Criteria for the Analysis of Gaps and Limitations of V&V Methods for Safety- and Security-Critical Systems

Enrico Ferrari[1], Rupert Schlick[2], Jose Luis de la Vara[3](✉), Peter Folkesson[4], and Behrooz Sangchoolie[4]

[1] Rulex Innovation Labs, Genoa, Italy
enrico.ferrari@rulex.ai
[2] Austrian Institute of Technology, Vienna, Austria
rupert.schlick@ait.ac.at
[3] Universidad de Castilla-La Mancha, Albacete, Spain
joseluis.delavara@uclm.es
[4] RISE Research Institutes of Sweden, Gothenburg, Sweden
{peter.folkesson,behrooz.sangchoolie}@ri.se

Abstract. As society increasingly relies on safety- and security- critical systems, the need for confirming their dependability becomes essential. Adequate V&V (verification and validation) methods must be employed, e.g., for system testing. When selecting and using the methods, it is important to analyze their possible gaps and limitations, such as scalability issues. However, and as we have experienced, common, explicitly defined criteria are seldom used for such analyses. This results in analyses that consider different aspects and to a different extent, hindering their comparison and thus the comparison of the V&V methods. As a solution, we present a set of criteria for the analysis of gaps and limitations of V&V methods for safety- and security-critical systems. The criteria have been identified in the scope of the VALU3S project. Sixty-two people from 33 organizations agreed upon the use of nine criteria: functionality, accuracy, scalability, deployment, learning curve, automation, reference environment, cost, and standards. Their use led to more homogeneous and more detailed analyses when compared to similar previous efforts. We argue that the proposed criteria can be helpful to others when having to deal with similar activities.

Keywords: Verification & Validation · V&V method · Gaps · Limitations · Analysis criteria · Safety-critical systems · Security-critical systems

1 Introduction

Safety- and security-critical systems such as industrial robots and connected vehicles with advanced driving support play a major role in society. They support many daily-life activities and we strongly rely on them. On the other hand, as the use and complexity of these systems are increasing, system manufacturers and component suppliers

M. Trapp et al. (Eds.): SAFECOMP 2022 Workshops, LNCS 13415, pp. 35–46, 2022.
https://doi.org/10.1007/978-3-031-14862-0_9

require methods that help them to confirm that safety, cybersecurity, and privacy (SCP) requirements are satisfied [8], i.e., V&V (verification and validation) methods.

V&V can be defined as the process of determining whether the requirements for a system or component are complete and correct, the products of each development phase fulfill the requirements or conditions imposed by the previous phase, and the final system or component complies with specified requirements [15]. This is necessary so that a safety- and security-critical system can be deemed dependable. From a general perspective, a method corresponds to a particular procedure for accomplishing or approaching something, especially a systematic or established one [24]. In this paper, we focus on methods for V&V of safety- and security-critical systems. Examples of these methods are fault injection [21] and model-based testing [17].

The new as well as automated features of safety- and security-critical systems, such as AI-based recognition, require that dedicated V&V methods are applied to them [8]. The methods must consider how to cope with the scale and complexity of the systems as well as their high level of inter-connection.

In addition, it is important to analyze if the methods present some gaps or limitations, such as scalability issues or high cost. However, the analysis and later selection and use of the methods are not always as systematic and thorough as they could. As we have experienced in various collaborative projects, e.g. [4, 14, 22], analysis criteria are usually not explicitly agreed upon, defined, and thus applied. This results in analyses that vary in scope and depth, ultimately hindering result comparison and comparison of V&V methods.

As a solution, we present a set of criteria for the analysis of gaps and limitations of V&V methods for safety- and security-critical systems. By gap we refer to an unexplored idea, i.e., some feature that nobody has ever implemented or studied in a V&V method. For example, nobody might have ever determined the extent to which a new method is applicable for compliance with a safety standard. On the other hand, by limitation we refer to a constraint that reduces the applicability of a method in certain situations. For example, formal verification methods often suffer from scalability issues. We consider both gaps and limitations in the scope of the V&V methods.

