Rail4Future

Projekttitel:	**Resilient Digital Railway Systems to enhance performance**
Start Datum:	**01/04/2021**
Durchlaufzeit:	**42 Monate**
Projektnummer:	**882504**
Ausschreibung:	**8. Ausschreibung COMET Projekte 2019**

Deliverable 1.3.1

Simulation Reliability Assessment Indices

Fälligkeitsdatum	
Einreichungsdatum	
Eingereicht von	

Version	**Datum**	**Bearbeitet von**	**Beschreibung**
0.1	16.05.2024	Stefan Reiterer	Erstellung
0.2	05.06.2024	Stefan Reiterer	Übersetzung
0.3	09.06.2024	Manfred Grafinger	Korrekturlesung
1.0	11.06.2024	Manfred Grafinger	Endkorrektur

Deliverable released

by Area 1 Manager
Dr. Michaela Haberler-Weber

Deliverable released

by Area 1 Scientific Lead
Dr. Manfred Grafinger

1 Executive Summary

Simulations can only be helpful if appropriate processes and metrics are available to assess their quality and determine their significance. If possible, these evaluations should be automated in order to seamlessly integrate the evaluation into processes so as not to create additional hurdles for developers, but still create value and support the evaluation of the available simulation results and data. These criteria are therefore also necessary to enable virtual certification.

2 Table of Content

3 Abbreviations and Akronyms

Abbreviations / Akronyms	Description
FMEA	Failure Mode and Effects Analysis
ASPICE	Automotive Software Process Improvement Capability dEtermination.
KPI	Key Performance Indicator

4 Problem Description / Objective / Purpose

4.1 Problem Description

To create value for engineers from simulations they must also be correct. This in turn requires clear criteria that can be tested automatically if possible and can be fully integrated into a development project to accompany the process and not hinder it. Criteria must therefore be available to determine whether the accuracy and quality of the simulation is given, or whether the accuracy is of a qualitative or quantitative nature. Simulations that are not exact, but from which qualitative information can be derived with great certainty and in a short time, create value for the users when it is provided to what extent they can trust the result.

4.2 Objective

The objective of this report is to identify the state of the art of existing methodologies and how they can be applied within the R4F project.

4.3 Purpose

To provide quality metrics and criteria to help decision makers evaluate and continuously improve their simulation. In addition, they are crucial to enable virtual approvals in early stages, as physical tests are often still mandatory due to a lack of (scientific) criteria and metrics.

5 Significance for the overall project

To make simulation development sustainable, there must be a process and metrics to ensure the quality of the simulation results and data in order to avoid errors in decisions and to improve continuously the quality of the simulations. These methods are also essential regarding potential virtual approvals and releases.

6 Description

6.1 State-of-the-Art

The development of credible simulation processes is not yet fully established in many industries. As a starting point for our work, we use findings from the automotive industry where corresponding initiatives such as the ITEA3 UPSIM project (https://upsim-project.eu/) are already being carried out to determine and continuously improve the quality of simulations. SetLabs has developed the SetLevel Credible Simulation Process. Methodologies for determining the required metrics are proposed there as well.

6.2 Credible Simulation Process

In this Section we present a summary of the Credible Simulation Process described in [1].

