Working draft text for the ADS GTR and UN Regulation – Testing provisions

Structure of the proposed regulatory text (updated following latest discussions) – New elements in red

Testing provinsions in the ADS Regulation (in bold the sections for which regulatory text proposals have been provided and are under discussion)					
Regulation Section	**Level 1**	**Level 2 (non exhaustive list)**	**Level 3 (non-exhaustive list)**	**Level 4 (examples - not exhaustive list)**	**Level 5 (examples - not exhaustive list)**
Assessment by [authority/assessor]	**Assessment of the safety case testing activities**	**Assessment of the test environment**	**Appraisal of the physical test environment**	**The assessor shall verify the suitability of the test track and the real-world environment selected for the tests…**	
			Appraisal of the credibility framework developed by the manufacturer for virtual testing	**The assessor shall verify that the credibility assessment has been properly carried out… Chapter 5**	
		Assessment of the scenarios and their management	**Assessment of the scenario generation process**	**…**	
			Assessment of scenario coverage and completeness	**…**	
			Assessment of scenario representativeness (selection and characterisation including behavioural competences)	**…**	
		Assessment of the processes in place for testing	**Assessment of processes for DDT testing**		
			Assessment of processes for ADS user testing	**…**	
		Assessment of DDT testing evidence	**General testing assessment**	**Assessment of the coverage of the test results (SAFE)**	
				…	
			Assessment of virtual tests	**The assessor shall verify that virtual testing have been carried out considering the uncertainty of the results**	
				The assessor may request the re-execution of some scenario to verify the stochasticy of the simulation toolchain(s)	
				…	
			Assessment track testing	**The assessor shall verify that track testing has been carried out…**	
			Assessment of real-worl testing	**The assessor shall verify that real-world testing have been carried out…**	
		Assessment of ADS user testing evidence	Assessment of the research performed in the user centred design process.	The assessor shall verify the user testing results	The assessor shall assess whether the applied testing procedures are in line with common scientific practice.
					The assessor shall assess whether the users are representative for the general driving population.
					The assessor shall assess the applied statistics including the power analyses.
					.....
	Confirmatory testing	**Testing by [authority/assessor]**	**Track tests**	**~~The assessor shall execute the mandatory tests foreseen by the regulation and compatible with the ODD of the ADS~~**	
				The assessor shall repeat a number of tests carried out by the OEM	**The assessor shall repeat a number of track-tests carried out by the OEM on track**
					The assessor shall repeat a number of virtual tests carried out by the OEM
				The assessor shall carry out additional tests with respect to those carried out by the OEM	**In case the assessor idetifies reasonably foreseable scenarios not tested by the OEM, it shall tests the system on a number of parameter variations for those scenarios check the relevance of this category (relevance to be checked)**
					In addition to the tests concerning new scenarios, the assessor should check different parameter variations of the scenarios tested by the OEM
			Real-world tests	**The assessor shall test the vehicle inside its ODD for a sufficient duration to cover different traffic conditions, different weather conditions, different lighting conditions…**	

The following table provides:

- In the first column, the reference used to draft the proposed text;
- In the second column, the proposed regulatory text in which the track changes (**new text** in bold, ~~deleted text~~ strikethrough, text in red for elements not discussed within the group, text highlighted in green agreed within the group, and highlighted text for the references to other parts of the regulation) outline the modification compared to the GRVA-20-36e (Guidelines and recommendations for ADS safety requirements, assessments and test methods to inform regulatory development);
- In the third column, proposals for revision and comments

Reference (GRVA-20-36e)	Proposed text (ADS-XX)	Proposals and comments
	Definitions	
	General Requirements	
	Requirements/Specifications	
	Compliance assessment	
	Assessment of the safety case testing activities	

NEW	**6.X.**	**The assessor shall verify that the approach to testing adopted by the manufacturer is suitable for the demonstration of the safety case and the compliance with performance/functional requirements**	
	Assessment of the test environment		
NEW	**6.X.1.**	**The assessor shall verify that the physical testing (proving ground and/or public road) facilities and environment used by the manufacturer in the assessment of the safety case are suitable for the tests that are being conducted.**	**OPI**: To check during phase 3 if any specific requirement can be developed to demonstrate the credibility of the physical testing (ref. Figure 1 in GRVA-20-36e)
NEW	**6.X.2.**	**The assessor shall verify that the simulation toolchain(s) used by the manufacturer in the assessment of the safety case is suitable for conducting virtual tests and in compliance with requirements listed in 5.8.1.**	**OPI**: to check cross reference after structure is established
Annex 5 – Appendix 1 – 4. The proposed credibility assessment framework provides a general description of the main aspects needed for assessing the credibility of an M&S solution together with guidelines of the role played by the relevant assessor in the validation process with respect to credibility. The assessor should investigate the documentation and evidence supporting credibility during the audit phase. It is understood that the actual validation tests will take place once there is sufficient evidence that a simulation tool or toolchain produces credible results.	**6.X.2.1**	**The assessor shall review the manufacturer's credibility framework to determine whether the simulation the toolchain(s) is suitable to undertake virtual testing.**	**OPI**: Developed in the Credibility Assessment sub-group
NEW	**6.X.2.2.**	**The assessor shall review the documentation and evidence supporting the manufacturer's claims.**	**OPI**: Developed in the Credibility Assessment sub-group

