**Credibility assessment framework**

|  |
| --- |
| Document History |
| Version | File | Inputs provided | Date |
| 1 | VMAD-SG2-16-02 rev2 Credibility Assessment.docx | EC, CLEPA on the basis of SG2 presentations | 22 April 2021 |
| 2 | VMAD-SG2-17-02 Credibility Assessment.docx | Continental, DE, NL, Safe, JP, BMW | 28 April 2021 |
| 3 | VMAD-SG2-18-02 Credibility Assessment.docx | PTB, Catarc, EC, CLEPA, RDW, SAFE | 12 May 2021 |
| 4 | VMAD-SG2-19-02 Credibility Assessment.docx | EC, RDW | 02 June 2021 |

**How the Credibility Assessment Fits With ADS Safety Requirements and Other Validation Methods**

**[**I think we need to explain how this all fits together. As I understand it: FRAV will produce ADS requirements and metrics. SG-1 will develop scenarios that can be used to validate if the ADS meets those requirements. SG-1 or VMAD as a whole will determine which validation methods are appropriate for testing each scenario. Some will require test track or on-road testing, but some will be validated only by virtual testing (most likely those involving crash avoidance or failure mitigation capabilities). Is SG-2 recommending that the type approval authority accept as fully validated those requirements for which the ADS’s performance has been validated by simulation IF the assessor has deemed the simulation toolchain credible? Or would the type approval authority re-run the simulations using the manufacturer’s toolchain? If the former, this would seem to be similar to self-certification of certain requirements based on credible virtual testing. If the latter, it’s not clear how the type approval authority would access the toolchain or why it would even want to repeat identical simulations. What this suggests is that any requirements that can be validated by credible simulation will not be subject to further testing by the type approval authority. If that’s the intent, it should be made clear.

Meanwhile, the audit pillar would seem to be the place at which this gets determined, and the extensive documentation of the sim tools’ credibility would be a huge part of that assessment.

Also, we need to address what the scoring criteria are for credibility (perfection on all points?) and what the result is if the sim tools do not pass the credibility assessment. Would the ADS manufacturer have a chance to challenge the assessor’s judgment? Would the manufacturer be given a chance to cure specific identified deficiencies in the toolchain? This will be incredibly important to get right because so much of ADS validation rests on simulation and incorrect judgment by an assessor could have enormous impact on the manufacturer. ]

***Credibility assessment for using virtual toolchain in ADS validation***

* ***Introduction, motivation, and scope***. The use of Modelling & Simulation (M&S) is becoming widespread thanks to the increasing computational capabilities, accuracy, usability, and availability of M&S software packages. M&S can be beneficial for ADS safety validation because it allows to overcome some real testing limitations and to increase the number of testing scenarios. Nonetheless, M&S can also lead to erroneous/seemingly correct results, especially in relation to complex simulations not adequately supported by robust practices addressing all M&S aspects beyond pure validation. Therefore, higher confidence in M&S outcome is needed to apply virtual testing instead of/in conjunction with the other NATM pillars. In other words, M&S can be used for virtual testing if an assessor is able to consider the simulation results *credible* enough to make sound decisions taking into account the potential uncertainties of M&S. The validation of M&S can be considered the hallmark of simulation credibility. However, the validation has some limitations, which include the limited scope of the validation tests and the difficulty in retrieving data supporting the validation procedures. The use of M&S requires more attention towards all factors influencing the quality and validity of M&S with aim at:
	1. identifying a common framework to determine, justify, assess and report the overall credibility of the M&S,
	2. getting confidence on the predictive capability of virtual testing for ADS validation,
	3. indicating the levels of confidence in results.

At the same time, this framework should be general enough to be used for different M&S types and applications. However, the goal is complicated by the broad differences across ADS features and the variety of M&S types and applications. These considerations lead to introduce a (risk-based/informed) credibility assessment framework relevant and appropriate to all M&S applications.

The proposed credibility assessment framework provides a general description of the main aspects considered for assessing the credibility of an M&S solution together with guidelines of the role played by 3rd parties assessors in the validation process with respect to credibility. Concerning the latter point, the assessor shall investigate the produced documentation supporting credibility at the audit phase, whereas the actual validation tests occur once the OEM has developed the integrated simulation systems.

Ultimately, the outcome of the current credibility assessment shall define the *envelope* in which the virtual tool can be used to support the ADS assessment.