The criteria for the analysis of gaps and limitations have been identified in the context of VALU3S (Verification and Validation of Automated Systems' Safety and Security) [1], a large-scale industry-academia project that aims to evaluate and improve state-of-the-art V&V methods and tools. Thirty-three organizations agreed upon the use of nine criteria: functionality, accuracy, scalability, deployment, learning curve, automation, reference environment, cost, and standards. Sixty-two people used the criteria to analyze 53 V&V methods, resulting in a more homogeneous and more detailed analyses when compared to similar previous efforts.

The result of this analysis is a comprehensive and sound classification of gaps and limitations that can help researchers in the field to better identify weak points in V&V methods and therefore to improve the techniques in a more precise way. This can support also large projects such as VALU3S in a uniform evaluation of reference, existing results, as well as of the obtained ones.

The rest of the paper is organized as follows. Section 2 presents the background of the paper. Section 3 describes the criteria for the analysis of gaps and limitations of

V&V methods for safety- and security-critical systems, whereas Sect. 4 describes the application of the criteria. Section 5 presents our main conclusions and future work.

2 Background

2.1 The VALU3S Project

Manufacturers of automated safety- and security-critical systems and of their components have been allocating an enormous amount of time and effort developing and conducting research on these systems. They need to make sure that the systems function in the intended way and according to specifications, which is not a trivial task. For example, system and thus V&V complexity rises dramatically the more integrated and interconnected these systems become with the addition of automated functionality and features to them. This also translates into an overhead on the V&V process, making it time-consuming and costly.

Within this context, the VALU3S project [1] aims to evaluate state-of-the-art V&V methods and tools, improve them, and design a multi-domain framework that provides a clear structure around the components and elements needed for V&V [2]. The main expected benefit is to reduce the time and cost needed for V&V of safety- and security-critical systems with respect to SCP requirements. This is done through identification, classification, and development of evaluation methods, tools, environments, and concepts for system V&V with respect to the mentioned requirements.

The consortium of VALU3S consists of partners from 10 different European countries, including 25 industrial partners, six research institutes, and 10 universities. Thirteen use cases with SCP requirements are studied in detail from six domains: aerospace, agriculture, automotive, healthcare, industrial robotics/automation, and railway.

One of the first tasks of the project dealt with the review of state-of-the-art and state-of-the-practice V&V methods [29]. The methods were planned to be applied in the project use cases with the intention to improve how SCP requirements were addressed, ensured, and confirmed. Fifty-three methods were reviewed and classified according to the categories shown in Fig. 1. A line between two categories indicates that some relationship was identified between V&V methods of the corresponding categories. The methods were then studied in more detail to identify their main gaps and limitations [30]. This paper presents the criteria selected for such an analysis of gaps and limitations, as well as its outcome. Next, effort was spent on addressing the gaps and limitations, thus on improving the methods [31].

2.2 Related Work

The criteria for the analysis of gaps and limitations of V&V methods for safety- and security-critical systems were referred to in a prior publication on workflow modelling by VALU3S partners [7]. However, the definition of the criteria, information about how they were applied, and application examples were not provided.

The selection and use of criteria to validate and evaluate V&V methods has been widely addressed. For example, new methods must show that they fulfil certain characteristics so that they can be regarded as effective and efficient V&V means. Examples of

