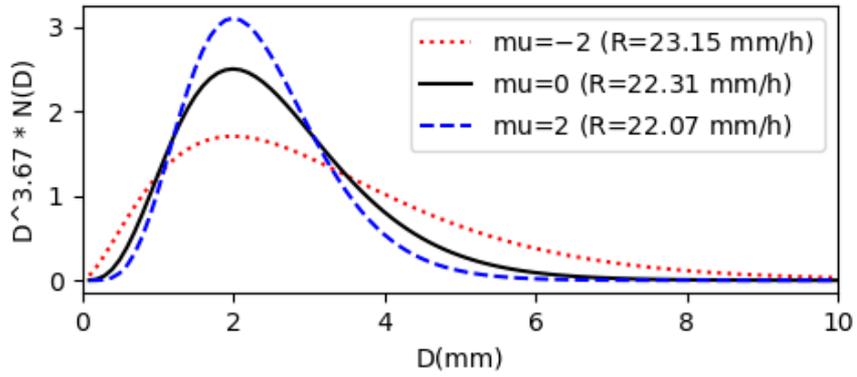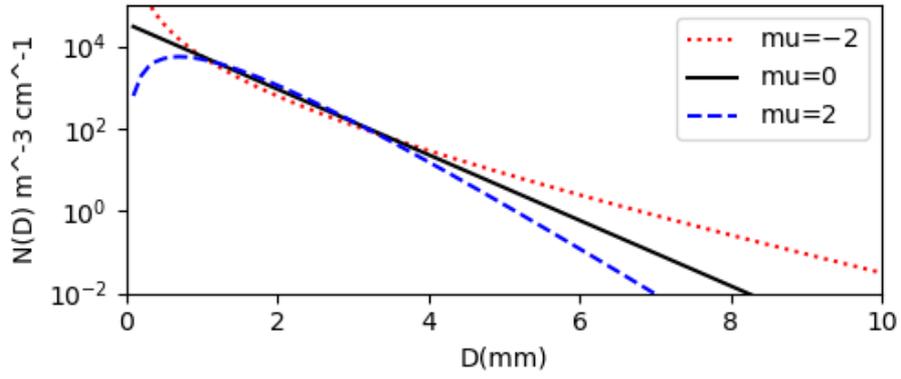
0.9 um
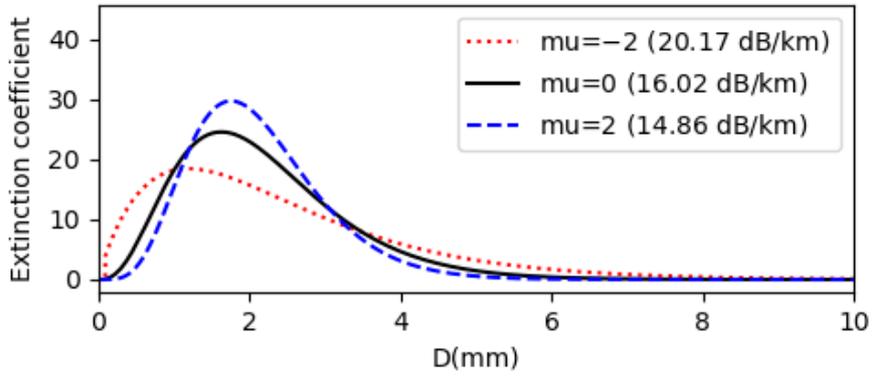
77 GHz

The gamma function parameters are influenced by the environmental factors relating to the rainfall production. It is useful to think of 3 broad categories of rainfall:

Frontal (or stratiform or dynamic) – produced by large scale weather systems (μ is mostly positive)
Convective (showers, thunderstorms) – produced by atmospheric instability. For more vigorous events, very large rain drops are possible with much fewer smaller ones (μ is usually in the range 0 to 2)
Orographic – produced or significantly enhanced by forced assent over hills and mountains with larger numbers of smaller drops (μ is negative)

The Marshall Palmer (1948) distribution sets μ=0 and was optimised to ensure good agreement between radar reflectivities and measured rainfall rate (and therefore troubled itself very little with the accurate representation of smaller droplets).

The following plots demonstrate the net impacts of the DSD choice when comparing the attenuation values for a lidar and mm-wave radar simultaneously. Firstly, Fig 4 shows the sensitivity of the calculated 2-way attenuation A for lidar and radar for a range of drop sizes derived from Ulbricht. For visual clarity this has only been done for a small set of rainfall values so that the spread of possible values of attenuation simply due to DSD can be appreciated.
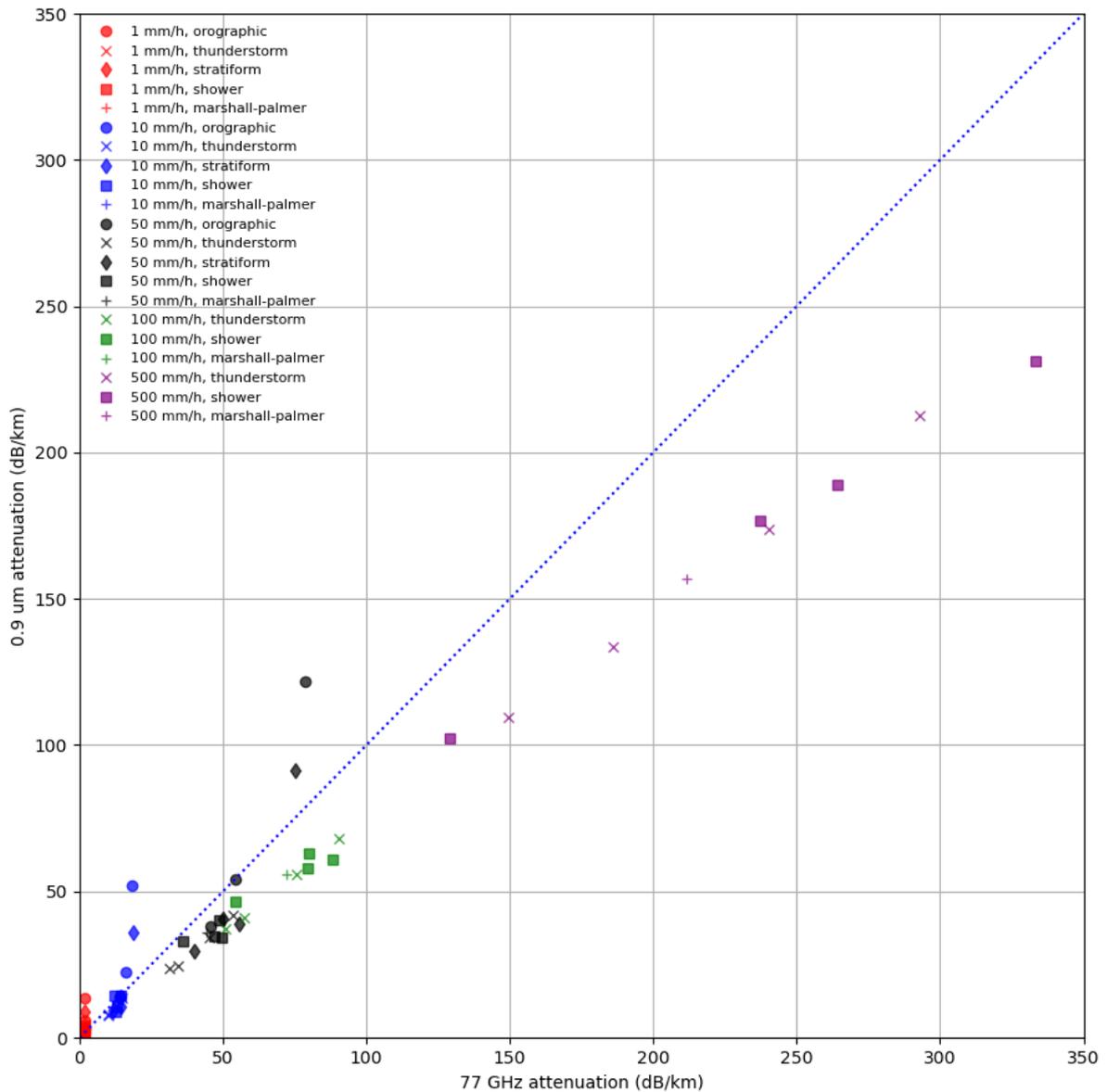
Figure 4: Scatterplot of 2-way attenuation for lidar (y-axis) and radar (x-axis). The plot shows how attenuation varies for 5 different rainfall rates (1, 10, 50, 100 & 500 mm/hr), with the scatter at each rain rate due to the assumption of DSD. The shapes refer to the type of rainfall, which will have a strong influence on the distribution of rain across smaller and larger sizes

What is immediately obvious is that a single rainfall rate is capable of generating a very wide range of attenuation values.

Fig 5 zooms in on this and highlights the spread of attenuations that are achievable with a single rain rate simply due solely to the redistribution of the rain mass across the range of droplet sizes.
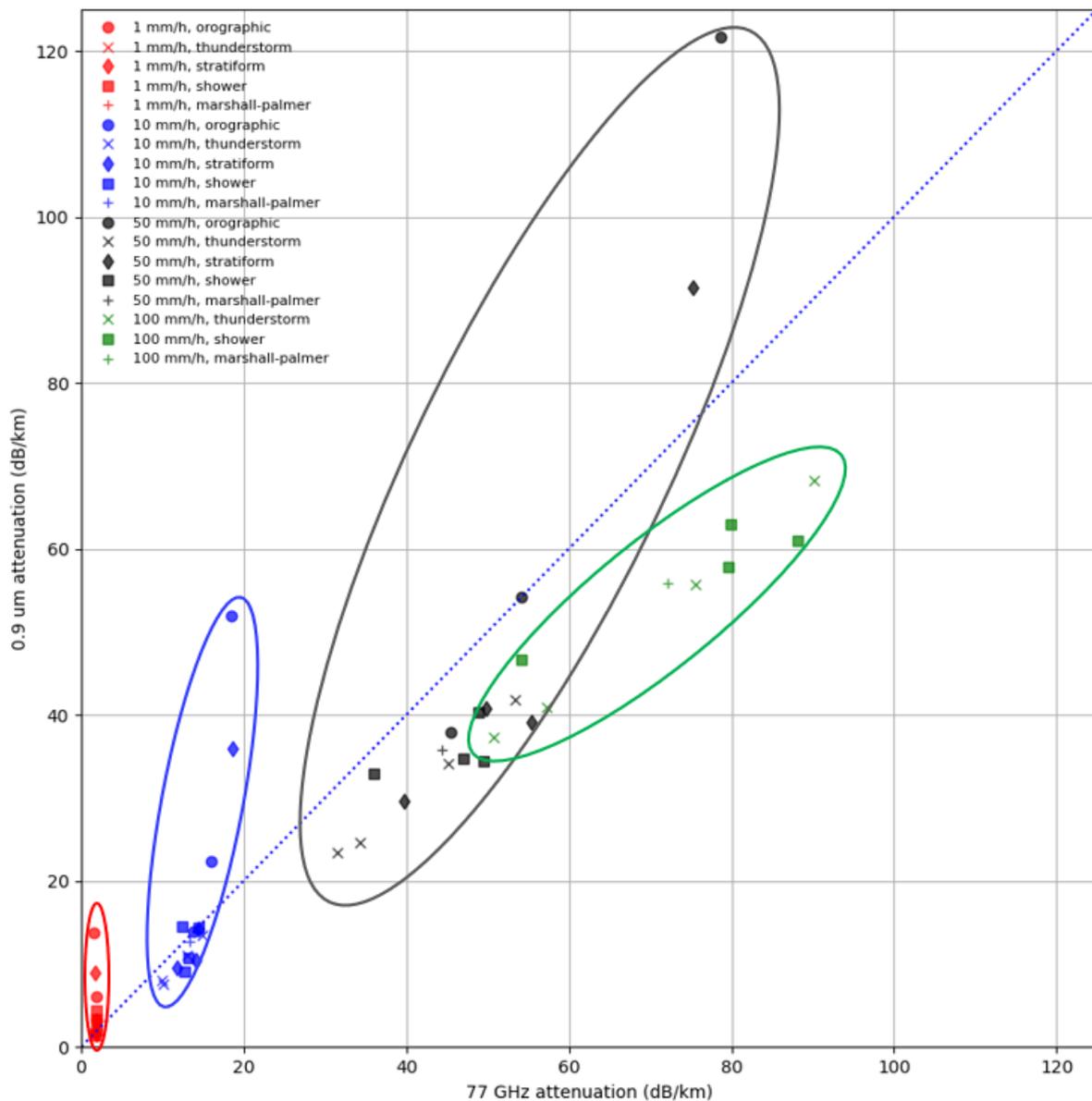
Figure 5: The same plot as previously but focussing on rain rates up to 100 mm/hr. The ellipses highlight the spread of points for a single rain rate due the change in drop size distribution

Some commentary is necessary at this point.

1. For the lowest three rain rates plots (1, 10 & 50 mm/hr), the size of the spread increases with rainfall rate but then appears to reduce at 100 mm/hr. This is because between 50 and 100 mm/hr we have removed the orographic and frontal (stratiform) rain from the list of DSDs that can be plotted, as they are unlikely to produce such high rain rates. There is therefore less variability available from the differing DSDs considered here.

2. The higher rainfall rates result in more larger droplets, which for the lidar wavelengths will result in a significant proportion of strong forward scattering. It is likely that this may give an overestimate of the effective attenuation for the lidar compared to the radar because the scattered energy is still available within the beam. However, this will dependent on the detail of individual sensor (such as its sensitivity to pulse spreading and its angular resolution). Nevertheless, the headline attenuation figures are a useful guide as to the relative magnitude of impact on sensors as a function of the DSD and the rainfall rate.

3. The plots give an insight into how uncertainties in the characterisation of the sensors in the test environment may be correlated between sensor types. Simplistically, if a central value from the ellipse in Fig 5 is used then the ellipse will be indicative of the correlated error spread if that value is assumed to be true. Careful inspection of the 100 mm/hr case reveals that the Marshall Palmer DSD is roughly central to the spread and is therefore used as the basis for the walkthrough of the golden thread.