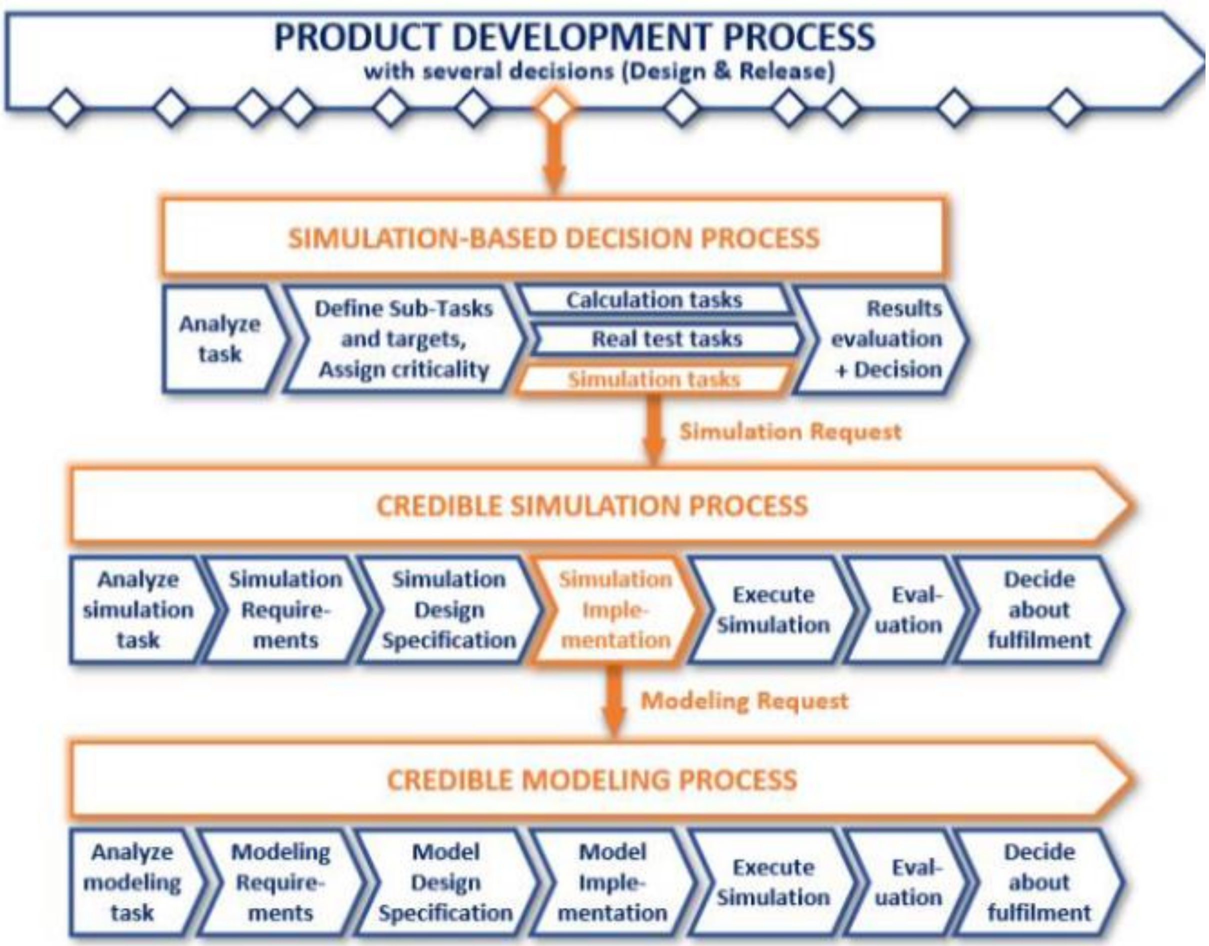

Figure 1: Credible Simulation Process [2]

Figure 1 shows the process schematically.
(https://setlevel.de/assets/forschungsergebnisse/Credible-Simulation-Process.pdf)

First, the task to be simulated is analyzed. Then the criticality of the simulation for the decision-making process is classified (Figure 2). The calculation is based on the calculation method of the Failure Mode and Effects Analysis (FMEA). Based on this analysis, the individual cases to be treated are divided into categories as shown in Figure 3. Subsequently, a required "Credibility Level" is determined (see Figure 4). In order to achieve these credibility levels, the processes must meet corresponding quality requirements, which must be represented by corresponding process capability levels (see Figure 5). It is important to note that this process is strongly based on A(automotive)SPICE, and therefore the procedures have been tried and tested over long time in analogous problems.

