

TOWARDS A MEASUREMENT TOOL FOR VERIFICATION AND VALIDATION OF SIMULATION MODELS

Zhongshi Wang

ITIS GmbH
Werner-Heisenberg-Weg 39
D-85577 Neubiberg, GERMANY

ABSTRACT

Model deficiencies, despite their negative influences on assessment of modeling and simulation (M&S) applications, carry a large amount of insightful information, which can be used to measure different aspects of the M&S development process and its verification and validation (V&V). Although there already exist various categorizations of model deficiencies, none of which can be used as a measurement tool to classify and analyze deficiency data collected in practice. This paper describes a framework for developing model deficiency classifications and the pilot application of the established classification scheme as a quantitative method to measure and control the V&V process being conducted. Investigations indicate that the proposed approach is capable of providing process diagnoses reflecting the real problems and identifying the improvement potentials. Based on the findings achieved in practice, a planning and tailoring concept is being developed for efficiently applying different V&V techniques in a simulation study.

1 INTRODUCTION

Development of modeling and simulation (M&S) applications involves human- and knowledge-intensive activities, in which errors, uncertainties, or inadequacies leading to quality deficiencies in models and simulation results are inevitable. Verification and validation (V&V) focuses on assessing the accuracy quality characteristic of an M&S application with respect to its objectives (Balci 2004, Shannon 1975), and is intended to ensure that only correct and suitable models and simulation results are used in practice.

Model deficiencies, on the one hand, due to their negative impact on the assessment issue, should be prevented from the development's point of view, or be identified and corrected from the V&V's point of view. On the other hand, a large amount of information is attached to the model deficiencies, characterizing the different aspects of model development and its V&V. In order to extract the semantic information of interest from the collected deficiency data, classification of model deficiencies is required. However, there is currently not sufficient research in this area.

In the M&S literature, the term "errors" is frequently used to refer to model deficiencies. Its meaning is sometimes, however, extremely different, such as the definitions of type I, II, and III errors (Balci and Nance 1985), and of acknowledged or unacknowledged errors (Oberkampf et al. 2000). Typically, model deficiencies are classified into several coarse-grained categories relating to the corresponding development phases in the M&S life cycle (Shannon 1975, Carson 2002, Schmeiser 2001), such as data model errors, logic model errors, programming errors, experimentation errors, or interpretation errors. Such taxonomies cannot be used as a basic instrument to measure and analyze the semantics included in model deficiencies.

In the software community, several defect classification schemes and even an IEEE standard (IEEE 1993) have already been developed. In publications, the most referenced classification approaches are the IBM Orthogonal Defect Classification (ODC) scheme (Chilarege et al. 1992) and the Hewlett-Packard (HP) scheme (Grady 1992). The basic concept of software defect classifications can be outlined as follows:

1. defining a classification scheme with different attributes;
2. using the defined classification scheme to extract the semantic information from individual defects;
3. analyzing and evaluating the aggregated defect data.

Findings of the investigation indicate that the software defect classifications (ODC and other schemes) can not be applied to analysis of model deficiencies because of the more complex application context in M&S (Wang 2010). The general classification principles, however, are still essential for developing an M&S-specific classification scheme. Because of the huge variety of M&S applications, an overall and complete scheme is under current practice impossible. Therefore, this paper presents a classification framework with four different attributes to capture the information used to measure and analyze the V&V process being conducted. This classification framework can be applied to any simulation study with structured model development and V&V processes.

The purpose of this paper is threefold: (1) to describe the development of the framework; (2) to illustrate the use of this framework to establish a classification scheme for analyzing model deficiencies; and (3) to demonstrate the application of the established classification scheme in a pilot project and investigate the findings.

2 RELATED WORK IN SOFTWARE DEFECT CLASSIFICATIONS

2.1 Orthogonal Defect Classification (ODC)

The IBM ODC scheme (Chilarege et al. 1992) is one of the most prominent measurement methodologies used to classify and analyze software defects. An essential criterion of this scheme is to categorize a software defect into distinct classes characterized by an orthogonal set of attributes. The term “orthogonal” means that both attributes and their values do not overlap and are mutually independent. This non-redundant nature not only reduces the human error in classification, but also enables the capture of unique and unambiguous information of interest for evaluating the current development and testing process.