The following concluding statements are made as a result of the above analysis and are included in the main body of the report.

1. **The drop size distribution (DSD) experienced by the sensors under test is such a strong driver of sensor performance that it must be measured accurately – it is not sufficient to merely characterise the rainfall in terms of its rate in mm/hr. As the DSD shape is strongly influenced by the type of rainfall, this study recommended that BSI PAS 1883 includes rainfall type in the ODD taxonomy.**

2. **It is necessary to expose sensors to a variety of DSDs in order to assess their performance fully. This implies a need to access a wide range of rainfall conditions, either naturally in the external environment and/or with a high level of control in a CETF.**

3. **It <u>cannot</u> be assumed from the outset that a CETF will be capable of producing the full range of naturally occurring DSDs. It is essential to have an external testbed, the purpose of which will be one or both of (a) verifying the realism of the CETF DSD and corresponding sensor response and (b) complementing the data from a CETF by combining the information in a Bayesian sense.**

The impact of non-spherical raindrops on the attenuation
The analysis to this point has made the assumption of spherical raindrops. In the real world, raindrops deform as they fall due the aerodynamic forces on the droplet and the effects of surface tension, with the drop flattening along the axis in the direction of fall. This distortion increases with droplet size until the droplets become too large and break up. Fig 6 shows two parameterisations from the literature for the variation in droplet aspect ratio with radius. Both assume the raindrop adopts the shape of an oblate spheroid, itself a simplification. While they differ in their detail at very large drop sizes, it should be noted that beyond 6 mm diameter, the numbers of droplets are so low that their contribution to total attenuation is small. Below this size the distortion is similar.
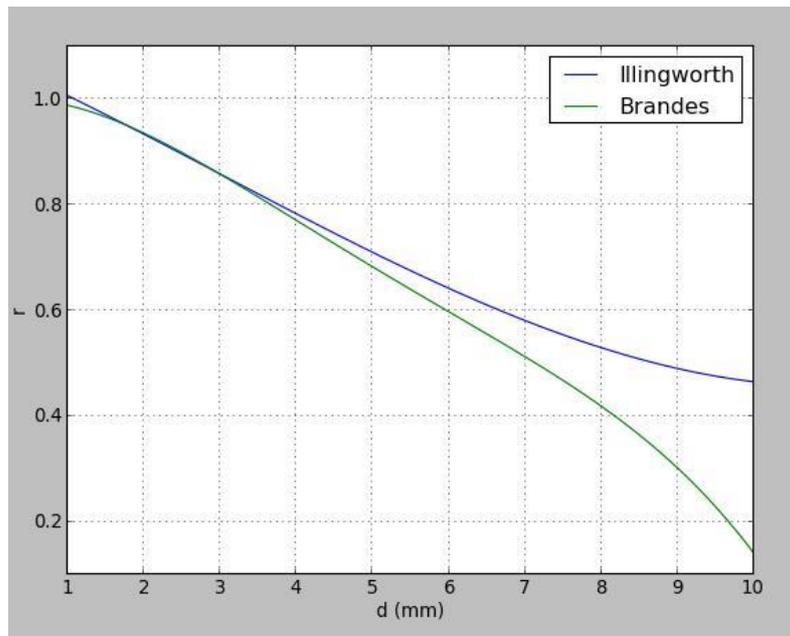
Figure 6: Axial ratio of raindrops versus diameter according to two parameterisations from the literature, Illingworth and Blackman[26] and Brandes et al[27]. While there is disagreement at the larger droplet sizes, the effect on extinction at these sizes will be relatively small due to the low numbers of droplets of this size

The droplet flattening means that the extinction of the wave becomes dependent on the polarisation of the incoming beam with respect to the droplet. Put (over)simply, as the drop is squashed in the vertical direction, radiation will "see" a smaller drop dimension in the vertical (V) polarisation than it will in the horizontal (H) and is consequently attenuated less. This can be demonstrated by using T-matrix code[28] for oblate spheroids instead of the spherical Mie calculation.

Fig 7 compares the 2-way attenuation as a function of rainfall rate for both spherical and oblate raindrops at 77 GHz. For the spherical case, V & H polarisations are identical. For oblate drops the V-polarisation undergoes less attenuation.

As drop distortion is an inherent part of the process reaching terminal velocity, this further highlights the need to pay particular attention to the height of the CETF if we aspire to maximise the realism of the generated rainfall. (The reader is reminded that this analysis only considers the attenuation of the beam – the apparent droplet size in the orientation of the sensor polarisation may also have an impact by presenting a larger backscattering cross-section. This may reduce the signal to noise ratio of the return from a target object with respect to its immediate environment.)

[26] Illingworth A.J. and T.M. Blackman, 2002: The Need to Represent Raindrop Size Spectra as Normalized Gamma Distributions for the Interpretation of Polarization Radar Observations, *J. Appl. Meteor.,* **41** (3), 286-297.

[27] Brandes, E.A, G. Zhang and K. Vivekananda, 2004: Drop Size Distribution Retrieval with Polarimetric Radar: Model and Application, *J. Appl. Meteor.,* **43** (3), 461–475.

[28] Waterman PC., 1971: Symmetry, unitarity, and geometry in electromagnetic scattering. *Phys Rev D* 1971;3:825–39. (Reprinted in: Kerker M, editor. Selected papers on light scattering (SPIE Milestone series, vol. 951). Bellingham, WA: SPIE Press; 1988. p.811–2)
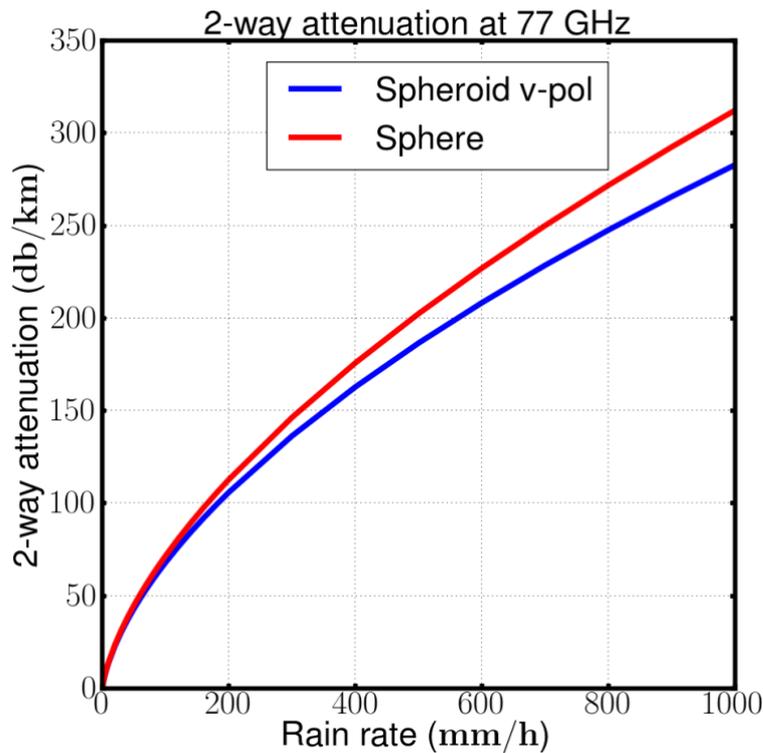
Figure 7: 2-way attenuation as a function of rainfall rate, assuming a Marshall-Palmer DSD, at 77 GHz. The red line is that calculated for spherical droplets, which is independent of polarisation. The blue is value for the V-polarisation if droplet distortion according to Illingworth and Blackman (2002) is factored in using T-matrix

The sensitivity to droplet shape is not pursued further in this report and spherical drops are assumed for remainder as this is a sufficiently robust assumption to explore the remaining effects. For completeness it is also noted that there is also a temperature dependence of the attenuation coefficient at 77 GHz due the variation of the dielectric constant of water with temperature. While this is relatively small compared to the variation due to DSD[29] exhibited in Figs 4 & 5, it is recommended that its contribution to the KPI uncertainty budget is explored in further work.

Recap: The impact of rainfall on CAV sensor channels is sensitive to the underlying detail of how that rainfall volume is distributed across possible droplet sizes. In the natural world this will be linked to the mechanisms that produced the rainfall in the first place. In a CETF this will be determined by the level of control over the generated DSD. In either case the DSD must be measured in the different test environments. It is unlikely that all relevant control parameters of the attenuation will be known when the CAV is on the road, and this must therefore be reflected in the error budget when determining if a CAV sensor is within its ODD.

The variability of rainfall rate (and by implication, DSD) at different scales
The preceding analysis demonstrates a sensitivity to the detail of the rainfall *assuming a single steady-state rainfall rate*. As the ultimate application of the sensor assurance process will be to determine if a CAV sensor is within its ODD in an operational sense, it is essential that this rainfall definition can

---

[29] *The effect of the built and natural environment on millimetric radio waves* (February 2018), Report for DCMS, published by Ordnance Survey, available from https://www.gov.uk/government/publications/ordnance-survey-5g-planning-and-mmwave-environment-reports

be interpreted unambiguously so that any 'safe thresholds' can be compared to real world values that are either measured or provided from an external source.

In this section the fractal nature of rainfall variability (both in time and space) is explored and its implications for rainfall definition are considered. Firstly the definition of rainfall rate is discussed.

Rainfall rate is usually expressed as an average intensity in the units of mm/h, corresponding to the depth of rain water collected over a known surface area divided by the time interval over which it is collected. However, this unit applies *regardless* of the size of the collection area and the time interval over which it is evaluated; the following average intensities all share the same unit:

- the 1-minute average rainfall intensity measured by a standard meteorological rain gauge
- the UK-wide average annual surface rain rate evaluated over a 10 year period
- the instantaneous rainfall rate observed by a meteorological radar pixel of the order of 1 km^2

i.e. the unit does not seek to capture the variability over the time interval and the area of evaluation; in particular it does not contain information regarding peak values of high intensity rainfall within the sample. This is discussed in depth in the report by Dixon (2018)[30].

When considering attenuation, for example, a CAV sensor will typically be concerned with near instantaneous rainfall distributed over a path length of order 10-200 metres. The degree of sensitivity to the precise detail of this rainfall over that path length will be dependent on the degree of non-linearity between attenuation and rainfall rate – a linear relationship will be less sensitive to using an average rainfall value to estimate path attenuation.

While there is no single conversion formula between rainfall measurements, much *is* understood regarding the relationship between the different scales and this is now applied to the CAV sensor characterisation challenge through the consideration of a simulated case study derived from a real-world event. As there are a number of important steps, it is worth mapping out the approach taken end-to-end first:

1. **Case study:** A significant rain UK event is identified from the Met Office's archive of operational high rainfall radar data.
2. **Scaling factors**: A statistical description of the variability within the radar picture is determined, yielding "universal" scaling parameters that describe the variability of the rainfall value at both higher *and* lower resolutions to that observed.
3. **Small scale simulation**: These factors are then used to *simulate* realistic rainfall variation at progressively smaller scales within a single observed 2x2km radar pixel in order to demonstrate the "hidden" variability that is relevant to a CAV sensor.
4. **Conversion to attenuation**: At each of these different scales, these rainfall values are converted to 77 GHz 2-way attenuation values, using the extinction coefficient calculations described previously. This provides insight into the possible differences between the real attenuation (i.e. CAV scale) and that which might be inferred from the rainfall rates provided by meteorological observing networks.
5. **Impact on the KPI**: By considering a range of sensor to target distances, the impact of these differences on the interpretation of the maximum range KPI are demonstrated.
6. **Practical approaches to resolution**: Two different approaches to managing the complexities of variability are introduced in brief.

Each step will now be described in turn.

---

[30] Dixon, J (2018) Report for Innovate UK *Short-period rainfall extremes in the UK*, available from
https://www.metoffice.gov.uk/services/transport/cav

Case study

As part of its role as the UK National Meteorological Service, the Met Office produces analyses of instantaneous surface rainfall over the UK every 5 minutes. These are derived primarily from operational C-band radar data, supplemented where necessary with satellite data. These data are archived, allowing access to recent historical real weather events. The UK is served by 15 radars, complemented by others in Ireland and the Channel Islands as indicated in Fig 8.
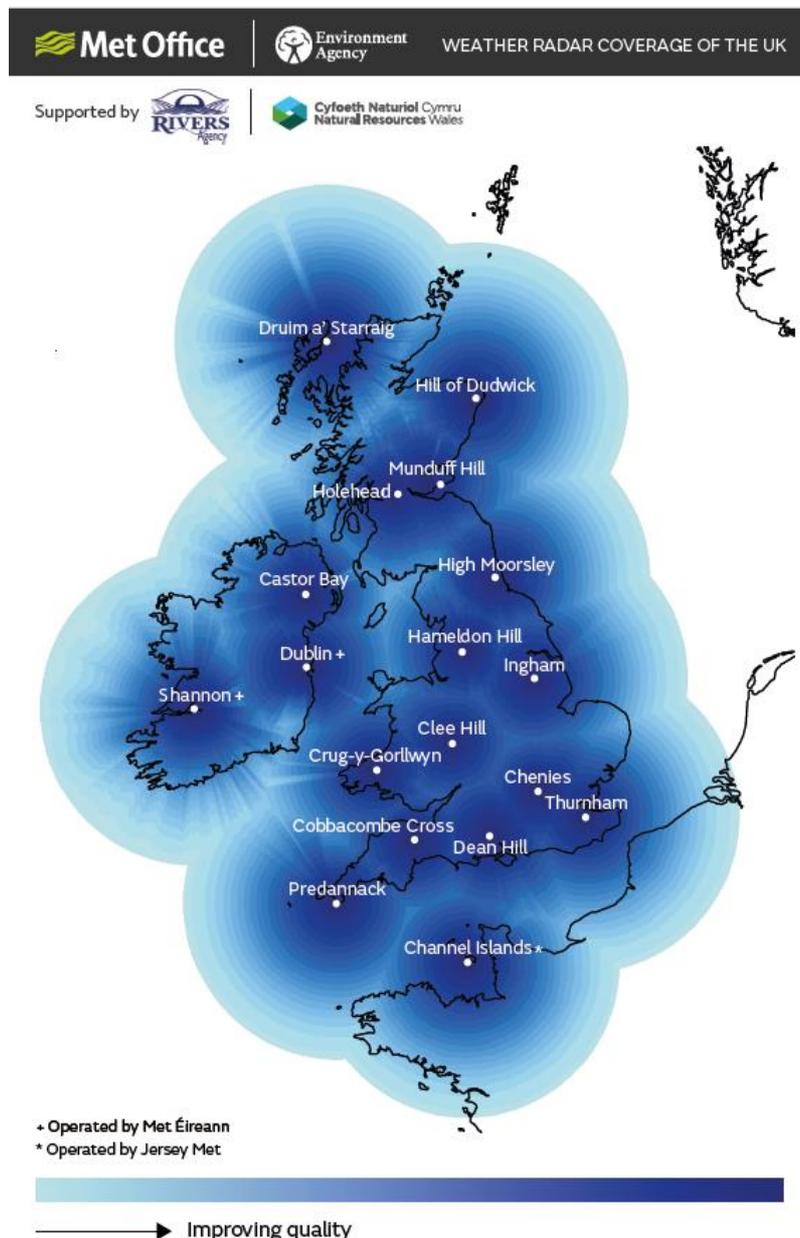

Figure 8: The UK weather radar network and its coverage quality. Observations taken nearer to the radar sites are typically closer to the ground and of higher implicit resolution, thereby improving the quantitative estimate of surface precipitation

The combination of scanning geometry, sampling regime and antenna pattern of the radar, means that the spatial resolution of the radar and the height at which the precipitation is sampled is a function of location of the measurement with respect to radar. It also implicitly is a measurement in polar coordinates. For that reason, products are generally mapped onto a Cartesian grid which are

available in real time at granularities of 500 m and 1 km. They are also archived for general use at a resolution of 2 km, which we take as our starting point for the spatial analysis of this study.

