Appendices

CATAPULT Connected Places

NPL

Appendix A

Sensor Capabilities Definitions

A definition is provided for the each of key capabilities of perception systems, differentiation for individual sensors types is provided where appropriate.

1 Sensor Capabilities Definitions

i. Angular/Spatial Resolution

Is the capability to discriminate between two adjacent targets, in conjunction with the range position provides the localisation of the target.

Radar
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.

Camera
Is the capability to discriminate between two adjacent targets, in conjunction with the range and contrast dictate minimal detectable target. It can vary in both Azimuth and Elevation axes.

Lidar
Depending on the lidar's scanning method, vertical resolution and horizontal resolution may differ; therefore, horizontal and vertical resolution should be measured separately.

Ultrasound
Since ultrasound sensors are uni-directional, a single sensor cannot detect multiple objects at different angles. This capability is therefore not relevant. However, with multiple sensors operating into each of the desired directions, coarse angular resolution can be obtained.

ii Range Resolution

Is the capability to discriminate two targets at different ranges.

Radar
In radar systems it is inversely proportional to the transmitted bandwidth and provides a fundamental capability for the radar sensor.

Camera
In stereo-camera set-ups it is proportional to the accuracy of determining disparity, and inversely proportional to the baseline separating the cameras, and the provides a fundamental capability for the camera system.

Ultrasound
A single sensor primarily responds to one target. Potentially, multiple targets on the same line of sight, resulting in distinct times of flight could be detected, but such functionality is not currently implemented in existing systems. This capability is therefore not currently relevant to ultrasound.

iii 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 perpendicular to the Line of Sight between the sensor system and the scene.

Radar
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.

Cameras
The speed resolution is proportional to the range resolution and the frame rate of the camera.

Lidar
Some lidar sensors, (i.e. those using FMCW lasers), are able to directly estimate speed. In other cases, speed of an object must be estimated from the object's tracking information. Accuracy of speed estimation can be derived from a similar experimental set up to that for analysing tracking capability.

Ultrasound
Time of flight operating mode is not able to resolve targets with different speeds. In other (non-automotive) applications ultrasound sensors can exploit doppler effect to assess target speed, which could be exploited in future applications.

iv Maximum Detectable Range

Is the maximum distance at which the sensor is able to detect a target.

Radar
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.

Camera
Depends on a range of parameters including: spatial resolution, dynamic range, and atmospheric environment. It is defined as the distance where there is a 50% chance of a black extend object on the ground being detect against a bright horizon, typical given as the point where contrast threshold C_{TH}=0.05.

Lidar
Must be defined for a target of specified size and reflectivity.

Ultrasound
Ultrasound in air attenuates rapidly with distance, with higher frequencies attenuating more strongly. However, long-range ultrasound sensors are emerging that use higher frequencies (up to 70 kHz). The other factor determining maximum detectable range is the sound pressure level that the sensor is capable of emitting. 100 dB is a typical level. Significantly higher levels may be difficult to achieve with current sensor geometries. It is supposed that the emerging higher frequency sensors achieve their longer-range performance through increases in sound pressure level capability. The other important consideration is that the time of flight for the given distance must be shorter than the repetition rate of the pulsed stimulus, to be detectable.

v Minimum Detectable Range

Is the minimum distance at which the camera system is able to detect a target. Sensor systems for autonomous systems may not have a nominal minimum detectable range so in principle it can be 0 m.

Radar
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.

Camera
Typical mono camera systems do not have a nominal minimum detectable range so it can be 0 m. However, this is not the case in stereo-camera setups as there must be significant overlap in the FoV.

Ultrasound
In theory the minimum detectable range is given by the half of the acoustic wavelength, which sets a theoretical limit of around 5 mm. However, as described above, the pulsed nature of the acoustic stimulus leads to ringing in the transmitter sensor and a delay before the sensor can switch to receiver mode to capture reflected energy. This ringing limits the minimum detectable range of practical systems to around 15 cm when the actual distance to the target is to be quantified, and as low as 3 cm when the simple presence of an object is to be detected.

vi Maximum Unambiguous Speed

Is the maximum speed that the sensor is able to measure unambiguously.

Radar
For Radar 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.

Camera
In camera systems this parameter depends on spatial resolution, shutter implementation and frame rate of the camera which defines the sampling frequency.

vii Minimum Detectable Speed

Is the minimum speed that the sensor is able to measure.

Radar
For Radar 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.

Camera
In Cameras this parameter depends on spatial resolution, shutter implementation and integration time

viii Update rate/Responsiveness

Responsiveness measures the speed with which the sensing system provides its outputs in terms of detections of objects, and estimative of their properties (such as movement/speed) etc. Metrics that relating to responsiveness are (i) “update rate” or “refresh rate”/”frame rate” of the sensor, (ii) “object revisit rate”, the time between two shots at the same location, which recognises that detection and tracking of objects requires the same point to be revisited multiple times, and (iii) “latency” the delay between first imaging of an object, to output of the desired outcome (detection of object), and may include elapsed time of signal processing of the sensor’s output.

Camera
In camera systems frame rate is the limiting case.

Ultrasound
The pulse repetition rate is the limiting factor. Typically 20 Hz or 50 ms.

ix Tracking Capability

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.

x Tracking Capacity

Is the number of targets that the sensor is able to track at the same time, in congested environments. As such for CAV it is an important feature to be able to track as many targets as possible.

xi Contrast

Is a measure of the variance in luminance between the target and the scene background

Camera
Contrast (C_O) is defined as:

$$C_0 = \left(\frac{|\Delta L|}{L_B}\right) = \left(\frac{|L_T - L_B|}{L_B}\right)$$

where L_T is the luminance of the target and L_B is the luminance of the background[23] ,[24]. It is assumed that both the target and back ground are lit by the same source. Contrast Threshold (C_{TH}) defines the limited of detectability where probability of detection equals 50%. Normally assumed to be C_{TH}=0.05. It's a key parameter in the ability to detect/track/recognise targets.

xii Sidelobe Levels

Radar
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.

xiii Integrated Sidelobe Ratio

Radar
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.

xiv False Alarm Probability

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

xv Detection Probability

Is the probability of the sensor to detect 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.

xvi Dynamic Range

Is the ratio between the maximum and the minimum value of the received signal power that the sensor can handle.

Camera
In camera systems a poor dynamic range means that the sensor would easily saturate in presence of bright light such as direct sunlight or vehicle head lights leading to blinding (or lack of ability to detect targets in low light).

23. P.C.D Hobbs, "Building Electro- Optical Systems", 2nd Ed., (John Wiley & Sons, 2008)
24. G.C. Holst & T.S. Lomhein, "CMOS/CCD Sensors and Camera Systems," (SPIE Press, 2007).

Ultrasound
Is the difference between the upper limit of linear operation (at some defined level of distortion, typically 3%) and the noise floor. As noted earlier, the signal received by the sensor is also generated by the sensor, so there is no possibility of overload or saturation. The minimum detectable signal will be governed by the sensitivity of the sensor in receiver mode and of the detector electronics. However, dynamic range is not normally specified by sensor producers. It has limited value in time of flight based systems.

xvii 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.

xviii Linear Range

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.

xix Linearity

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.

xx Antenna Patterns

Radar
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.

xxi Field of View

The angle through which the sensor can detect electromagnetic radiation, it can vary significantly between the azimuth and elevation.

2 Advanced Sensing Capabilities:

A definition is provided for each advanced capability of perception systems, differentiation for individual sensors types is provided where appropriate.

i Artificial Intelligence (AI)

Are advanced processing frameworks that exploit AI to extract targets' information. Their use needs to have extensive validation and the outputs must be investigated in details and reproducibility must be tested.

ii Clustering of Extended Targets

Is the capability to group together multiple detections from the same object. High range resolution sensors would provide multiple detections in an area that a sensor would be able to group together in order to enable more reliable target recognition and tracking.

iii Target Recognition / Classification Capabilities

Is the capability of the sensor to discriminate different classes of targets, the set of interest depends on the application domain. It is often associated with a confidence level for each recognition output, [9] such as Targeting Task Performance.[25]

3 Operational Capabilities:

A definition is provided for each operational capability of perception systems, differentiation for individual sensors types is provided where appropriate.

i Interference Mitigation

Is the capability of the sensor to mitigate the interference from other active sensors and/or the environment. Waveform diversity and frequency division multiplexing (automatic frequency allocation) are standard approaches to mitigate interference.

ii Effect of Sensor Mechanical Instability

Is the effect on the sensing capabilities of the mechanical instability of the sensor when fitted on the vehicle, it is important on both short time-scale (in operation) and long-term time scale (over the period of use of the sensor between servicing).

iii Multipath Mitigation

Is the capability of the sensor to deal with the electromagnetic multipath that might occur, especially in the congested/urban environment. Poor multipath mitigation might lead to the detection of ghost targets and false tracks.

iv Robustness in Non-Ideal Weather Conditions

Is the capability of the sensor to maintain the KPI within the design range when the sensor is operating in non-ideal weather conditions, such as: rain, fog, snow, ice, hail, sandstorms, surface water, water sprays. Assessment of the KPI should be performed at various ranges of operations and with various level of weather severity.

Camera
Additionally, for cameras attention should be paid to the performance of systems in low light at high temperatures where the level of dark current and its effect is maximised.

25. S. Blackman, "Design and Analysis of Modern Tracking Systems," (Artech House, 1999)

Ultrasound
Sensor functionality is largely unaffected by weather conditions. Ultrasound will propagate through rain and fog and is unaffected by levels of light. However, the sensor sensitivity depends on temperature so detections requiring operation close to the noise floor may be affected. In addition, the speed of sound is a function of temperature and relative humidity, therefore these factors will influence the time of flight estimates and therefore the distance-to-target estimates, unless additional compensation is made. The packaged sensors are generally sufficiently robust to withstander most weather conditions, although a build-up of snow or ice may create a barrier that impedes ultrasound propagation.

v Robustness to Physical Impairments

Is the capability of the sensor to maintain the KPI performance when operating in environments presenting physical impairments, such as dirt, mud, snow and ice on the aperture of the sensor.

4 Vertical Capabilities:

A definition is provided for each vertical capability of perception systems, differentiation for individual sensors types is provided where appropriate.

i Interfaces/API for Integrability

Is the provision of standard interfaces/API to ensure the integration of the sensor within the electronics and software on board of the vehicle.

ii On-board Processing Validation

Is the validation of all the sensor KPI once the sensor is fitted and integrated with the other sensors. The validation should be performed in realistic operational environments.

iii Self-assessment/Calibration& Fault detection

Is the capacity of the sensor to perform autonomously calibration and assessment of the performance in order to provide a fault detection capability.

iv Frequency of Servicing

Is the frequency that the sensor would need to be tested and calibrated, important features are also how easy is the sensor to service, including provision of sensor testing/ calibration tools.

5 Electromagnetic Compatibility and Safety

A definition is provided for each electromagnetic capability of perception systems, differentiation for individual sensors types is provided where appropriate.

i Electromagnetic Compatibility and Radio Spectrum Matters

Are the characteristics related to the electromagnetic emissions of the sensor.

ii Eye Safety

Lidar
To guarantee eye safety, lidar sensors should only use lasers certified as Class 1 Eye-Safe per IEC-60825-1. No additional testing of eye safety is then necessary.

Appendix B

Sensor Capabilities Testing Methodologies

A method is suggested for testing the each of the key capabilities of perception systems, differentiation for individual sensors types is provided where appropriate.

1 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[26] describes tests for multi-sensor implementations for covering key zones at different azimuths around the vehicle.

ii 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.

26. ISO 17386:2010 Transport information and control systems — Manoeuvring Aids for Low Speed Operation (MALSO) — Performance requirements and test procedures

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.