2.1.1 ODC Attributes

In the ODC scheme, the following eight different attributes are defined (Chilarege et al. 1992; Butcher, Munro, and Kratschmer 2002):

- Activity refers to the testing activity being performed at the time the defect was discovered, i.e., when was the defect found?
- Trigger represents the environment or condition (during the testing activity) that had to exist for the defect to surface, i.e., how was the defect discovered?
- Impact describes the effect on the customer if the defect had not been found, i.e., nature and degree of pain.
- Target indicates the high-level identity of the entity that was fixed, i.e., what gets fixed?
- Defect Type specifies the fix that was made to resolve the defect, i.e., what had to be fixed?
- Qualifier captures the element of either a nonexistent or wrong or irrelevant implementation.
- Source identifies which part of the work product the defect target belongs to, e.g., in-house, out-sourced, library reuse and so on.
- Age shows the developmental history of the defect target, e.g., new, old, rewritten, or refixed.

More detailed description of the eight attributes and their values defined in the ODC scheme can be found in the ODC website (IBM Research 2010).

2.1.2 Analysis of Classified Data

While all of the ODC attributes capture the meaningful information from a defect and therefore are valuable to analyze, the full scope of the data is not always necessary for practical use (Chillarege 2007). Depending on the objective of an analysis, the data of one certain attribute alone or in combination with other defect attributes can be classified and applied.

When alone, the trend of the defect type distribution can be used as measurement to estimate the progress of the software product through the development process. For example, the function defects should be found typically in the early phase like, design inspection, and decreasingly in the late phases such as unit test, integration test, and system test. On the contrary, the timing and serialization defects should be found mainly at system test. Thus, when compared with other defect types, a decreasing trend of the function defects should be observed from phase to phase. If however, the percentage of the function defects from a current development is lower in the early phase and much higher in the testing phases than it should be, such a distribution deviation rather indicates an inadequate design inspection. In this context, a repeated process for design review seems more effective than the intensive testing (Chillarege et al. 1992).

On the other hand, if some defect attributes are applied in combination, then an analysis of multiple perspectives is possible. As defined in the ODC scheme, the defect type attribute addresses the semantics of defect correction, while the defect triggers are referred to the catalysts which allow the defect to surface, namely the facilitators that force a fault to a failure (Chillarege and Bassin 1995, Chillarege and Prasad 2002). The attribute trigger provides insight on the defect detection process, indicating how defects are revealed. Thus, when a relationship between defect types, triggers, and activities is established, it is possible to investigate what kind of detection technique finds which type of defects for each detection activity. In practice, applications of ODC for different purposes are reported, such as for enhancing the root cause analysis, quality management, and process improvement (Schulz 1999, Chillarege 2007).

2.2 Other Software Defect Classifications

The Hewlett-Packard (HP) approach (Grady 1992) is another industrial classification scheme applied frequently in practice. Compared to the ODC scheme, only three attributes: Origin, Type, and Mode are defined in the HP approach. The defect origin indicates the software development phase, in which the defect was introduced (not where it was found). Unlike ODC, the HP defect type is referred to the particular area of an origin, which is responsible for the defect. In addition, the mode attribute describes the reason of a defect, i.e., whether the information was missing, unclear, or wrong etc.

Thus, by means of the close relationship between the three defect attributes, the HP scheme focuses rather on the process improvement, namely identifying the error-prone areas for future development or testing, based on the analysis of the associated defect origins and types (Huber 2000). However, the issue on how to detect defects such as the activities and triggers in ODC is not an emphasis discussed in the HP scheme.

The focus of the IEEE Standard Classification for Software Anomalies (IEEE 1993) is on providing a general classification process to establish a defect classification scheme for the needs of users. The process is consisted of four sequential steps: Recognition, Investigation, Action, and Disposition. According to these process steps, a classification scheme is categorized into several dimensions which contain the different attributes characterizing the work flow from the recognition of a defect through to its closure. For example, the attributes activity, phase, and symptom are assigned to Recognition, the defect cause and type appear in Investigation, and the defect correction is found in Action. However, with respect to practical application, case studies on this standard classification are not sufficiently reported (Wagner 2008).

3 A CLASSIFICATION FRAMEWORK FOR ANALYZING MODEL V&V

As already discussed in (Wang 2010), software defect classifications cannot be used in the M&S context directly. This section describes the development of a classification framework for measuring model deficiencies collected by conducting V&V.

3.1 Software Defect vs. Model Deficiency

In the software community, despite some different terminologies, the relationship between the cause of a defect and its consequences is consistently specified. In general, as defined in an IEEE Standard (IEEE 1990), an error is a human oversight or wrong decision, which results in a fault, or a defect within the software (including requirements specifications, design, and code). When the software is executed, a fault, or a combination of faults, may (or may never) cause a failure. In many cases, the term “defect” is used in a generic manner referring to a fault, a failure, or even an error, such as in the ODC, and HP schemes as well as in this paper so far.