The case study chosen is the large scale convective event of the 23rd June 2016, which brought significant disruption to London and SE of England as it was already used to demonstrate the impact of rainfall on 5G networks in a joint study with Ordnance Survey and 5G Innovation Centre[31], which relates closely to this challenge as 5G wavelengths are close to those of CAV sensor.

Fig 9 below provides one frame of the radar data from this case
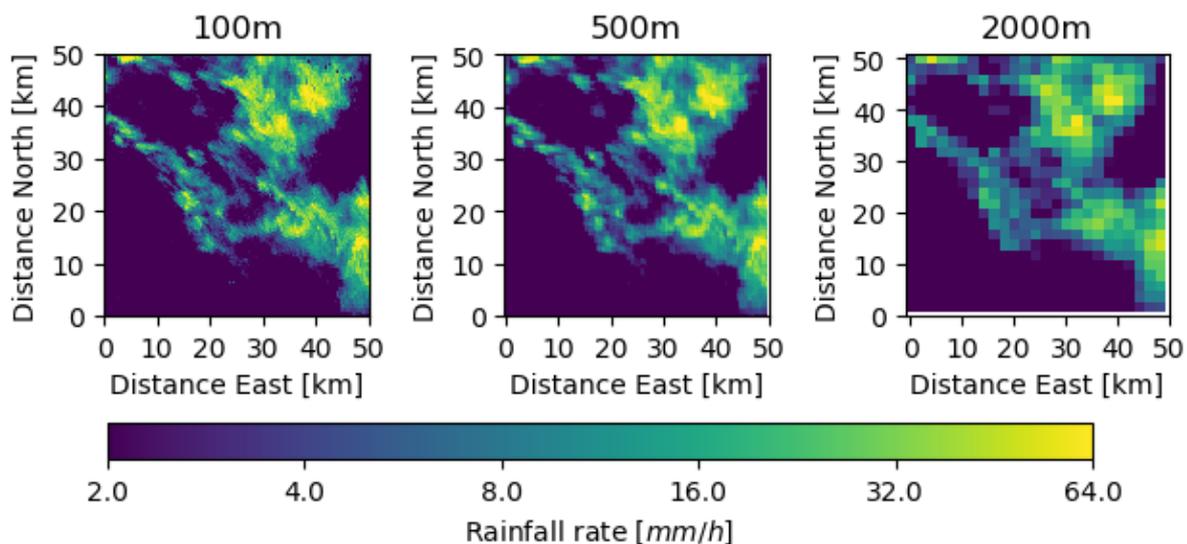

Figure 9: Radar rainfall rates for the 23rd June case for a 50x50 km box. Due to the proximity of the nearest radar, it is possible to generate a meaningful 100 m resolution mapping as well as those at 500 m and 2000 m. Careful inspection shows how the larger averaging area removes detail and extreme rainfall values

Scaling factors

Starting from a standardised gridded rainfall product, it is then possible to infer the small scale structure of relevance to the CAV through the consideration of the fractal nature of turbulence in the atmosphere.

Turbulent wind flow structures exist over a wide range of scales in the atmosphere, from the planetary ~10,000 km scale to the 1 km scale, and even down further to the millimetre scale. These fluctuations in wind velocity, when viewed as "spectra" in the frequency domain, can usually be characterized by simple "scaling laws"; typically the wave power decreases exponentially with decreasing length scale, as described by the Kolmogorov (1941)[32] theory.

Rainfall intensity is a "passive tracer of turbulence", that is it is affected by the turbulent motion of atmospheric eddies but has no effect on the turbulence itself. As a result it can also have scaling

---

[31] *The effect of the built and natural environment on millimetric radio waves* (February 2018), Report for DCMS, published by Ordnance Survey, available from https://www.gov.uk/government/publications/ordnance-survey-5g-planning-and-mmwave-environment-reports

[32] Kolmogorov, A.N., 1941. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. In *Dokl. Akad. Nauk SSSR*, **30**(4), pp. 301-305.

behaviour; as demonstrated recently at scales down to ~1 m by Lovejoy and Schertzer (2008)[33]. The scaling factors for this case study were determined, thus allowing the simulation of spatially realistic rainfall fields at scales both larger and smaller than those observed by the radar. Figure 10 shows simulated rainfall fields equivalent to those in the previous figure.
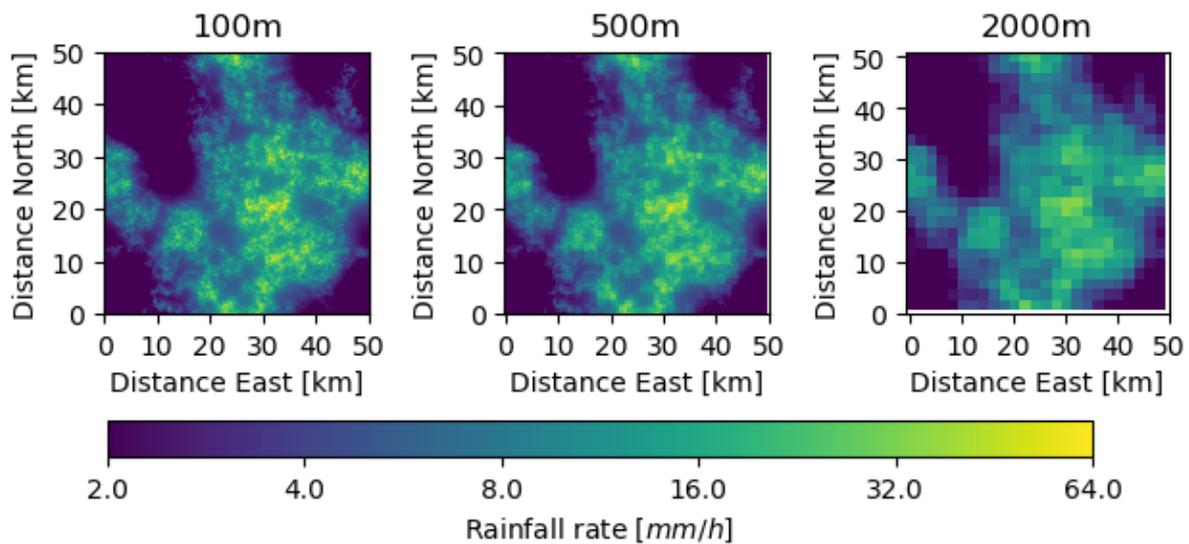

Figure 10: Simulated radar rainfall rates for the 23rd June case for a 50x50 km box using the scaling factors derived from the real radar data. The two figures should not be expected to look the same, however the variability of rainfall rate at the different scales are statistically very similar. The simulation is there considered to be a well understood version of real rainfall for the purposes of this study

Small scale simulation

Now that the scaling factors are known, it is possible to explore the variability of the rainfall on increasingly smaller scales. In this study we start with a single radar-derived rainfall intensity in a 2x2 km pixel (Note there is a degree of arbitrariness about this, it could equally have been 1x1 km). The important point is that it represents a sensible "observed scale" that is significantly larger than the path length of interest to a CAV.

<u>**Important note:**</u> **From this point on the colour scales for the plots move from a log-scale to a linear scale and are not directly comparable with the previous ones.**

The starting point for the analysis is a "simulated" 2x2 km pixel (Fig 11) with an average rainfall intensity of 50 mm/h, which is considered to be very heavy rainfall over this averaging area.

---

[33] Lovejoy, S. and Schertzer, D., 2008: Turbulence, raindrops and the l1/2 number density law. *New Journal of Physics*, **10**(7), p.075017.
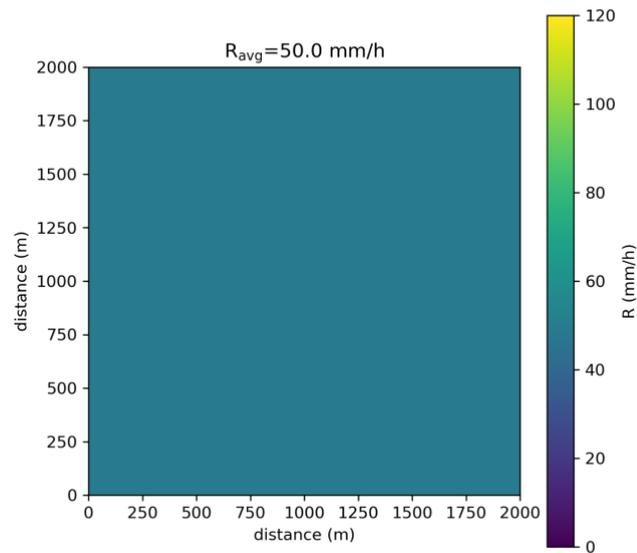
Figure 11: A 2x2km radar pixel with an average rainfall intensity of 50 mm/hr

The scaling factors are then used to generate a statistically valid representations of rainfall at progressively finer resolutions, by subdividing the pixel into smaller areas, all of which would combine to a 2x2 km average of 50 mm/hr [34]. Fig 12 shows a number of these realisations.

---

[34] Strictly speaking, the very highest resolution version (corresponding to 1 m pixels) was generated first and all intermediate resolutions produced by progressively averaging over larger areas, however this has no bearing on the discussion.
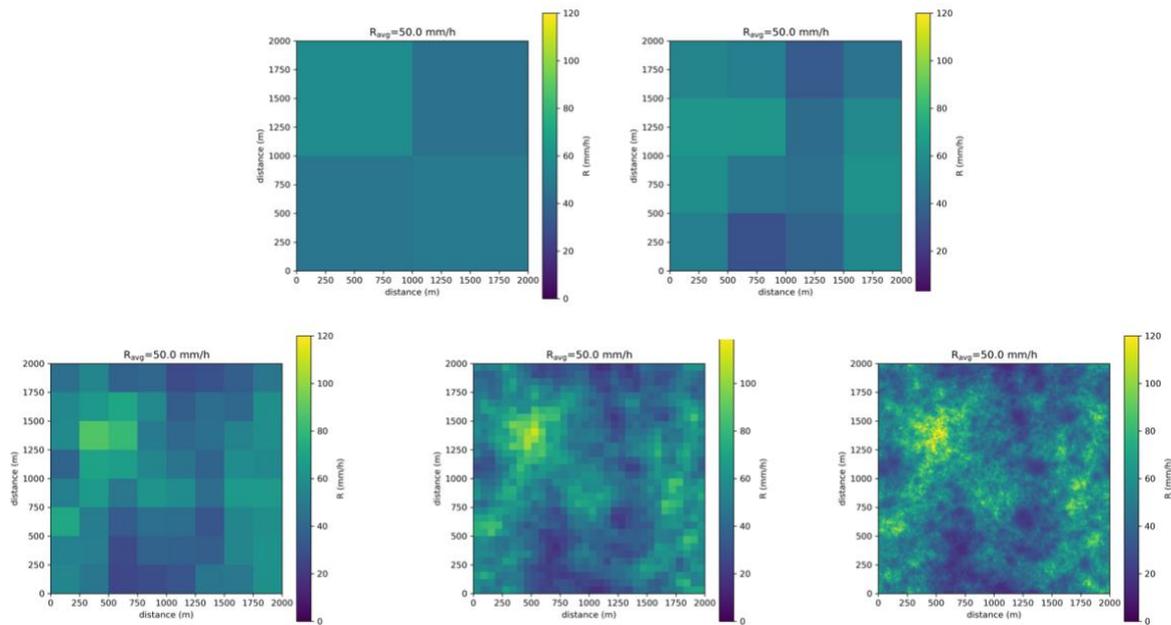
Figure 12: The same 2 km rainfall pixel divided into smaller subpixels, all with a 2x2 km average intensity of 50 mm/h. The approx. pixel sizes are (top left to right) 1 km & 500 m and (bottom left to right) 250 m, 62 m & 1 m

For completeness, it is noted that this scaling relationship approaches its limits at scales below 1 m. This is because the inertia of individual raindrops means that they do not follow the small scale turbulence, which are the driver of the scaling laws. Also at these smaller scales, the time variability of the rainfall (as they fall as individual droplets through a smaller surface area) also becomes an issue. This has to be borne in mind when considering the extent to which these techniques are used to provide simulated weather in a virtual test environment. As before, at this stage these simulations are not offered as fit for the purpose for use in a simulator without further investigation.