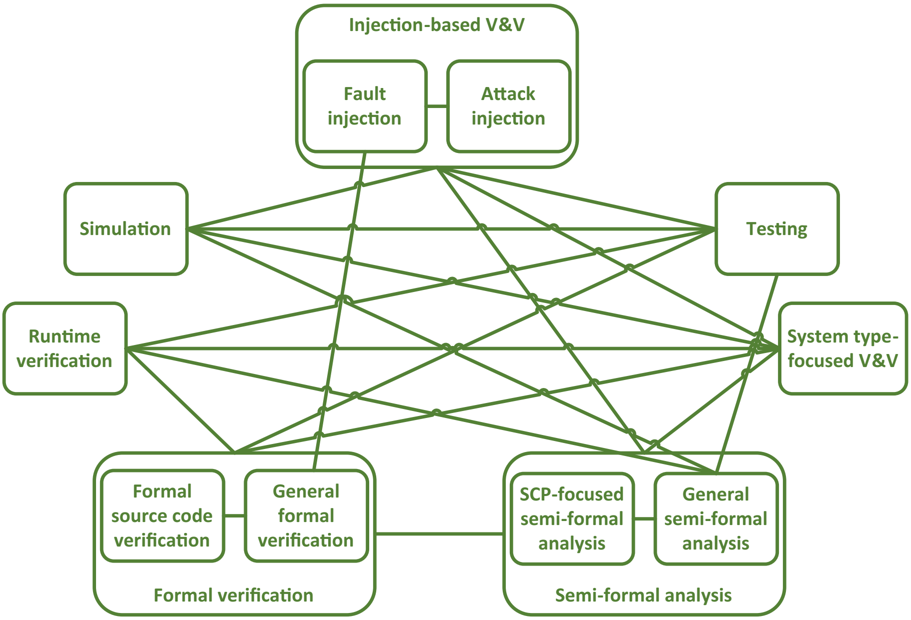

Fig. 1. Categories of V&V methods analyzed in VALU3S

such criteria include accuracy [20], cost [25], and scalability [18], among others. It is easy to find publications that have used our criteria for the analysis of gaps and limitations, or similar criteria, when studying and discussing V&V methods. What distinguishes this paper from this kind of prior publications is that: (1) we present a set with nine different criteria, not only one or a few; (2) we explicitly focus on the analysis of gaps and limitations as a way to identify improvements; and (3) we present criteria selected and used for analyses of tens of V&V methods in the scope of a large collaborative effort between industry and academia. The latest requires an agreed, clear definition of the criteria so that they are homogeneously applied. In addition, publications that have paid attention to a reduced set of criteria have resulted in narrower analyses.

Other publications have proposed criteria and metrics for characterization and evaluation of engineering products and processes, e.g., software ones [10, 16], as well as of V&V methods, e.g., [19, 23, 27]. On the one hand, the depth of these pieces of work is different, as we do not deal with definition of detailed metrics and measurement procedures for the different criteria presented. The analyses in VALU3S have mostly been qualitative. Quantitative information was based on prior studies. On the other hand, the breadth of our set of criteria is larger, as we consider further criteria.

Regarding prior research projects on V&V, it is common that they reviewed V&V methods, typically at the beginning of the projects, e.g., [3, 6]. It is also common that the projects evaluated the methods developed at later stages, e.g., [5, 9]. For the first effort, we are not aware of any other large-scale industry-academia project that has explicitly defined and agreed on a set of criteria so that all the partners use them. For the latter, the differences with this paper are the same as those publications referred in the previous paragraph: different depth and different breadth.

In summary, prior work has proposed and used criteria for evaluation of V&V methods. However, its scope and breadth do not match the needs identified in VALU3S for analysis of gaps and limitations, thus the needs of similar efforts.

3 Criteria for the Analysis of Gaps and Limitations of V&V Methods

To better identify gaps and limitations of V&V methods, we propose a set of criteria that aims to cover the main aspects to consider, ranging from more functional issues to more operational ones. The criteria are proposed to guide further development in V&V methods. There could be overlaps between criteria, but we decided to leave a loose boundary between them in order to make them applicable to different methods.

The nine criteria proposed for analysis of gaps and limitations are defined below. The examples provided have been extracted from the deliverable of the VALU3S project that reports the analysis of gaps and limitations conducted [30].