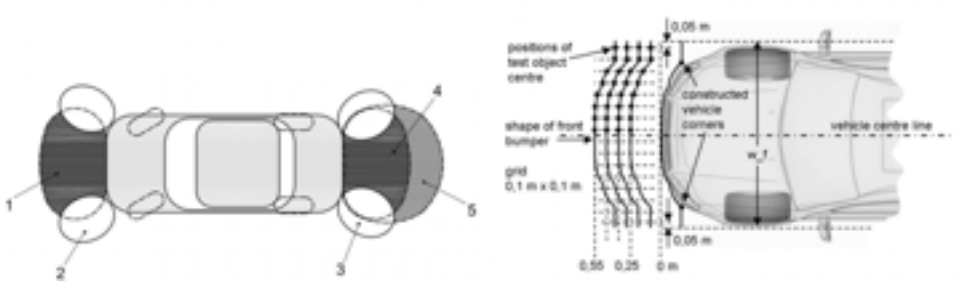


Figure 1: ISO 17386 Monitoring zones and example grid pattern identifying test locations at the front of the vehicle.

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.

iii 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.

iv 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 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 of 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 characterisation depending on the target's size and material.

Camera
In order to test the maximum detectable range an extended target black should be positioned against a bright horizon, at a range 1m 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 of 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 metre 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 metre 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.

v 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 (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.

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.

vi 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.

vii 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.

viii 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 rates.

ix 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.

x 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.

xi 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.

xii 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.

xiii 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.

xiv 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.

xv 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.

xvi 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_{max}) 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 obtained 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.

xvii 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.

xviii 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 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[27]) 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

27. 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)

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.

xix 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.

xx Antenna Patterns

Radar
In order to test the antenna pattern, standard antenna measurements procedure should be used[28].

xxi 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.

2 Advanced Sensing Capabilities:

Recommendations are provided for testing the each of the advanced capabilities of perception systems, differentiation for individual sensors types is provided where appropriate.

i Artificial Intelligence (AI)

The use of needs to have extensive validation and the outputs and explainability must be investigated in details and reproducibility must be tested. Generative Adversarial Networks (GAN) in conjunction with real world sensor data of determined uncertainty, could be used provide broad coverage of supervised learning, testing and fuzzing.

ii Clustering of Extended Targets

Moving and static varying geometry targets could be used to test the ability of high range resolution sensors to provide multiple detections and to group these together in order to enable more reliable target recognition and tracking.

iii Target Recognition / Classification Capabilities

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

In order to test the target classification capabilities extensive tests should be performed in all the operational scenarios. Depending on the actual nominal recognition capabilities and application of the sensor the tests should be performed using classes of different targets in different ranges, different aspect angles and speeds. The accuracy of the classification and the confidence level obtained during the testing should be measured.

3 Operational Capabilities:

Recommendations are provided for testing the each of the operational capabilities of perception systems, differentiation for individual sensors types is provided where appropriate.

i Interference Mitigation

Radar
In order to test the interference mitigation capabilities an interfering source that emits a frequency within the receiver bandwidth of the sensor under test should be used. The interferer should be located at different ranges and angles. The emitted power of the interferer should be known and no targets should be present in the scene. Noise floor and background should be removed from the sensor's output. The residual power will represent the residual interferer power (P_{IntRes}) that compared with the expected power from the interferer at the receiver (P_{Int}) will indicate a measure of interference mitigation.

The ratio between the two powers ($\frac{P_{intRes}}{P_{int}}$ expressed in dBs) will represent the interference mitigation capability of the sensor. If it is not possible to access the sensor's received power or received signal, then the interference mitigation would need to be assessed through replicating the tests depicted in Appendix B.1 for the capabilities IV, V, IX, X, XI, XII, XIII, XIV and in Appendix B.2 for the capabilities I, II and III in presence of an interferer and assessing the effect of the interference on these capabilities.

Lidar
In order to test the interference mitigation capabilities an interfering source that emits a frequency within the receiver bandwidth of the sensor under test should be used. Tests should be conducted with the interferer located at a variety of ranges and angles. The emitted power of the interferer should be known, Interference mitigation should be assessed through replicating the relevant tests of Appendices B.1 and B.2 in the presence of an interferer, and assessing the effect of the interference on the results.

Ultrasound
The main sources of interference are from extraneous environmental noise and ultrasonic systems on other close-by vehicles. Environmental noise is generally considered not include ultrasound components. However, there is an increasing number of emerging sources of ultrasound to be found. The most relevant to automotive applications and the most prevalent, are ultrasonic deterrents for birds, rodents and other animals. These operate in a similar frequency band to automotive ultrasound sensors and are seeing widespread outdoor use in urban areas.

The testing principle is similar to that outlined for Radar where, in the absence of a target, phantom detections due to the interfering signal can be evaluated. This will most likely need to be conducted in a laboratory environment where other sources of background disturbance can be controlled.

ii Effect of Sensor Mechanical Instability

In order to assess the effect of mechanical instability on the sensor capabilities, the sensor once fitted on the designed vehicle component (i.e. car bumper) should be tested. A set of controlled mechanical stimuli should be used to emulate realistic mechanical vibrations and oscillations. The effect of the mechanical instability would need to be assessed through replicating the tests depicted in Appendix B.1 for the capabilities IV, V, IX, X, XI, XII, XIII, XIV, XV and in Appendix B.2 for the capabilities I, II and III in presence of the mechanical stimuli and assessing their effect of the sensor capabilities.

iii Multipath Mitigation

In order to test the multipath mitigation capabilities of the sensor the tests depicted in Appendix B.1 for the capabilities IV, V, IX, X, XI, XII, XIII, XIV and in Appendix B.2 for the capabilities I, II and III should be replicated in a cluttered environment (e.g.: between buildings) assessing the effect of the residual multi-path on the sensor's capabilities.

iv Robustness in Non-Ideal Weather Conditions

To test the robustness against physical impairments the tests the tests depicted in Appendix B.1 for the capabilities IV, V, IX, X, XI, XII, XIII, XIV, XV and in Appendix B.2 for the capabilities I, II and III should be replicated at the presence of physical impairments, such as dirt, mud, snow and ice occluding the sensor aperture at a distance equivalent to the distance where these impairments would build up once the sensor would be fitted on the vehicle. The effect of the impairments on the sensor's capabilities should be assessed in these conditions.

Camera
Additional testing of camera lowlight sensing capability should be conducted at elevated temperature to determine the impact of dark current on detection ability.

Ultrasound
There are no additional requirements to those stated above for ultrasound sensor. Naturally, only those tests relevant to ultrasound sensors need be carried out.

v Robustness to Physical Impairments

To test the robustness against physical impairments the tests the tests depicted in Appendix B.1 for the capabilities IV, V, IX, X, XI, XII, XIII, XIV and in Appendix B.2 for the capabilities I, II and III should be replicated at the presence of physical impairments, such as dirt, mud, snow and ice occluding the sensor aperture at a distance equivalent to the distance where these impairments would build up once the sensor would be fitted on the vehicle. The effect of the impairments on the sensor's capabilities should be assessed in these conditions.

4 Vertical Capabilities:

Recommendations are provided for testing the each of the vertical capabilities of perception systems, differentiation for individual sensors types is provided where appropriate.

i Interfaces/API for Integrability

In order to test Interfaces and APIs for integrability the interfaces should be standardised and provide access to the sensor's measurement. APIs should be able to provide all the information about the operating functions of the sensor and allow for enabling/disabling capabilities in order to enable testing of individual and combination of sensor modes.

ii On-board Processing Validation

The on-board processing validation should be performed by testing all the sensor's KPIs once the sensor is fitted on the vehicle.

iii Self-assessment/Calibration& Fault detection

In order to test the robustness in non-ideal weather conditions the sensor should be tested in the following weather conditions: rain, fog, snow, ice, hail, sandstorms, surface water, water sprays. Additionally, for cameras, attention should be given to performance with temperature. The tests the tests depicted in Appendix B.1 for the capabilities IV, V, IX, X, XI, XII, XIII, XIV, XV and in Appendix B.2 for the capabilities I, II and III should be replicated in these weather conditions and the effect of the different weather conditions on performance of the sensor should be assessed. The self-assessment, calibration and fault detection capabilities (were available/claimed) should be tested through the interface/API of the sensor. The test should be performed while performing the tests for the sensor's capabilities described in Appendices B.1 and B.2. Special cases, such as sensor blinding with a close and strong interferer as well as a physical occlusion should be evaluated assessing the self-assessment, calibration and fault detection capabilities of the sensor.

iv Frequency of Servicing

This capability should be indirectly assessed based on the other KPIs tested. In particular the effect of mechanical stability will provide an estimation of how often servicing would be required.

5 Electromagnetic Compatibility and Safety

Recommendations are provided for testing the each of the Electromagnetic Compatibility of perception systems, differentiation for individual sensors types is provided where appropriate.

i Electromagnetic Compatibility and Radio Spectrum Matters

Aspects relative to limits, testing and calibrations are described in the standard ETSI EN 301 091-1 V2.1.1 (2016-11) [7].

ii Eye Safety

Lidar
Use of Class 1 Eye-Safe lasers as per IEC-60825-1 to avoids the need for additional testing.

Appendix C

Market Analysis Detail

Interviews have been conducted from a cross-section of the CAV Sector, including the categories:

- OEMs
- Sensor Developers and Manufacturers
- ADS Developers
- Testing - experts and facility providers
- Simulation Tool providers
- Government policy and standards bodies

Fifty potential organisations were approached for interview and 18 were able to participate. Interviews were designed to elicit responses which provide us with a basic understanding of the state-of-the-art; current challenges with automotive sensor testing; where the gaps are; and what are the barriers to creating new methods, services and facilities to plug these gaps. *Some organisations have expressed caution about participating from a competitive standpoint as either a) they feel they are able to undertake their own sensor validation; or b) they are concerned that this type of activity has the potential to interfere in their business models.*

The information below constitutes a cross-section of highlights from the interviews undertaken so far. This information should be treated as official sensitive.

Standards

- The UK CAV Standards Program has identified a series of standards to be created to assist in the development of Connected and Autonomous Vehicles.

- Standards which consider the underlying sensor performance currently do not exist. There is some evidence that standardisation of performance metrics would facilitate the understanding of sensor behavior, thereby enabling sensor models which are more accurate and representative of real-world limitations. Feedback from stakeholders has highlighted a need for standards which:

    - Consider the integration aspects of physical sensors in vehicles and sensor models in simulation and some standardisation of data outputs between components.
    - Are technology agnostic and do not mandate a particular design, but which focus on the KPI metrics which assess the perception performance of the sensor system as a whole rather than the fidelity of individual sensors.

- Define a standardised way of characterising and measuring weather conditions and other environmental factors which affect sensor system performance.

Sensors: challenges

The underlying sensor performance standards do not currently exist. There was support for an aspect of a future competition which would look to understand what shape such standards should take. At this current stage reference is made only to related standards for example: for example, referenced from LiDAR module suppliers - IP69K, Vibration: SAE J1211, Shock:IEC 60068-2-27, ETSI EN 300 019-2-5. Auto operational temp range -40 to +80°C.

A common issue noted by respondents is the constraints preventing commercial organisations to share information or standardise interfaces. This would seem to be motivated by concerns about giving away IP or confidential / proprietary data. This has translated into limited availability of performance specification at the level of detail required by ADS developers and Simulation Environment operators.

Respondents expressed differing levels of confidence with regards under which conditions each sensor type would be operating outside its performance limitations. Lack of common approaches or reliable data for modelling conditions - in particular weather - or specifying performance is considered an issue across respondents.

There is evidence of divergence between the Tier 1 / OEM sectors where the focus has so far been on ADAS integration; and the newer entrants into the market for ADS development and sensor design, who are focused on level 3 and above solutions. The latter placed a stronger emphasis on needing to understand the performance characteristics of their sensors and how they degrade under different conditions. Many agreed that what can be bought off the shelf is not good enough for L4/5 and probably L3: increased performance and functionality is required; and better-quality performance data. A regular response from Vehicle Manufacturers is that they want to be assessed on whether the outcome of the data fusion system is capable of delivering the desired performance; they are not interested in indicating the fidelity of each sensor type.

There is a challenge transitioning from Level 2/2+: currently ADAS sold as a whole package (hard to analyse the sensor performance): i.e. one company provides the package of sensor, software and decision / instruction. One example of concern voiced by respondents is that different radar systems can produce different numbers of targets.