Conversion to attenuation

The previous section discussed how attenuation values can be estimated for a given rainfall intensity by assuming a functional form of the drop size distribution (DSD) with rainfall rate[35]. The grids of rainfall rates above were therefore straightforwardly converted to attenuation values assuming the Marshall-Palmer DSD. This was chosen as a DSD which is considered to the 'middle of the pack'. Others will yield different absolute values. Also it is acknowledged the exact functional form of DSD can varying across a single weather system. Care is taken therefore to ensure the conclusions drawn from this simulation, namely the need to pay careful attention to spatial scales, are robust to these assumptions. Fig 13 shows the attenuation values corresponding to the rainfall plots above for a 77 GHz automotive radar.

---

[35] It is stressed again here, that these calculations will produce an attenuation that ignores the fact that some of the signal scattering may be in a strongly forward direction and may therefore represent an overestimated effective attenuation compared to real world, most particularly for lidar and visible wavelengths. This was a factor in deciding to focus on 77 GHz radar.
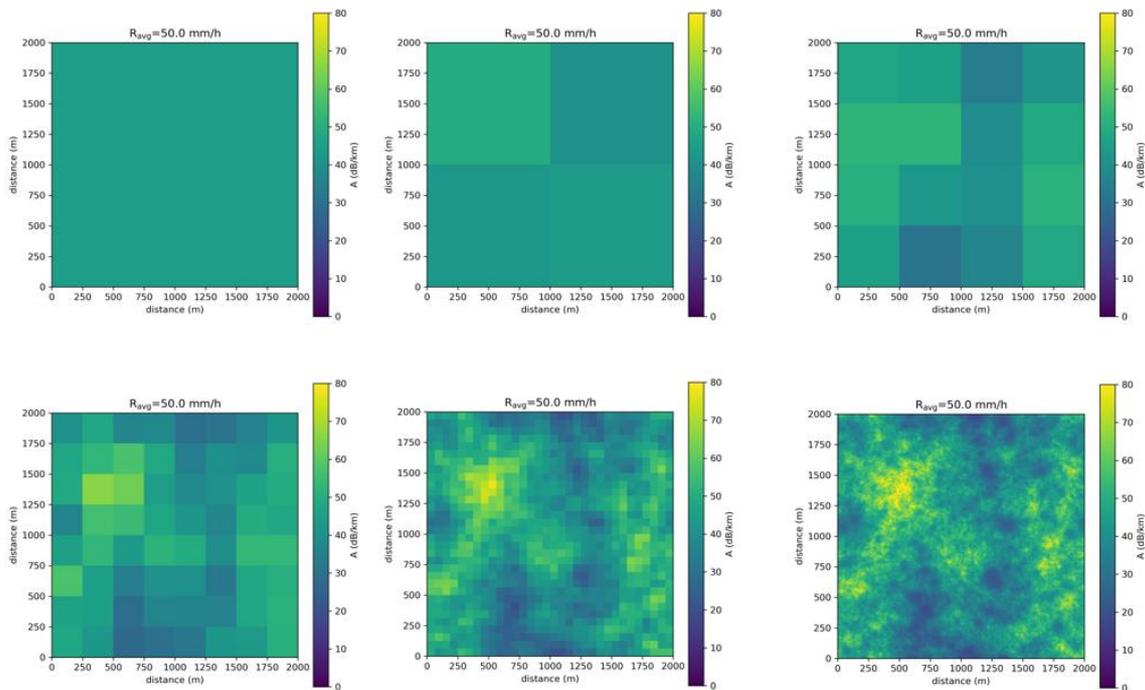
Figure 13: 77GHz 2-way attenuation values within a 2 km area, corresponding to the previous figures. The approx pixel sizes are (top left to right) 2km, 1 km & 500 m and (bottom left to right) 250 m, 62 m & 1 m

It is clear from the figure above that, depending on where within the original 2x2 km a CAV is situated, the low resolution "observed" rainfall rate (and therefore calculated attenuation) may be significantly higher or lower than that actually encountered. This is demonstrated by the next figure.

Impact on the KPI

Consider a CAV travelling from the left to right across the example 2x2 km square, through the most intense part of the rainfall (centred approximately 500 m East and 1400 m North of the origin). Fig 14 demonstrates how the rainfall rate in the location of the CAV (and therefore the 2-way attenuation at 77 GHz) can significantly exceed that which might be inferred from the lower granularity "observation". A worst case 200m section is demonstrated through a local peak of rainfall.
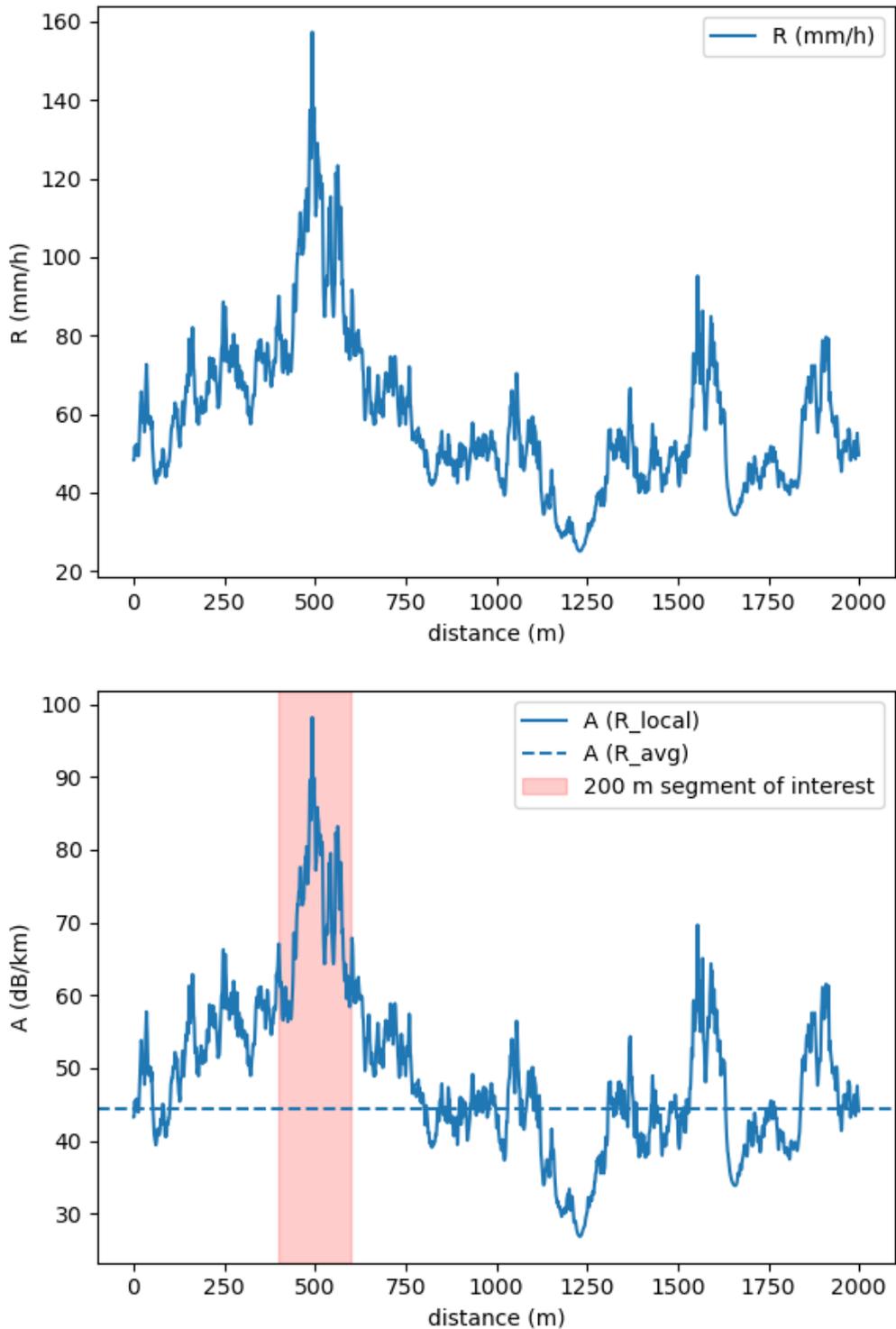
Figure 14: (Top) A section of high resolution rainfall rates representing a the section along '1400m North' through the 2x2 km sample area. This represents "truth" in simulation
(Bottom) The calculated 2-way attenuation values from those rainfall values. The dashed blue line is the attenuation value calculated from the 2x2 km average of 50 mm/hr. The shaded pink area represents a 200 m section through this peak, corresponding to a reasonable sensor-to-target viewing range

The implications of this can be interpreted as follows:

1. Consider a 77GHz radar mounted on CAV driving along a road during a heavy rainfall event. The CAV's ADS system receives an external feed of radar-derived rainfall data at a resolution of 2x2 km, indicating that the car will experience an average rainfall intensity of 50 mm/h.
2. At its current speed CAV is deemed to be safe in conditions where the maximum range KPI for the radar is greater than 200 m.
3. Using a combination of external testbeds and controlled environmental tests facility measurements, the 77 GHz radar has been demonstrated to be able to function normally when the total 2-way attenuation due to rainfall is 9 dB or less. The 2-way attenuation through a uniform rate of 50 mm/h over 200m is **8.88dB** and so the CAV is deemed to be within its ODD at 50 mm/h.
4. Within the 2x2 km area, the rainfall demonstrates significant variability, and the CAV experiences a local peak in intensity.
5. As the CAV encounters the high rainfall area, it experiences spot rainfall rates exceeding 150 mm/hr. The maximum total attenuation over the 200 m range was **13.90 dB**.
6. The CAV exceeded its ODD.

It is essential therefore that when comparing the rainfall intensities experienced in the testing process, which will inherently be small scale, with those observed/estimated/forecast in the *operational* environment that these spatial effects are properly accounted for, either explicitly or within the error budget. **This has been reflected explicitly in the description of rainfall in PAS 1883**.

A few notes about temporal variability
1. The example uses weather radar (or equivalently any grid of spatially averaged intensities) to demonstrate the need for care when evaluating the ODD. Another standard form of meteorological measurement is the rain gauge. As discussed in Dixon (2018), this samples average rainfall intensity over the order of a minute (or more) but over a much-reduced sampling area of order ~0.1 m^2. It is possible to construct linkages between radar rainfall values and rain gauge values (for example by advecting high-resolution spatial simulations over a virtual rain gauge, or by comparing rainfall and radar accumulations over longer periods such as an hour), however care must be taken as ultimately this will depend entirely on the rate of motion (and speed of development) of the weather system producing the rain.
2. This example above explores the issues around spatial variability where the meteorological rainfall monitoring system is observing on a large spatial scale. For completeness, it is noted that there is an inherent assumption that the CAV radar is "seeing" a constant (in time) sample of this rainfall i.e. its beam is sufficiently broad and its integration time sufficiently long that the attenuation does not vary in time. However the real rainfall will manifest itself as a sequence of samples from an average DSD and the exact nature of the impact of this will be sensor specific. Nevertheless the example is a robust demonstration of the need for careful handling of spatial scales.
3. This time/space effect must be carefully considered in the context of the sensor tests themselves. Depending on the integration time and beam width (or field of view) of the sensor under test, at any instant in time the sensor will "see" a different sample of raindrops to that being measured by any associated meteorological observing kit, even if it is very closely located; the CAV sensors and meteorological observations are simply sampling from the same statistical distribution which comprises random individual events. Comparison will therefore have to be made over integration times that allow both to sample the distribution adequately, which will be a necessary step in order to link sensor performance to the measurable parameters of the ODD. This will be explored further as part of any external testbed demonstration.

Practical approaches to resolution

This section demonstrates that the complexities associated with spatial and temporal variability can be managed. There are fundamentally two different approaches:

- *'ODD' space*: By adopting a more conservative ODD rainfall threshold to that established by testing
- *'Sensor' space*: By adding a safety margin onto the 2-way attenuation determined by testing for a given rainfall rate

No recommendation is made here, however it is instructive to consider the challenge from both angles.

ODD space

In the above example, the spatial variability of rainfall within the 2x2 km pixel led to a maximum 200 m 2-way attenuation of 13.90 dB, compared to a value of 8.8 dB calculated from the pixel average rainfall of 50 mm/hr. Assuming the spatial variability remains the same, it is possible to identify the average 2x2 km pixel rainfall intensity that yields a peak value of 200 attenuation equal to the safe 9 dB limit. This value is 27 mm/hr.

In summary, if the test environment had demonstrated the 77 GHz would operate nominally at a rainfall rate of 50 mm/hr, the rain fall ODD should be specified as *"Rain rate less than or equal to 27 mm/hr as defined by a 2x2 km resolution rainfall product"*.   Clearly there are a number of caveats around this (including assumptions of rainfall type), however it is broadly indicative of how the spatial variability might be managed in "ODD space" and would readily support the addition of confidence margins associated with the sensor assurance measurement uncertainty.

Sensor space

An alternative perspective is to consider the impact of the rainfall variability on the sensor itself i.e. ask the question "What error in calculated attenuation might the sensor experience if the operationally determined rainfall estimate is taken at face value?"

To explore this question requires an assumed 'truth', which is the highest resolution simulation of rainfall within the 2x2km pixel. In this case this when the 2x2 km pixel is divided into a 2048x2048 grid of subpixels, each roughly 0.5 m square[36]. A range of interest is then chosen; here 250 m is selected as it is close to the 200 m used earlier[37].

The whole 2x2 km area can be subdivided into a total 16,384 lengths of 250 m; 8 in the x-direction by 2048 in the y-direction.   For each sample length, the 2-way path integrated attenuation (PIA) can be calculated by summation of the individual pixels along that length, using attenuations derived from any underlying resolution of rainfall data.   This is explained fully in Fig 15.

---

[36] Although in practice, such a small subdivision is not required.