1. **Functionality**, in relation to the capabilities and features of a V&V method, and to its practicality in general, as well as to the range of V&V activities that it supports. It could be determined that a method could better address some V&V need. Note that, this evaluation is done considering what is in the scope of the method: a functionality unrelated to the method should not be considered a gap.

   Example: *Test Oracle Observation at Runtime* only performs analyses on an individual behaviour and it is not exhaustive.
2. **Accuracy**, regarding whether the outcome from using a V&V method conforms to an expected correct value or a quality level, corresponding to a qualitative assessment of correctness or freedom from error. It is important that a method is reliable enough, especially when employed for some critical application. Notice that this includes both the accuracy of a single instance (i.e., could we trust that if a method says that a situation is safe, it is actually safe) and the statistical accuracy (i.e., how far estimates of V&V are from the actual values).

   Example: *Behaviour-Driven Formal Model Development* requires input from both formal methods and domain experts, so its accuracy depends on the quality of the communication and understanding between them.
3. **Scalability**, to analyse if a V&V method requires too many computational resources (time or memory) and therefore can be applied only to limited or simplified scenarios. This narrows the applicability of a method.

   Example: High accuracy of *CPU Verification* demands high computation power and limits the testing capacity for software.
4. **Deployment**, for consideration of possible problems of a V&V method when put into operation in real-world contexts. For instance, issues in integrating a method with others and a lack of proper tool support can negatively impact deployment.

   Example: *Human Interaction Safety Analysis* currently lacks tool support for efficient use.

5. **Learning curve**, to evaluate the expected progress by a person in gaining experience or new skills to successfully use a V&V method. For cost-effective use, a method might require high-level technical knowledge or skills.

   Example: *Fault Injection in FPGAs* requires comprehensive knowledge in FPGAs.
6. **Automation**, regarding the degree of automatic operation of a V&V method. For instance, if a method is not fully automated and requires large intervention, such as tuning by human users, the corresponding V&V process could become long and error prone. When dealing with large-scale complex systems, automatability of the whole V&V process could become a key feature.

   Example: For *Model-Based Mutation Testing*, building test models is not easily automatable.
7. **Reference environment**, to assess whether a V&V method works only for some settings and conditions, e.g., in a simulated environment. As a consequence, there might be no warranties that findings are still valid for other environments, e.g., for operation in the real world.

   Example: *Vulnerability and Attack Injection* requires a prototype or a real system.
8. **Cost**, considering if using a V&V method requires significant investments in terms of, e.g., hardware, software, time, or human resources. In general, the lower the costs, the better.

   Example: *Simulation-Based Robot Verification* may require a large amount of hardware resources depending on the number of tests to be performed.
9. **Standards**, which indicate the extent to which aspects related to regulations and standards, and the compliance with them, are taken into consideration by a V&V method. This is especially important for critical domains such as avionics and railway, in which compliance with standards and certification are required so that a system is allowed to operate. It can also be determined that a method is against some recommendation in a standard. For example, Fault correction with artificial intelligence is not recommended for railway (EN 50128 standard).

   Example: No explicit and direct link with compliance has been established between *Knowledge-Centric System Artefact Quality Analysis* and most assurance/engineering standards.

4 Application of the Criteria for the Analysis of Gaps and Limitations of V&V Methods

The criteria defined in Sect. 3 have been applied for identifying the gaps and limitations in the set of methods introduced in [8]. This contributes to the validation of the criteria. The activity was performed within the VALU3S project by 62 people, who analyzed 53 V&V methods adopting the criteria. In this section, two different analyses are reported. In Subsect. 4.1, two methods are analyzed in detail in order to better show the criteria and how they could be applied. In Sect. 4.2, the main results obtained applying the criteria to all the methods are synthetically reported.

4.1 Application of the Criteria to Two Methods

The results of the application of the criteria to two methods are reported in detail in this section. The two methods have been selected to ensure that various types of methods are covered and that a wide range of criteria is used. The first method (Model-implemented Fault Injection) is an experimental one, i.e., it is based on the observation of the behaviour of the system under certain circumstances. On the other hand, the second method (Knowledge-Centric System Artefact Quality Analysis) is an analytical one, since it is based on the analysis of defined metrics that does not require system execution.