Another area of measurement which several respondents felt was critically important to be correct on revolved around the fusion of sensor outputs - for example, matching imaging data and synchronising with 3D data provided by LiDAR, proximity data provided by RADAR.

The key requirements in determining which sensors to use included:

- **Price** - Cost was a significant determinant. At the same time consideration is given by some developers to whether to spend more on a better sensor with more information.
- **Reliability**
- **Performance** - what does it do?

Sensors are relatively expensive, and the supplier base is small, although this is showing signs of growth. The sensor supplier industry structure and approaches used are fragmented. There are implications on quality / conformity / reliability as costs are driven down. Data from manufacturer varies, but it is not considered sufficient, for example with LIDAR there are no suitable standards (e.g. for range).

Feedback from one respondent was that in order to meet cost targets of ~$1,000/LiDAR module then there needs to be a move to commoditise CS detection systems as there has been in the optoelectronics communications industry and with respect to Si semiconductor detection systems (this comes hand in hand with the need to move to longer operational wavelengths >1,400 nm to achieve greater range and remain eye-safe). Solid state beamsteering solutions based on Si Photonics may also provide a pathway to scalable solutions for hybridised systems.

The primary use designated for different sensors was varied and this seemed to be guided by both modality of the vehicle and also the known performance of the sensor.

Sensor manufacturers rarely, if at all, warranty the performance of their products for safety critical applications.

With regards Sensor types there is no common approach on sensor mix, which could partially be explained by different modalities/ODD for vehicles. For example:

- None of the ADS developers interviewed incorporated Ultrasonic in their sensor suite for AV – some considered this not compatible with the operating parameters of the vehicle.
- LiDAR was not used by some ADS developers due to cost issues, while others prioritised LiDAR.

Some Respondents indicated that further testing was undertaken against sensor manufacturer's stated performance, either through stand-alone testing, or on-vehicle testing. Other respondents indicated they wanted more control over sensors than OEM's could offer. In other instances, testing is currently empirical, based on post-test review of recorded data taken from the physical and simulated environments. Approaches included:

- Using a test vehicle (not autonomous) which collects data from the roads. Data is then assessed against benchmark data sets and fed into their test environment. Using GNSS and other ground truth sensors for reference. Issues with the test vehicle approach, as it is impossible to create every scenario and use case.
- Several ADS developers used bench testing - test for distance and data rates. (Would value the ability to do more bench test on a sensor before deciding to use it, however this is limited as there are no standard targets.)
- Respondents indicated they couldn't test for reliability, vibration and temperature.
- Some respondents developing test methods to identify edge-cases, identify performance limitations and provide in-situ testing to assess performance during temperature cycling and vibrational stress.
- Development of photon link budget models that incorporate effects of turbulence, validated by test data.

At the moment the market doesn't have a collective view/understanding regarding which aspects of measurement are of critical importance, the following were highlighted:

- Weather conditions and testing with realistic objects at range is critical and is the biggest question mark. Do not have a good understanding of how certain sensor types behave in different weather and environmental conditions, nor how to characterise certain weather events in terms of the effect on the sensor.
- Other environmental effects are also important.
- Installation and packaging have an influence on sensor performance.
- System must be able to localise (bearing and velocity also).
- Effects of installation e.g. vibration etc is important but is more within respondents' control.
- Latency when looking at milliseconds, or performances.
- There is some evidence industry has only limited appreciation of the challenges of complex electromagnetic environment i.e. multipath return signals, interference, sensor "blinding." Some academics are working on it and previous work has been done in Defence Sector for EM-Warfare, but this was for £ xxxK units not 10p commodity products.

Respondents had differing levels of confidence with each sensor type with regards the conditions it will be operating outside its performance limitations.

- Many agreed that what can be bought off the shelf is not good enough for L4/5 - More performance and functionality required; and better-quality data. e.g. for LIDAR: 3d Point cloud and angular resolution: IS velocity/direction of each pixel relevant.
- Weather is the key priority to characterise and test for. Along with the associated environmental effect It is not clear how weather is defined. Not strong evidence that the sensor manufacturers have this information.
- For weather definitions there is also concern about the level and relevance of detail: e.g. factors such as droplet size. Recreating such conditions is a real challenge.
- Limited access to fog chambers of significant length.
- The following risks in the testing processes need mitigating; Redundancy, resilience to cyber-attacks, performance continuation at edge-cases.

The use of ML in the sensors themselves presents challenges as this makes it very hard to determine whether any two different sensors are identifying the same objects in the same way. A structured test process accompanied by performance standards would be welcomed by some respondents.

Investigation into the processes currently being considered to validate the outputs from Sensor data fusion systems, indicated a lack of conformity and a reluctance for any single organisation to accept liability in the validation process.

Sensors: identifying solutions

Better collaboration would be useful to help understand sensor performance criteria and a common, standardised language / interface for defining this. The industry will always be interested in common metrics/semantics and reference points which help with how they validate their own systems. There was consensus on the value of a common approach on the characterisation of different conditions affecting sensors: in particular, it would be useful to have definitions of weather types, and how that affects sensor performance. A sensor characterisation

call would be welcome by the industry - a funded CR&D project along this theme which helps commercial organisations to de-risk this, could be beneficial.

Respondents have highlighted that the UK has different combinations to other countries (and vice versa). Therefore, any sensors in the UK market need to be able to handle UK conditions, while for any UK test facility to attract international clients it needs to be able to replicate combinations of conditions available in different locations. This highlights the importance of developing standardised approaches to weather condition models, which can be re-used for local weather conditions around the world.

Sensor positioning on different platforms is considered key for different users and for overall reliability and is affected by known performance: recommended this as an item for validation.

There was interest in both smaller test chambers for reliable, repeatable testing of individual sensors; as well as a large scale facility, for testing sensor systems, navigation systems and whole vehicles, where objects can be controlled, and weather conditions simulated.

With regards the gaps in the current testing facilities for CAV sensors, general response was that the industry is reliant on the UK weather to provide 'variation' in weather, which they need to test with (not currently underpinned by common definitions). To become commercially viable, any large-scale facility would need to be able to emulate environmental conditions on a repeatable basis. For some this is not necessary at a whole vehicle level and could be done on a sensor level - it would still require a fairly large facility to get the several tens / hundreds of meters range that modern sensors can detect.

All respondents felt there is definitely a case for a facility where a manufacturer brings a sensor / sensor set to run a standardised set of tests which produce results / a set of data in a standardised way. There also needs to be test capabilities at all levels of integration from wafer through to chip, components and system and effective and agreed standardisation of the testing protocols.

Simulation and sensor models: challenges

Feedback from organisations developing / operating simulation environments was that Tier 1 OEM's are providing minimal data on their sheets, therefore, not enough to create models to test with. Respondents indicated that the traditional Tier 1 suppliers have so far engaged less than some of the smaller sensor manufacturers. One respondent observed that Tier 1's are only just getting involved with Simulation.

Fundamental challenge is how to replicate sensors in SIM environment. It is necessary to have sufficient information from sensor suppliers to understand how the sensor works, however the data supplied is usually insufficient. In particular, correctly defining noise and performance levels for sensor models and quantifying fidelities. The lack of standardised interfaces between automotive sensors and other components is seen as a significant issue.

Agreement from respondents that simulation must play a role in both dev/testing and accreditation. But this raises questions such as what those requirements for accreditation of tools are? Other activities focused on standards e.g. scenario standards, file formats, interfaces between components such as virtual environment to CAV > OSI standard for transferring data; industry standards for imagery (streaming).

Sensor models: for most of the industry, focus is at level 2 and 2+. (ADAS) What this means is that currently, Tier 1 provide measurement interfaces, then Tier 1 meets requirements which are written by OEM. A major concern for Tier 1 (and all sensor manufacturers) is that the internal information of sensor represents a big IP concern. European OEM/Tier 1's more straightforward to work with.

Industry as a whole overly dependent on "idealised photorealistic" camera-based simulations - not electromagnetically valid for all (in some cases any) interrogating wavelengths. Do not account for weather conditions in any complete or meaningful way i.e. physically/ electromagnetically valid. Sensor characterisation / models for conditions: these are considered essential for test scenarios (by the respondents) to move on from idealised tests.

Respondents who manage or develop simulation test environments agree with the need to understand better when/how sensors fail; this includes issues such as vibration and sensor misalignment; Installation and packaging also have an influence on sensor performance.

There have been numerous claims in company sales pitches that sensor models are validated, but this does not bear up under examination.

Sensor characterisation / models for conditions: these are considered essential for test scenarios (by the respondents) as currently too idealised.

Sensor suppliers are not providing the models but are helping to understand how they work. It is not critical to get the core models for the sensors but what is more important is having the engagement from the supplier to understand and interpret the data. (Simulation Environment operators also want to avoid the risk of 'baking' sensor manufacturers IP into their models). More likely to get this from the smaller suppliers compared with the big Tier 1s. (This is potentially a big advantage to the smaller, more co-operative suppliers.

Respondents don't want the model to provide too "clean" an output to ADS, so want to create noise such as imperfections caused by weather conditions and perhaps vibration.

Sensor positioning on a platform is also key; i.e. need to focus on what the sensor model looks like, both in terms of the physical geometry as well as its performance.

There were different views in determining which environmental conditions affecting sensor response to take into account in simulation. Not considering radar scatter at the moment, as it's hard to set up a test. Not sure how to recreate LIDAR interference.

With weather conditions and models, observations include:

- Understanding performance metrics is very important to respondents and is an area to be considered in the near term, but some indicated they had to first focus on sensor models.
- How to correlate O/S conditions in SIM environment to real weather. What material properties are changing? I.e. how is sensor response affected by a wet vs dry target.
- Work planned in the future on pressure and temperature interfaces.
- Some have used a weather station nearby to record data for emulation in the simulation; e.g. pressure, temperature, light levels.
- v. hard to test weather conditions in a simulation model / environment. Confronted by the challenge of how to externally evaluate these? The sensors work on/at their boundaries. - in this instance, the models become v. complex or too complex.

When building sensor models for virtual simulation (e.g. physics based, interpolation of testing data, assumption of perfect response based on manufacturer's specification, other):

- Different respondents used different approaches, or have not decided on an approach, to validate the reliability of sensor models and simulation environment stimuli; e.g. validating against data collected in a real car (collected data acts as ground truth); validating the system (sensor model + environment model + AI) as a whole; validating individual parts.
- Some respondents indicated that they used physics-based sensor models or are seeking to develop these. There was indication of a mixed approach between physics-based model; Ideal ground truth and probabilistic model (e.g., for Radar) where specific targets are parameterised; and data-based models which come from publicly available data sheet.
- Examples of respondents taking data RADAR and LIDAR sensor mounted on rig (different azimuth, length, weather etc). Then running test procedure to measure sensor performance. Building up database to understand what and how to correlate against certain conditions.
- With cameras one challenge identified was how to get data from graphics card to AI processor.
- One respondent created adaptors / gateways to interface with sensors in order to convert data (software based).
- In one example, sensor models were created on the basis of providing object list to test HIL ADAS. This involves no physical sensors (except ground truth sensors); only software, not the sensors are therefore tested in ADAS and therefore the focus is on creating viable Artificial or Synthetic data.
- More than one respondent is intending to run tests in real world on real vehicle which will then be compared with the simulated environment to correlate.

An identified challenge was the need to generate arbitrary scenes and work out how to feed this into simulation.

Raw v processed data is another challenge: Raw data is important to some AI developers (some don't like sensors doing object detection; some want more processing at the edge). Some developers indicated that the sensor models need to generate the raw detections (rather than e.g. processed target information). One proposed resolving this issue of post processing sensors in a simulation environment by separating out raw data and processing into a sequence.

Some respondents trying to work out what low and high fidelity actually means in terms of the quality of response of the sensor model to the stimuli.

Simulation/sensor models: identifying solutions

It has also been noted that respondents see benefits to linking a physical testing facility with simulation environments, and a desire to loop sensor hardware into the simulator in order to feed it synthetic data.

It has been discussed in several interviews that a large testing facility should be able to test the impacts of a combination of influencing factors, e.g. rainfall, leaves, loss of GPS coverage etc., and in particular replicating weather conditions as they impact CAV sensors. This could be a source of differentiation in the international market.