	a) A successful outcome of the assessment will be a confirmation that the claims of the manufacturer about the capability of the simulation toolchain(s), including its scope, are correct and that it can be used to perform the virtual testing as part of the ADS assessment. **b) The simulation toolchain(s) can only be used to undertake virtual testing once the credibility of the same has been established.**	
Annex 5 – Appendix 1 – 43 (g). The assessor should audit the documentation provided by the manufacturer and may carry out tests of the complete integrated tool. If the output of the virtual tests does not sufficiently replicate the output of physical tests, the assessor may request that the virtual and/or physical tests to be repeated. The outcome of the tests will be reviewed and any deviation in the results should be reviewed with the manufacturer. Sufficient explanation is required to justify why the test configuration caused deviation in results.	**6. X.2.3.** **The assessor shall audit the information provided by the manufacturer and may request or carry out additional tests of the simulation toolchain(s) or physical tests. The outcome of the tests shall be reviewed and any concerns or discrepancies shall be raised and reviewed with the manufacturer.** **The manufacturer shall provide an explanation of the discrepancies in the results. If the results from the simulation toolchain(s) do not sufficiently replicate the output of physical test or does not have sufficient scope the assessor shall inform the manufacturer.** **The manufacturer shall conduct extra validation activity and resubmit their information for further assessment.**	
	Assessment of the scenarios and their management	
NEW – Based on Annex 3 text	**6.X+1.2** **The assessor shall verify that the manufacturer has used suitable and documented processes to derive behavioural competencies and scenarios that are ODD-relevant and are relevant to the ADS safety case**	**OPI**: To check with SMS.

NEW – Based on Annex 3 text	**6.X+1.2.1**	**The assessor may refer to the methodology outlined in the Annex [X] as a suitable approach against which to review the approach adopted by the manufacturer.**	**OPI**: To check relevance of information from the Annex 3 of the Integration Document
NEW – Based on Annex 3 text	**6.X+1.3**	**The assessor shall verify that the manufacturer's approach and processes to identify and generate scenarios:**	
	6.X+1.3.1	**covers the necessary nominal, critical and failure scenarios.**	
	6.X+1.3.2	**takes into account data driven, knowledge driven and stochastic approaches to systematically identify hazardous events and other occurrences used to develop scenarios.**	
	6.X+1.3.3	**properly maps and characterises the behaviour of all the elements included in the scenarios.**	
NEW – Based on Annex 4 and linked to comments during Testing Session #4	**6.X+1.4**	**The assessor shall verify that the set of scenarios resulting from the manufacturer scenario generation and identification process is suitable to demonstrate the ADS safety case and are able to cover the space of reasonably foreseeable situations and conditions that the ADS will encounter during its real-world operations. In particular the assessor shall verify that the set of scenarios selected as evidence to support the ADS safety case includes:**	
NEW – Based on Annex 4 and linked to comments during Testing Session #4	**6.X+1.4.1**	**a sufficient number of situations in which the ADS approaches and crosses its ODD boundaries.**	
	6.X+1.4.2	**reasonably foreseeable scenarios that are not deemed to be preventable by the ADS (e.g. related to unsafe behaviour by other road users or to inappropriate infrastructural elements)**	

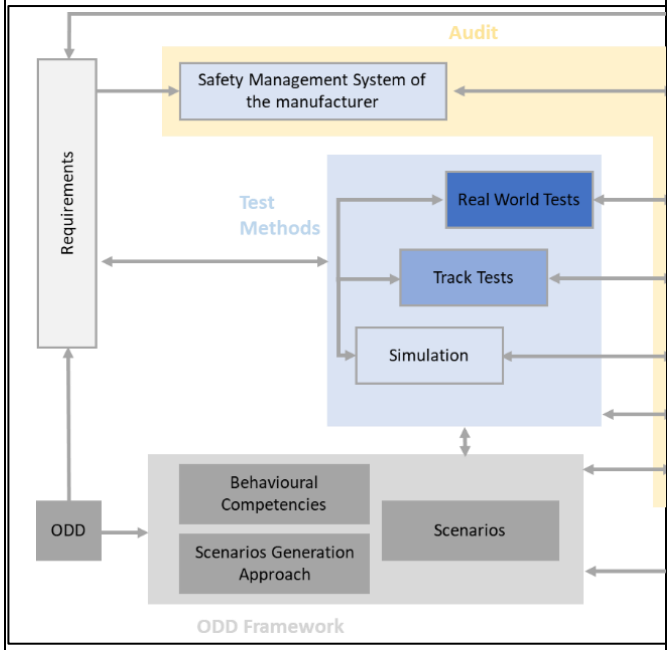
	6.X+1.5	**The assessor may refer to the methodology outlined in the Annex [XX], including the provided scenario template, as a suitable approach against which to review the approach adopted by the manufacturer.**	
Annex 4.17 It is recommended that sampling techniques be used when selecting parameters to be used in creating logical and concrete scenarios for ADS validation for a particular ADS and its ODD to avoid the ADS being optimized for a set of known test cases. Using a maxim number of random samples is clearly preferable from a credibility perspective, it is recognized that this can place a greater burden on manufacturers and the relevant authority (e.g. technical service). This should be considered when determining the volume of tests to be conducted when using the random sampling. It is assumed that for simulation/virtual testing the burden of random sampling is less and therefore maximizing the number of random samples for this facet of the testing is more feasible.	**6.X+1.6**	**The assessor shall verify** that **the manufacturer has used** sampling techniques ~~be used~~ when selecting parameters to be used in creating logical and concrete scenarios **used as evidence to supporting the** ADS **safety case** ~~validation for a particular ADS and its ODD~~ to avoid the ADS being optimized for a set of known test cases. ~~Using a maxim number of random samples is clearly preferable from a credibility perspective, it is recognized that this can place a greater burden on manufacturers and the relevant authority (e.g. technical service). This should be considered when determining the volume of tests to be conducted when using the random sampling. It is assumed that for simulation/virtual testing the burden of random sampling is less and therefore maximizing the number of random samples for this facet of the testing is more feasible.~~	
	Assessment of the processes in place for testing		
NEW	**6.X+2**	**The assessor shall verify that the manufacturer has suitable processes, resources and competent personnel in place for the testing that has been undertaken to demonstrate the ADS safety case**	**OPI**: to check with SMS.
4.6. This framework focuses on subjecting the ADS to these scenarios and assessing the behavioural competencies demonstrated by the ADS under each scenario against requirements for performance of the Dynamic Driving Task (DDT). These requirements	**6.X+2.1**	**The assessor shall verify that the manufacturer has suitable processes and competent personnel** ~~This framework focuses on subjecting the ADS to these scenarios and assessing~~ **to assess** the behavioural competencies demonstrated by the ADS under each scenario against requirements for	**OPI**: To check with SMS.