[37] As can be seen from the figures, 250m corresponds exactly to a factor-2 subdivision of our original pixel, which significantly aids the analysis.
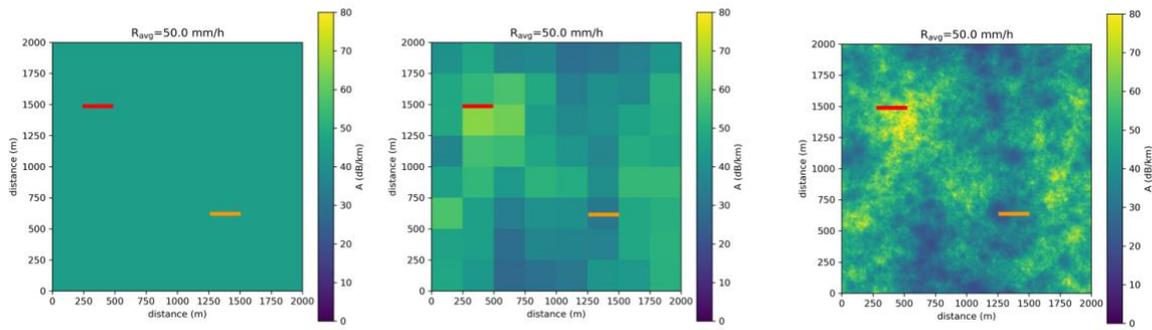
Figure 15: Exploring the effect of the resolution of the underlying rainfall dataset. From left to right the panels represent 2-way attenuation (dB/km) for the 2x2 km area at resolutions of 2 km, 250 m and ~1 m i.e. the values that would be inferred if the rain were observed/measured at that resolution. The far right panel is the 'truth' for this exercise.
The red line is an example 250 m range over which the total 2-way path integrated attenuation (PIA) is calculated. The orange line is a second example path. A total of 16,384 PIAs are calculated.

In the figure for each sample, the 'real' 250 m 2-way attenuation is that which is calculated from the rightmost panel. The other two represent the values if calculated directly from rainfall values measured at a *lower* resolution. Visual inspection suggests that it is perfectly possible for the attenuation inferred from a rainfall measurement that is significantly more coarse than the scale of interest to CAV to be an underestimate as well as an overestimate of the actual attenuation. The former is likely to have the greater safety implication.

Fig 16 demonstrates this at summary level as the probability (and cumulative) distribution functions of PIA 'error' from all of the 16,384 paths in the 2x2 km area. Of these, the cumulative distribution function suggests a mechanism for how to manage the uncertainty around ODD exceedance that arises from suboptimal measurement of the weather when the CAV is on the road.

Consider again the case where the test environment had characterised the 2-way attenuation at 77 GHz in a rainfall rate of 50 mm/h, this time over 250 m. When the CAV is deployed operationally it enters an area where the 2x2 km rainfall measurement suggests an intensity of 50 mm/h. If this rainfall rate is assumed to be representative of the path directly in front of the CAV, then (by eye), the CDF value at 'zero error' is approximately 0.6 i.e. there is a 40% chance that the real attenuation will exceed this. In order to be more confident of our estimate we would have to add a 'safety margin'. The construction lines in the figure suggest that our confidence that the CAV is <= our estimated attenuation can be increased to 95% (CDF=0.95), by adding 5dB onto the attenuation. If the rainfall is observed at a higher granularity a smaller safety margin is required because the observation has captured more of the CAV-scale variability. So if the rainfall measurement is made at 500x500 m, only 3 dB has to be added to the estimated attenuation to achieve a 95% confidence.
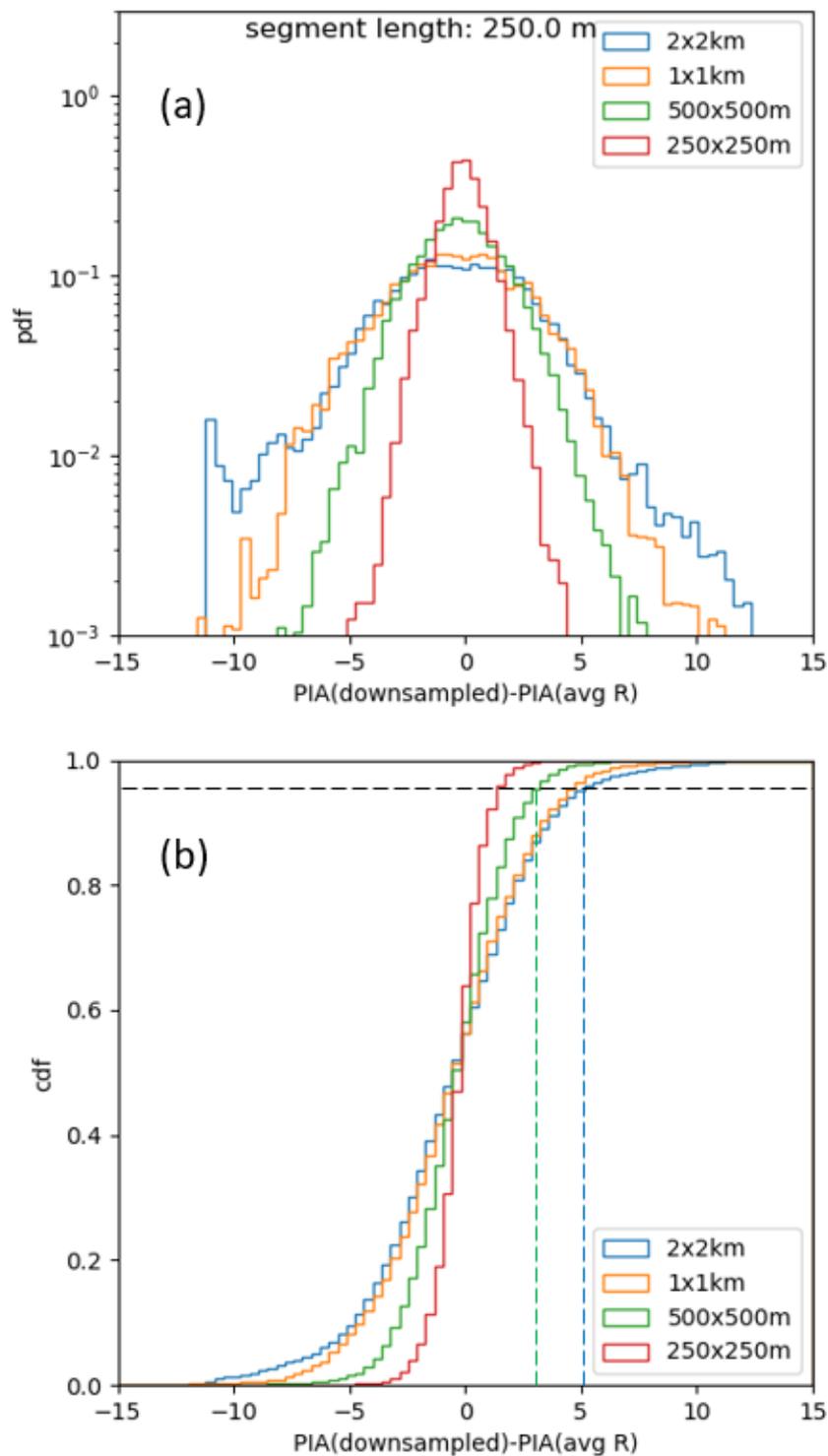
Fig 16: (a) Probability distribution functions for the 'error' in 2-way path integrated attenuation (PIA) when calculated directly from rainfall rate 'measured' at different spatial resolutions, all consistent with a 2x2 km average rainfall intensity of 50 mm/h.
(b) The corresponding cumulative distribution functions. While the data are all simulated, it is clear that the spread of error values can be reduced significantly via improved granularity of the rainfall measurement. The dashed lines refer to values associated with a 0.95 CDF value (see text)

The message is clear – greater confidence in assessing whether the CAV sensor is within its ODD comes with higher resolution measurement of the ODD parameter (rainfall in this case). However it is

possible to mitigate suboptimal measurement of the ODD through the adoption of confidence safety margins.

For completeness, the above exercise can be repeated for other, typically shorter, path lengths. Fig 17 shows the error the distribution of PIA with range length of interest. There are two notable trends in these plots. Firstly, the error spread in calculated attenuation increases with range (as can be seen by looking at any single figure). This is simply due to the cumulative errors over longer path lengths. Secondly, *between* figures the spread for a single target range of interest decreases as the resolution of the underlying meteorology improves, as discussed earlier. (In fact, Fig 17 was effectively created by using the 250 m lines from each of the plots in Fig 16).
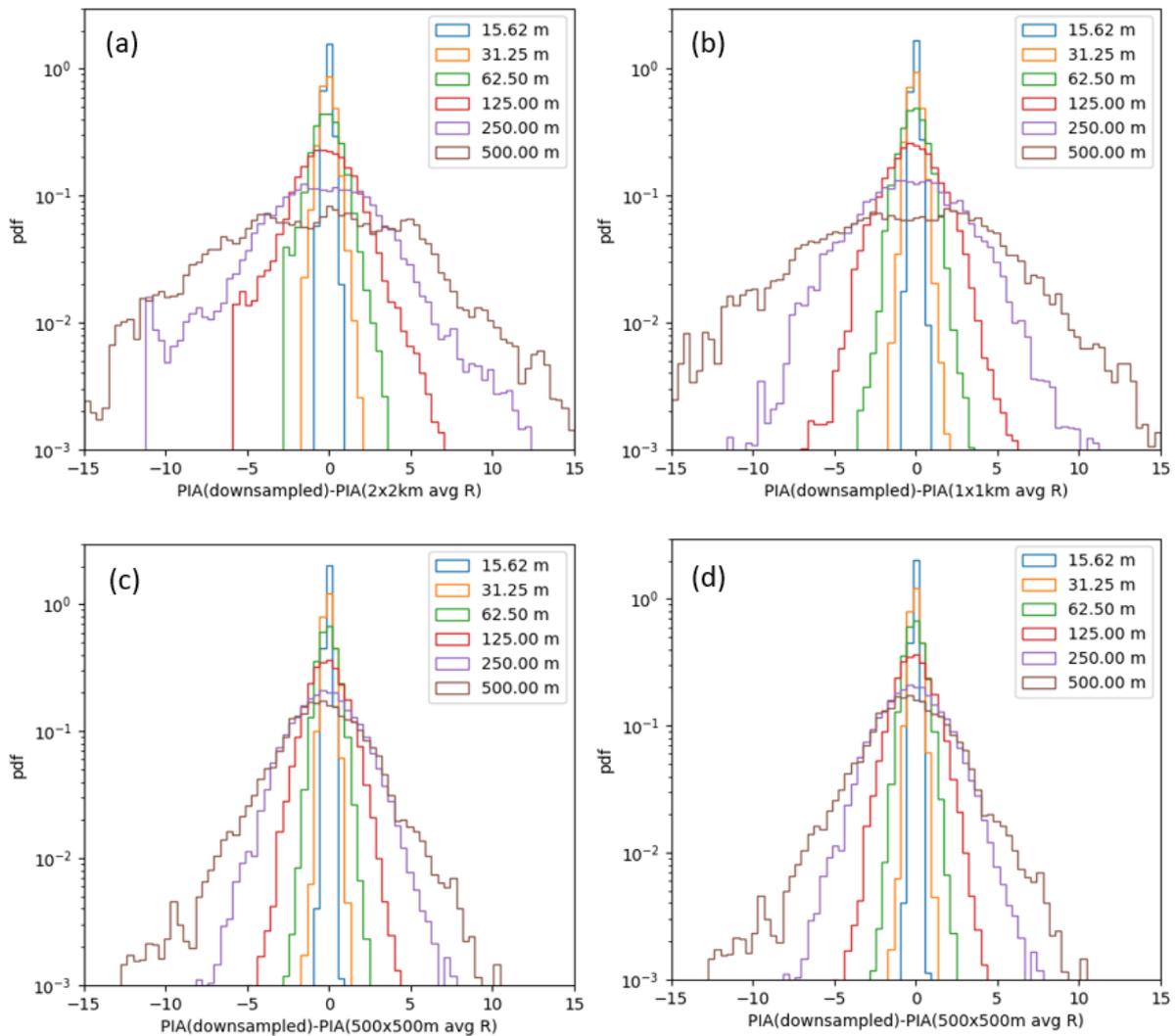
Figure 17: Probability density functions of the simulated error in 2-way PIA (in dB) resulting from calculation directly from a low spatial resolution rainfall field for a selection of target ranges. The resolution of the 'measured' rainfall field increases from 2 km [(a) top left)] to 250 m [(d) bottom right] , with an associated reduction in spread as the meteorological measurement approaches the CAV-relevant scales. (b) is 1 km and (c) 500 m. A positive x-axis value indicates the 'true' attenuation is greater than that inferred from the measured rainfall. Note the logarithmic y-scale which means that the extreme errors may appear exaggerated

"ODD space" vs "Sensor space"

Neither of these options, "ODD space" or "Sensor space", are offered as recommendations at this stage, not least because they only explore a small part of the error budget and the case studies are highly idealised. However, both demonstrate how uncertainties associated with meteorology can begin to be factored into the management of uncertainty in the sensor assurance framework and the application of it outputs in operation.

A systems thinking approach to managing uncertainty

The sections above demonstrate that even in a simplified and idealised case, there are a number of mechanisms through which sensor characterisation (in the test environment) and performance assessment (in the operational environment) are impacted by uncertainty.

It is worth briefly considering the relationship between these two, again in a simplified context. Consider a CAV sensor that has been characterised in test environment with respect to a given ODD attribute (e.g. rain).   The KPI (e.g. maximum ranges) reaches a minimum acceptable value at a given rain rate threshold *T*.   This threshold has an estimated uncertainty value of *ΔT*. In operations, the CAV receives updated information about its operating environment, in particular the actual rainrate in its location, *A*. For the reasons above, A is not known exactly and its estimated uncertainty us correspondingly *ΔA*. *A* and *T* are in the same units and a critical value is the *difference* between the *A* and *T*, i.e. the proximity to the ODD edge *(A-T)*. While the measurement challenges associated with determining *A* and *T* may share some common underlying drivers, for the calculation of *(A-T)* when the CAV is on the road they are independent. The uncertainty *E* in determining the proximity to the ODD edge may therefore simply be written:

$$E = \sqrt{\Delta A^2 + \Delta T^2}$$

Fig 18 shows how these errors combine using arbitrary units. It is clear that while reducing both *ΔT* and *ΔA* will result in lower uncertainty (increased confidence) in determining the proximity to the ODD edge, the confidence budget is dominated by the largest source of uncertainty.