In **Model-implemented Fault Injection**, separate blocks modelling faults are injected into a model of the System Under Test (SUT) [28]. MATLAB and LabVIEW are examples of tools used to build such system models. This method is used to verify and validate the system's capability to handle faults. The fault handling includes attributes such as fault detection, correction, or fallback with or without the fault handling mechanisms implemented. This type of fault injection method is used for a system's evaluation at early design stages [11].

- *Functionality.* (i) The method can be improved by adding techniques such as pre-injection analysis and post-injection analysis [12] to reduce the number of the tests and still get the same or improved results in terms of time, cost, and effort. Pre-injection analysis is done before any fault injection experiments are performed while post-injection uses the results of previous fault injection experiments. (ii) Adding more fault models will increase the functionality of the method.
- *Accuracy.* The accuracy of the method depends on the accuracy of the modelled faults and systems. Since the model of the system might not accurately represent the real system in a real environment, supplemental V&V activities (e.g., acceptance tests) are recommended to be performed at later development stages.
- *Scalability.* Exhaustive fault injection or full system monitoring may require a lot of computational resources depending on the complexity of the target system and its environment.
- *Deployment.* (i) The model-implemented fault injection method is not feasible for final implementations of systems. (ii) The method must be adapted to the simulation tool environment used, e.g., MATLAB toolboxes and MATLAB versions used.
- *Learning curve.* The method requires knowledge and skills regarding the simulation tool environment, e.g., MATLAB/SIMULINK skills.
- *Automation.* The configuration of fault injection campaigns and result analysis are done manually.
- *Reference environment.* This method is only applicable for the simulation environment.
- *Costs.* (i) Software such as MATLAB/SIMULINK is not open source and needs investments. (ii) There is also some cost involved in terms of time when conducting model implemented fault injection. For example, exhaustive fault injection or full system monitoring increases V&V cost.
- *Standards.* No relevant gap or limitation has been identified. Examples of standards including requirements which this method may fulfil are ISO 26262, IEC 62061, ISO 13849, and IEC 61508.

Knowledge-Centric System Artefact Quality Analysis is a method to assess the quality of systems artefacts, such as textual requirements specifications and system models, by exploiting knowledge bases, e.g., an ontology [26]. The assessment is quantitative according to different artefact characteristics (correctness, consistency, and completeness) and to different metrics (e.g., based on the number of elements with a given property in an artefact, such as the number of vague words in a requirement).

- *Functionality*. The amount of model-specific quality analysis means is currently limited. Most of the available support focuses on textual requirements.
- *Accuracy*. A detailed study of quality analysis accuracy has not been conducted.
- *Scalability*. Issues can arise with large and complex system artefacts. Tool solutions have nonetheless been developed to mitigate it.
- *Deployment*. Connectors with the system artefact sources are required, i.e., means to connect with tools (for requirements management, system modelling, etc.) or files to get system artefact data.
- *Learning curve*. There is a barrier in the need for knowing how to create and properly manage ontologies.
- *Automation*. Creation and management of ontologies, as well as of connectors to system artefact sources, would benefit from automation support.
- *Costs*. Creation and management of ontologies is mostly a manual effort that can require significant time.
- *Standards*. No explicit and direct link with compliance has been established for most assurance/engineering standards. Nonetheless, the method (i) has been applied for many systems under regulatory requirements, and (ii) supports INCOSE rules for writing requirements [13], among other reference documents.

4.2 Application Results

The analysis of 53 V&V methods led to identification of 400 gaps and limitations, which corresponds to about 7.5 gaps or limitations for each method. The methods were selected according to the needs and challenges of VALU3S industrial use cases.