	focus on desired driving capabilities and outcomes. The requirements intentionally avoid technical specifications and performance limits because each traffic situation requires a response appropriate to its combination of elements, risks, and available options.	performance of the Dynamic Driving Task (DDT). ~~These requirements focus on desired driving capabilities and outcomes. The requirements intentionally avoid technical specifications and performance limits because each traffic situation requires a response appropriate to its combination of elements, risks, and available options.~~	
4.8.	However, defining performance criteria in critical scenarios might prove difficult, especially under conditions where requirements must be prioritised. In these cases, the framework proposes the use of appropriate safety models to enable assessment of ADS performance within the limits of the safety model(s). For example, it is recognised that an ADS might not be able to avoid a collision, so the ADS performance needs to be compared with safety-model performance to set the threshold where avoidance is required and that where avoidance is not feasible, and if mitigation may be possible.		**OPI**: not suitable for regulatory purposes
4.9.	In cases where the behavioural competency demonstrated by the ADS involves such exceptions, the framework relies on safety models to determine whether the exceptions are justified. For example, an ADS might violate a lane restriction in order to avoid a collision. The safety model enables determinations on the collision risk, the ADS response, and the necessity of the traffic-rule violation.		**OPI**: not suitable for regulatory purposes
4.10.	Failure scenarios address situations where the ADS performance of the DDT has been		**OPI**: not suitable for regulatory purposes

	compromised by a system fault. Unless a fallback user manages the response to the fault, the ADS is expected to bring the vehicle to a safe, stopped condition (i.e., a minimal risk condition). However, depending on the severity of the fault, the safety requirements allow the ADS to adapt its performance of the DDT to the nature of the fault. This tolerance permits an ADS where possible to mitigate risks while reaching a safe location to stop the vehicle.		
4.11.	The guidelines recommend consolidation of these scenarios into a scenario catalogue that may be used under the NATM to systematically validate the safety of an ADS.		**OPI**: not suitable for regulatory purposes
4.12.	These guidelines address the safety of ADS vehicle users via sets of requirements aligned with the relationships that users might have with a given ADS during use of the ADS vehicle. These relationships can vary depending on whether a user is located inside or outside the ADS vehicle, the degree(s) of control that a user may exercise over the vehicle during a trip, and whether a user has a one-to-one relationship with a single vehicle or may be performing functions relative to multiple vehicles.		**OPI**: initial text revision to fit regulatory purposes
4.13.	Regardless of any assistance systems, drivers perform the DDT until they activate an ADS feature. An ADS feature is specific to an ODD. Activation of an ADS feature initiates ADS performance of the tactical and operational functions required to perform the entire DDT within the ODD of the feature. In		**OPI**: not suitable for regulatory purposes

	the context of the driver relationship, the vehicle is moving (i.e., the user is driving the vehicle) and the activation involves a transition of control over vehicle operation from the driver to the ADS.			
4.14.	Upon activation of a feature, the ADS performs the entire DDT necessary to operate the vehicle within the ODD of the feature. The driver, therefore, shifts to the role of fallback user or passenger. Some ADS designs may initiate a system-initiated deactivation of the ADS (i.e., fall back to the user) in the event that the ADS can no longer perform the DDT (e.g., prior to reaching the boundary of the ODD of the feature in use).			**OPI**: not suitable for regulatory purposes
4.15.	A passenger has no capabilities to perform the DDT. Nonetheless, passengers require means to select destinations, routes, and stops and therefore have necessary interactions with the ADS.			**OPI**: not suitable for regulatory purposes
4.16.	These guidelines propose principles and specifications to ensure the safety of users and their use of ADS vehicles across these relationships. The guidelines recognise that additional relationships might need consideration in the further development of such safety requirements.	**6.X+2.2**	**The assessor shall verify that the manufacturer has suitable processes and competent personnel** ~~These guidelines propose principles and specifications~~ **to assess the capability of the ADS** and to ensure the safety of users and their use of ADS vehicles~~ across these relationships~~. ~~The guidelines recognise that additional relationships might need consideration in the further development of such safety requirements.~~	
		Assessment of DDT testing evidence		
4.17.	The assessment of an ADS for compliance with these safety recommendations rests on five validation pillars:	**6.X+3.**	**The assessor shall review ~~assess~~ the evidence produced by the manufacturer ~~in assessing~~ in demonstrating the ADS safety case and in meeting**	