M&S Results Influence						
	5: Controlling	(G)	(Y)	(R)	(R)	(R)
	4: Significant	(G)	(Y)	(Y)	(R)	(R)
	3: Moderate	(G)	(Y)	(Y)	(Y)	(R)
	2: Minor	(G)	(G)	(G)	(Y)	(Y)
	1: Negligible	(G)	(G)	(G)	(G)	(Y)
		I: Negligible	II: Minor	III: Moderate	IV: Significant	V: Catastrophic
		Decision Consequence				

Figure 1: NASA M&S Critically Assessment Matrix (vgl. [3])

	Level	Label	Consequence for Product and/or Humans and/or Business in case of failure
Safety	10	Fatal	**Product:** Fatal impact on user and uninvolved persons. Causes the operational safety to be violated without a prior warning, so that government regulations will be violated. 100% scrap, no rework possible. **Human:** Fatal injury that leads to permanent disability or death. **Business:** Fatal detriment to organizational business. Causes major project cancellation and/or has significant influence on the overall organizational business.
	9	Severe	**Product:** Severe impact on user and uninvolved persons. Causes the operational safety to be endangered with a prior warning, so that government regulations will be violated if the bad event cannot be prevented after the warning. 100 % scrap, no rework possible if bad event will not be prevented. **Human:** Severe injury that may lead to temporal disability or death. **Business:** Severe detriment to project and overall organizational business. Causes major project reductions.
	8	Serious	**Product:** Serious deterioration of product safety. Even may call for total rework, up to 100% scrap possible. However, no violation of regulation or government norms. **Human:** Serious injury or severe occupational illness leading to hospitalization and typically associated with claims for compensation. **Business:** Serious detriment to full completion of project. Typically, a significant scope of the project will be cancelled for being able to deploy a distinctive sub-scope of the project.
Functionality	7	Significant	**Product:** Significant deterioration of product functionality, including non-functioning or inoperable. Serious disruption of work. Some part may completely scrap. **Human:** Significant injury or significant occupational illness that makes a doctor's visit necessary immediately after the event. **Business:** Significant detriment to project execution. Requires cancellation of smaller sub-projects or partial scopes of the project.
	6	Moderate	**Product:** Moderate deterioration of product functionality. May be noticed by nearly all users (more than 95%) with annoyance and discomfort. Some part may call for rework, scrap possible. **Human:** Moderate injury or little occupational illness that makes a doctor's visit necessary within one week after the event. **Business:** Moderate business detriment. Requires the project's business plan to be refined.
	5	Minor	**Product:** Minor deterioration of product functionality. May be noticed by most of users (around 80%). Unscheduled rework. **Human:** Minor injury or minor occupational illness without the need for a doctor's visit. **Business:** Minor business detriment, but outside calculations.
Performance	4	Notable	**Product:** Notable faults to detriment of product performance. May be noticed by around the half of all users. Also, minor rework may be called for. **Human:** Notable human detriment like significant discomfort. **Business:** Notable business detriment, but within calculations.
	3	Inconsequential	**Product:** Inconsequential faults to detriment of product performance. May be noticed by some users (around 20%). **Human:** Inconsequential human detriment like significant annoyance and/or little discomfort. **Business:** Inconsequential business detriment.
	2	Negligible	**Product:** Negligible faults to detriment of product performance. May only be noticed by attentive users (less than 5%). **Human:** Negligible human detriment like little annoyance. **Business:** Negligible business detriment.
	1	Nonexistent	No human, product or business detriment.