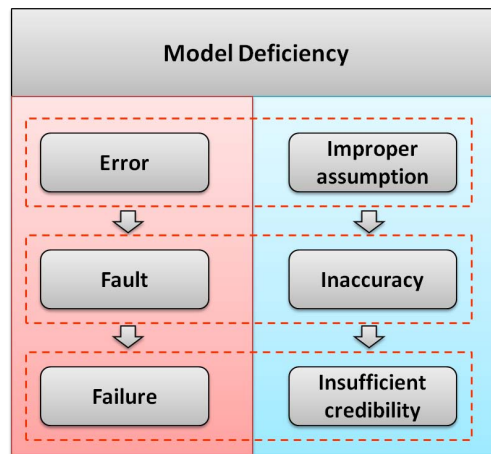


Figure 1: M&S deficiencies.

Compared to a software defect, the meaning of what is wrong or inaccurate in a model appears more complex, just as the well-known statistician George Box stated: “All models are wrong, some are useful”. From the V&V viewpoint, two sources of M&S deficiencies (shown in Figure 1) can be identified within the M&S life cycle. On the one hand, since M&S requires intensive investment of human efforts, and includes software development, the causal structure of human errors, faults, and failures also exists in simulation models.

On the other hand, due to lack of knowledge, e.g., uncertainty (Oberkampff et al. 2002) in the different M&S development phases, or for the purpose of model simplification, assumptions and approximations must be made. Whether they are accurate enough or actually acceptable can only be estimated with respect to the specified objectives of the model by particular V&V techniques such as face validation (Shannon 1975, Balci 1995, Law and Kelton 2007). Thus, inaccuracies are those deficiencies in the model, which result from improper or unacceptable assumptions and approximations made.

In some cases, although the individual assumptions and approximations are estimated as acceptable, the effect of their aggregation could still cause unacceptable inaccuracies. In the course of M&S development, improper assumptions and approximations cause further inaccuracies which can eventually lead to insufficient model credibility, i.e., the model is not useful.

Thus, both sources of M&S deficiencies must be handled by the V&V activities. In the following, the term “deficiency” is used in a generic manner referring to both defects and inaccuracies in a model.

3.2 Development of the Model Deficiency Classification

Due to the huge variety of M&S applications, and their corresponding development and V&V processes, it is under current practice not possible to develop an overall and complete classification scheme applicable to any simulation study. Therefore, the aim of this work is to propose a deficiency classification framework with definitions of the determined attributes. The individual values of each attribute are not specified in the framework. They will be determined according to the structure of the concrete M&S processes used for a simulation study in practice. For developing the classification framework, two issues are crucial:

- the main objective of the classification is to measure model V&V being conducted and provide a process diagnosis for identifying improvement potentials;
- the classification framework should be applicable to any simulation study with well defined and structured model development and V&V processes.

As shown in Figure 2 (Wang and Lehmann 2010), when assessing the V&V effort, the information about in which V&V activity a model deficiency is detected by using which V&V technique, is necessarily required. Therefore, the attributes Activity and Detection are defined in this framework.

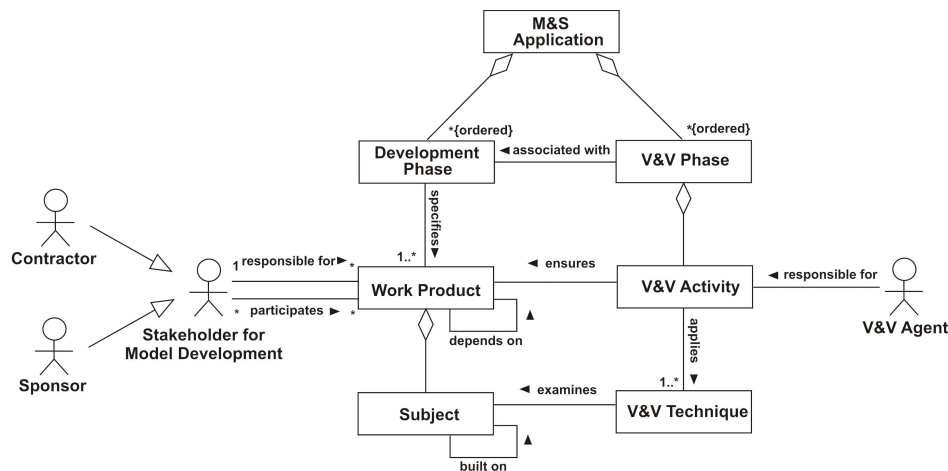


Figure 2: Structured M&S and V&V.