1. Documentation and audit 2. Virtual testing 3. Track testing 4. Real-world testing 5. In-service monitoring and reporting.		**~~for what concerns~~ the requirements ~~for~~ to perform~~ance of~~ the Dynamic Driving Task (DDT). The assessor shall review the manufacturer's use of the different** ~~The assessment of an ADS for compliance with these safety recommendations rests on five validation pillars~~ **testing methods**: ~~1. Documentation and audit~~ 2. Virtual testing 3. Track testing 4. Real-world testing ~~5. In-service monitoring and reporting.~~	
NEW	**6.X+3.1**	**The assessor shall verify the suitability of the set of tests carried out as evidence to support the safety case, in particular in the terms of coverage and relevance**	
NEW	**6.X+3.1.1**	**The assessor shall assess the results of the tests for meaningfulness and consistency.**	**OPI**: consider to add guidance in the interpretation document
NEW	**6.X+3.1.2**	**The assessor shall verify that the results of the tests are able to demonstrate the behavioural competencies of the ADS when performing the DDT. In particular the assessor shall verify that the test results are able to demonstrate the performance of the ADS:**	
NEW	**6.X+3.1.2.1**	**in nominal, critical and failure scenarios;~~.~~**	
NEW	**6.X+3.1.2.2**	**while approaching and crossing the ODD boundaries; and**	
NEW	**6.X+3.1.2.3**	**in the case that collisions with other road users are not deemed to be preventable**	
Annex 2 - Matrix of DDT and User requirements with applicable test pillars			**OPI:** not suitable for regulatory purposes. Move to interpretation document

		**OPI**: not suitable for regulatory purposes
4.18.	These pillars are intended for use in combination(s) to produce an efficient, comprehensive, and coherent assessment of ADS compliance with the guidelines on safety performance. Each of the testing methodologies possesses its own strengths and limitations, such as differing levels of environmental control, environmental fidelity, scalability, and cost, which should be considered. In some cases, the application of more than one method could be necessary to assess the capability of an ADS to cope with range of situations that can arise in real-world	**OPI**: not suitable for regulatory purposes

	traffic. The use of multiple methods allows for flexibility in the composition, sequencing, and application of testing across the diversity of ADS while avoiding unnecessary redundancies and overlaps. Figure 1 above illustrates relationships across the ADS safety requirements, ODD analysis and scenario generation, and the validation pillars.			
4.19.	The pillars concern Audit, Test Methods, and In-Service Monitoring and Reporting.			**OPI**: not suitable for regulatory purposes
		Assessment of virtual tests		
4.21.1.	Virtual testing provides means to assess ADS performance across a wide range of traffic scenarios efficiently. These guidelines recommend procedures for evaluating the reliability of the manufacturer's virtual testing tool chains and methodologies. This credibility assessment enables confidence in applying these tools and methods, and the evidence they generate, to the assessment of ADS safety (see Annex 5).	**6.X+3.2**	**The assessor shall review ~~assess~~ the evidence from virual testing that is provided by the manufacturer to support ~~build~~ the ADS safety case. In particular the assessor shall verify that the evidence from virtual testing shows that the ADS complies with the requirements laid down in the DDT section. ~~by using~~** virtual testing. ~~provides means to assess ADS performance across a wide range of traffic scenarios efficiently. These guidelines recommend procedures for evaluating. This credibility assessment enables confidence in applying these tools and methods, and the evidence they generate, to the assessment of ADS safety (see Annex 5)~~.	**OPI**: Verify the reference
NEW		**6.X+3.2.1**	**The assessor shall verify that the manufacturer's virtual testing has been carried out incorporating proper consideration of the assumptions, accuracy and uncertainty in the simulation toolchain(s) and hence in the the results in line with the requirements laid down in 5.8.1.**	
NEW		**6.X+3.2.2**	**The assessor shall verify that any virtual test using simulation toolchain(s) containing stochastic**	

			elements has taken account of possible uncertainty in the results.	
4.21.1.1.	Virtual testing uses different types of simulation toolchains to assess compliance of an ADS with safety requirements across a wide range of traffic scenarios, including some of which would be difficult (if not impossible) to reproduce in physical settings.			**OPI**: not suitable for regulatory purposes
4.21.1.2. (a) (b) (c) (d) (e)	The toolchain methodologies include (but are not necessarily limited to): Model in the Loop (MIL); Software in the Loop (SIL); Hardware in the Loop (HIL); Vehicle in the Loop (VIL); Driver in the Loop (DIL).			**OPI**: not suitable for regulatory purposes
4.21.1.3.	Virtual testing enables efficient assessment across nominal, critical, and failure scenarios and ranges of parameters within scenarios relevant to the ADS configuration, intended uses, and limitations on use, including determination of the boundaries between collision avoidance and crash mitigation. Virtual testing also enables assessment of compliance with safety requirements relevant to user interactions, especially through DIL and similar "user in the loop" methodologies.			**OPI**: not suitable for regulatory purposes
4.21.1.4.	Virtual testing may be more suitable when there is a need to vary test parameters and a large number of tests need to be carried out to support efficient scenario coverage (e.g., for path planning and control, or assessing perception quality with pre-recorded sensor data).	**6.X+3.2.3**	**The assessor shall verify that virtual testing has been used** ~~Virtual testing may be more suitable when there is a need~~ to vary test parameters and **perform** a large number of tests ~~need to be carried out~~ to support efficient scenario coverage **including critical and unpreventable scenarios as well as low probability events** ~~(e.g., for path planning and control, or~~	