* ***Components of the credibility assessment framework***. M&S can be used for virtual testing if its credibility is established by evaluating the fitness of M&S for the intended purpose. The credibility can be achieved by investigating and assessing four M&S properties:
	1. Capability – what the M&S can do, and what are the risks associated;
	2. Accuracy – how well M&S does reproduce the target data;
	3. Correctness – how sound & robust are M&S data and algorithms;
	4. Usability – what training and experience is needed and what quality of the process applied to it.

Therefore, credibility requires a unified method to investigate these properties and get confidence in the M&S results. The Credibility Assessment framework introduces a way to assess and report the credibility of M&S based on quality assurance criteria that allow indicating the levels of confidence in results. In other words, the credibility is established by evaluating the following M&S influencing factors that are considered as main contributors for M&S properties and therefore for the overall M&S credibility: M&S management, team's experience and expertise, M&S analysis and description, data/input pedigree, verification, validation, uncertainty characterization. Each of these factors indicates the level of quality achieved by M&S, and the comparison between the obtained levels and the required levels leads to consider the M&S credible and fitness to use for virtual testing.



* + ***M&S (Models and Simulation) Management***. The M&S lifecycle is a dynamic process with frequent releases that shall be monitored and documented. Management activities shall be established to support the M&S in a work product management fashion. Relevant information on the following aspects shall be included in this section:
		- **M&S management process:** shall:
			* describe the modifications within the releases,
			* designate the corresponding software (e.g., specific SW product and version) and hardware arrangement (e.g., XiL configuration),
			* record the internal review processes that accepted the new releases,
			* be supported throughout the full duration of the virtual model utilization
		- **Releases management:** the M&S toolchains used for ADS virtual testing purposes are, in general, complex multi-tool virtual environments. As such, the user(s)/developer(s) shall be able to detail the updates which have led to the tool’s official version used for validation purposes. The development of the virtual models constituting the testing toolchain shall be documented in terms of the corresponding validation methods and acceptance thresholds to support the overall credibility of the toolchain.
		- **Quality check of virtual data:** data completeness, accuracy, and consistency shall be ensured throughout the releases and lifetime of an M&S toolchain to support the verification and validation procedures.
		- **Criticality assessment:** the simulation models and the simulation tools used in the overall tool-chain shall be investigated in terms of their **responsibility in case of a safety error in the final product**. The proposed approach for criticality analysis is derived from ISO 26262, which requires qualification for some of the tools used in the development process. In particular, ISO 26262 sets 3 levels of Tool Confidence Level (TCL) qualification which are related to: "the possibility that the malfunctioning software tool can… …fail to detect errors in a safety-related item” [ISO 26262-8, 11.2].
			* **TCL1: lowest tool confidence level**. The tool does not play an important role regarding the safety of our final product. Therefore, there is no need to have a high confidence in the correct tool behavior from an ISO 26262 view. A tool qualification is not needed.
			* **TCL2/3: This corresponds to a medium/high tool confidence level**. The tool plays an important role regarding the safety of our final product, so we need to have a certain level of confidence and therefore need to perform a tool qualification to demonstrate the reliability of that tool.
		- **Releases archive:** any M&S toolchain’s version used to release data for certification purposes shall be stored. The developer shall enforce a method to trace generated data to the corresponding M&S version.
* ***Team's Experience and Expertise***. Even though Experience and Expertise (E&E) are already covered in a general sense within organization, it is important to get evidence on the specific experience and expertise for M&S activities. In fact, the credibility of M&S is subjected to the E&E of the personnel involved in the M&S activities and this factor can play an important role in producing credible results. Team’s Experience and Expertise include two levels:
1. **Organizational level:**

The credibility is established by setting up processes and procedures to identify and maintain skills, knowledge, and experience to perform M&S activities. The following processes should be established, maintained and documented:

* 1. Process to evaluate the individual’s competence and skills;
	2. Process for training competent personnel to perform M&S-related duties.
1. **Individual level:**

The credibility is established by documenting the education received, the number of years of experience, the training received by the personnel involved in the M&S activities. At least, the following aspects should be documented and justified in order assess that M&S team has adequate Experience and Expertise for performing M&S activities in a credible manner.

* 1. Educational Background;
	2. Experience in M&S Discipline;
	3. Training for the Specific M&S;
	4. Experience with the Specific M&S.