Furthermore, the issue of deficiency propagation is taken into account. Model deficiencies, if not detected in time, propagate through the whole M&S life cycle. The later they are detected, the more problems they cause, and the more expensive and time-consuming their diagnosis and correction become. Hence it is also important to investigate by which development activity a deficiency is introduced in the model, in addition to its detection. Consequently, the attribute Introduction is defined.

Finally, the attribute Type of a deficiency is defined to indicate which subject of the model is inaccurate or unacceptable due to the existence of the deficiency. This definition is quite different from the one of certain software defect classifications, e.g., the ODC scheme. In ODC, Defect Type is defined as what was fixed when correcting the defect. This means that (Kelly and Shepard 2001):

- the ODC defines the attribute Type from the viewpoint of the software developer or programmer;
- the ODC Type can be determined only at the time the software developer corrects the defect.

In the context of this framework, the definition of Type should however reflect the facts as a V&V agent performs his assessment of the model: Is there something incorrect with respect to the intended uses? If true, which subject of the model is affected because of the deficiency? e.g., if a deficiency causes an

incorrect (missing, or unclear) item in formulating M&S objectives for a model, the Type “M&S objective” should be assigned to the deficiency. Certainly, this judgment has been made before the deficiency is corrected by the model developer in a simulation study. Figure 3 describes the four defined attributes in more detail.

Attribute Name	Attribute Meaning	Description
Activity	When was the deficiency identified?	The V&V activity being performed at the time the deficiency was identified. The values of this attribute are the individual activities defined in the V&V process, which is integrated into the underlying model development process.
Detection	How was the deficiency identified?	The V&V technique or the combination of V&V techniques which is used to reveal the deficiency. Examples are face validation, inspections, semantic analysis, testing techniques, proof methods etc.
Introduction	What task causes the insertion of this deficiency?	The first phase of the M&S development life cycle, in which the deficiency could have been prevented. Values of this attribute are the different phases of model development, such as Problem Definition, System Analysis, Model Formalization, Model Implementation, and Model Application.
Type	What is incorrect in the model?	The subject of a model (or an intermediate product), which is no longer sufficiently accurate because of existence of the deficiency. For example, possible values for the intermediate product Structured Problem Definition (SPD) are M&S objectives, M&S requirements, M&S constraints, and Acceptability criteria.

Figure 3: The deficiency classification framework.

4 AN IMPLEMENTATION EXAMPLE

When implementing the framework in a practical context, a set of values for each deficiency attribute is to be determined on basis of the M&S and V&V processes being applied. Two issues are to be considered (Freimut 2001):

- The values for each attribute should be orthogonal (or distinct), i.e., only one attribute value is appropriate for a particular model deficiency;
- The values for each attribute should be complete. This means that each deficiency could obtain an appropriate attribute value.

In the following, as a guidance on using this framework to establish a complete classification scheme, the selection of the attributes values is demonstrated with an example.

Figure 4 illustrates a structured V&V process with two parts of closely associated V&V activities: model V&V and data V&V (Brade 2000, Wang and Lehmann 2007). This V&V process is integrated in M&S process including the development phases (depicted as black boxes): Problem Definition, System Analysis, Formalization, Implementation, and Experimentation. The corresponding results of these phases are referred to as intermediate products, namely Structured Problem Description (SPD), Conceptual Model (CM), Formal Model (FM), Executable Model (EM), and Simulation Results (SR). The V&V activities are organized as a triangle-like matrix. The columns of the matrix represent the V&V main phases (numbered 1 through 5), which are associated with the intermediate products; while intersections between the columns and rows split the V&V main phases into V&V sub-phases.