			~~assessing perception quality with pre-recorded sensor data)~~.	
4.21.1.5.	Virtual testing enables identification of scenarios that result in exceptions to nominal DDT performance requirements (e.g., deviation from traffic rules, evasive manoeuvres, collision outcomes) for assessment based on safety models.			**OPI**: not suitable for regulatory purposes
4.21.1.6.	Methods of randomization of parameters and scenario composition enable ADS performance assessments under critical scenarios, including low probability events.			**OPI**: not suitable for regulatory purposes
4.21.1.7.	Virtual testing enables the identification of high-value scenarios that can be applied to track testing. After ADS deployment, virtual testing can contribute to the analysis of ADS behaviours inconsistent with behavioural competencies demonstrated during the original assessment.			**OPI**: not suitable for regulatory purposes
		Assessment of track testing		
4.21.2.	Track testing concerns the physical assessment of ADS performance under controlled conditions on closed-access grounds. For these reasons, track testing may be best suited to assessment of ADS performance under scenarios that entail significant safety risks in case of failure to meet the requirements, where performance can be assessed through a discrete number of physical tests, and where testing benefits from the capacity to control conditions (e.g., for HMI and fallback responses, under critical scenarios).	6.X+3.3	**The assessor shall review the evidence from track-testing that is provided by the manufacturer to support the ADS' safety case. In particular the assessor shall verify that the evidence from track testing shows that the ADS complies with the requirements of the DDT section.** ~~For these reasons, track testing may be best suited to assessment of ADS performance under scenarios that entail significant safety risks in case of failure to meet the requirements, where performance can be assessed through a discrete number of physical tests, and where testing benefits from the capacity to control~~	

		~~conditions (e.g., for HMI and fallback responses, under critical scenarios).~~	
NEW	6.X+3.3.1	**The assessor shall verify that the manufacturer has suitable processes in place to identify the set of scenarios to be tested via track-testing.**	
NEW from 4.2.1.2	6.X+3.3.2	**The assessor shall verify that at least part of the scenario tested via track-testing includes critical scenarios possibly resulting into a collision**.	
NEW	6.X+3.3.3	**The assessor shall verify that the results of track testing are consistent with the results of virtual testing executed considering the same scenarios in similar conditions.**	**OPI: text proposed following the Testing session #6**
	Assessment of real-world testing		
4.21.3. Real-world testing assesses the capability of the ADS to perform the DDT and its interactions with its user(s) while in operation on public roads under real-world traffic conditions. Real-world testing may be more suitable to ensure a level of fidelity that might not be represented virtually or on a test track (e.g., interactions with other road-users and perception capabilities).	6.X+3.4	**The assessor shall review the evidence from real-world testing that is provided by the manufacturer to support the ADS safety case. In particular the assessor shall verify that the evidence from real-world testing shows that the ADS complies with the requirements laid down in the DDT section.** ~~Real-world testing assesses the capability of the ADS to perform the DDT and its interactions with its user(s) while in operation on public roads under real-world traffic conditions. Real-world testing may be more suitable to ensure a level of fidelity that might not be represented virtually or on a test track (e.g., interactions with other road-users and perception capabilities).~~	
4.21.3.1. The primary aim is to verify compliance with safety requirements for DDT performance under normal operational and road conditions and for nominal ADS interactions with its user(s).			**OPI**: not suitable for regulatory purposes

4.21.3.2.	While this method provides a high degree of environmental fidelity for testing an ADS, constraints on time, cost, controllability, reproducibility, and safety assurance limit the feasibility of covering traffic scenarios in the strict sense.			**OPI**: not suitable for regulatory purposes
4.21.3.3.	Therefore, this method requires attention to designing test routes that capture predictable aspects of the ODD (e.g., road types and geometries), elements found in the related nominal scenarios (e.g., other road users, signs, and signals), and typical dynamic conditions (e.g., high/low traffic densities). The test routes should also enable verification of nominal requirements for the safety of user interactions, including prior to, at the time of, and after entering and exiting the ODD of an ADS feature.	6.X+3.4.1	**The assessor shall verify that the manufacturer has suitable processes in place to identify** ~~Therefore, this method requires attention to designing~~ test routes that capture predictable aspects of the ODD (e.g., road types and geometries), elements found in the related nominal scenarios (e.g., other road users, signs, and signals), and typical dynamic conditions (e.g., high/low traffic densities). The test routes shall also enable verification of nominal requirements for the safety of user interactions, including prior to, at the time of, and after entering and exiting the ODD of an ADS feature.	
4.21.3.4.	To the extent that an ADS encounters critical or failure situations during a real-world test drive, the response of the ADS, including exceptions to the nominal performance requirements, should be considered in conjunction with the outcomes of track and virtual testing.	6.X+3.4.2	To the extent that an ADS encounters critical or failure situations during a real-world test drive, the response of the ADS, including exceptions to the nominal performance requirements, **shall** ~~should~~ be considered **by the assessor** in conjunction with the outcomes of track and virtual testing.	**OPI**: initial text revision to fit regulatory purposes
NEW		6.X+3.4.3	**The assessor shall verify that the results of real-world testing are consistent with the results of virtual testing and track-testing executed considering the same scenarios in similar conditions.**	
		Assessment of ADS user testing evidence		
		Verification testing		