		Error in **characterising** CAV performance with respect to ODD attribute **ΔT**						
		1	2	3	4	5		
Error in measuring ODD attribute **ΔA**	1	1.4	2.2	3.2	4.1	5.1		≤ 3
	2	2.2	2.8	3.6	4.5	5.4		≤ 4
	3	3.2	3.6	4.2	5.0	5.8		≤ 5
	4	4.1	4.5	5.0	5.7	6.4		> 5
	5	5.1	5.4	5.8	6.4	7.1		

Figure 18: Values of uncertainty in the proximity to an ODD edge (in arbitrary units) given uncertainties in determining a critical ODD threshold value T in the test environment and the uncertainties in measuring the actual value of the ODD attribute A in operations. Colours are used to highlight different bands of total uncertainty

This leads to an important consideration – the sensor assurance framework is part of a larger **system** involving both sensor characterisation with respect to the weather *and* the realtime determination of the weather in the location of the CAV on the road. Should one source of uncertainty dominate, then investments in the other may not fully pull through into the ability of the CAV to operate confidently within the widest possible ODD. This should inform both the level of investment in each part of the system and the measurement protocols developed in the sensor test ecosystem.

For completeness, two further points should be noted:

- The degree of the impact of the aforementioned total *uncertainty* in the proximity to the ODD edge is a function of the how far away from the edge the CAV sensor i.e. it becomes a consideration only when the uncertainties are comparable to, or greater than, the separation. While a statement of the obvious, this has implications for ODD measurability because as the separation becomes much greater than the uncertainty, it shifts the emphasis from *accurate* measurement of the ODD parameter towards a more *categorical* demonstration that the environment is well away from the ODD edge, which may be easier to deliver operationally.

- The term *measurability* is increasingly used to describe our ability to characterise the relevant ODD parameters. In the meteorological context, it is not always necessary to 'physically' measure

the parameter; a forecast or nowcast of the ODD parameter may be sufficient. For example, if a CAV sensor has been demonstrated to work up rainfall rates that would only be experienced in a highly convective shower (e.g. a thunderstorm), a forecast that indicates 0% chance of thunderstorm is an acceptable measurement for quantifying the ODD.

The above discussion leads to the *"Linkage to the Operational Design Domain (ODD)"* recommendation in the critical success factors of the main report.

Appendix G: Indicative meteorological measurements at the external testbed

Indicative meteorological equipment list for comprehensive external test bed		
Item/height	Quantities	Justification
10m		
Sonic anemometer, PRT, humidity sensors	Winds, turbulence, heat and moisture fluxes, T, RH,q	10m is standard height for wind measurement. Effect of turbulent fluxes on rainfall can be assessed. Effect of Humidity on vehicle sensors can be assessed.
2m		
As 10 m	As 10m	As 10m but at car height. Allows cross-checking with 10m data.
Radiation: Up and downwelling radiometers/pyrgeometers	Up and down-welling short and longwave radiation. Grass tip temperature.	Radiation levels add information regarding cloud and visibility, plus aid assessment of vehicle cameras (?)
Disdrometer (5)	Network over 100x100m of rainfall and rain drop size distribution. Precipitation type.	Network allows some small-scale assessment of rain to be made. Direct assessment of rain on vehicle sensors possible
Precipitation particle spectrometer	size distribution, fall velocity, and rain rate of droplets from 50 µm to greater than 6.4 mm	Reference standard for assessing limitations of cheaper disdrometers (especially larger droplets)
Fog spectrometer	LWC, mean radius, droplet spectra (2-50 microns dia)	Assessment of fog on vehicle sensors
Aerosol spectrometer	Aerosol size distributions, optical depth (0.25-30 microns dia)	Quantify effect of hydrating aerosols on vehicle sensors
Grass level		
Aspirated shield with T & RH	T and RH	Allow assessment of dew probability and surface icing.
Rain gauges	Event-led rainfall rate	'Reference' for disdrometers, plus during high rain rates may provide more temporal resolution.
Other		
Ceilometer	Cloud base, gradients in aerosol content	Assessment of cloud base on vehicle sensors.
Web Cam	Visual images	General assessment of weather and security

Appendix H: Global rainfall distributions

Preamble

With regard to the external test beds, it is obvious that these will contribute most to the assurance process when their exposure to the weather element of interest is maximised. In short, in the context of this proof of concept study, it is desirable to locate the testbed(s) where there is frequent heavy rain.

This appendix describes a short exploratory study of readily-available literature, to identify where such regions might be. It is stressed here that, with a focus on extreme instantaneous rainfall intensity, it is not fully comprehensive, focussing mainly on the convective rainfall case. Further consideration must be given to the orographic and dynamic cases as follow up work.

Summary

Those areas where the most intense, sub-daily rainfall will occur most often are identified as being in the Tropics, between approximately 23° of latitude north and south of the Equator. The lack of *sub-daily* rainfall data to analyse for these areas led to a literature search, where the world of telecommunications proved a rich source of useful information. Telecommunication designers seek to minimise the attenuation of electromagnetic (typically radio) waves due to various confounders, one of which is rainfall. To assist with this an International Telecommunications Union standard has focussed upon the 1-minute average rainfall rate with a risk of annual exceedance of 0.01% in any year. The standard contains methods to calculate this value for any location and also a global map. This *'near extreme'* 1-minute rainfall rate approaches the relevant timescale for CAV, however as indicated in Appendix F, great care is required to link these values to those measured in the testbed environment.

Using the above definition, the highest *near-extreme* 1-minute rainfall rate derived from the standard for the Tropics is of the order of 100 mm/h. Various pieces of independent research suggest that the standard tends to underestimate conditions, with their results suggesting values for the Tropics of the order of 150 mm/h or higher.

Background

An initial Met Office report for Innovate UK (Dixon, 2019)[38] considered extreme rainfalls in the UK using both observed data and published design guidelines. It concluded that, for the UK, the climatology of summertime, extreme convective rainfall in SE England provides an appropriate worst case for the whole of the UK. Building design guidelines suggested 2-minute storms with rates varying from 80 mm/h (a twice a year event) through to 320 mm/h (a 1 in 500-year event) in the London area. This study extended that work to consider rainfall rates around the world.

Focus of this study

The aim of this study was to identify the global distribution of impactful rainfall rates, over very short periods (minute or less), to inform the possible locations of future external CAV

---

[38] Dixon, J., 2019: Short-period rainfall extremes in the UK, *Met Office Report for Innovate UK*, available from
https://www.metoffice.gov.uk/services/transport/cav

sensor testbeds. The aim was not specifically to seek the maximum values achievable, although some commentary is given for context.

As the Met Office archives for overseas data have, at best, only 3-hourly totals of rainfall (most locations have 6-hour totals), their use in establishing sub-hourly rainfall rates is limited. Consequently, no climatological data have been analysed here. This study contains a brief description of global rainfall, the identification of relevant areas and the results of a literature search for near-extreme global, sub-hourly, rainfall rates.

Global Rainfall Regime

The highest intensity rainfall rates are most likely to be experienced in convective rainfall, i.e. rain that is caused by the vertical motion of an ascending mass of air that is warmer than its environment; the horizontal dimension of such an air mass is generally of the order of 15 km or less and forms a typical cumulonimbus (Cb)[39] cloud. Convective rain is sometimes accompanied by thunder and hail.
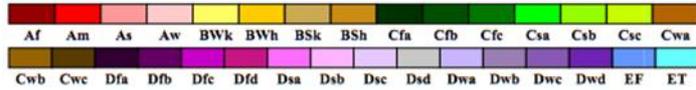
The climate of a region is mainly defined by its latitude, distance from the sea, prevailing winds and altitude. The most well-known global climate classification is that due to Köppen[40], see Figure 1 for an updated version of this map. Those areas of the world where the ideal combination of heat and moisture combines to produce the most frequent and intense convective rainfall described in the previous paragraph will be within the Tropics, approximately 23° north and south of the Equator. These areas coincide with the red/pink regions of Köppen's map and are defined as *Tropical* with sub-groups depending upon whether the region is humid all year, only in the winter, only in the summer or experiences Monsoon conditions. Such areas include, Malaysia, the Philippines, Indonesia, parts of central Africa and much of the Amazon basin. The darkest red regions, described as 'Equatorial Fully Humid' in Figure 5, will be the areas where convective rain will be more frequent and more intense. Annual average rainfalls in excess of 2,500 mm are likely in these areas.

---

[39] http://www-web/glossary/C/CUMULONIMBUS.html

[40] Kottek,M., Grieser,J., Beck,C., Rudolf,B. and Rubel, F.(2006): 'World Map of the Köppen-Geiger climate classification updated' Meteorologische Zeitschrift, Vol. 15, No. 3, 259-263 (June 2006), c by Gebrüder Borntraeger 2006. Open access: https://www.schweizerbart.de/journals/metz
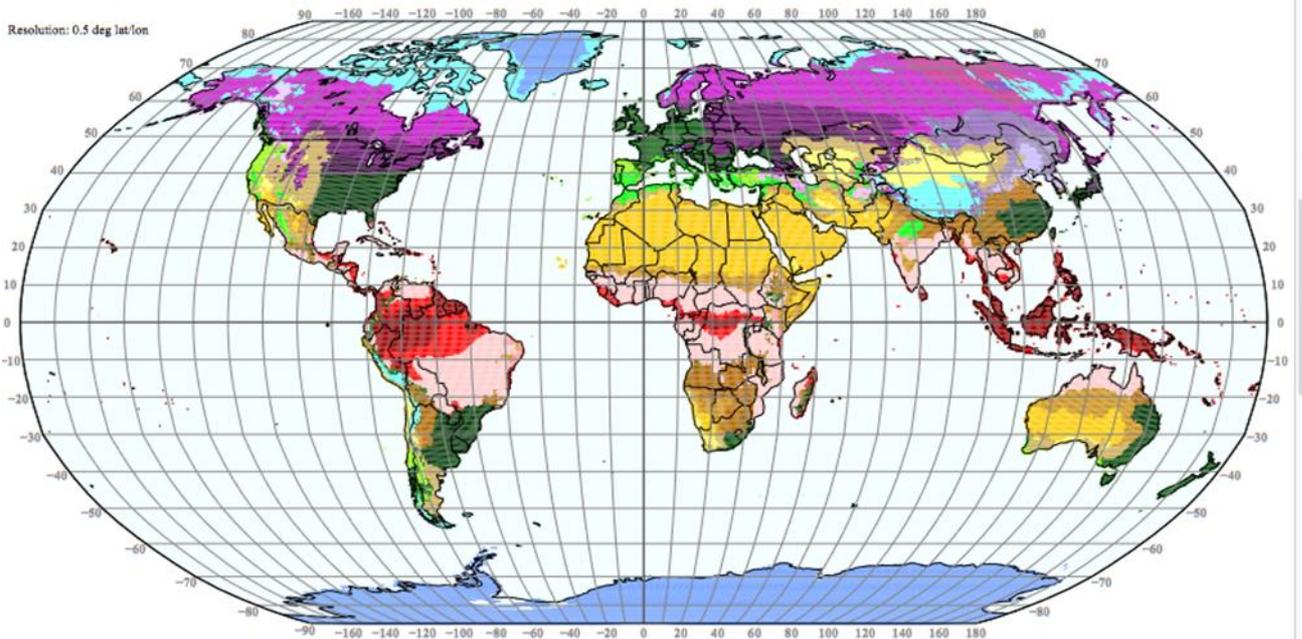https://www.weather.gov/media/jetstream/global/Koppen-Geiger.pdf

Figure 5:
Kottek et al (2016) updated Köppen-Geiger's climatological regions of the world.
Thanks and acknowledgements to Meteorologische Zeitschrift for granting permission to
reproduce this figure. www.borntraeger-cramer.de/journals/metz

Figure 6 is a global, annual average rainfall map showing the variation in rainfall across the globe based upon data for the period 1991-2000[41]. Of note here is that, in addition to the tropical areas, there are a number of mid-latitude regions (on the western sides of the landmasses) that also have significant rainfall totals (including the UK). These correspond to areas that are exposed to midlatitude depressions (dynamic rain), often with rainfall intensity enhancement due to air being forced vertically by mountains (orographic rain).