With regards weather models, respondents indicated the need for: 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 .

Facilities which are able to create repeatable conditions in which to test the effects of weather and environmental variation on RADAR sensor properties, and correlate against the results from models and simulations, are needed.

The developers of military simulation environments have looked into some of these issues and there are some standards for effects caused by weather etc such as HLA, an object model which represents weather as a set of objects which can be used in simulation.

Appendix D

Sensor models for virtual testing

Sensor models for virtual testing are different from many applications of modelling, because the user of the models generally cannot change the operation of sensor (other than by exchanging it for another sensor), and so does not care why or how it is doing what it is doing, whereas many models are used either to design an object or to control a process so that understanding the link between cause and effect is a lot more important. This difference affects the model types that are of most use for sensor modelling for virtual testing.

The aim of the model is solely to reproduce the behaviour of the true sensor, rather than to explain that behaviour or to provide guidance on how to achieve a particular performance level.

Virtual testing of autonomous vehicles is clearly a safety-critical application, and so every aspect of the modeling and software associated with the application must be tested and documented to the highest standards. For physics-based models, this means a complete mathematical specification of the equations being solved, a list of the assumptions made in their derivation, any input parameters, and a description of the methods used to solve the equations. For data-driven models it means that mathematical form of the model and all parameters must be specified, and a specification of the measurement data on which the model is based, and the method used for parameter estimation should also be supplied. In both cases a test specification and associated results should be supplied.

Definitions and vocabulary

A **model** is an algorithm that takes in a set of physically meaningful inputs and calculates a set of outputs. A good model generates outputs that correlate well with the behaviour of the real sensor.

An **input** to a model is a quantity that the model uses to calculate results, and that is expected to be varied from model run to model run. An example would be the intensity of the received scattered signal for a lidar model. Model inputs are typically the quantities that are varied during the testing used to generate data.

A **parameter** or **hyperparameter** of a model is a quantity intrinsic to the model (and not necessarily meaningful outside of the model) whose value is calculated during the model building process, and whose value will not change from model run to model run. An example is the slope of a straight line fit to data. If the same approach is used to model several different sensors of the same type, it would be expected that each sensor would have its own parameter values.

An **output** of a model is a calculated result that describes a quantity of interest to the model user. For sensor models, the outputs are the sensor responses, ideally provided in a form that is easy to interface with the vehicle AI under test.

Literature survey

Current literature seems to focus largely on problems of systems implementation and the differences between vehicle in the loop, hardware in the loop, and software in the loop approaches. A thorough overview is given in [1], where a list of simulation tools that include models of the most commonly used AV sensor systems is given. As many of these tools are proprietary it is not easy to work out whether the associated sensor models are data-driven or physics-based.

Many of the papers that consider simulation of sensing systems take a purely geometric approach [2, 3, 4] and do not consider uncertainty or noise. Some of the research [5] goes to great lengths to ensure accurate reproduction of the physics of complex phenomena such as the scattering of light by water that could perhaps be more efficiently considered as a source of uncertainty, but still assumes perfect sensor response rather than capturing real sensor behaviour.

Two papers have been found [6, 7] that do consider noise and uncertainty. Work from Old Dominion University [6] reports the results of combining real (i.e. physical) and virtual (i.e. perfect) sensor data and adding "distortion". The approach does not explain how the distortion is defined, but the approach is an interesting way to combine the idealised and real sensors.

A technical report from the University of Wisconsin-Madison [7] discusses an autonomous vehicle simulation framework including LIDAR and camera sensors that includes explicit discussion of noise. In particular it provides a description of the read and photon noise distributions for camera sensors, and the various sources of error and noise associated with LIDAR. The work is presented as a work in progress, so no results are shown.

Fitting the sensor model into the testing

Sensor models for virtual testing essentially act as an interface between the vehicle AI and the artificial testing environment. The typical information flow in shown in figure 1. The vehicle AI sends requests for information to the sensor model (this process may not exist if the sensor operates in “always on” or “push” mode), the sensor model probes the testing environment, and receives a response which it then converts into a signal which is sent to the vehicle AI. If testing is to be set in a regulatory environment, each of the information flows within this figure will need to be in a standardised form so that the system for test can easily be integrated into the test environment.

A complication to this flow is the distinction between the vehicle artificial intelligence (AI) and the sensor, in terms of where functionality such as object recognition takes place. At present, many sensor manufacturers are selling their products as advanced driver assistance systems (ADAS). These systems often have object recognition and/or tracking built into them so that the driver can be given information in an easily digestible form. These algorithms are almost always “black box” and are often based on machine learning or similar approaches for object recognition.

Two problems arise if fully autonomous vehicles deploy this kind of sensor. The first is that there is a potential conflict between the vehicle AI, which receives information from multiple sensors (some of which may not include object recognition), and the sensor AI. The second is that in order to correctly reproduce the sensor response, the sensor AI would have to be wrapped into the virtual testing environment directly and connected to a model of the non-AI parts of the sensor. This requirement would mean that any autonomous vehicle to be tested virtually would have to have access to a form of the AIs of any sensors separately from the sensor, and would have to have a characterisation of the sensor without the AI independently so that the sensor response can be modelled and thus can communicate with the virtual environment. This approach is sketched in figure 1B. It seems likely that vehicle developers would prefer to wrap all object recognition etc. into their systems as this will simplify testing and internal quality assurance.

Another potential complication is the interactions between the vehicle and the sensor. These effects include the method of mounting, including any covers or paint over the sensor and possible damage or drift of the mounting over time; the vehicle motion causing vibration or similar local motion of the sensor; and possible sensor damage or occlusion due to impacts or dirt. In general, sensor characterisation data will be captured from an unmounted static sensor in perfect condition. It is expected that vehicle models within the virtual environment will make it possible to consider many of these aspects so they will not be discussed in more detail here.

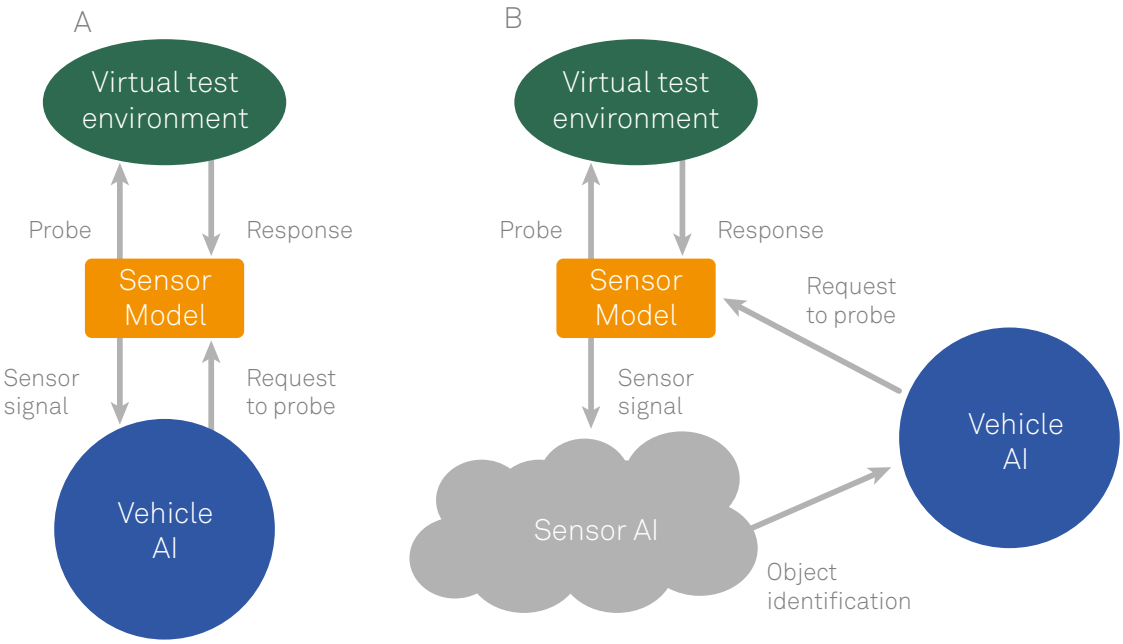


Figure 1: Simplified information flows in a virtual testing environment without (A) and with (B) a sensor AI. In the real version each vehicle AI will exchange information with multiple sensor models.

The inputs to the sensor model and the outputs of the testing environment need to be complementary. As an example, consider the situation shown in figure 2A, where a LIDAR sensor is operating on a foggy day. If a data-driven approach is used and data has been gathered that describes the sensor's operation under foggy conditions, then the sensor model will treat "fogginess" as an input quantity and will vary its response accordingly (as illustrated in figure 2B), and the testing environment will supply "fogginess" as one of its outputs on being probed by the sensor model. If that data is not available, but the response of the sensor to an incoming laser-generated signal of known power is available, then the testing environment would need to calculate the effects of fog on signal power and return a suitably attenuated signal to the sensor model so that the sensor response can be evaluated (as illustrated in figure 2C). This kind of potential variation in testing environment definition will need to be reflected in the variation of data format standards for sensor models and testing environments, so that users can understand what sort of data and sensor models are required for different virtual testing environments.

This report will not discuss virtual environments in detail, but there are important points to be noted regarding specification of each environment's capabilities. Testing environments aim to simulate in faster than real time, and this may limit the level of detail available in the environment. In addition, collecting the data to populate the environment has an associated cost, and characterising every object for every sensor type may not be possible. It will therefore be necessary for every testing environment to specify any gaps or simplifications in its scope. In order for such gaps to be easily identified, it may be useful to define what information a perfect virtual environment would include, such as material properties for each object, weather condition information, road roughness, and ambient temperature and pressure.

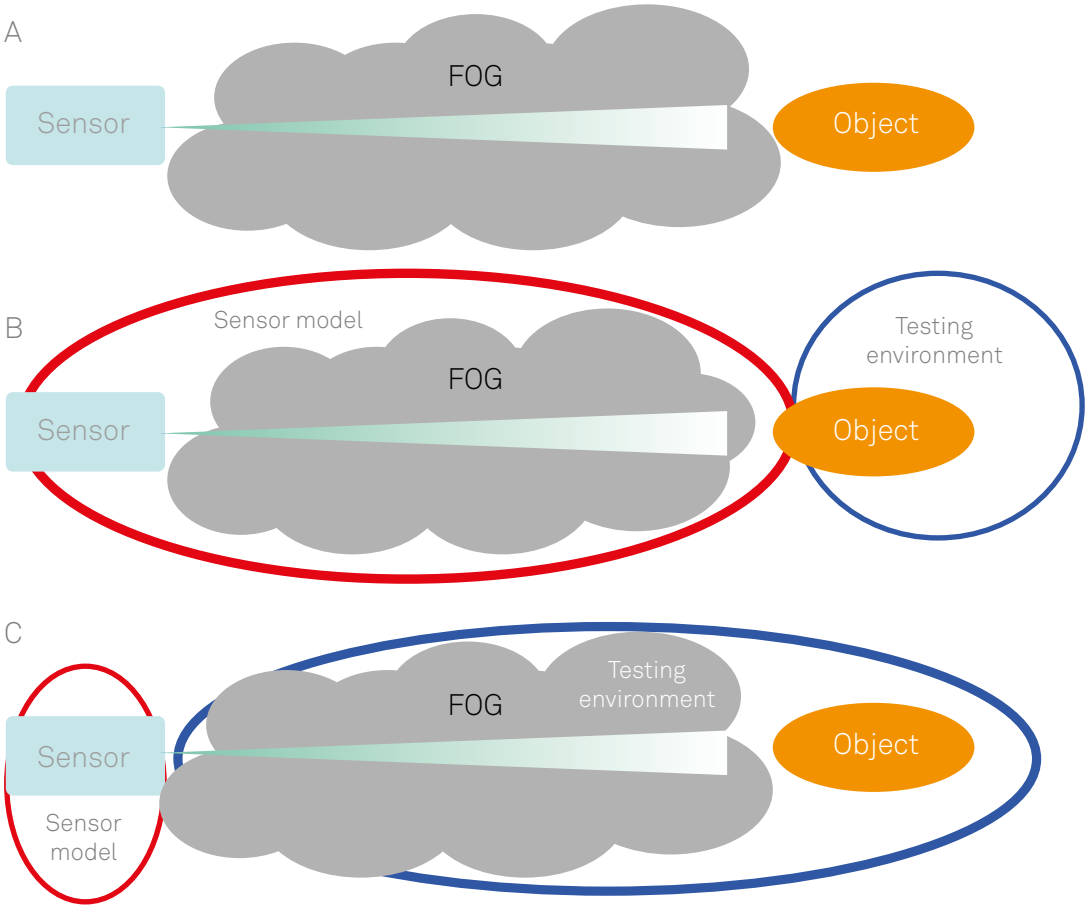


Figure 2: Example of different separations between sensor model and testing environment for the same scenario.