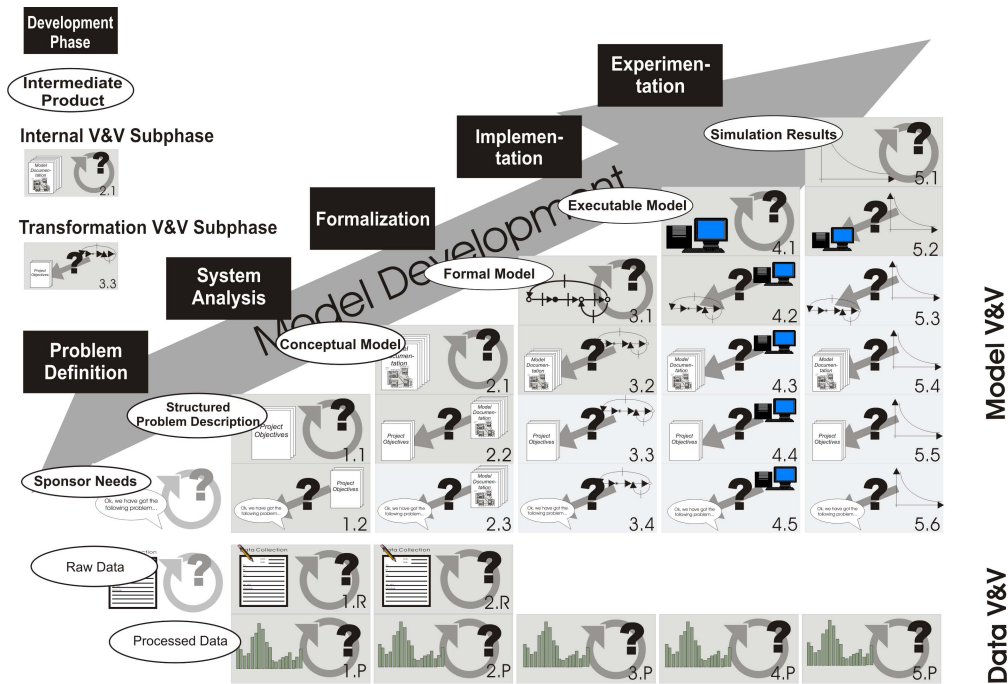


Figure 4: Example of a structured V&V process.

In addition, a documentation guideline (Lehmann et al. 2005; Wang, Kibner, and Siems 2009) was developed to facilitate model development and V&V. As shown in Figure 5, this guideline defines a hierarchical structure of model documentation and provides practical templates for documenting each intermediate product, which in turn consists of several well defined subjects. This structure is used as a baseline to classify the attribute Type.

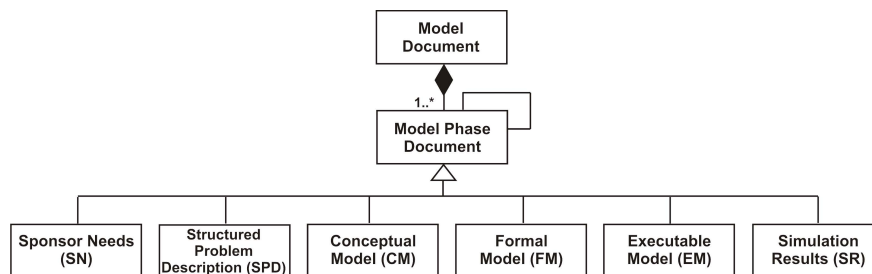


Figure 5: Templates of the documentation guideline.

Thus, according to the definitions of the Framework, the values of each attribute can be determined as follows:

- Activity
The values are the individual V&V activities numbered from 1.1 to 5.P in Figure 4.
- Detection
To determine the values of the attribute Detection, Balci's taxonomy (Balci 1998) is considered, in which the most (approx. 80) V&V techniques applicable to M&S applications are categorized into four groups: informal, static, dynamic, and formal. Those classified V&V techniques can be used as attribute values for Detection.

- Introduction
The values are clearly the different phases defined in the M&S process, i.e., Problem Definition, System Analysis, Formalization, Implementation, and Experimentation.
- Type
As mentioned above, the values of Type are the individual subjects of each intermediate product structured in the documentation guideline. By means of a two dimensional matrix, these values are described in Figure 6.

SPD	CM	FM	EM	SR
<ul style="list-style-type: none"> • M&S objectives • Model boundary requirements • Model structure and behaviour requirements • Data requirements • Experimental framework requirements • M&S constraints • Acceptability criteria 	<ul style="list-style-type: none"> • System environments and subsystems • Model input and output parameters • Model structure specification • Model behaviour specification • Interactions of submodels • Input data analysis 	<ul style="list-style-type: none"> • Identification of formalisms and tools • Formalized model boundary description • Formalized model structure • Fomalized model behaviour • Fomalized dynamic interactions between submodels • Input model • Formalized experimental framework 	<ul style="list-style-type: none"> • Simulation environments, language, and libraries • High level design • Low level design • Simulation infrastructure • Runtime environment • Interface • Structure and internal behaviour • Interaction of submodels • Data 	<ul style="list-style-type: none"> • Design of experiments • Tactical planning • Input data • Execution environment • Execution of experiments • Output documentation and presentation • Results analysis • Results interpretation

Figure 6: Description of the attribute values of Type

Since the deficiency classification scheme is established exactly based on the well defined M&S and V&V processes, especially the structured documentation guideline, the completeness and the orthogonality of Activity, Introduction, and Type are obvious. With respect to Detection, the V&V techniques summarized by Balci can be regarded as (quasi-)complete from the practical point of view. Since each V&V technique reveals a distinct difference, the attribute Detection is also orthogonal.