---

[41] https://opendata.dwd.de/climate_environment/GPCC/html/gpcc_normals_v2020_doi_download.html

GPCC Precipitation Normals in mm/year (Version 2020) per 0.25 degree grid for YEAR
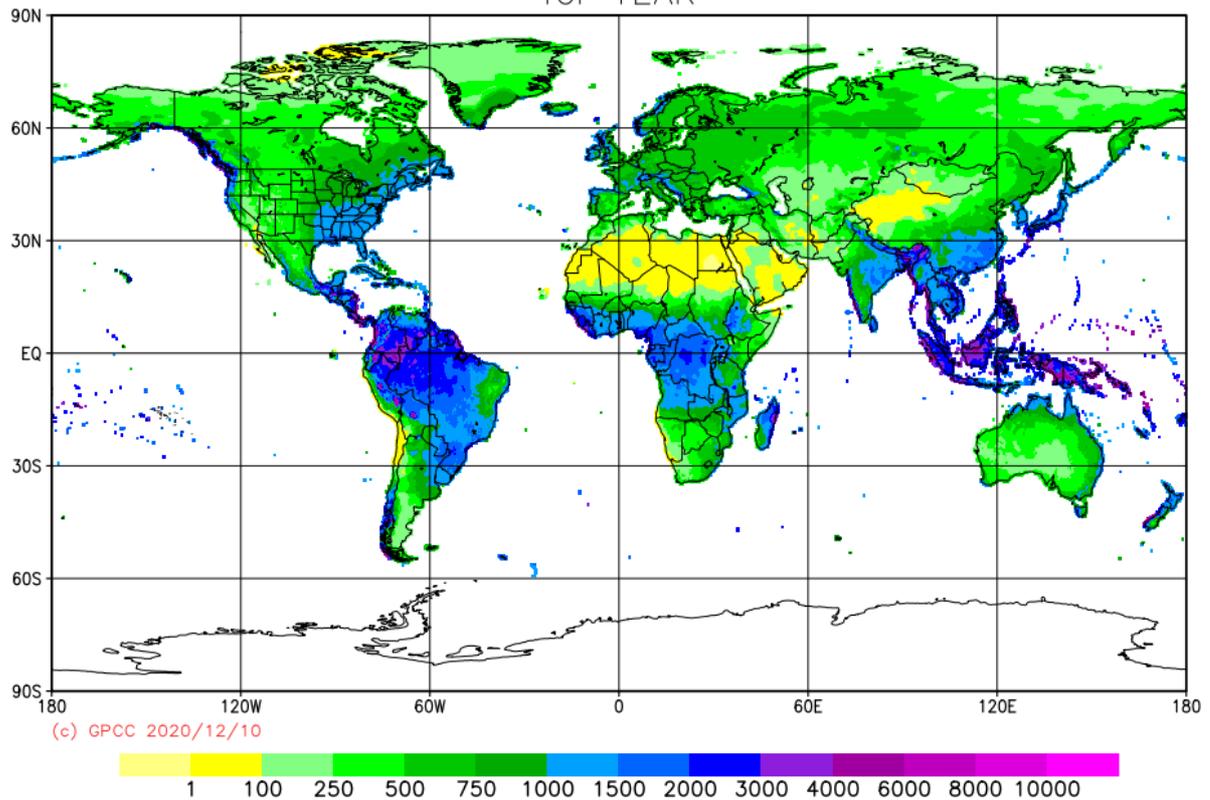
(c) GPCC 2020/12/10

1 100 250 500 750 1000 1500 2000 3000 4000 6000 8000 10000

Figure 6: Map of annual average rainfall, 1951-2000.
Source*:*
https://opendata.dwd.de/climate_environment/GPCC/html/gpcc_normals_v2020_doi_download.html
Thanks and acknowledgement to the Global Climatology Centre (GPCC) a part of the German meteorological service (Deutscher Wetterdienst DWD) for permission to reproduce this map.

The tropical rainfall areas identified by Köppen, will be those parts of the world where convective rainfall (and therefore thunder and lightning) are more frequent. This is confirmed in the red and dark red areas of the global map of the number of lightning flashes per km² per year in Fig 3. These coincide with the red areas of Köppen's map but with some variations including central USA reflecting the preferred hurricane tracks along the east coast of the USA and tornado tracks in central USA, which result in intense rainfall conditions and associated thunder and lightning, occurring further north, beyond the Tropics.
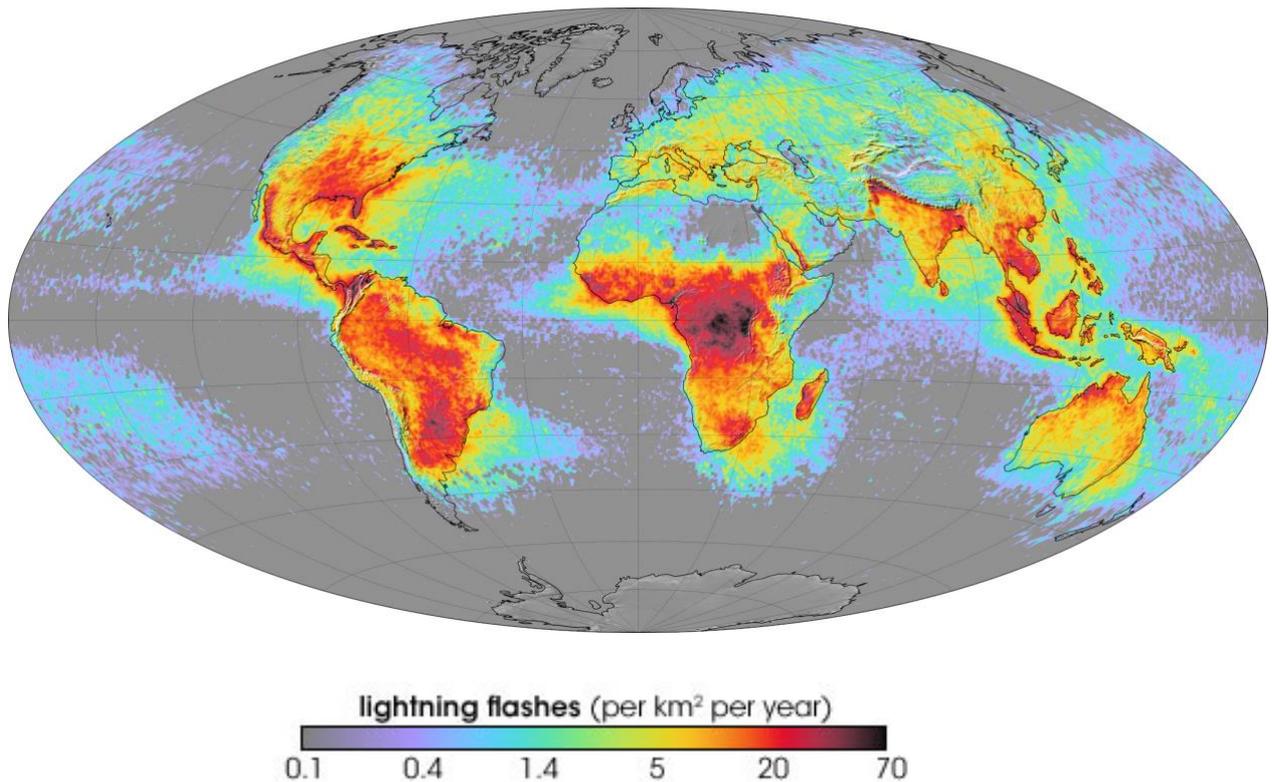
Figure 3: World Lightning Map: The map above shows the average yearly counts of lightning flashes per square kilometre based on data collected by NASA's Lightning Imaging Sensor on the Tropical Rainfall Measuring Mission satellite between 1995 and 2002. Places where less than one flash occurred (on average) each year are grey or light purple. The places with the largest number of lightning strikes are deep red, grading to black.
Source: https://earthobservatory.nasa.gov/images/6679/patterns-of-lightning-activity
Thanks and acknowledgement to NASA's Lightning Team for allowing reproduction of this figure.

The areas identified, where convective rain occurs most often on a daily basis are within the Tropics, however this does not preclude more infrequent but equally intense events occurring in other parts of the world. For example, the official record, recognised by the World Meteorological Organisation (WMO), for **the greatest amount of rain falling in 1 minute is 31.2 mm (1872 mm/h)** recorded on 04/07/1956 at Unionville, Maryland, USA, which is approximately 50 km north-west of Baltimore and about five thousand kilometres north of the Equator (WMO website https://wmo.asu.edu/). A WMO publication[42] (WMO, 1994) had suggested an even larger amount of 38 mm in 1 minute at Barot, Guadeloupe (Caribbean) but this appears to have subsequently been rejected. Other sub-hourly records included in that WMO publication include those in Table 1.

---

[42] WMO, 1994. Guide to Hydrological Practices, WMO no. 168, 5th Edition. World Meteorological Organisation (WMO), Geneva, Switzerland.

Duration (mins)	Rainfall (mm)	Location	Date	Av Rainfall Intensity (mm/hr)
5	62	Port Bells, Panama	29 November 1911	744
8	126	Fussen, Bavaria	25 May 1920	945
15	198	Plumb Point, Jamaica	12 May 1916	792
20	206	Curteade Arges, Romania	7 July 1889	618
40	235	Guinea, Virginia, USA	24 August 1906	353
42	305	Holt, USA	22 June 1947	436

Table 1: Highest sub-hourly rainfall totals accepted by the WMO.
Source: https://wmo.asu.edu/

The rainfall record at Holt is also the WMO's official highest hourly total. Holt, Missouri, is approximately 40 km northeast of Kansas City, which places it in the infamous Tornado Alley of the southern USA. Rakhecha and Singh (2009) also provide significant rainfall occurrences for various countries including India, China, Australia and Japan. For example, the maximum hourly rainfall total for a selection of station include 129 mm at Mumbai, 88mm at both Darwin and Brisbane and for Japan 187 mm at Do.

There are methods available that attempt to calculate the maximum probable precipitation (PMP) that could fall from a column of air, knowing its meteorological characteristics (e.g. WMO, 2009[43]). However, neither calculating this nor knowing the very worst extreme events will not be relevant for this report; near-extreme conditions will be more relevant as they relate more closely to climatological expectation to inform the possible siting of testbeds. Near-extreme conditions can be defined in any number of ways depending upon its application but one sector where heavy rainfall has a significant impact and, fortuitously apposite for CAV work, is the world of telecommunications. The next section considers this in more detail.

Standards for Telecommunications

The closely related discipline of telecommunications offers a means to access global precipitation information, that may be closer to being fit the for the CAV application. In particular, the International Telecommunications Union (ITU) standard, '*Recommendation ITU-R P.618-13(12/2017). Propagation data and prediction methods required for the design of Earth-space telecommunication systems P Series Radiowave propagation*.' (ITU-R,2017a[44]), which includes a consideration of rainfall. ITU have chosen a 1-minute integration time as

---

[43] WMO, (2009). Manual on Estimation of Probable Maximum Precipitation (PMP). WMO No. 1045, Geneva.

being the most desirable for attenuation prediction (Mandeep and Hassan, 2007[45]) and an annual probability of exceedance of 0.01%, taken to mean that there are only approximately 53 minutes worse than this in the whole year of 525,600 minutes). For brevity within this report, this rate is termed the 'near-extreme rate'.

A supplement to the afore mentioned standard, '*Recommendation ITU-R P.837-7(06 /2017) Characteristics of precipitation for propagation modelling*' (ITU, 2017b[46]) contains a world map of these values of these 'near-extreme rates based upon monthly mean rainfall for the 50 years' (1951-2000) of data from GPCC Climatology (V2015) database over land and from 36 years' (1979-2014) of the European Centre of Medium-Range Weather Forecasts (ECMWF) ERA interim data over water. This map is reproduced in Figure 4.
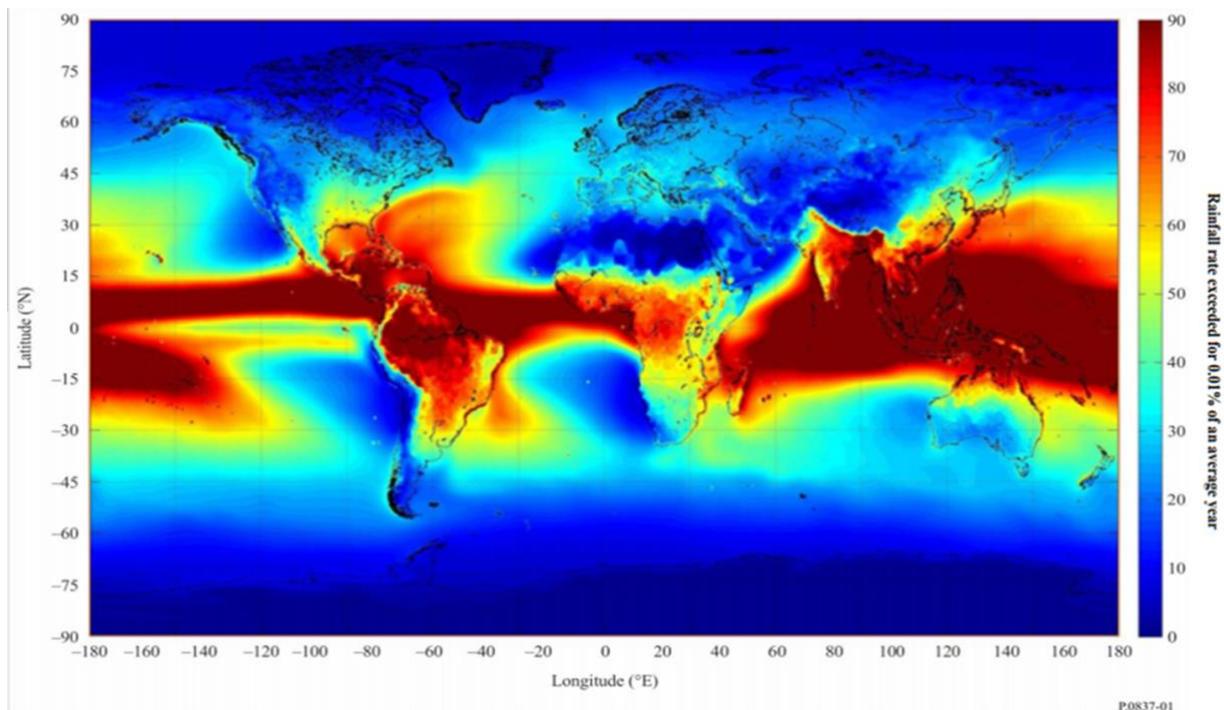

Figure 4: Global map of 'near-extreme' rainfall rates (0.01%) based upon monthly mean total rainfall derived from 50 years (1951-2000) of data from the GPCC Climatology (V 2015) database over land and from 36 years (1979-2014) of the European Centre of Medium-range Weather Forecast (ECMWF) ERA Interim data over water. (ITU, 2017b). Thanks and acknowledgement to the ITU for granting permission to reproduce this map in this report. Note authorisation was granted on a non-exclusive basis and is non-transferable to third parties.

https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.618-13-201712-I!!PDF-E.pdf

The Supplement shows that areas experiencing the most intense rainfalls are concentrated in the Tropics and approximately align with the 'Tropical' areas defined by Köppen, (shown in

---

[45] Mandeep JS, Hassan SIS. 2008. '60 to 1-minute rainfall rate conversion: comparison of existing prediction methods with data obtained in the Southeast Asia region'. *Journal of Applied Meteorology and Climatology* **47**: 925-930.