Table 1 shows the number of gaps for each criterion, which ranges from 16 to 69. The criterion with more gaps or limitations is Functionality; this is understandable since many functionalities could be added to each method. For the other criteria, on average, about one gap or limitation has been defined for each method, except Standards and Reference Environment, which have less than 0.5 gaps/limitations for each method. This limited number is probably due to the fact that Reference Environment and Standards are as relevant for some methods as for others. For example, some methods could be applied in contexts where no clear standard is defined or they could be natively defined in a reference environment, so no issue regarding this is envisaged.

Regarding the number of gaps for each method, the ones with the highest number of gaps/limitations are Model-based Testing and Penetration Testing, with 15 gaps/limitations each.

Referring to the categories of methods illustrated in Fig. 1, the average number of gaps/limitations per method is reported in Table 2. The category with the highest number of gaps/limitations is Attack Injection, with about 11.3 for each method. On the other side, only 5.8 gaps/limitations were found on average for Testing methods.

Figure 2 shows the average number of gaps/limitations for each criterion and method category. A detailed analysis of this plot can help to qualitatively understand which types of gaps are more frequent in each category. For example, for the Testing methods, a higher number of gaps or limitations connected to Accuracy have been identified, while fewer limitations were pointed out as regards to Functionality. Understanding if some methods share the same type of gaps and limitations and which type of gaps/limitations are more frequent could clarify the direction where the improvements should be carried out.

Table 1. Number of gaps and limitations for each criterion.

Gap/limitation criterion	# of gaps & limitations
Functionality	69
Accuracy	56
Scalability	49
Deployment	47
Learning curve	45
Reference environment	25
Costs	50
Automation	43
Standards	16
Total	400

Table 2. Average number of gaps and limitations for each category of V&V methods.

Gap/limitation criterion	Average # of gaps & limitations
Attack injection	11.3
Fault injection	8.1
Simulation	7.8
Testing	5.8
Runtime verification	8.3
Formal source code verification	9.0
General formal verification	6.4
SCP-focused semi-formal analysis	6.9
General semi-formal analysis	7.0
System-type-focused V&V	10.0

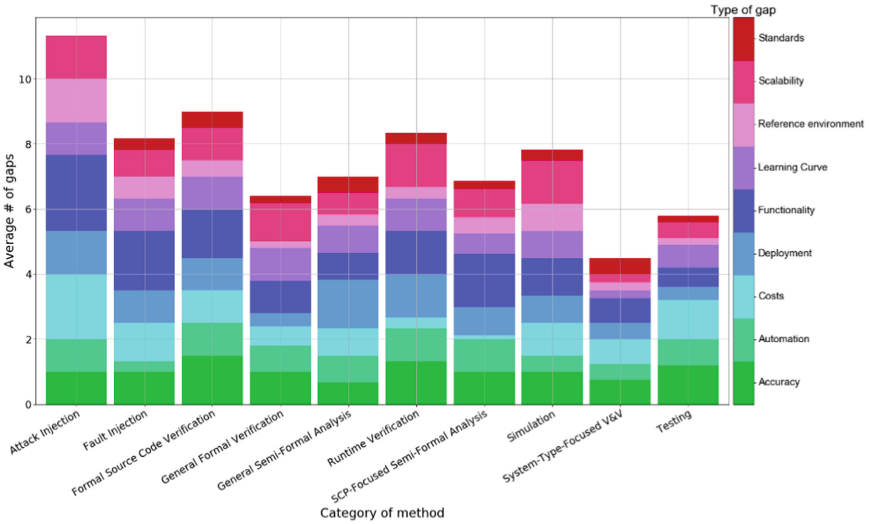

Fig. 2. Average number of gaps and limitations for each criterion and category of methods.

4.3 Discussion

This section complements the insights provided in the previous section by discussing some relevant general aspects of the criteria for analysis of gaps and limitations of V&V methods and of their application.