5 PILOT APPLICATION AND INVESTIGATION

With practical examples, this section demonstrates the use of the implemented classification scheme and its ability to perform useful analyses. The examples are based on a real simulation project with a duration of 12 months. For reasons of confidentiality, only exemplary deficiency data of the intermediate products Structured Problem Description (SPD), Conceptual Model (CM), and Formal Model (FM) are available for the pilot application.

First, the scheme was used to classify the collected deficiencies documented in the V&V reports, i.e., for each model deficiency to select an appropriate value for the attributes Activity, Detection, Introduction, and Type. Due to the well defined documentation structure, the required information could be found in the V&V reports directly. Thus, the classification effort expended was insignificant.

Based on the information extracted from the classified data, different assessments can be performed. As an example, the propagation of the model deficiencies is investigated. To do so, the relationship between the injection phase of a deficiency and its detection in the model needs to be considered. As shown in the left panel of Figure 7, a considerable number of deficiencies (based on the data collected in the first five months of the project), which were introduced by the first development phase Problem Definition, could be detected only in the later phases. This also indicates that ca. half of the CM deficiencies and one third

of the FM deficiencies were caused due to the inaccuracies in the Problem Definition phase. The right panel of Figure 7 shows the type distribution of the SPD deficiencies. While the detected deficiencies were almost in the M&S objectives and requirements, there were hardly any deficiencies of acceptability criteria. That is because at that time, the rational and feasible criteria could not be determined yet. These observations reflected the real project properties well: 1.) Since this simulation project was defined by an urgent decision, the formulated Sponsor Need was quite vague; and 2.) There were a large number of uncertainties in the real system to be modeled. As a consequence, two suggestions were proposed by the V&V team, and approved by the project sponsor and the model developer:

- With respect to the model development, a special workshop involving subject matter experts (SME) should be held for developing the next version of the SPD;
- From the V&V view point, more focus should be placed on V&V of the SPD and the CM.

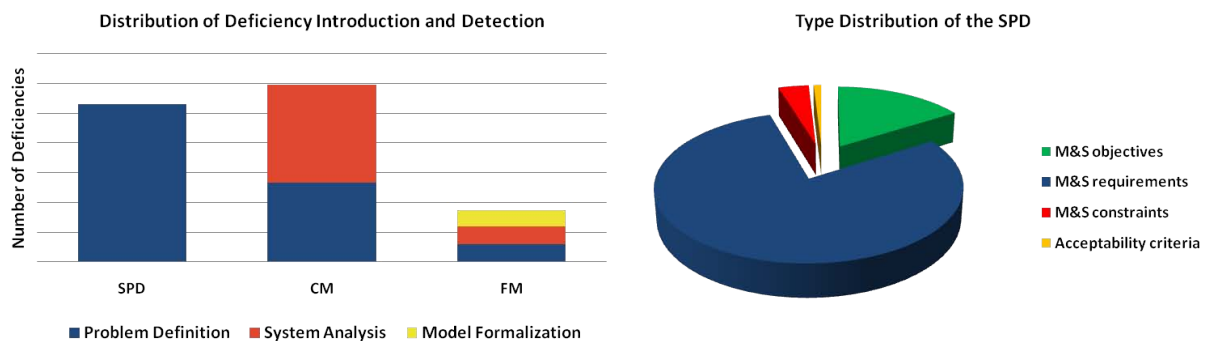


Figure 7: Analysis of deficiency propagation.

When assessing the model deficiencies, the V&V team (including three V&V agents) also identified a weakness of the conducted V&V: Desk checking (also known as self-inspection) (Balci 1998) was the only V&V technique applied so far. For the purpose of improving the effectiveness of the V&V efforts, application of additional techniques was strongly recommended. Therefore, two Inspections were planned in connection with the experts workshop. In addition, some essential model structures and dynamic behaviors were demonstratively realized by means of the computer animation. Thus, a combination of the V&V techniques Inspections, Face Validation, and Visualization/Animation was applied.

The left panel of Figure 8 shows the comparison of the deficiencies detected in the first five months and during the Inspections. This result confirms that on the one hand, the technique Inspections is highly effective, on the other hand, its application is costly and complex. Simply because of the appointments difficulties between the participants, further Inspections could not be organized in the scope of this project. This fact motivates us to improve the current quality assurance process. A planning and tailoring concept is being developed for application of different techniques in the entire V&V process.

