**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 |
| 5 | VMAD-SG2-20-02 rev2 Credibility Assessment.docx | SAE, BMW | 09 June 2021 |
| 6 | VMAD-SG2-21-02 Credibility Assessment.docx | Japan, UK and RDW | 16 June 2021 |
| 7 | VMAD-SG2-21-02 rev 2 Credibility Assessment Clean.docx |  | 17 June 2021 |
| 8 | VMAD-SG2-22-02 Credibility Assessment.docx |  | 23 June 2021 |
| 9 | VMAD-SG2-26-02 Credibility Assessment.docx | RDW | 23 July 2021 |
| 10 | VMAD-SG2-27-02 Credibility Assessment.docx | EC/CLEPA – clean version + new ‘team experience’ | 1 September 2021 |
| 11 | VMAD-SG2-28-02 Credibility Assessment.docx | EC/CLEPA – new ‘criticality assesment’ | 8 September 2021 |
| 12 | VMAD-SG2-28-02rev2 Credibility Assessment.docx | Clean version for VMAD | 10 September 2021 |
| 13 | VMAD-SG2-30-02 Credibility Assessment.docx | Input from SAE, CAN, J, NL, SAFE | 1 October 2021 |
| 14 | VMAD-SG2-32-02 Credibility Assessment.docx | Clean version for VMAD  | 20 October 2021 |
| 15 |  |  |  |

***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 credibility 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. indicating the levels of confidence in results from the validation phase.

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[[1]](#footnote-1) in the validation process with respect to credibility. Concerning the latter point, the assessor should investigate the produced documentation supporting credibility at the audit phase, whereas the actual validation tests occur once the ADS manufacturer has developed the integrated simulation systems.

Ultimately, the outcome of the current credibility assessment should 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 five 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.
	5. Fit for Purpose – how suitable the M&S is for the ODD and ADS assessment.

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. A graphical representation of the relationship among the components of the credibility assessment framework is reported in Figure 1.



**Figure 1. Graphical representation of the relationships between the components of the credibility assessment framework**

* + ***M&S (Models and Simulation) Management***. The M&S lifecycle is a dynamic process with frequent releases that should be monitored and documented. Management activities should be established to support the M&S in a work product management fashion. Relevant information on the following aspects should be included in this section:
		- **M&S management process:** should:
			* 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:** Any M&S toolchain’s version used to release data for certification purposes should be stored. The virtual models constituting the testing toolchain should be documented in terms of the corresponding validation methods and acceptance thresholds to support the overall credibility of the toolchain. The developer should enforce a method to trace generated data to the corresponding M&S version.
		- **Quality check of virtual data:** data completeness, accuracy, and consistency should be ensured throughout the releases and lifetime of an M&S toolchain to support the verification and validation procedures.
* ***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 establish the basis for confidence on the specific experience and expertise for M&S activities.

In fact, the credibility of M&S depends not only on the quality of the simulation models but also on the E&E of the personnel involved in the validation and usage of the M&S. For instance, a proper understanding of the limitations and validation domain will prevent from possible misuse of M&S or from misinterpretation of its results.

In this perspective, it is important to establish the basis for the ADS manufacturer’s confidence on the experience and expertise of:

* the Teams that will validate the simulation toolchain and,
* the Teams that will use the validated simulation for the execution of virtual testing with the purpose of validating the ADS

Thus, Team’s E&E increase the level of confidence on the credibility of M&S and its outcomes by ensuring that the human factors behind the M&S are taken into consideration and any possible human component risk is controlled as expected in any suitable Management System.

If the ADS manufacturer's tool chain incorporates or relies upon inputs from organizations or products outside of the manufacturer's own team, the ADS manufacturer will include an explanation of measures it has taken to support its confidence in the quality and integrity of those inputs.

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

Once a M&S has been finalized, its credibility is mainly dictated by the skills and knowledge of the individual/team that will validate the M&S Toolchain and will use the M&S for the validation of ADS. The credibility is established by documenting that these Teams have received adequate training to fulfil their duties.