The following scale should be used as a reference to assess the credibility for E&E. (derived from NASA-STD-7009)

|  |  |  |
| --- | --- | --- |
| **Level** | **E&E** | **Evidence** |
| Level 4 | Extensiveexperience  | * Engineering or science degree or extensive M&S experience in development and use of M&S
* Extensive M&S experience in development and use of M&S under evaluation
* Experience in the development and use of recommended practices (if identified)
 |
| Level 3 | Advanced experience | * Engineering or science degree or extensive work experience in M&S,
* Extensive work experience in development and use of M&S,
* Training and experience with the specific M&S
 |
| Level 2 | Formal experience  | * Formal technical education,
* M&S development experience,
* Training and experience with the specific M&S
 |
| Level 1 | Basic Experience | * Basic technical education, training, and experience related to M&S are documented
 |
| Level 0 | Insufficientevidence | Insufficient evidence |

Ultimately, the goal of this section it to provide evidence on the ability of the team and the organization to develop and use M&S

* + ***M&S Analysis and description***. The M&S analysis and description aim to define the whole M&S and identify the parameter space that can be assessed via virtual testing. It defines the scope and limitations of the models and toolchain and the uncertainty sources that can affect its results.
		- General description:
			* OEM should provide a description of the complete toolchain along with how the simulation data will be used to support the ADS validation strategy.
			* The OEM should provide a clear description of the M&S toolchain’s objective.
		- Assumptions, known limitations and uncertainty sources:
			* The OEM shall motivate the modelling assumptions which guided the design of the M&S toolchain
			* The manufacturer-defined assumptions play a major role in defining the limitations of the toolchain
			* Assumptions of the real world defined in IEEE P2846 may be used to support this
			* Different degrees of fidelity may be required for each of the model's
			* Simulation fidelity is dependent on the input data and how the data is used to support the ADS validation.
			* The OEM should provide [information on] any data and scenarios used for virtual testing toolchain validation.
			* The OEM should define a minimal tolerance of sim-real correlation
			* In addition to the assumptions used in developing the M&S, known limitations define conditions for which the virtual toolchain or one of its component is not valid and which the model cannot be used for
			* Finally, this section shall include information about the sources of uncertainty in the model. This will represent an important input to final uncertainty analysis, which will define how the model outputs can be affected by the different sources of uncertainty of the model used.
		- Scope (what is the model for?). It defines how the M&S is used in the ADS validation.
			* The credibility of virtual tool shall be enforced by a clearly defined scope of utilization the developed models.
			* The matured M&S shall allow a virtualization of the physical phenomena to a degree of accuracy which matches the fidelity level required for certification. Thus the M&S will act as a “virtual proving ground” for ADS testing.
			* Simulation models need dedicated scenarios and metrics for validation. The scope includes the list of scenarios, among those needed in the ADS validation, that M&S will allow to execute together with the corresponding parameters’ limitation.
			* ODD analysis is a crucial input to derive requirements, scope, effects that the M&S must consider in order to support ADS validation.
			* Parameters generated for the scenarios will define extrinsic and intrinsic data for the toolchain and the simulation models.
	+ ***Data/Input pedigree***. The data/input pedigree contains a record of traceability from the OEM’s data used in the development of the M&S.
		- Description of the data used for the M&S
			* The OEM shall document the data used to develop the model and note important quality characteristics
			* The OEM shall provide documentation showing that the data used to develop the models covers the intended functionalities the toolchain aims at virtualizing
			* The OEM shall document the calibration procedures employed to fit the virtual models’ parameters on the collected input data
		- Effect of the data quality on model parameters uncertainty
			* The quality of the data used to develop the model will have an impact on model parameters’ estimation and calibration. Uncertainty in model parameters will be another important aspect in the final uncertainty analysis.
	+ ***Data/Output pedigree***. The data/output pedigree contains a record of the signals selection that the M&S allows investigating.
		- Description of the data generated by the M&S
			* The OEM shall document the exported data and note important quality characteristics
			* The OEM shall trace a M&S output to the corresponding simulation setup
		- Effect of the data quality M&S credibility
			* The M&S output data shall be sufficiently wide to ensure the correct execution of the validation computation
			* The output data shall allow consistency/sanity check of the virtual models via possibly exploiting redundant information
		- Managing stochastic models
			* Stochastic models shall be characterized in terms of their variance
			* Stochastic models shall be ensured the possibility of deterministic re-execution
		- ***Verification***. The verification of an M&S deals with the analysis of the correct implementation of the conceptual/mathematical models building up the M&S toolchain. The verification contributes to the M&S’s credibility via providing assurance that the M&S will not exhibit unrealistic behavior for a set of input which cannot be tested. The procedure is grounded on a multi-step approach which includes code verification, calculation verification and sensitivity analysis.