	Testing by assessor	
NEW	**6.N.1. The assessor shall review the documents and the evidence provided by the manufacturer to support their safety case claims. Once the assessor is satisfied with the safety case and the supporting evidence they shall undertake their own testing using the various methods to confirm that the evidence provided by the manufacturer is representative. ~~The assessor shall complement the documentation review of the safety case with scenarios based testing aimed at verifying/confirming the safety case claims~~**	
Taken from Annex 6 of GRVA-20-36e	**Track testing by assessor**	
A.6.1 Track testing occurs on a closed-access testing ground that uses real obstacles and obstacle surrogates (e.g., vehicle crash targets, etc.) to assess the safety requirements of an ADS (e.g., human factors, safety system). This testing approach allows for the ADS of physical vehicles to be validated through realistic scenarios by evaluating either sub-systems or the fully assembled system. The external inputs and conditions can be controlled or measured during a test.		**OPI**: to be added to annex or interpretation document Check whether this sentence needs to be extended. To combine virtual testing with track testing. We may want to include virtual obstacles next to physical ones.
A.6.2 Track testing is suitable for assessing the ADS capabilities in nominal scenarios, critical scenarios, and failure scenarios. It can also be used to verify the performance of the vehicles regarding human factors or fall-back in these scenarios. However, operating on test tracks can be resource intensive.		**OPI**: to be added to annex or interpretation document
A.6.3 It is recommended that:		**OPI**: not suitable for regulatory purposes

A.6.3.(a)	track testing be used to assess the performance of ADS in a number of selected important nominal, critical, and failure scenarios, notably given that, unlike real-world testing, track testing can accelerate exposure to known rare events or safety critical scenarios, and in a more controlled and safer environment	**6.?.1**	**The assessor shall use** track testing to assess the performance of the ADS in a number of selected important nominal, critical, and failure scenarios. ~~, notably given that, unlike real-world testing, track testing can accelerate exposure to known rare events or safety critical scenarios, and in a more controlled and safer environment.~~	**OPI:** to come up with the proposal to highlight that this is part of the work carried by the assessor
A.6.3.(b)	track testing is conducted on a testing ground that is part of, or suitably represents, the ODD of the ADS. This excludes track tests where the objective is to assess compliance with non-ODD or extra-ODD related requirements, e.g. tests verifying that the ADS safely responds to crossing ODD boundaries, where applicable	6.?.2	Track testing **shall be** ~~is~~ conducted on a testing ground that is part of, or suitably represents, the ODD of the ADS including its physical boundaries to verify that the ADS safely responds to [crossing] ODD boundaries, or that the ADS cannot be activated outside its ODD, where applicable.	**OPI:** to work on a revised text divided in bullets
A.6.3.(c)	a test on public roads that are closed to other road users shall be considered a track test.			**OPI**: to be added to annex or interpretation document **OPI**: To check/add/amend the definition of test-track
A.6.3.(d)	real-world variation is included in the test parameters instead of limiting the test parameters to standardised parameters, standardised test objects and standardised test environments. The test parameters should therefore go beyond available standards but should remain within the ODD of the ADS. It is recommended to develop a harmonized method for selecting parameters	6.?.3	real-world variation **shall be** ~~is~~ included in the test parameters instead of limiting the test parameters to standardised parameters, standardised test objects and standardised test environments. The test parameters **shall** ~~should~~ therefore go beyond available standards but shall remain within the ODD of the ADS. ~~It is recommended to develop a harmonized method for selecting parameters~~**.**	**UK:** This seems an odd way to describe track testing. All physical testing has some natural variability. It's unclear what this is asking beyond that and what an assessor should be doing to "set this up".
NEW		6.?.3.1	The test-track, the test environment and the test objects may also be virtual elements part of a simulation toolchain, provided that the assessor is able to guarantuee their credibility is in line with the requirements laid down in 5.8.1. The ADS or the	**OPI**: added to account for the possibility that also the assessor may use virtual testing. However only HIL and VIL shall be considered as the assessor shall test real components/vehicles

			component being tested shall not be virtual elements or part of a simulation toolchain.	
A.6.3.(e)	with regard to d), the test equipment, the test set-up, and the test environment, as well as alterations made to those, are recorded at such a detail that ensures replication of the specific test.	6.?.4	with regard to 6.?.3, the test equipment, the test set-up, and the test environment, as well as alterations made to those, **shall be** ~~are~~ recorded with sufficient ~~at such a~~ detail to allow ~~hat ensures [replication] of the~~ the ~~specific~~ tests to be replicated.	**OPI**: Is replication the right word? Since you can never replicate the test exactly. **UK:** Attempt to improve the wording.
A.6.3.(f)	the selection of scenarios to be conducted on a test track is appropriate to the ODD, where possible. Track test environments allow for controllability and assurance that specific parameters that can vary in the ODD can be delivered during physical testing.	6.?.5	the selection of scenarios to be conducted on a test track **shall be** ~~is~~ appropriate to the ODD, where possible. ~~Track test environments allow for controllability and assurance that specific parameters that can vary in the ODD can be delivered during physical testing.~~	
A.6.3.(g)	the behaviour of the ADS towards other road users is verified on a test track using several scenarios.	6.?.6	the behaviour of the ADS towards other road users **shall be** ~~is~~ verified on a test track using several scenarios.	
A.6.3.(h)	with regards to human factors, the human machine interaction is tested with the ADS user under different scenarios to ensure safe use of the ADS.	6.?.7	with regards to human factors~~,~~ the human machine interaction **shall be** ~~is~~ tested with the ADS user(s) under different scenarios to ensure safe use of the ADS.	
A.6.3.(i)	for track testing a protocol is developed containing minimum requirements that standardise how for the test relevant data are to be collected and analysed (e.g., how the data is recorded, how measurements are derived from the recorded data, and how the measurements are analysed).	6.?.8	for track testing a protocol **shall be** ~~is~~ developed containing minimum requirements that standardise how for the test relevant data are [to be] collected and analysed (e.g., how the data is recorded, how measurements are derived from the recorded data, and how the measurements are analysed).]	**OPI**: the current wording suggests a standardisation. If so then this text can better be moved to the interpretation document. However a description of how data is recorded, measurements derived and analysed should be part of the test protocol.
A.6.3.(j)	to develop the track tests in line with the approach set out in Appendix 1 to this Annex.	6.?.9	~~to developed~~ the track tests **shall be executed** in line with the approach set out in Appendix 1 to this Annex.	**OPI**: to be checked and probably be moved to the interpretation document
A.6.4.	It is acknowledged that pass-fail criteria depend on the specific scenario tested, and that the selection of scenarios depends on the			**OPI**: to be added to annex or interpretation document