The output of a sensor depends on the sensor type. For radar and lidar sensors, the output is a specification of the nearest point that has scattered or reflected the signal in a given direction, whereas a camera provides a response to the distribution of scattered light from multiple undefined sources. For virtual testing purposes, the sensor models need to produce the same output as the sensor in the same format as the sensor based on the model response to the contents of the testing environment.

Data driven vs physics based

There are two main approaches to model construction: physics-based modelling and data-driven modelling. The two are not mutually exclusive, and a blended approach is often used. In general, models that answer, "how does this work?" questions need to include a significant amount of physics, because only physics can explain why things happen. Data-driven models can help to identify trends and correlations, and can reproduce the response of real-world objects, but correlation and causation are not the same.

For both approaches, a good model encapsulates all of the available knowledge about the system and imposes as few extra constraints on the system as possible. "Constraints" here includes extra parameters and over-fitting. As an example, consider figure 3. The figure shows the same set of data fitted with two different functions. The left-hand plot shows a straight line fit, and the right-hand plot shows a 10th order polynomial that passes through every data point. Unless there is a good physical reason to justify use of the high order polynomial fit, the straight-line fit is preferred as it is a simpler function with fewer parameters. Occam's razor applies: simpler theories and models are preferable to more complex ones because they are more testable.

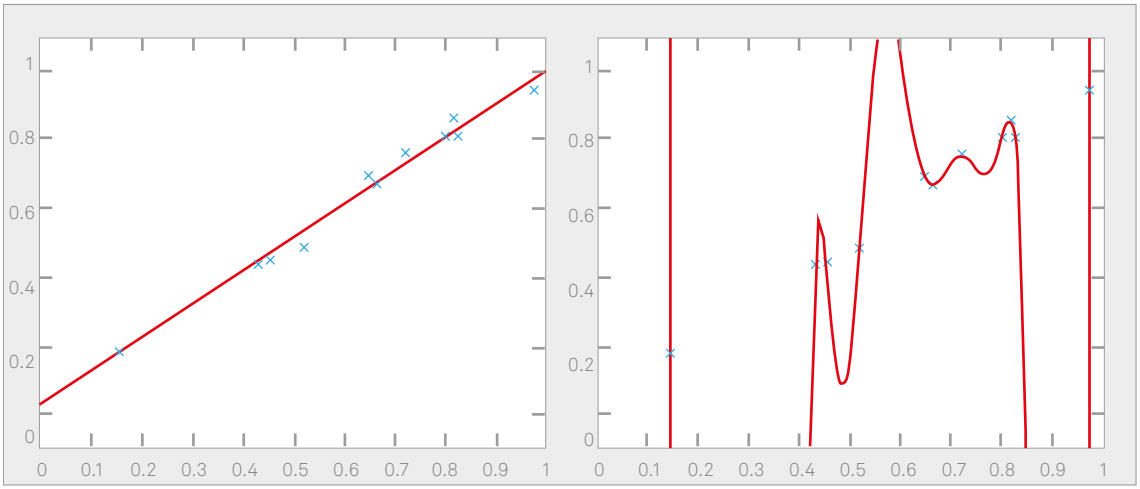


Figure 3: A data set fitted with two different functions. The axes on the right-hand plot are chosen to show the detail of the function near the points, with the result that the line goes off of the axes in some regions.

Physics-based models are often better placed to make predictions than data-driven models because data-driven models are effectively bound by the data used to create them, so for instance a physics-based model would be able to take the effect of fog on LIDAR into account by considering the light-scattering properties of water droplets, whereas a data-driven model that has only been given data for clear days will not be able to capture response in fog correctly.

This potential limitation shows the importance of careful planning: if the data-driven model is planned carefully then the test design used to obtain the data will include measurement in fog, and if the physics-based model is planned carefully it will take the scattering and absorption properties of the air into account. Any model development should be preceded by a thorough scoping exercise. This aspect is particularly important for safety critical systems (such as autonomous vehicles) where aspects that are beyond the control of the user (such as weather conditions) play a significant part in the system response.

This scoping exercise will feed into the test design as well as the model design. It should take into account physical insight and expert knowledge about the sensor as well as consideration of all possible operating conditions. For instance, the scoping would ensure that:

- the correct inputs are varied from test to test; for instance, it may be known that a particular sensor type is very temperature sensitive, in which case temperature will be treated as a variable quantity and the sensor response under different temperature conditions will be measured and used as a model input.

- the values of the inputs are chosen so that any complicated behaviour is reflected in the test results. Most types of sensor have a region of operation within which their response is linear or near-linear, but the sensor response outside of these regions can be strongly non-linear and the testing needs to reflect that by including enough points to capture the nonlinearity. In other cases, the sensor may be known to saturate above a certain level of stimulus, so fewer tests at high stimulus levels are needed.

- the full range of objects that the sensor is likely to encounter can be included; for instance, a shiny object will scatter electromagnetic waves differently from an object with a rough surface.

- enough tests are carried out that a set of validation data, entirely separate from the data used to build the model, is available.

The testing could also potentially look at effects of the mounting of the sensor on the vehicle, such as any objects that lie between the sensor and the environment, effects of angle and height, and potentially the effects of vibration from the road if the suspension system does not cushion the sensor adequately. The drawback with this is that it ties the sensor model to a specific vehicle, but it potentially improves model fidelity (provided these aspects are captured in the model).

Data-driven models are often quicker to evaluate than physics-based models once the appropriate parameters for a given data set have been obtained. Many physics-based models require numerical solution of partial differential equations, which is generally a time-consuming process. This disadvantage is significant for applications such as virtual testing where the goal is to have faster than real time simulation and becomes even more important when uncertainty evaluation using sampling methods such as Monte Carlo simulation is considered.

It would be possible to construct a surrogate model that runs quickly based on a physics-based model by applying the data-driven models outlined below to model results as though they are data. As with the data used for data-driven models, the model results must span the full range of operational situations that the sensor will see and should have associated uncertainties. This approach introduces errors, but provided they can be quantified and controlled this is a valid and effective way of addressing the computational expense problem for physics-based models.

Physics-driven models generally require the developer to have a detailed description of the construction of the system of interest, including numerical descriptions of the key features of the system such as geometry and material properties. This requirement is likely to be restrictive when modelling sensors. In general, the detailed nature of the elements within the sensor and how they are connected together is not available to the end user of the sensor, usually because it is proprietary information. It would therefore be difficult to construct a reliable model of the system that was purely physics-based. It would be possible to construct a physics-based model of a sensor that performs in a perfect manner, but obviously this would not reflect the real performance, and is arguably likely to be less accurate in the most dangerous situations.

Physics-based models will not be discussed in detail as it is unlikely that a typical sensor user will have access to the information required to develop a reliable physics-based model. A typical sensor would contain some or all of the following elements:

- some form of reactive element that responds to an external stimulus (e.g. a pixel in a camera responding to light) and typically produces an electrical signal;
- a system to enable the stimulus to reach the reactive element (e.g. a camera lens);
- a system to probe the environment if the stimulus to the reactive element comes from a scattered or reflected signal (e.g. a laser in a lidar system);
- some form of electronics to process the signal and output the processed results;
- other elements such as a power supply, temperature regulation, casing and mounting; etc.

The physics-based model would have to include descriptions of all of these features, unless they can be shown to have no effect on the sensor response. Hence the model would need to be a combination of an electric circuit model, a thermal model (unless temperature effects can be shown to be negligible), and a model describing the physical stimulus to which the reactive element reacts. In many cases the physics of the reactive element can be simplified, so for instance in general the physics associated with a laser would be described using a ray-tracing approach rather than by solving Maxwell's equations to obtain the electromagnetic waves.

The thermal aspects are likely to be particularly challenging to model because the response of individual electronic components to a change in temperature is likely to be difficult to obtain or estimate. Similarly any consideration of mechanical response to road roughness or potholes is likely to be difficult as it may require coupling to a full vehicle model to calculate how the sensor position changes.

Building a data-driven model

Data-driven modelling seeks to create a function that can take in any physically valid set of input values and return a corresponding output value by interpolating or extrapolating a set of test data (sometimes known as the training data set). The complexity of the interpolating function can range from a straight line or "nearest neighbour" approach to machine learning methods that generate functions with hundreds or thousands of internal parameters. The more complex the model, the more data is required to build the model and to validate it.

The first stage in developing a model based on data is always to plot the data. Plotting the data is often the easiest way to "sanity check" the test results, and to check the results against the physical understanding used to design the test plan. In general, it is better to start by producing a large number of plots with a small amount of data than one large plot with all the data on it, particularly for test results with a large number of varying "control" quantities. It is often easier to identify common features from a large number of small plots than it is to pick out individual details from a single comprehensive plot. These small plots can then often help the user to identify larger scale plots that can give a better overall understanding.

When looking at plots before designing the model, the following points should be considered:

- Is this data set sufficient? If there are areas of large change in the outputs that do not contain many data points, or large changes in output value between points that are adjacent in the input space, it may be worth obtaining more data.
- Inputs: do some of the inputs affect the sensor response significantly more than others? Does the model need to take all of the inputs into account?
- Outputs: do the measured outputs directly correspond to the outputs of the model or does the data need to be transformed?
- Limits: how does the sensor response behave at very low and very high levels of the variable quantities?
- Smoothness: does the output vary smoothly with all of the inputs, or is the response discontinuous?
- Linearity: is there a clear linear response region?
- Uncertainty: are the uncertainties larger for some values of the inputs than others?

If the plotting does not show any clear trends or patterns, the data collection process should be reassessed to double check that the data is appropriate for the application.

Prior to model fitting, it is sensible to split the data into a set that will be used to build the model and a set that will be used to validate the model. Both of these sets should cover the full ranges of the input quantities. As mentioned above, ideally the test plan that generated the data will have considered this requirement.

Another step that can improve the model building process is data scaling. Many algorithms used in model fitting perform best when the input data and the output data all lie in the range [0, 1]. If all quantities x are rescaled such that

$$X = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where xmin is the minimum value of x and xmax is the maximum value of x, the values will all lie within this range, making the fitting process more likely to succeed. The main drawbacks of this approach are that the parameters obtained from the model fitting become more difficult to interpret physically and that models of the same sensor based on different data sets may be difficult to compare directly because the scalings on input and output values may be different.

In general, it is best to start simple and add complexity, so for instance fit a single low order polynomial initially rather than a multi-stage high-order function. Once an initial fit has been made, it is often useful to calculate the differences between the data points and the model predictions and plot those, as sometimes these differences can indicate what sort of function should be fitted next.

The fitting process should also generate uncertainties associated with the model parameters. In some cases (for instance polynomials) it may be possible to determine the associated uncertainties analytically, but for other models the best approach is to estimate the uncertainties by generating a large number of perturbed data sets by sampling from the uncertainty distribution associated with the measured values and fitting the model to each of the perturbed sets. This approach generates a distribution for the model parameters that characterises their uncertainties. Note that it is very likely that the parameters will be correlated. The uncertainty generation process can also provide information about model suitability: if one of the model parameters has a very large associated uncertainty, it is likely that the parameter is not playing a significant role in the model and the model can therefore be simplified.

The following subsections provide some guidance on models that may be suitable for fitting to measured data. The subsections assume that all of the inputs are numerical and continuous, so for instance categorical inputs such as “quite rainy” and “very rainy” are not included, and nor are quantities that can only take integer values.