		- Code verification is concerned with the execution of test demonstrating that no numerical/logical flaws affect the virtual models
			* The OEM shall document the execution of proper code verification techniques
			* The OEM shall provide documentation showing that the exploration in the domain of the input parameters was sufficiently wide to identify parameters’ combination for which the M&S shows unstable or unrealistic behavior
			* The OEM shall adopt sanity/consistency checking procedures whenever data allows
		- Calculation verification deals with the estimation of numerical errors affecting the M&S
			* The OEM shall document numerical error estimates (e.g. discretization error, rounding error, iterative procedures convergence)
			* The numerical errors shall be kept sufficiently bounded to not affect validation
		- Sensitivity analysis aims at quantifying how model output values are affected by changes in model input values and pointing out the input factors having the greatest impact on the model results
			* The OEM shall provide supporting documentation that demonstrate sufficient robustness of the models
			* Ultimately, the sensitivity analysis results will also help defining the inputs and parameters whose uncertainty characterization needs particular attention in order to properly define the uncertainty of the simulation results.
	+ ***Validation***. The quantitative process of determining the degree to which a model or a simulation is an accurate representation of the real world from the perspective of the intended uses of the M&S.
		- Measures of Performance (metrics)
			* The performance metrics are defined during the M&S analysis.
			* Metrics for validation may include:
				+ Discrete value analysis e.g. detection rate, firing rate.
				+ Time evolution e.g. positions, speeds, acceleration.
				+ Flow of actions based analysis e.g. distance/speed calculations, TTC calculation, brake initiation.
		- Goodness of Fit measures
			* The analytical frameworks used to compare real world and simulation metrics. They are generally KPIs indicating the statistical comparability between two sets of data.
			* The validation should show that these KPIs are met.
		- Validation methodology
			* The OEM should define the concrete scenarios used for virtual testing toolchain validation. They should be able to cover to the maximum possible extent the domain of usage of virtual testing for ADS validation.
			* The exact methodology depends on the structure and purpose of the toolchain. The validation may consist of one or more of the following:
				+ Validate Subsystem models e.g. environment model (road network, weather conditions, road user interaction), sensor models (RADAR, Camera, LIDAR), vehicle model (steering, braking, powertrain)
				+ Validate vehicle system (vehicle dynamics model together with the environment model)
				+ Validate sensor system (sensor model together with the environment model)
				+ Validate integrated system (sensor model + environment model with influences form vehicle model)
		- Accuracy requirement
			* Accuracy requirement is defined during the M&S analysis. The validation should show that these KPIs are met.
		- Validation scope (what part of the toolchain to be validated)
			* A toolchain consists of multiple tools, and each tool will use a number of models. The validation scope includes all tools and their relevant models.
		- Internal validation results
			* The documentation should not only provide evidence of the simulation model validation but also used to obtain sufficient information related to the processes and products that provide overall credibility of the toolchain used.
			* Documentation/results may be carried over from previous credibility assessments.
		- Independent Validation of Results
			* The assessor shall audit the results of the OEM by carrying out physical tests of the complete integrated tool
	+ ***Uncertainty characterisation***. This section is concerned with characterizing the expected uncertainties of the virtual toolchain results. It is composed by two phases. In a first phase the information collected the M&S Analysis and Description section and the Data/Input Pedigree are used to characterise the uncertainty in the input data, in the model parameters and in the modelling structure. Then, by propagating all the uncertainties through the virtual tool-chain, the uncertainty in the model results is quantified. Depending on the uncertainty in the model results, proper safety margins will need to be introduced in the use of virtual testing of ADS validation.
		- Characterisation of the uncertainty in the input data
		- Characterisation of the uncertainty in the model parameters (following calibration)
		- Characterisation of the uncertainty in the M&S structure
		- Characterisation of aleatory vs. epistemic uncertainty
		- Combined effect of inputs/parameters/model uncertainty on model outputs
	+ ***Documentation structure***. This section will define how the aforementioned information will be collected and organised in the documentation provided by the OEM to the relevant authority.
		- The OEM shall produce a document (a “simulation handbook”) structured after the present outline providing evidence for the topics presented
		- The documentation shall be delivered together with the corresponding release of the M&S and related produced data
		- The OEM shall provide clear reference that allows tracing the documentation to the corresponding M&S/data.
		- The documentation shall be maintained throughout the whole lifecycle of the M&S utilization
	+ ***Interdependences with VMAD SG1 and SG3.***
		- VMAD SG1’s developed scenarios are the input of the M&S toolchain
		- The credibility analysis can be exploited to support industry audit’s procedures established in VMAD SG3