First of all, the criteria were agreed upon and selected in the scope of a given project, VALU3S, with a specific purpose and specific needs. The criteria fit these purpose and needs, such as the identification of improvement opportunities in V&V methods for safety- and security-critical systems, focusing SCP requirements, to later realize the corresponding improvements. Although other efforts and projects might differ in scope and purpose, we are still confident that the criteria presented are general enough to apply to many similar situations. The overall characteristics to assess of a V&V method are the same regardless their application context, although characteristics some might be more relevant than others. For example, standards can be regarded as a more important criterion for safety-critical systems than for other system types.

Regarding the completeness of the criteria, we acknowledge that other researchers might decide that different criteria need to be considered or that the set of criteria proposed needs to be refined. For instance, and considering again the example of safety-critical systems, it could be valuable to explicitly consider tool qualification aspects when analyzing the automation and standards, or as a complement to them. It is important to ensure that, e.g., tools work as expected, identify errors effectively, and do not introduce errors according to the requirements of safety standards. Our main argument is that the set of criteria has been suitable for VALU3S, thus we consider that it could also be for others. Indeed, we plan to propose the use of the criteria in similar future efforts, such as new large-scale industry-academia projects that plan to review V&V methods.

As indicated above, one of the main benefits that we have found in the definition of an explicit and agreed set of criteria for analysis of gaps and limitations is that it has resulted in a more homogeneous assessment of V&V methods. In other projects, we have experienced that different people considered different aspects when reviewing V&V methods and to a different extent, leading to a more difficult comparison and more limited identification of improvement opportunities. A simple broad, but still precise, comparison such as the one shown in Fig. 2 could not be provided for the prior projects in which we have been involved.

5 Conclusion

As we increasingly rely on safety- and security-critical systems, it is essential that their dependability is confirmed by using adequate V&V methods. To this end, the possible gaps and limitations of the methods must be analyzed. Explicitly defined criteria can aid in making these analyses more precise, homogeneous, and comparable.

In this paper, nine criteria for classifying gaps and limitations have been proposed: functionality, accuracy, scalability, deployment, learning curve, automation, reference environment, cost, and standards. The criteria have been applied in the VALU3S project on a set of 53 V&V methods belonging to different categories and application fields. The outcome helps in addressing the efforts in improving available V&V methods. We also consider that the use of the criteria aids in obtaining more homogeneous analyses of gaps and limitation of V&V methods, also contributing to a more accurate comparison of V&V methods and method types.

As future work, the criteria will guide the development of new V&V methods to overcome existing gaps and limitations, as well as the improvement of existing methods. Moreover, the impact of the use of criteria for addressing current gaps and limitations is planned to be evaluated.

Acknowledgments. The research leading to this paper has received funding from the VALU3S (H2020-ECSEL grant agreement no 876852; MCIN/AEI ref. PCI2020-112001; NextGen.EU/PRTR), iRel4.0 (H2020-ECSEL grant agreement no 876659; MCIN/AEI ref. PCI2020-112240; NextGen.EU/PRTR), ETHEREAL (MICINN/AEI ref. PID2020-115220RB-C21; ERDF), and Treasure (JCCM SBPLY/19/180501/ 000270; ERDF) projects, and from the Ramon y Cajal Program (MICINN RYC-2017-22836; ESF). We are also grateful to all the VALU3S partners that have provided input and feedback for the selection of the criteria and that applied them.

References

1. Agirre, J., et al.: The VALU3S ECSEL project: verification and validation of automated systems safety and security. Microprocess. Microsyst. **87**, 104349 (2021)
2. Aguirre, J., et al.: Multidimensional framework for characterizing verification and validation of automated systems. In: EDCC (2022)
3. Amalthea4public project: D3.1 - Analysis of state of the art V&V techniques (2015)
4. AMASS project: https://cordis.europa.eu/project/id/692474
5. AMASS project: D1.7 - AMASS solution benchmarking (2019)
6. AMASS project: D3.1 - Baseline and requirements for architecture-driven assurance (2018)