The right panel of Figure 8 illustrates the deficiencies detected by the respective V&V activities (shown in Figure 4). The most deficiencies were detected by performing the activities V&V 1.1 and V&V 2.1, whose objectives are to ensure completeness and internal consistency of the SPD and the CM. Normally a large number of deficiencies should also be detected by V&V 1.2, whose objective is to ensure consistency between the SPD and the Sponsor Needs (SN). The reason V&V 1.2 hardly detected deficiencies in this project was that the document SN was, due to its vague formulation, completely revised in the SPD with involvement of the project sponsor and SMEs. Thus, the V&V effort based on the direct comparison of the SPD and the original SN documents was no longer effective.

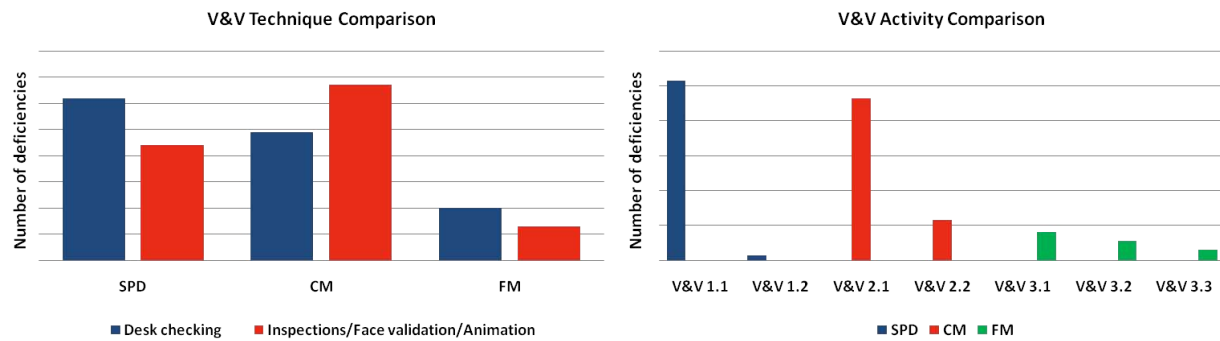


Figure 8: Analysis after inspections.

6 CONCLUSION

This paper presents a strategy of using the model deficiency classification as a quantitative method to measure the results of the V&V effort. Instead of an all-embracing and detailed scheme, a classification framework is introduced. Four attributes are defined: 1.) Activity shows when the model deficiency was identified; 2.) Detection specifies how the deficiency was detected; 3.) Introduction refers to the activity causing the insertion of the deficiency into the model; and 4.) Type indicates what is incorrect in the model. Since this framework has an open structure, new attributes for additional measurement aspects can be easily added. For example, a new attribute Impact, which refers to the severity level of the deficiency for accuracy assessment, is conceivable.

When applying the classification framework to a concrete V&V process, an application-specific scheme can be determined. Its measurement ability is however heavily dependent on whether the underlying M&S and V&V processes are well defined and structured. In addition, model documentation is of crucial importance for both the classification and analysis effort.

Findings achieved in the pilot project indicate that applying the classification scheme provides the quantitative evidence about the status of the V&V being conducted, and particularly due to the objective observations, the proposed improvement actions gain a common understanding and acceptance by the project management.

REFERENCES

- Balci, O. 1995, December. "Principles and Techniques of Simulation Validation, Verification, and Testing". In *Proceedings of the 1995 Winter Simulation Conference*, edited by C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 147–154. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Balci, O. 1998. "Verification, Validation, and Testing". In *Handbook of Simulation*, edited by J. Banks, Chapter 10, 335–393. John Wiley & Sons.
- Balci, O. 2004, December. "Quality Assessment, Verification, and Validation of Modeling and Simulation Applications". In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 122–129. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Balci, O., and R. E. Nance. 1985. "Formulated Problem Verification as an Explicit Requirement of Model Credibility". *Simulation* 45 (2): 76–86.
- Brade, D. 2000, December. "Enhancing Modeling and Simulation Accreditation by Structuring Verification and Validation Results". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 840–848. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