It is also assumed that the amount of data available is such that construction of a machine learning-based model is not feasible. This approach would add further complications for reliability and testing even if enough data were available and is best avoided for this application.

Finally, it should be noted that this section considers how to model one output at a time. So, for instance the approach would be used to develop a model of the response of a LIDAR sensor for a single direction, so that the model would be called repeatedly to build up a point cloud. Similarly, the response of an individual pixel in a camera would be modelled using this approach, with the overall image being built from multiple routine calls.

Nearest Neighbour

Nearest neighbour modelling carries out no interpolation of the measured results. For each set of input values used to run the model, the set of input values used in the test data set that are closest to the input values of interest is identified, and the model returns the output value from that member of the test data set. An example in two dimensions is shown in figure 4. The dots indicate the test data input values. There is exactly one dot within each shape defined by the red lines. A new set of input values would generate the output value associated with the dot that lies within the same red line shape. The red lines do not need to be generated during the solution (they are the Voronoi tesselation of the points); it is sufficient to calculate the distances from the input values of interest to each of the test data points and select the shortest distance.

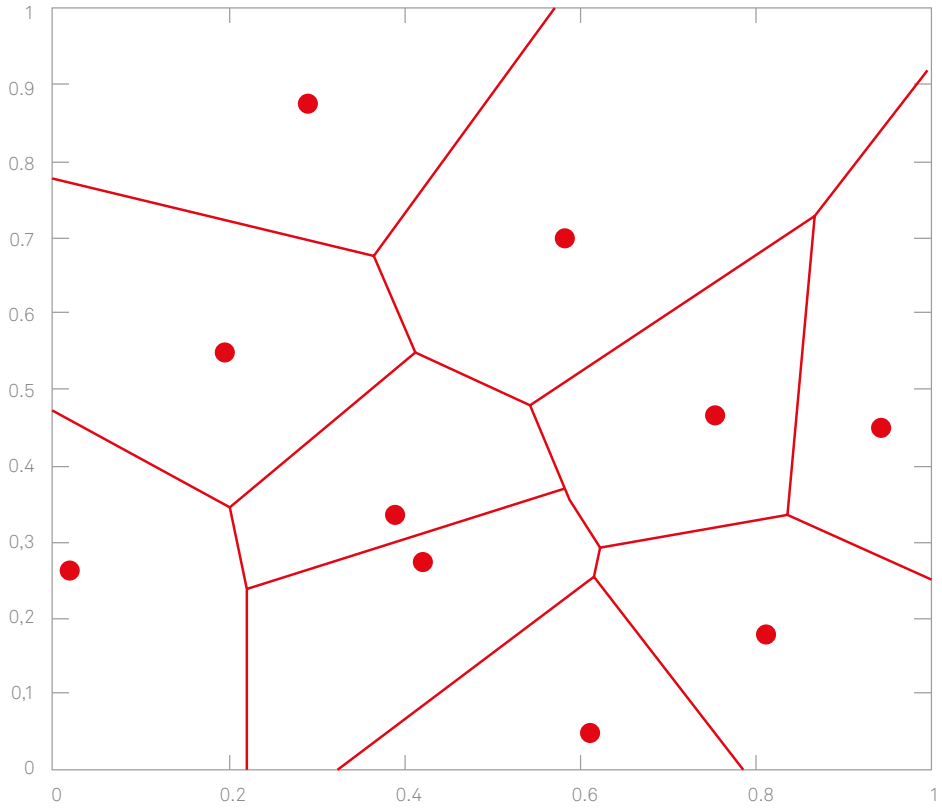


Figure 4: Illustration of regions produced by nearest neighbour interpolation. Each region returns the same function value, that of the data point within the region.

Data scaling is particularly important for nearest neighbour interpolation, particularly for models whose inputs are of very different orders of magnitude. Scaling the data means that all inputs are of equal importance when the closest point is identified, whereas if one input takes much larger values it may dominate the definition of "closest".

The main advantage of nearest neighbour interpolation is that it is easy to implement. It is likely to be less accurate than most other approaches, and it leads to a discontinuous response which is usually physically unreasonable.

Mesh-based interpolation

Mesh-based interpolation is similar to nearest neighbour in its geometric approach, but it leads to a piecewise continuous response. This approach involves defining a set of geometric objects that have corners at the measured points, and interpolating the measured quantity within those geometric objects. This approach is similar to that used for the approximation of variables within a mesh in finite element analysis.

The geometric objects are typically a set of k-simplices in a k-dimensional space. A k-simplex is a convex geometric object that has k+1 corners. In two dimensions it is a triangle and in three dimensions it is a tetrahedron. Delaunay triangulation is a reliable algorithm for constructing this set such that no shape in the set contains any of the other points, so that the interpolation of the data can be uniquely defined. Multidimensional equivalents of Delaunay triangulation exist. This approach leads to a piecewise continuous function that reproduces the experimental data exactly. It is fairly simple to implement and many mathematical software packages have Delaunay triangulation functions built in. For areas of rapid change it may not describe the output well, and for sets including many points it may not be computationally efficient since evaluation requires identification of the k-simplex that contains the point of interest.

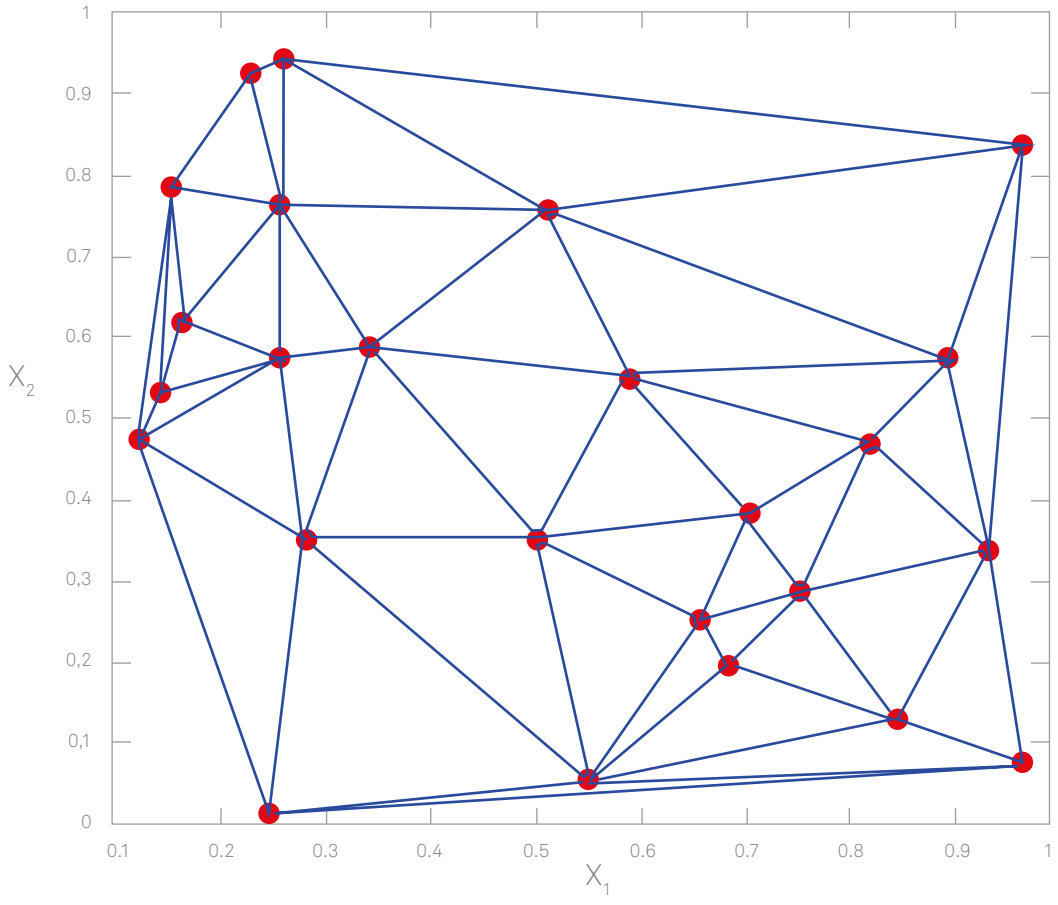


Figure 5: Delaunay triangulation of a set of points in two dimensions.

Figure 5: Delaunay triangulation of a set of points in two dimensions.

Polynomial fitting

For sensors whose outputs vary fairly smoothly and evenly across the full range of operating conditions, fitting a polynomial in multiple dimensions may be appropriate. The most general form of polynomial has the form

$$f(\mathbf{x}) = b_0 + \sum_{n=1}^{N} b_n x_n + \sum_{n,m=1}^{N} b_j x_n x_m + \sum_{n,m,p=1}^{N} b_j x_n x_m x_p + \cdots$$

where $\mathbf{x}$ = {xi, i = 1, 2, ..., N} is a vector of inputs, the quantities b_j are constants found from fitting, with *j* giving appropriate non-repeating indices. Fitting is a process of minimisation of the error

$$S^2 = \sum_{m=1}^{M} \left(y_m - f(\mathbf{x_m})\right)^2$$

where the y_m are the measured values being fitted and the x_m are the associated values of the inputs. Finding the minimum of the sum by differentiating with respect to each of the b_j in turn and setting the result equal to zero leads to a set of linear equations in the unknown quantities.

Whilst linear equations are nominally simple to solve, the use of high powers of the x_i in the polynomial can lead to very small or very large terms, which can make the system of equations ill-conditioned. For functions of one variable it is good practice to fit the function using Chebyshev polynomials as they have better stability, conditioning and convergence properties. The extension of this approach to more than one input is not well established. Normalisation of variables, as discussed above, can help with calculation stability.

It is also important to note that the number of measurement points (M above) must exceed the total number of unknown coefficients, otherwise the coefficients cannot be uniquely determined.

Polynomial approximation will generally lead to a continuous function that may not pass through all of the measured points. Extrapolation of polynomials beyond the range of fitted data should only be undertaken with care because polynomials are likely to increase without limit outside of this range.

Another possible approach is to divide the full input space into regions and fit a low order polynomial within each region, ideally fitting all of the polynomials simultaneously. This approach can be helpful if most of the sensor response is linear but it becomes non-linear at extreme values. Extra constraints can be applied to the fitting to ensure that the function value and its derivative are continuous across the boundary between regions (although these constraints may become difficult to implement for high-dimensional spaces).

Gaussian process modelling

Kriging was developed as a geostatistical estimator for single realisations of random fields that infers the value of the field at unobserved locations from samples. The theory was developed by Georges Matheron and Danie G. Krige. It is a specific application of a broader class of surrogate models known as Gaussian process models. Technical details will not be given here as the approach is mathematically involved, but an excellent guide is available in [8], a summary is given in [11], and many software implementations of the method are available.

A Gaussian process is a (possibly infinite) set of random variables, any finite number of which have a joint Gaussian distribution. A Gaussian process model maps between such a collection of random variables and the inputs of the measurements. The model is a random function that predicts the distribution of the function value for any new set of input quantity values, quantified as a mean and a standard deviation.

The standard deviation is extremely useful in identifying the regions where the model is least reliable. The reliability can be improved by adding more data in these regions, so Gaussian process modelling can help to define which measurements should be made next.

Gaussian process modelling in its simplest form is optimal interpolation based on regression. The predicted value of the function for a new set of input variable values is a weighted sum of the values of the measument points. The weights in the sum determining the mean and variance of the function at a new set of input variable values are determined by a covariance function that depends on the distances between the new set of variable values and the training input values.

The covariance function includes a set of hyperparameters used to fit the model to the data. Different covariance functions can be used depending on the features of the data, but a common approach is to define the correlation function between two sets of input values as

$$C(\mathbf{x}^1;\, \mathbf{x}^2) = \sigma^2 \prod_{i=1}^{N} \exp\left(-\left(\frac{x_i^1 - x_i^2}{\ell_i}\right)^2\right)$$

where the hyperparameters and li, i = 1, 2, ..., N, are found by fitting to the data. This approach gives an infinitely differentiable function that makes points less correlated the further apart they are.