[46] ITU (2017b) - 'Recommendation ITU-R P.837-7(06 /2017) Characteristics of precipitation for propagation modelling' (ITU, 2017b). https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.837-7-201706-I!!PDF-E.pdf

the blue areas of Figure 1) in particular south-east Asia, northern Australasia, central and western Africa, northern South America and the Caribbean. Values range from 90 mm/h in these highlighted tropical areas, through 50mm/h in sub-tropical regions (aligned approximately with the Koppen's sub-Tropical regions coloured orange in Figure 1) and < 20mm/h for North Africa, Middle East, Europe and Canada in the Northern Hemisphere and all areas south of 45° in the Southern hemisphere plus the western coast of South America south of the Equator and west of the Andes.

The ITU standard gives guidance on how to use suitable local observations, if applicable, but in general these will not be available. If that is the case, the standard uses monthly rainfall and monthly temperature data using 50 years (1951-2000) of the Global Precipitation Climatology Centre (GPCC: operated by the German National Meteorological Service [Deutscher Wetterdienst, DWD] under the auspices of the World Meteorological Organization (WMO)), Climatology (V 2015) data over land and 36 years (1979-2014) of the European Centre of Medium-range Weather Forecast (ECMWF) ERA Interim data over water. Using these monthly values and various statistical techniques eventually an estimated 1-minute rainfall rate is arrived at. There is an accompanying spreadsheet containing monthly data and conversions factors to derive the 1-minute value, which appears to be based upon the method proposed by Segal (1986)[47]. These methods are discussed later in this section. The map in the standard (**Error! Reference source not found.**) provides an accessible means for communication engineers to arrive at the 1-minute value without having to go through the calculations. The standard suggests that the values derived from the map are within 0.3 mm/h of a calculated value for 99.9% of the world.

There is a spreadsheet accompanying the Supplement, which has calculated the near extreme rate for a selection of latitudes and longitudes. The highest value in the spreadsheet is 99 mm/h for Kuala Lumpur, Malaysia. This is higher than the map value but it should be noted that the spreadsheet was released in 2019 and may contain more up-to-date information and so some differences are likely. Looking at Kuala Lumpur on the map, this is where the very darkest red colour occurs; similar colours and hence similar values would occur in most of Indonesia, the Philippines, the coastal strip between Bangladesh and Vietnam, the coastal area around the Amazon River estuary and parts of Central America.

The ITU spreadsheet also contains rainfall rates for other annual probabilities at Kuala Lumpur and the same information for a few other locations:- Southern Egypt, Miami, New Delhi, Rome and London plus some marine sites in the mid-Atlantic and the Mediterranean (between Malta and Libya). Table 2, provides an extract from this comparing the values from Kuala Lumpur with those from London.

[47] Segal, B., 1986: The influence of raingauge integration time on measured rainfall-intensity distribution functions. J. Atmos. Oceanic Technol., 3, 662–671. https://journals.ametsoc.org/doi/pdf/10.1175/1520-0426%281986%29003%3C0662%3ATIORIT%3E2.0.CO%3B2

Location	Lat	Long	Annual probability (%)				
			0.01	0.1	0.15	0.3	0.35
Kuala Lumpur	3.133	101.7	99.2	34.6	27.8	18.3	16.5
London	51.5	-0.14	26.5	9.0	7.2	4.7	4.2

Table 2: 1-minute rainfall rates (mm/h) exceeded for 0.01, 0.1, 0.15, 0.3 and 0.35% of an average year based for Kuala Lumpur and London. (ITU 2017b accompanying spreadsheet).

The method used to produce these results is based upon one devised by Segal (1986), one of several techniques that have been published, for converting readily available longer-period rainfalls (daily, monthly or annual) to the required but scarcer sub-daily or sub-hourly values. Mandeep and Hassan (2008) assessed five different techniques for doing this and a brief description of each is given below.

Segal (1986) –analysed ten years of high-resolution tipping-bucket precipitation records for 45 locations in **Canada** to yield empirical conversion factors appropriate for five- or ten-minute sampling times. The locations were chosen to provide a sampling of different climatological and physiographical regimes. This method is used as part of the process of producing the ITU map in Error! Reference source not found.**.**

Burgueno et al. (1988)[48]- analysed 49 years of rainfall rate measurements from various sites in **Barcelona**, Spain, to produce cumulative distributions for different effective gauge sampling intervals. They sought to develop a direct and universal expression between 1- and x-min integration rainfall rates.

Chebil and Rahman (1999)[49] - proposed an empirical method for approximating the rainfall-rate conversion factor from a 60- to a 1-min integration for 82 locations in **Malaysia**. Hourly and annual long-term rainfall data (1991-8) were recorded using tipping bucket rain gauge (sensitivity of 0.5 millimetres per tip). The Chebil and Rahman method was developed based on seasonal and diurnal variations in **convective storm rainfall**. Stratiform rainfall was not considered.

Moupfouma and Martin (1995)[50]- their method for integration times, ranging from 1 to 60 minutes was based on rainfall-rate measurements taken at **Chilbolton, United Kingdom**, by the Rutherford Appleton Laboratory. Rain events from 1985 to 1992 were recorded using a tipping-bucket rain-gauge.

---

[48] Burgueno, A. M., M. Puigcever, and E. Vilar, 1988: Influence of raingauge integration time on the rain rate statistics used in microwave communication. *Ann. Telecomm.*,**10**,522–527

[49] Chebil, J., and T. A. Rahman, 1999: Rain rate statistical conversion for the prediction of rain attenuation in *Malaysia. Electron. Lett*., **35**, 1019–1021. https://www.crossref.org/iPage?doi=10.1049%2Fel%3A19990685

[50] Moupfouma, F., and L. Martin, 1995: Modelling of the rainfall rate cumulative distribution for the design of satellite and terrestrial communication systems. *Int. J. Satellite Commun*., **13**, 105–115.

Joo et al. (2002)[51]- this method, based upon only 2 years' worth of rainfall-rate measurements from Korea, produced rainfall rate cumulative distributions for different integration times (1, 10, 20, 30, and 60 min). Rainfall-rate data with various time integrations were collected using an optical rain gauge.

Deriving the near-extreme 1-minute value is a two-part process that involves a) creating distribution functions of short-term rainfall rates (sub-daily) and establishing relationships between 1- minute and 60-minute totals (correlation coefficients unique to each location/study) then, b) using a conversion factor (Segal or other techniques) to convert an x-minute rainfall rate distribution into its equivalent 1-minute distribution. All involve power-law relationships, very much along the lines of the spatial power laws in Appendix F, which Mandeep and Hassan suggest means that they can be used in both tropical and temperate regimes (see the individual articles for more details). Note, no further investigations were conducted to test the validity of these methods in this report.

Mandeep and Hassan's conclusion was that the method devised by Segal (1986), was the best method (lowest overall percentage error) and they recommended its use in tropical areas, which is perhaps surprising as the results were based upon research throughout Canada. They did note, however, that the method tended to *underestimate* the 0.01% values with standard errors creeping up to 10% from well below 5% for the less rare annual probabilities. However, the method by Chebil and Rahman (1999), which used the Moupfouma and Martin (1995) method for a) and Segal's method for b), produced smaller errors of up to 6% for the 0.01% estimates but was less accurate at 0.1 and 1% probabilities. Similar comments about the errors were made about the Moupfouma and Martin (1995) method.

Based upon these conclusions, it suggests that if 0.01% estimates are required for tropical areas, as required in this report, then the Chebil and Rahman (1999) approach might be a better choice if any future investigations into this area are required, using appropriate local data.

Consequently, these methods open-up the possibility of estimating 1-minute values (not always globally measured but more relevant for this report) from hourly- or longer period-values, contained in the Met Office archives or other sources, for a variety of annual probabilities of exceedances. Such work has already been carried out by various authors and all of these studies (based in the Tropics) indicate that the mapped and spreadsheet ITU recommended values are all lower than their calculated values. It is to be expected that site-specific measurements and analysis would produce different and likely more accurate results to those derived from the monthly data and the 'down-scaling' techniques (to arrive at shorter time periods methods) used to produce the smoothed global map. Noting that the sign of these differences is consistently the same, means that the mapped values should be viewed as more conservative estimates (less extreme). The results of some of these studies are discussed and the ITU values quoted are from the 2017 standard:

---

[51] Joo, H. L., S. K. Yang, H. K. Jong, and S. C. Yong, 2002: Empirical conversion process of rain rate distribution for variousintegration time.Proc. URSI Commission F Wave Propagation and Remote Sensing, Maastricht, Netherlands, URSI,1450–1454

- Suryana et al (2005)[52] provide annual 0.01% rainfall rates for 24 Indonesian locations, with values ranging from **109 to 159 mm/hr**. The ITU map value – 90 mm/h; ITU spreadsheet value 99 mm/h (Kuala Lumpur)
- Mandeep and Hassan (2008) analysed observed sub-hourly rain data from seven stations in Southeast Asia (Malaysia, Thailand, Indonesia, Philippines, Fiji, Singapore and Papua New Guinea). The results estimated rainfall rates of between **90 and 130 mm/hr** for the 0.01% one-minute estimate. Papua New Guinea and Philippines at the lower end and the others clustered at the higher end. ITU map value - 90 mm/h; ITU spreadsheet value 99 mm/h (Kuala Lumpur).
- Rashid and Majumder (2011) calculated one-minute 0.01% rainfall rate estimates based upon annual average rainfall for 34 meteorological sites in Bangladesh using the Chebil and Rahman method, which produced values between **109-148 mm/h (**23 of the 34 sites were between **115 and 130 mm/h).** The ITU map value 90 mm/h; ITU spreadsheet value 99 mm/h (Kuala Lumpur); 63.6 mm/h for new Delhi.
- TRMM satellite data analysed by Omotosho et al (2013)[53] for 57 sites in Malaysia suggest 1-minute rates with 0.01% exceedance of **85-154 mm/h** in Eastern Malaysia and **82-144 mm/h** in Western Malaysia; the 0.001% values exceeded 200 mm/h. The deduced one-minute rainfall rates correlated fairly well with those obtained from the previous work carried out in Malaysia by Chebil and Rahman (1999), with correlation coefficients of 0.7 in all the 57 locations.

It is pleasing to note that despite different methods, different data, different periods of analysis, the results show a similar range of values generally in the range 85-145 mm/h for 1-minute rainfall rates with a probability of occurrence of 0.01% in tropical areas. As already noted, these estimates are all higher than those suggested by ITU, although the maximum ITU value lies within this range but nearer the lower boundary. Omotosho et al (2013) comment in their paper that the 'ITU-R SG3 database (2009) predictions for the 0.01%, 1-minute rainfall was found to be valid for only limited areas in Malaysia and underestimates the remaining regions'. From the dates of the research, the results must have been compared to an earlier version of the standard, possibly ITU (2007)[54]. However, even the updated information in the 2017 standard map (figures quoted above), still appears to underestimate the values produced by the independent pieces of research discussed. Therefore, this underestimation needs to be born in mind when using the ITU mapped information.

Summary
- Those areas where the most intense, sub-daily rainfall will occur most often have been identified as being in the Tropics, between approximately 23° north and south of the

[52] Suryana J, Utoto S, Tanaka K, Igarashi K, Iida M. 2005. 'Study of prediction models
compared with the measurement results of rainfall rate and ku-band rain attenuation at Indonesian tropical
cities'. 5th International Conference on Information, Communications
and Signal Processing, Bangkok, Thailand.

[53] Omotosho, T. V., Mandeep J. S, , Abdullah M., and Adediji, A. T. (2013): 'Distribution of one-minute rain rate in Malaysia derived from TRMM satellite data'. *Ann. Geophys*., **31**, 2013–2022, 2013

[54] ITU (2007): Rec. ITU-R P.837-4 1 RECOMMENDATION ITU-R P.837-5. 'Characteristics of precipitation for propagation modelling (Question ITU-R 201/3) (1992-1994-1999-2001-2003)
https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.837-4-200304-S!!PDF-E.pdf

equator. On average, 150 thunderstorms days per year are likely and over 70 lightning strikes per km^2 could occur in these areas. Other areas of the world could experience equally extreme rainfalls but far less frequently.

- Identifying the *absolute* extreme rainfall intensities is not really relevant for the purposes of this report because the intent of the study to begin to inform locations with increased likelihood of high intensity events. (Near-extreme events will be more useful.)

- Time scales of a minute, begin to approach the timescales of direct relevance to the CAV case, however there are no data on this time scale available for the (global) areas of interest in the Met Office archives; three- or six-hourly totals are the norm. Therefore, no analysis of data has been carried out in this study.

- A literature review has discovered a useful source of near-extreme (0.01% annual probability of exceedance) 1-minute rainfall rates. These are required by the telecommunications industry to mitigate against the attenuation of radio waves (and others) by rainfall. International standards (International Telecommunications Union - ITU) exist containing global maps of this statistic (also methods to estimate the value from longer sampling time intervals) and confirm that the Tropics is the area most likely to provide a reliable number of high intensity rainfall cases.

- The ITU map is based upon global model data and various techniques to convert monthly rainfall statistics to rainfall rates on a sub-hourly basis for a variety of annual probabilities of exceedance. The highest values derived are in the order of 100 mm/h for areas such as Malaysia.

- Work by independent researchers suggest that the ITU values tend to underestimate those produced by analysis of on-site observations. Results from a variety of locations across the Tropics suggest the highest 0.01% 1-minute rainfall rates are generally between 85-145 mm/h.

- Note, while the Tropics is mostly likely to provide sites that will allow more frequent access to intense convective rainfall event, it may well be possible to find locations to sample intense orographic rainfall within the UK. This was not explicitly addressed in this study so further work is required in this respect.

Recommendations

Due to the lack of sub-hourly rainfall data for the Tropics available to analyse, it would appear that the procedures and values accepted by the global telecommunications sector represent the most accessible without further study. The ITU standard's maximum 1-minute rainfall rate is of the order of 100 mm/hour but further independent research might suggest higher values up to 150 mm/h.