	Level	Label	Probability that event from decision consequence happens at least once during lifetime
Likely	10	Certain	Event occurrence is almost certain (p ≥ 1:2)
	9	Frequently	Very high number of event likely (p ≥ 1:3)
	8	Repeatedly	High number of event likely (p ≥ 1:8)
Occasionally	7	Periodically	Moderately high number of event likely (p ≥ 1:20)
	6	Notable	Medium number of event occurrence likely (p ≥ 1:80)
	5	Erratically	Occasional event likely (p ≥ 1:400)
Unlikely	4	Scarce	Very low likelihood of event (p ≥ 1:2K)
	3	Rarely	Rare likelihood of event (p ≥ 1:15K)
	2	Negligible	Extremely low chances of event occurrence (p ≥ 1:150K)
	1	Impossible	Event will practically not occur or will be occuring very rarely (p ≤ 1:1.5M)

	Level	Label	Influence on decision of higher-level engineering task
Controlling	10	Critical	The simulation task is critical for the decision, as it has prevailing influence, meaning that it can veto the contribution of other sub-tasks for the decision.
	9	Dominating	The decision is highly dependent on the simulation task, meaning that it is one of the dominating sub-tasks for the decision, but without the ability to veto the decision.
	8	Governing	The decision is dependent on the result of the simulation task, meaning that it is among the most important sub-tasks for the decision, but without dominating character.
Tantamount	7	Important	The simulation task will have important influence, meaning that it is among the important sub-tasks for the decision, but not among the most-important sub-tasks.
	6	Significant	The simulation task will have significant influence, meaning that it will be taken into account for the decision with significant importance compared to other sub-tasks in the decision-making process, without belonging to the important sub-tasks.
	5	Evident	The simulation task will have evident influence, meaning that it will be taken into account for the decision without significant importance compared to other sub-tasks in the decision-making process.
Supporting	4	Considerable	The simulation task will only have supporting character, but has considerable supporting influence, meaning that it can support the other sub-tasks in the decision-making process as one of the most important supporting tasks.
	3	Notable	The simulation task will only have supporting character and has notable supporting influence, meaning that it can support the other sub-tasks in the decision-making process as one of the moderate important supporting tasks.
	2	Negligible	The simulation task will only have supporting character and it has negligible supporting influence, meaning that it can support the other sub-tasks in the decision-making process as one of the less important supporting tasks.
	1	Insignificant	The decision will be made exclusively based on results from other tasks. The simulation task will only have confirming character.

Figure 2: UPSIM Criticality Indicator Determination Scales (vgl. [3])

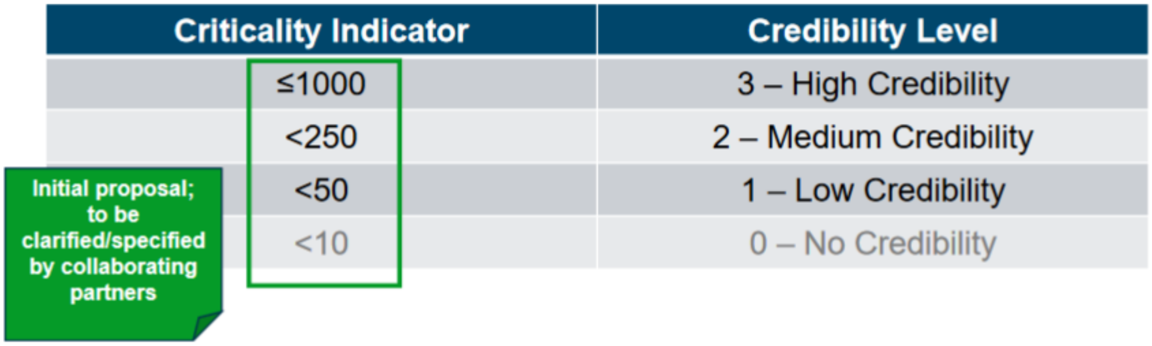

Figure 3: Determination of the Credibility Level's Criticality Indicator (cf. [3])