An example of a data set fitted with a Gaussian process model is shown in figure 6. The red line is the original generating function, the red points are the data used to fit the model, the black line is the fitted model, and the blue dashed lines show a 95% confidence interval for the function value. These dashed lines suggest that the function is least accurately determined at the ends of the interval, where there are no data points, and that the estimate could be improved by having points there. Note that this interval is an estimate of the model fitting error rather than an estimate of the overall uncertainty: kriging models can be used to evaluate the uncertainty dues to propagation of the uncertainties associated with the inputs but this requires a separate calculation.

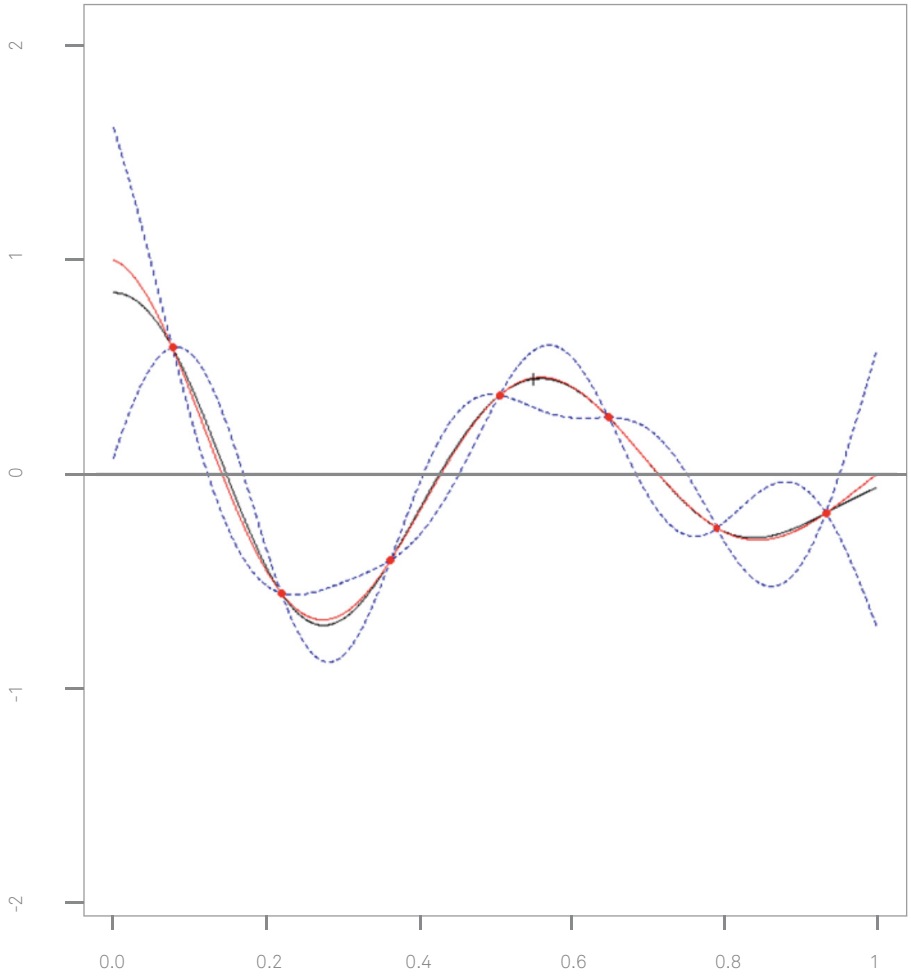


Figure 6: Data from a single input fitted using Gaussian process modelling

Gaussian process modelling is an extremely flexible method that imposes minimal assumptions on behaviour. Further flexibility is given by the wide range of covariance functions available. The functions can be tailored to fit particular features of a training data set, see section 5.4.3 of [8] for an example.

The technique can be applied to problems with many inputs (examples with up to 50 input quantities appear in the literature). The performance is good and comparable to second order polynomial regression models.

Drawbacks of the method include the difficulty of choosing an appropriate set of measurement data, particularly since closely-spaced points can lead to ill-conditioning of the fitting problem. The minimisation process used to determine the most likely hyperparameter values can be difficult to code and may have multiple local minima. The model construction is rather time consuming and the calculations are sophisticated, although software packages are available to carry out many of the common calculations. Addition of extra training points requires recalculation of all terms in the predictor and this can be time-consuming and potentially unstable.

Validation

Validation is the process of checking that the results of a model are in sufficiently good agreement with measurements of the reality the model is intended to simulate (the process of checking that the model is mathematically correct, verification, will not be discussed here). The definition of "sufficiently good" is an application-specific decision, but for testing of safety-critical systems such as autonomous vehicles, the level of agreement needs to be high. Furthermore, the level of agreement between model and reality needs to be high for all scenarios that the sensor might encounter in reality, and in particular for the scenarios where timely and correct decision-making is critical to occupant safety. A key example of such a situation is heavy rain where the atmospheric conditions not only affect sensor performance but also affect the car's response to steering and speed changes.

The measured data used to validate the model must be entirely independent of the data used to calibrate or fit the model during development. This division ensures that the validation is not a self-fulfilling prophecy: a well-constructed model will always agree well with the data used to obtain its parameters. This division also applies to physics-based models: if a data set has been used to obtain some parameter used in the model then it should not be used for validation.

The data set used for validation therefore needs to span the full set of operational circumstances without reusing data. These needs mean that when the data is first collected, it should be divided into two distinct sets, with each measurement being a member of one set or the other but not both: one for model construction and one for model validation.

At its strongest, validation should treat both the validation data and the model result as probabilistic quantities based on the associated uncertainties, and should calculate the overlap in probability distributions as shown in figure 7(a). This approach gives a quantified level of confidence in the validation. See below for a discussion of uncertainty propagation through models.

If uncertainties associated with the model results are not available, the available results can be compared to the distribution associated with the measurement result and a confidence level assigned based on distance from the mean, as shown in figure 7(b). Note that this does not give an absolute level of confidence but it does give a grounds for comparison.

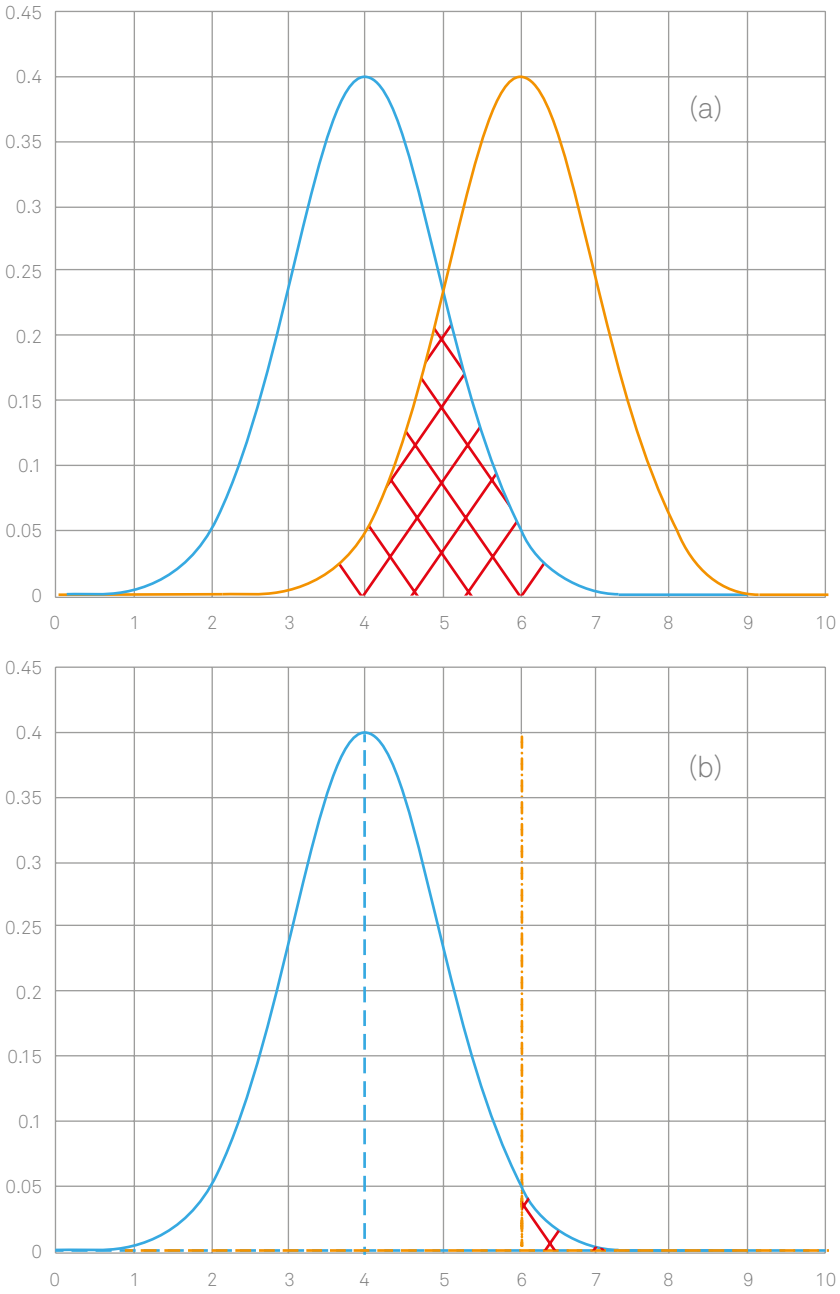


Figure 7: Comparisons for validation.

The least reliable form of validation is simple comparison between two numbers. This approach lacks any consideration of the likely variability of the values, and hence cannot generally have a level of confidence associated with it. As figure 8 shows, the same difference between two single numbers can give very different actual levels of agreement as the variability changes.

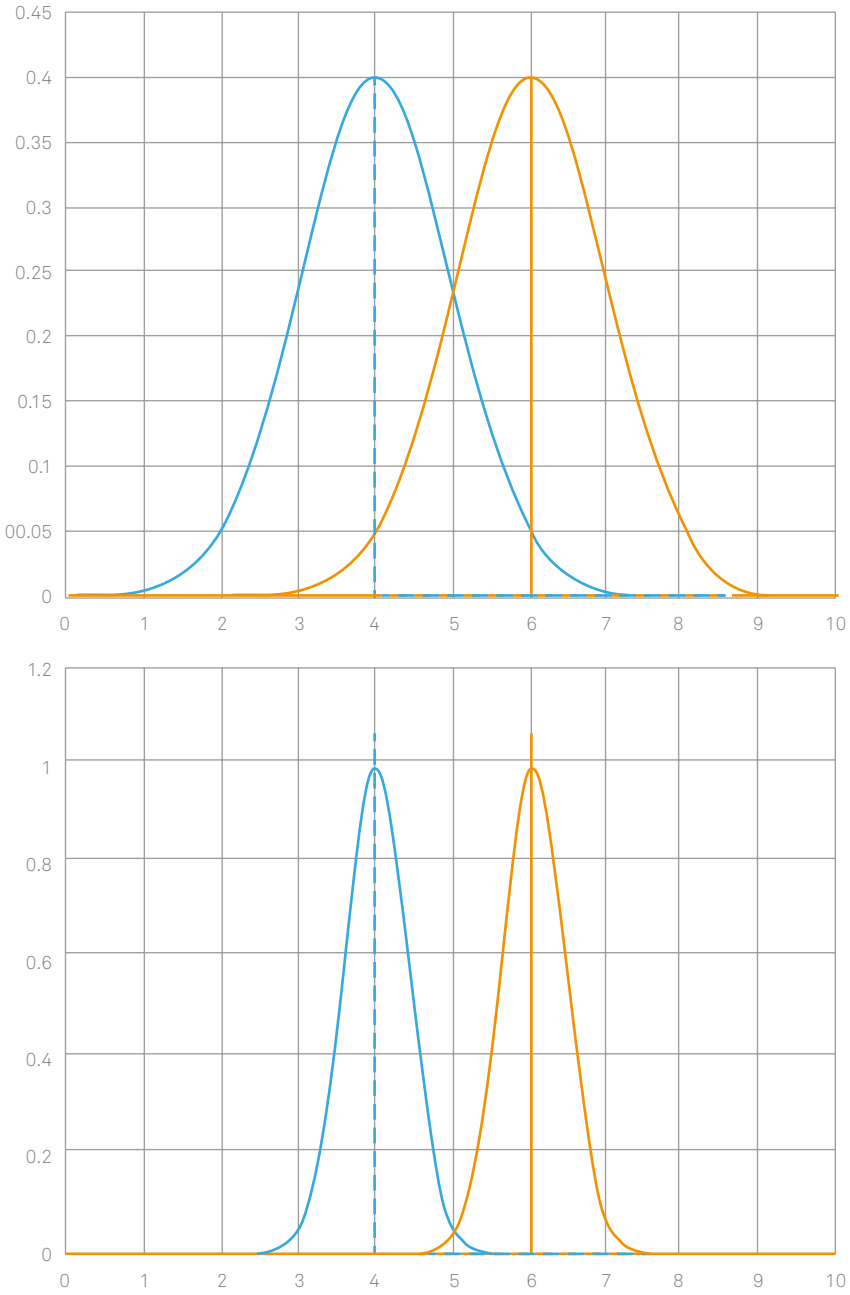


Figure 8: Variation in level of agreement between two numbers as the associated uncertainty varies.