7. Bauer, T., et al.: Cross-domain modelling of verification and validation workflows in the large scale European research project VALU3S. In: Orailoglu, A., Jung, M., Reichenbach, M. (eds) Embedded Computer Systems: Architectures, Modeling, and Simulation. SAMOS 2021. LNCS, vol. 13227. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-04580-6_25
8. de la Vara, J.L., et al.: A proposal for the classification of methods for verification and validation of safety, cybersecurity, and privacy of automated systems. In: QUATIC (2021)
9. ElasTest project: D7.3 - Public demonstrator artifacts (2019)
10. Fenton, N.E., Pfleeger, S.L.: Software Metrics - A Rigorous and Practical Approach, 3rd edn. CRC Press, Boca Raton (2015)
11. Folkesson, P., Ayatolahi, F., Sangchoolie, B., Vinter, J., Islam, M., Karlsson, J.: Back-to-back fault injection testing in model-based development. In: Koornneef, F., van Gulijk, C. (eds.) SAFECOMP 2015. LNCS, vol. 9337, pp. 135–148. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24255-2_11
12. Grinschgl, J., et al.: Efficient fault emulation using automatic pre-injection memory access analysis. In: 2012 IEEE International SOC Conference, pp. 277–282. Niagara Falls, NY (2012)
13. INCOSE: Guide for Writing Requirements (2019)
14. iRel40 project: https://cordis.europa.eu/project/id/876659
15. ISO: ISO/IEC/IEEE 24765: Systems and software engineering - Vocabulary (2017)
16. Kan, S.: Metrics and Models in Software Quality Engineering, 2nd edn. Addison Wesley, Boston (2002)
17. Kramer, A., Legeard, B.: Model-Based Testing Essentials-Guide to the ISTQB Certified Model-Based Tester: Foundation Level. Wiley, Hoboken (2016)
18. Ma, T., Ali, S., Yue, T.: Testing self-healing cyber-physical systems under uncertainty with reinforcement learning: an empirical study. Empir. Softw. Eng. **26**(3), 1–54 (2021). https://doi.org/10.1007/s10664-021-09941-z
19. Miller, L., et al.: Guidelines for the verification and validation of expert system software and conventional software. US Nuclear Regulatory Commission (1995)
20. Moreno, V., Génova, G., Parra, E., Fraga, A.: Application of machine learning techniques to the flexible assessment and improvement of requirements quality. Softw. Qual. J. **28**(4), 1645–1674 (2020). https://doi.org/10.1007/s11219-020-09511-4
21. Natella, R., et al.: Assessing dependability with software fault injection: a survey. ACM Comput. Surv. **48**(3), 44 (2016)
22. OPENCOSS project: https://cordis.europa.eu/project/id/289011
23. OPENCOSS project: D1.3 - Evaluation framework and quality metrics (2013)
24. Oxford UK Dictionary: Method. https://www.lexico.com/definition/method (2021)
25. Panesar-Walawege, R.K., et al.: Supporting the verification of compliance to safety standards via model-driven engineering: approach, tool-support and empirical validation. Inf. Softw. Technol. **55**(3), 836–864 (2013)
26. Parra, E., et al.: Advances in artefact quality analysis for safety-critical systems. In: 30th International Symposium on Software Reliability Engineering (ISSRE) (2019)
27. Roza, M.: Verification, validation and uncertainty quantification methods and techniques. NATO (2014)
28. Svenningsson, R., Vinter, J., Eriksson, H., Törngren, M.: MODIFI: a MODel-implemented fault injection tool. In: Schoitsch, E. (ed.) SAFECOMP 2010. LNCS, vol. 6351, pp. 210–222. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15651-9_16
29. VALU3S project: D3.1 - V&V methods for SCP evaluation of automated systems (2020)
30. VALU3S project: D3.3 - Identified gaps and limitations of the V&V methods listed in D3.1 (2021)
31. VALU3S project: D3.6 - Final description of methods designed to improve the V&V process (2022)