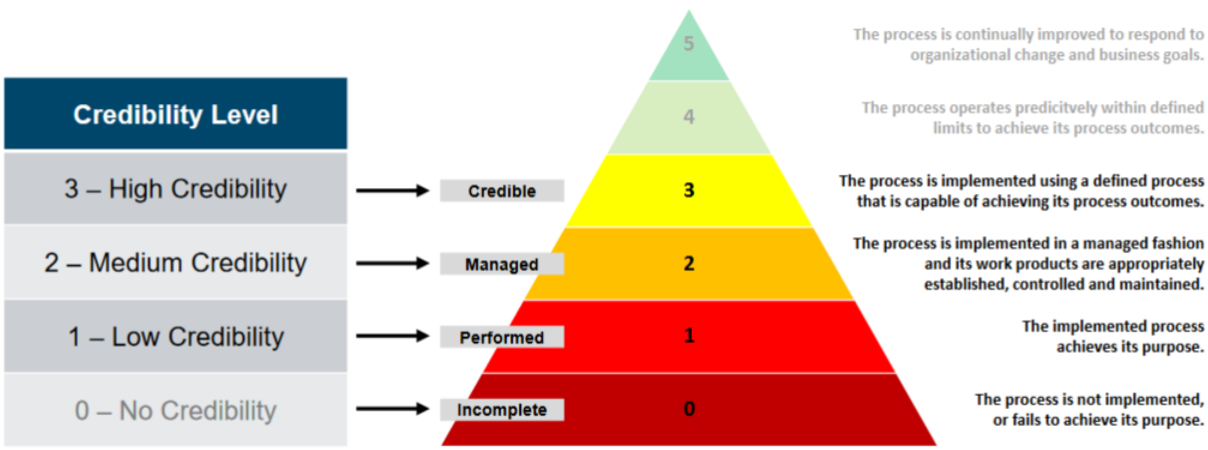

Figure 4: M&S SCPICE Process Capability Levels (ISO 2019; cf. [3])

6.3 Statistical Methods & Tools

In order to enable a quality control process for simulations, appropriate tools are required that can be used automatically if possible. These are based on statistical methods or the calculation of corresponding KPIs. The following methods can be considered in the R4F project:

- Comparison with KPIs (see e.g. [6] for an example from the automotive sector that could also be used in the rail sector)
- Sensitivity analysis (e.g. with the open tool Persalys [4]) to test the plausibility of models, especially of neural networks.
- Bayesian Model Selection to compare competing models. (See e.g. [5])

All of these methods can be implemented as software, e.g. in Python, and processed in a post-processing step. This makes it possible to apply them automatically.

At this point, however, it must also be noted that the determination and calculation of KPIs from simulation data in particular is a separate subject of research and therefore a general approach cannot be given. Apart from a few exemplary cases, it is therefore not possible to deal with the problem in general within the scope of this report, as KPIs are always dependent on the domain to be applied and often require their own theory to determine a good measure. Therefore, as described in the process above, the determination of credibility should always be accompanied by the definition of corresponding KPIs for each (sub)model to be developed. Depending on the problem, this may also require time-consuming calculations.

7 Summary

In this report, an ISO standardized process is presented that can also be used analogously for the evaluation of simulations in the context of R4F. Furthermore, some common methods for the evaluation of simulation were listed and discussed.

8 References

[1] SetLevel Credible Simulation Process, URL (abgerufen am 24.05.2024): https://setlevel.de/assets/forschungsergebnisse/Credible-Simulation-Process.pdf

[2] Ahmann, Maurizio, et al. "Towards Continuous Simulation Credibility Assessment." *Modelica Conferences*. 2022.

[3] Eichenseer, Frank, et al. "Modeling & Simulation SPICE: Assessing the Capability of Credible Simulation Processes." *INCOSE International Symposium*. Vol. 33. No. 1. 2023.

[4] Julien, Claire-Eleuthèriane Gerrer1 Hubert Blervaque, and Schueller1 Daniel Bouskela Sylvain. "Analysis and reduction of models using Persalys."

[5] Ando, Tomohiro. *Bayesian model selection and statistical modeling*. CRC Press, 2010.

[6] de Winkel, Ksander N., et al. "Standards for passenger comfort in automated vehicles: Acceleration and jerk." *Applied Ergonomics* 106 (2023): 103881.