The ADS manufacturer should then:

* Provide the basis for the ADS manufacturer’s confidence in the Experience and Expertise of the individual/team that validates the M&S Toolchain
* Provide the basis for the ADS manufacturer’s confidence in the Experience and Expertise of the individual/team that uses the simulation to execute virtual testing with the purpose of validating the ADS

The ADS manufacturer’s demonstration of how it applies the principles of ISO 9001 or a similar best practice or standard with regard to the competence of its M+S organization and the individuals in that organization will provide the necessary basis for this determination. The assessor may not substitute its judgment for that of the ADS manufacturer with regard to the experience and expertise of the organization or its members.

* + ***Data/Input pedigree***. The data/input pedigree contains a record of traceability from the ADS manufacturer’s data used in the validation of the M&S.
		- Description of the data used for the M&S
			* The ADS manufacturer should document the data used to validate the model and note important quality characteristics
			* The ADS manufacturer should provide documentation showing that the data used to validate the models covers the intended functionalities the toolchain aims at virtualizing
			* The ADS manufacturer should document the calibration procedures employed to fit the virtual models’ parameters on the collected input data
		- Effect of the data quality (e.g. data coverage, signal to noise ratio, and sensors’ uncertainty/bias/sampling rate) 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 ADS manufacturer should provide [information on] any data and scenarios used for virtual testing toolchain validation.
			* The ADS manufacturer should document the exported data and note important quality characteristics e.g. using the correlation methodologies as defined Annex II.
			* The ADS manufacturer should trace a M&S output to the corresponding simulation setup
		- Effect of the data quality M&S credibility
			* The M&S output data should be sufficiently wide to ensure the correct execution of the validation computation. The data should sufficiently reflect the ODD relevant to the virtual assessment of the ADS.
			* The output data should allow consistency/sanity check of the virtual models via possibly exploiting redundant information
		- Managing stochastic models
			* Stochastic models should be characterized in terms of their variance
			* Stochastic models should be ensured the possibility of deterministic re-execution
	+ ***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:
			* ADS manufacturer should provide a description of the complete toolchain along with how the simulation data will be used to support the ADS validation strategy.
			* The ADS manufacturer should provide a clear description of the test objective.
		- Assumptions, known limitations and uncertainty sources:
			* The ADS manufacturer should motivate the modelling assumptions which guided the design of the M&S toolchain
			* The ADS manufacturer should provide evidence on:
				+ How the manufacturer-defined assumptions play a role in defining the limitations of the toolchain
				+ the level of fidelity required for the simulation models
			* The ADS manufacturer should provide justification that the tolerance for sim-real correlation is acceptable for the test objective
			* Finally, this section should 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 should be enforced by a clearly defined scope of utilization the developed models.
			* The matured M&S should 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 scenario selection used for validation should be sufficient such that there is confidence that the toolchain will perform in the same manner in those scenarios outside of the validation scope.
			* ADS manufacturers should provide a list of validation scenarios 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.
		- **Criticality assessment:** the simulation models and the simulation tools used in the overall tool-chain should 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 order to derive how critical the simulated data is, the criticality assessment considers the following parameters:
* the consequences on human safety e.g. severity classes in ISO 26262
* the degree in which the simulated results influence’s the ADS

The table below provides a sample criticality assessment matrix to demonstrate this analysis. ADS manufacturers may adjust this matrix to their particular use case.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Influence on ADS | Significant | N/A | Perform degraded mode within reduced system constraints | Create a collision free and lawful driving plan | Correctly execute and actuate the driving plan |
| Moderate | Determine its location | Predict the future behaviour of other actors | Perceive relevant static and dynamic objects in the proximity of the ADS |
| Minor | Strategic control of the ADS by the User | Communicate and interact with other road users | Safe management of transitions of control | Determine if specified nominal performance is not achieved |
| Negligible | User interaction with HMI | User informed about operational status | N/A |
|  | Negligible | Minor  | Moderate | Significant |
| Decision consequence |

From the perspective of the criticality assessment, the three possible cases for assessment are:

* Those models or tools that fall within the red boxes are clear candidates for fully following the credibility assessment
* Those models or tools that fall within the yellow boxes may or may not be candidates for fully following the credibility assessment at the discretion of the assessor
* Those models or tools that fall within the green boxes are not required to follow the credibility assessment
	+ - ***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 ADS manufacturer should document the execution of proper code verification techniques, e.g. static/dynamic code verification, convergence analysis and comparison with exact solutions if applicable[[2]](#footnote-2)
			* The ADS manufacturer should 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. Coverage metrics of parameters combinations may be used to demonstrate the required exploration of the models behaviours.
			* The ADS manufacturer should adopt sanity/consistency checking procedures whenever data allows
		- Calculation verification deals with the estimation of numerical errors affecting the M&S
			* The ADS manufacturer should document numerical error estimates (e.g. discretization error, rounding error, iterative procedures convergence)
			* The numerical errors should be kept sufficiently bounded to not affect validation
		- Sensitivity analysis aims at quantifying how model output values are affected by changes in the model input values and thus pointing out the parameters having the greatest impact on the simulation model results. The sensitivity study also affords determining the extent to which the simulation model satisfies the validation thresholds when it is subjected to small variations of the parameters, thus it plays a fundamental role to support the credibility of the simulation results.
			* The ADS manufacturer should provide supporting documentation demonstrating that the most critical parameters influencing the simulation output have been identified by means of sensitivity analysis techniques such as by applying a perturbation of the model’s parameters;
			* The ADS manufacturer should demonstrate that robust calibration procedures have been adopted while identifying and calibrating the most critical parameters to the end of increasing the credibility of the developed toolchain.
			* 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 ADS manufacturer should define the abstract / logical scenarios used for virtual testing toolchain validation. They should be able to cover to the maximum possible extent the ODD 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
			* Requirement for the correlation threshold is defined during the M&S analysis. The validation should show that these KPIs are met. e.g. using the correlation methodologies as defined Annex II.
		- 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 should audit the documentation provided by the manufacturer and may carrying out physical tests of the complete integrated tool
	+ ***Uncertainty characterisation***. This section is concerned with characterizing the expected variability of the virtual toolchain results. The assessment should be made up of 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 by the ADS manufacturer in the use of virtual testing of ADS validation.
		- Characterisation of the uncertainty in the input data
			* The ADS manufacturer should demonstrate to have opportunely estimated critical model’s inputs by means of robust techniques such as providing multiple repetitions for the assessment of the quantity;
		- Characterisation of the uncertainty in the model parameters (following calibration)
			* The ADS manufacturer should demonstrate that critical model’s parameters that cannot be estimated identically are characterized by means of a distribution and/or confidence intervals;
		- Characterisation of the uncertainty in the M&S structure
			* The ADS manufacturer should provide evidence that the modeling assumptions are given a quantitative characterization of the generated uncertainty (e.g. comparing the output of different modeling approaches whenever possible);
		- Characterisation of aleatory vs. epistemic uncertainty:
			* The ADS manufacturer should aim to distinguish between the aleatory component of the uncertainty (which can only be estimated but not reduced) and the epistemic uncertainty deriving from the lack of knowledge in the virtualization of the process.
	+ ***Documentation structure***. This section will define how the aforementioned information will be collected and organised in the documentation provided by the ADS manufacturer to the relevant authority.
		- The ADS manufacturer should produce a document (a “simulation handbook”) structured after the present outline providing evidence for the topics presented
		- The documentation should be delivered together with the corresponding release of the M&S and related produced data
		- The ADS manufacturer should provide clear reference that allows tracing the documentation to the corresponding M&S/data.
		- The documentation should be maintained throughout the whole lifecycle of the M&S utilization. The assessor may audit the ADS manufacturer through assessment of their documentation and/or by conducting physical tests.
	+ ***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

**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. ]

1. For type-approval the manufacturer produces the whole documentation upfront with the authority requested to study it and provide its assessment. In the self-certification the same may be done by the assessor during market surveillance [↑](#footnote-ref-1)
2. Roy, C. J. (2005). Review of code and solution verification procedures for computational simulation. *Journal of Computational Physics*, *205*(1), 131-156. [↑](#footnote-ref-2)