	ODD of the ADS under test. Moreover, it is acknowledged that a proportion of the required pass-fail criteria are not yet available, and they, or some methods to derive such pass-fail criteria, still need to be developed or parts of them could remain subjective.			
A.6.5.	As performing assessments in crowded areas could be challenging on test tracks, it is recommended such assessments to be performed in real world tests instead. Such assessments should not cover safety critical scenarios.			**OPI**: to be added to annex or interpretation document
A.6.6.	Information generated during the track test can be used as additional data to validate the virtual tests by comparing an ADS' performance between a virtual test and a test track on the same scenario. For instance, track testing can be used as an additional tool/method to validate the quality/reliability of the virtual toolchain.	6.?.10	Information generated during the track test ~~can~~ **shall** be used as additional data to validate the virtual tests by comparing an ADS' performance between a virtual test and a test track on the same scenario **in line with the requirements laid down in 5.8**. ~~For instance, track testing can be used as an additional tool/method to validate the quality/reliability of the virtual toolchain.]~~	
Taken from Annex 6 of GRVA-20-36e		**Real-world testing by assessor**		
A.6.7.	Real-world testing uses public roads to test the capabilities and compliance with safety requirements (e.g., human factors, safety system) of a vehicle with an automated driving system (ADS) in real-world traffic. It therefore provides an opportunity to validate the safety of the ADS within its true operating environment.			
A.6.8.	It is acknowledged that also for real-world tests pass-fail criteria depend on the specific scenarios tested and encountered, and that the pre-selection of scenarios depends on the			**OPI**: to be added to annex or interpretation document

	ODD of the ADS under test. Moreover, it is acknowledged that a proportion of the required pass-fail criteria are not yet available, and they, or some methods to derive such pass-fail criteria, still need to be developed or parts of them could remain subjective.			
A.6.9.	It is recommended that:			
A.6.9.(a)	real world testing assess ADS in nominal scenarios. It is acknowledged that critical and/or failure scenarios may occur during real-world testing, but they generally should not be tested on purpose. In case such scenario would occur, it shall not be excluded from the assessment;	6.??.1	Real world testing **shall** assess ADS in nominal scenarios. It is acknowledged that critical and/or failure scenarios may occur during real-world testing, but they generally shall should not be tested on purpose. In case such scenario would occur, it shall not be excluded from the assessment;	**OPI**: to be added to annex or interpretation document
A.6.9.(b)	real world testing is done safely. It is therefore recommended, if applicable to the ADS use case, that the test supervisor has the possibility to end the real world test at any point. In addition, it is also recommended that any inappropriate behaviour observed and/or the reason for the forced end is investigated in detail later;	6.??.2	real world testing ~~is~~ **shall be** done safely. It is therefo **rerequired** ~~recommended~~, if applicable to the ADS use case, that the test supervisor has the possibility to end the real world test at any point. In addition, it is also **required** ~~recommended~~ that any inappropriate behaviour observed and/or the reason for the forced end is investigated in detail later;	
A.6.9.(c)	real world testing is only conducted if a minimum level of safety of the other road users on public roads and of in-vehicle users of the ADS can be ensured by considering the validation methods of simulation, audit, and track testing as well as the manufacturer's prior real-world testing of the ADS;	6.??.3	real world testing ~~is~~ shall only be conducted if a minimum level of safety of the other road users on public roads and of in-vehicle users of the ADS can be ensured by considering the validation methods of simulation, audit, and track testing as well as the manufacturer's prior real-world testing of the ADS;	**OPI**: Second part to be included here or in the interpretation document or Annex.
A.6.9.(d)	real world testing is always conducted with other road users. Tests on public roads that are closed to other traffic should be considered as track tests;	6.??.4	real world testing **shall** ~~is~~ always be conducted with other road users. Tests on public roads that are closed to other traffic ~~should~~ **shall** be considered as track tests;	