- Butcher, M., H. Munro, and T. Kratschmer. 2002. "Improving Software Testing via ODC: Three Case Studies". *IBM Systems Journal* 41 (1): 31–44.
- Carson, J. S. 2002, December. "Model Verification and Validation". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C. H. Chen, J. L. Snowdon, and J. M. Charnes, 52–58. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Chillarege, R., I. S. Bhandari, J. K. Chaar, M. J. Halliday, D. S. Moebus, B. K. Ray, and M.-Y. Wong. 1992. "Orthogonal Defect Classification – A Concept for In-Process Measurements". *IEEE Transactions on Software Engineering* 18 (11): 943–956.
- Chillarege, R. 2007. "ODC Measurement and Analysis – Industry Applications". Technical report, Chillarege Inc.
- Chillarege, R., and K. A. Bassin. 1995. "Software Triggers as a Function of Time - ODC on Field Faults". In *Dependable Computing for Critical Applications 5*, Dependable Computing and Fault-Tolerant Systems Vol. 10. Washington, DC: IEEE Computer Society.
- Chillarege, R., and K. R. Prasad. 2002. "Test and Development Process Retrospective – A Case Study Using ODC Triggers". In *Proceedings of the 2002 International Conference on Dependable Systems and Networks*, 669–678. Washington, DC: IEEE Computer Society.
- Freimut, B. 2001. "Developing and Using Defect Classification Schemes". Technical report, Fraunhofer IESE.
- Grady, R. B. 1992. *Practical Software Metrics for Project Management and Process Improvement*. Prentice-Hall Inc.
- Huber, J. T. 2000, March. "A Comparison of IBM's Orthogonal Defect Classification to Hewlett Packard's Defect Origins, Types, and Modes". In *Proceedings of International Conference on Applications of Software Measurement*. San Jose, California.
- IBM Research, Center for Software Engineering 2010. "Details of ODC v 5.11". <http://www.research.ibm.com/softeng/ODC/DETODC.HTM>.
- IEEE 1990. *IEEE Standard Glossary of Software Engineering Terminology*. IEEE std 610.12-1990.
- IEEE 1993. *IEEE Standard Classification for Software Anomalies*. IEEE std 1044-1993.
- Kelly, D., and T. Shepard. 2001. "A Case Study in Use of Defect Classification in Inspections". In *Proceedings of the 2001 Conference of the Centre for Advance Studies on Collaborative Research*, edited by D. A. Stewart and J. H. Johnson: IBM Press.
- Law, A. M., and W. D. Kelton. 2007. *Simulation Modeling and Analysis*. Fourth ed. McGraw-Hill.
- Lehmann, A., S. Bel-Haj-Saad, M. Best, A. Köster, S. Pohl, J. Qian, C. Waldner, Z. Wang, and Z. Xu. 2005. "Leitfaden für Modelldokumentation, Abschlussbericht". Technical report, ITIS e.V. an der Universität der Bundeswehr München.
- Oberkampf, W. L., S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin. 2000. "Estimation of Total Uncertainty in Computational Simulation". Technical report, Sandia National Laboratories.
- Oberkampf, W. L., S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin. 2002. "Error and Uncertainty in Modeling and Simulation". *Reliability Engineering and System Safety* 75 (3): 333–357.
- Schmeiser, B. W. 2001, December. "Some Myths and Common Errors in Simulation Experiments". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 39–46. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Schulz, C. 1999, February. "Orthogonal Defect Classification (ODC)-based Test Planning and Development". In *Proceedings of the International Conference on Applications of Software Measurement*. San Jose, California.
- Shannon, R. E. 1975. *Systems Simulation – the Art and Science*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Wagner, S. 2008, July. "Defect Classification and Defect Types Revisited". In *Proceedings of the International Workshop on Defects in Large Software Systems*, 39–40. New York, NY: Association for Computing Machinery, Inc.

- Wang, Z. 2010, October. “Evaluating the Potential of Orthogonal Defect Classification for Verification and Validation of Modelling and Simulation Applications”. In *Proceedings of the 2010 European Simulation and Modelling Conference*, edited by G. K. Janssens, K. Ramaekers, and A. Caris, 113–120. Hasselt, Belgium.
- Wang, Z., H. Kißner, and M. Siems. 2009, June. “Applying a Documentation Guideline for Verification and Validation of Simulation Models and Applications: An Industrial Case Study”. In *Proceedings of the 7th Industrial Simulation Conference*, edited by D. B. Das, V. Nassehi, and L. Deka, 26–30. Loughborough, UK.
- Wang, Z., and A. Lehmann. 2007. “Verification and Validation of Simulation Models and Applications: A Methodological Approach”. In *Recent Advances in Modeling and Simulation Tools for Communication Networks and Services*, edited by A. N. Ince and A. Bragg, Chapter 11, 227–240. New York: Springer.
- Wang, Z., and A. Lehmann. 2010. “Quality Assurance of Models and Simulation Applications”. *International Journal of Modeling, Simulation, and Scientific Computing* 1 (1): 27–45.

AUTHOR BIOGRAPHIES

ZHONGSHI WANG is a Research Associate at ITIS GmbH at the University of the Federal Armed Forces in Munich, Germany. His research interests include verification, validation, and accreditation (VV&A) of simulation models, model documentation, and tailoring of modeling and simulation processes. His email and web addresses are zhongshi.wang@unibw.de and <http://www.unibw.de/zhongshi.wang>.