The approach outlined above is simple to implement when the model results are a single quantity. For some measurement systems, such as cameras, the sensor response may be made up of many individual measurements, such as pixels, that are very likely to be highly correlated as they are generated by the same device. In such cases it is better to consider individual value comparisons in a probabilistic fashion rather than attempting to combine all pixels into a single comparison. So for instance it would be psosible to state “80% of pixels had a confidence level of at least 90% in their agreement” by comparing pixel-wise and treating model result and measurement as uncertain.

Uncertainty propagation

Measurement uncertainty is a statement of confidence in the value of a measured quantity. Uncertainty can never be eliminated altogether, but it can be reduced and managed. Uncertainty encapsulates natural variability, such as sample to sample variation of a material or electrical noise, and unknown biasses such as drift and temperature sensitivity.

Any model of a system making measurements must therefore have an associated uncertainty. Within a virtual testing environment, the output of the sensor model should be treated as a random variable so that the AI receives signals that are statistically equivalent to what it would receive in reality. A system that has been tested virtually using sensor models that do not include uncertainty has not experienced real-world conditions, and this may lead to bad decisions in safety-critical contexts. Virtual testing therefore needs to include uncertainty in its sensor models.

Propagation of uncertainty through sensor models generally requires a numerical approach rather than the analytical approach recommended by the Guide to the Expression of Uncertainty in Measurement [9], because the relevant analytical information is not generally available. The most general numerical approach, Monte Carlo sampling (MCS), is described in full in GUM Supplement 1 [10]. It assembles a cumulative density function for the output of interest by repeatedly evaluating the model with randomly chosen values of the inputs. For a reliable evaluation of uncertainty MCS requires thousands of model evaluations, which may be a problem if the model is computationally expensive. Alternative sampling methods can reduce this expense, as can the use of surrogate models. Both of these options are discussed in more detail elsewhere [11].

It should be noted that the uncertainty propagation needs to consider the uncertainties associated with the model parameters as well as the model inputs. As was noted in the previous section, the model fitting process should ideally generate the uncertainties associated with its parameters, so it should be possible to propagate all values through the model.

Uncertainty propagation within a virtual testing environment as a whole will not be discussed here, but the linked environment and AI can be regarded as a model that takes in inputs (i.e. the scenario definition including vehicle specification, known objects, and weather conditions) and produces an output (i.e. a series of manoeuvring decisions). The description of sampling approaches for uncertainty propagation for models can equally be applied to this highly complex situation.

References

[1] "A systematic review of perception system and simulators for autonomous vehicles research", F Rosique, P J Navarro, C Fernandez and A Padilla. Sensors, 19, p648-676, 2019, doi:10.3390/s19030648.

[2] "Analysis of automatic sensing model of an autonomous vehicle", A A K Singh, A Negi, and S Mudali, International Conference on Inventive Systems and Control 2017.

[3] "Fast generic sensoir models for testing highly automated vehicles in simulation", M Stolz, G Nestlinger, Electrotechnik & Informationstechnik, 135(4-5), 365-369, 2018, doi: 10.1007/s00502-018-0629-0.

[4] "The CAT vehicle testbed: a simulator with hardware in the loop for autonomous vehicle applications", R K Bhadani, J Sprinkle, M Bunting, 2nd International Workshop on Safe Control of Autonomous Vehicles (SCAV 2018), EPTCS 269, 2018, pp32-47, doi:10.4204/EPTCS.269.4.

[5] "High-fidelity autonomous surface vehicle simulator for thu maritime robotX challenge", P Smith and M Dunbabin, IEEE Journal of Oceanic Engineering 44(2), 2019, doi:10.1109/JOE.2018.2875571.

[6] "Integrating virtual and augmented reality based testing into the development of autonomous vehicles", T J Laverghetta, JF Leathrum Jr, N Gonda, MODSIM World 2019.

[7] "Virtual sensing for autonomous vehicle simulation in Chrono", A Elmquist and D Negrut, Technical report TR-2016-13.

[8] “Gaussian Process for Machine Learning”. C Rasmussen and C Williams, MIT Press, 2006.

[9] “Evaluation of measurement data - Guide to the expression of uncertainty in measurement”. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML. Joint Committee for Guides in Metrology, JCGM 100:2008, 2008.

[10] “Evaluation of measurement data - Supplement 1 to the Guide to the expression of uncertainty in measurement - Propagation of distributions using a Monte Carlo method”, BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML. Joint Committee for Guides in Metrology, JCGM 101:2008, 2008.

[11] “Novel Mathematical and Statistical Approaches to Uncertainty Evaluation: Best Practice Guide to Uncertainty Evaluation for Computationally Expensive Models”, K. Rasmussen et al, available from ***http://www.mathmet.org/publications/guides/index.php#expensive.***

Appendix E

Informal commentary on facility capabilities: Jules Verne and JARI

Author: Dr Dave Jones, Met Office

The definition of heavy rain. 80 mm/hr certainly counts as heavy rain in the sense of meteorological observing. The screen shot below is from our Innovate UK report.

Description of rainfall intensity	Rainfall rate associated with steady rainfall (mm/h)	Rainfall rate associated with showers (mm/h)
Slight	<0.5	<2.0
Moderate	0.5-4.0	2.0-10.0
Heavy	>4.0	10.1-50.0
Violet	-	>50.0

Table 1: Definitions of rainfall intensity (steady rain and showery/convective rain) taken from The Observer's Handbook (1982). Note that these definitions were established to assist the description of the rainfall by an observer during the typical six-minute observation 'window'.

However, these definitions relate to average intensities over a typical 6 minute period (the Innovate report further goes on to discuss intensities averaged over 1 and 2 mins) and not the instantaneous short duration, highly localised values that are likely to be more relevant to CAV case. At the CAV scale - instantaneous over a path length of 10s of metres - a heavy convective shower may yield peak intensities of the order of 300 mm/hr. It may the case that the JARI-generated rain rates may fall into the edge cases for the CAV sensors (TBD from the programme outlined above), but they certainly don't represent reasonable worst case even for the UK.

Size distribution. This comment may also apply to the fog generation too (we would look at this as a follow-on to the rain work). While we still have to look into how droplet/drop interaction actually manifests itself at the CAV sensor (explanatory comment* follows), we do know that there can be considerable sensitivity with regard to how you distribute your mass of falling rain (or fog) across the possible ranges of sizes of drops. For mm-wave attenuation, two perfectly plausible drop size distributions for the same rain rate can give you factor of two difference in attenuation (this comes from our 5G work). While this might be easy to handle for a single sensor, when combining multiple sensors (camera, lidar, mm-wave, far IR, IR...) this might create significant problems if one wished to demonstrate multi-sensor fusion in a test facility.

What it means. It may not be possible to control the drop size distribution sufficiently well to fully simulate what rain/fog looks like at all the sensor bands simultaneously (or even singly). This doesn't mean that it's worthless though - it may simply be necessary to measure the actual size distribution in the chamber (with the sorts of sensors that will propose in the testbed) and then create a link between real-world/emulated/simulated rain another way i.e. you don't invest lots of R&D in generating natural rainfall, you just measure what you've made. This might then mean "spoofing" some of the multi-sensor outputs or adding additional attenuation in the loop for some channels in order to make up for the deficiencies of the chamber. In my personal view, while these are all open questions, there is reason to be optimistic that there is an approach that uses the strengths of the different test environments and, by working smart, millions of pounds could be saved in the creation of the facilities.

Explanatory comment*: *If we were to assume that all the drops/droplets are perfect spheres, we can calculate very easily how much attenuation a single drop will cause. In traditional meteorological applications we then make an assumption of the functional form of the drop size distribution (DSD) and assume that radar/light waves pass through a long enough path length that the DSD is stable with time, so that we can simply add up attenuation across the DSD. We can then believe our attenuation value and its uncertainty. In the CAV case, the path lengths are short and the rain "looks" more noisy (over time/distance it averages out to the DSD but at any one moment it's just a random sample). What this means is that the received signal by the sensor is fluctuating rather than a smooth function. We have to make the link between the CAV scale and the meteorological observation scale.*

Appendix F

Glossary of Terms

Author: Dr Dave Jones, Met Office

API: Application programming interface.

AI: Artificial intelligence.

CAV: Connected and autonomous vehicles.

CPI: Coherent pulse interval - Coherent time used to improve the signal to noise ratio and perform Doppler processing.

dBs: Decibels - Is a logarithmic scale.

Doppler (Frequency): Indicates the change in frequency on the received signal compared to the transmitted due to the sensor's and target's relative motion.

Doppler processing: Estimation of Doppler frequency components in the received signal;

ETSI: European telecommunications standards institute.

FMCW: Frequency modulated continues waveform - A category of radar waveform designs used generally in short range-range measurements. The most common type of FMCW is automotive applications is the Linear FMCW.

KPI: Key performance indicator.

LO: Local oscillator.

Model: an algorithm that takes in a set of physically meaningful inputs and calculates a set of outputs.

MOT: Ministry of transport.

Multipath: Is the propagation phenomenon resulting in signals reaching the receiver antenna by more than one paths.

NPD: Nominal probability of detection.

NPFA: Nominal probability of false alarm.

PRF: Pulse repetition frequency - Is the frequency in which same pulses are transmitted.

Range Doppler map: 2D intensity map used map used to depict the energy distribution in different ranges and Doppler shifts.

Raw data: Unprocessed or low-level data.

RCS: Radar cross section - Is the measure of a target's ability to reflect radar signals in the direction of the radar receiver.

Resolution: Is the smallest distinguishable difference in values.

RF: Radio frequency.

Sidelobes: Are undesired secondary lobes other than the main lobe.

Tracking: Ability of a radar to predict and associate returns of one or multiple targets with previous observations;

VCO: Voltage-controlled oscillator

Notifications

This report was produced by National Physical Laboratory and the Connected Places Catapult 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:	• Richard Holland, Principal Engineer, Connected Places Catapult • Andre Burgess, MBA, Strategy Directorate, National Physical Laboratory • Dr. John Molloy, Senior Research Scientist, National Physical Laboratory
Contibuting Specialists:	• Louise Wright, Head of Science for Data Science, National Physical Laboratory • Dr. Tony Mansfield, Principal Research Scientist, National Physical Laboratory • Philip Moorcroft, Business Analyst, Connected Places Catapult • Dr. Carmine Clemente, Senior Lecturer, Chancellor’s Fellow Centre for Signal and Image Processing, Department of Electronic and Electrical Engineering, University of Strathclyde • Dr Richard Barham, Director, Acoustic Sensor Networks • Dr Alan Peters, Solution Architect, Connected Places Catapult • Lovedeep Brar, Senior Systems Engineer, Connected Places Catapult • Dr Dave Jones, Principal Consultant (Observational Meteorology), The Met Office
Reviewed by:	• David Webb, Head of Innovation, CCAV