A.6.9.(e)	real world testing be considered for assessing aspects of the ADS performance related to its capability to drive in real traffic conditions, such as: i. behavioural competencies; ii. interaction with other road users; iii. safe and anticipatory behaviour; iv. smooth driving; v. capability to deal with dense traffic; vi. maintaining flow of traffic; and being considerate and courteous to other vehicles;	6.??.5	real world testing **shall** be considered for assessing aspects of the ADS performance related to its capability to drive in real traffic conditions, such as: i. behavioural competencies; ii. interaction with other road users; iii. safe and anticipatory behaviour; iv. smooth driving; v. capability to deal with dense traffic; vi. maintaining flow of traffic; and vii. being considerate and courteous to other vehicles	
A.6.9.(f)	real world testing be considered for assessing aspects of the ADS performance at some ODD boundaries (nominal and complex scenarios), i.e. is the system triggering transition demands to the driver when it is supposed to (e.g. end of the ODD, weather conditions). The same testing could be used to confirm the performances related to human factors under these conditions;	6.??.6	real world testing **shall** be considered for assessing aspects of the ADS performance at some ODD boundaries (nominal and complex scenarios), i.e. is the system triggering transition demands to the driver when it is supposed to (e.g. end of the ODD, weather conditions). The same testing **shall** ~~could~~ be used to confirm the performances related to human factors under these conditions	**OPI**: Also mentioned by test track.
A.6.9.(g)	real world testing be considered for detecting issues that may not be well captured by track tests and simulation, such as perception quality limitation (e.g. due to light conditions, rain, etc.);	6.??.7	real world testing **shall** be considered for detecting issues that may not be well captured by track tests and simulation, such as perception quality limitation (e.g. due to light conditions, rain, etc.);	
A.6.9.(h)	real world testing be considered for assessing aspects relating to human factors, such as user-initiated deactivation, system-initiated deactivation (not leading to a minimum risk condition), audibility of messages in real world conditions, if applicable to the ADS.	6.??.8	real world testing shall be considered for assessing aspects relating to human factors, such as user-initiated deactivation, system-initiated deactivation (not leading to a minimum risk condition), audibility of messages in real world conditions, if applicable to the ADS.	
A.6.10.(a)	It is recommended that: (a) the environment and conditions of the selected test routes	6.??.9	the environment and conditions of the selected test routes **shall** reflect the applicable ODD's	**OPI**:

	reflect the applicable ODD's environment and conditions. In addition, the selected test routes should ensure that the ADS under test is expected to experience complex scenarios;		environment and conditions. In addition, the selected test routes ~~should~~ **shall** ensure that the ADS under test is expected to experience complex scenarios	Second part to be made separate requirement?
A.6.10.(b)	It is recommended that: (b) real world testing is developed in line with the approach set out in Appendix 1 to this Annex.	6.??.10	real world testing ~~is~~ **shall** be developed in line with the approach set out in Appendix 1 to this Annex.	**OPI**: to be checked and probably be moved to the interpretation document
A.6.10.(c)	It is recommended that: (c) for real world testing a protocol is developed containing minimum requirements that standardise how for the test relevant data are to be collected and analysed (e.g., how the data is recorded, how measurements are derived from the recorded data, and how the measurements are analysed).	6.??.11	for real world testing a protocol ~~is~~ **shall** be developed containing minimum requirements that standardise how for the test relevant data are to be collected and analysed (e.g., how the data is recorded, how measurements are derived from the recorded data, and how the measurements are analysed).	**OPI**: the current wording suggests a standardisation. If so then this text can better be moved to the interpretation document. However a description of how data is recorded, measurements derived and analysed should be part of the test protocol.
A.6.11	While the ADS is designed to perform the DDT only within the conditions represented by its ODD, it is recommended that real world testing assess the ADS both within its ODD and outside its ODD (e.g. to determine the ADS's appropriate recognition and response when not in its ODD) on public roads.	6.??.12	While the ADS is designed to perform the DDT only within the conditions represented by its ODD, ~~it is recommended that~~ real world testing **shall** assess the ADS both within its ODD and outside its ODD (e.g. to determine the ADS's appropriate recognition and response when not in its ODD) on public roads.	
A.6.12	Although it may not be possible to encounter all traffic scenarios during a real-world test, the likelihood of covering specific complex scenarios could be increased by selecting a specific type of ODD (e.g., highway) and examining when and where specific elements (e.g., high- or low-density traffic) typically occur.	6.??.13	Although it may not be possible to encounter all traffic scenarios during a real-world test, the likelihood of covering specific complex scenarios ~~could~~ **shall** be increased by selecting a specific type of ODD (e.g., highway) and examining when and where specific elements (e.g., high- or low-density traffic) typically occur.	
A.6.13.	Specific infractions identified during real-world testing may be reviewed and/or assessed by evaluating the data gathered	6.??.14	Specific infractions identified during real-world testing ~~may~~ **shall** be reviewed and/or assessed by evaluating the data gathered during that test and any	**OPI**: Part of SMS / Safety Case?

	during that test and any data gathered during additional virtual, track and real-world testing.		data gathered during additional virtual, track and real-world testing.	
A.6.14.	Data generated during real-world testing may be used as additional data to validate whether portions of a virtual and/or track-testing environment were modelled properly by comparing an ADS' performance within a simulation and/or track test with its performance in a real-world environment when executing the same test scenario.	6.??.15	Data generated during real-world testing ~~may~~ **shall** be used as additional data to validate whether portions of a virtual and/or track-testing environment were modelled properly by comparing an ADS' performance within a simulation and/or track test with its performance in a real-world environment when executing the same test scenario.	**OPI**: Part of SMS / Safety Case?
A.6.15.	It can also be used to support the development of new traffic scenarios for track and virtual testing, allowing for the identification of edge cases and other unanticipated hazardous situations that could challenge the ADS.	6.??.16	~~It can also~~ Data collected during real world tests **shall** be used to support the development of new traffic scenarios for track and virtual testing, allowing for the identification of edge cases and other unanticipated hazardous situations that could challenge the ADS.	**OPI**: Part of SMS / Safety Case?
A.6.16.	The information gathered from real world testing may also support improvements in the hazard and risk analysis and to the design of ADS.	6.??.17	The information gathered from real world testing ~~may~~ **shall** also support improvements in the hazard and risk analysis and to the design of ADS.	**OPI**: Part of SMS / Safety Case? 6.??.14 – 6.??.17 can be combined in one and then to added to SMS / Safety case?