Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

ENHANCING MODELING AND SIMULATION ACCREDITATION BY STRUCTURING VERIFICATION AND VALIDATION RESULTS

Dirk Brade

Institut für Technik Intelligenter Systeme (ITIS e. V.)
Universität der Bundeswehr München
D-85577 Neubiberg, Germany

ABSTRACT

Model Verification, Validation and Accreditation (VV&A) is as complex as developing a Modeling and Simulation (M&S) application itself. For the purpose of structuring both *Verification and Validation (V&V) activities* and *V&V results,* we introduce a refined V&V process. After identification of the major influence factors on applicable V&V, a conceptual approach for subphase-wise organization of V&V activities is presented. Finally a hierarchical presentation of V&V results is shown which addresses different people involved in use or in accreditation of simulation models.

1 INTRODUCTION

Validity of a simulation model is nothing absolute. Whenever simulation results are used to draw some conclusions for "reality", there always is the risk that simulation does not reflect all of the relevant aspects of the real world properly, or more simply, that the simulation does not fulfill its intended purpose. The intention of doing VV&A is to reduce the risk incident to simulation use. However, people using an accredited model must be aware of the fact that there is always some risk left, as only in very few cases any model malfunction or non-suitability can be definitely excluded.

Accreditation is based on credibility. If the accreditation agent decides that the model concept and the model behavior is sufficiently credible, the simulation model can be accredited. Credibility again should be based on V&V results. The more V&V tests the simulation model passes, the more credible it is (appropriate test selection assumed).

Thus, the question "how much V&V shall be applied to the simulation model?" arises. V&V costs time and money, and both should not be wasted. We think that the amount of necessary V&V is risk-dependent. If simulation faults result directly or indirectly in, e.g., lethal damage, expensive but useless investments, or the loss of valuable material, the risk incident to the use of the model (and the results it provides) is considered as high. If unsuitable or wrong simulation results just cause, e.g., marginal damage, or a minor investment in something useless, the risk incident to the use of the simulation model is considered to be low. Chris Mugridge (1999) introduces risk identification and describes risk levels which categorize the risk incident to the use of a simulation model or model results. It should be noted that the possible impact of unsuitable model results on "reality" makes an excellent indicator of the amount of *necessary* V&V.

After V&V were done, it still must be evident, why which test with which data was executed. It will be difficult to document numerous test results. It will be even impossible to derive the amount and type of applied V&V from an overall accreditation result like "sufficiently valid" or "not sufficiently valid". To increase trustworthiness, the accreditation result should state, "how much" V&V and which V&V techniques were actually applied to the simulation model. Thus, it is helpful to structure V&V results under consideration of several influence factors, among them, e.g., the specific model content examined and the comparison data. This calls for categorization of V&V results and is the subject of this paper.

In Section 2, we motivate the need for a measure of validity and identify several influences on both the V&V process and the overall V&V result. In Section 3, a refined V&V process and its phases and subphases are introduced and the impact of the refined process on V&V planning and structuring of V&V results is explained. Section 0 describes a proposed multi-layer approach for V&V result presentation, and section 5 summarizes results and future research goals.

2 DEGREE OF VALIDITY

When accrediting that a simulation model is appropriate for the intended purpose, the accreditation result should reflect the effort spent on V&V. This is not possible if the accreditation result just states "valid" or "not valid". The intensity of applied V&V should be expressed, too. Thus, a "degree of validity" states how much the model user can trust the accredited simulation model. The maximum achievable degree of validity depends on several influences summarized below.

2.1 Influences and Aims

There are several factors which influence simulation model V&V. We identify the following four being most important:

- available information about the modeled system,
- required resolution or "closeness to reality" of the simulation model which makes up the validation criteria,
- required reliability of the V&V results which depends on the risk incident to simulation model use, and
- available information about the simulation model itself.

The extend of the *available information about the real system* being modeled varies with the knowledge about the real system. For several reasons the information about the real system may be incomplete or fragmentary, among them:

- The real system does not exist.
- The real system is under development and only partially exists.
- The real system is not explorable or simply too complex.

During V&V, the model is judged under consideration of the data "representing" reality. The less is known about the real system, the less confidence in the correctness of the model can be reached by comparing the simulation model to the real system.

There already is a categorization scheme for assessment of comparison data. The CLIMB process (Confidence Levels in Model Behavior, (Flight Mechanics Panel Working Group WG-12 on Validation of Missile Simulation 1985)) categorizes measurements of real system data in five levels: basic conceptual information, theoretical data gained from science or other models, real subsystem data, hardware in the loop (HWIL) data, and real system operation data. If we just review the model concept, only CLIMB level 1 can be reached. If we evaluate the model concept, compare the model results to theoretically gained data and to real subsystem data from the lab, CLIMB level 3 may be achieved, and so on. Unfortunately this approach focuses on model results mainly and gives little importance to model concept V&V.

The *required resolution,* or *"closeness to reality"* of the simulation model strongly depends on the aim of the simulation experiment. The acceptability criteria are derived from the aim of the simulation experiment, and state the "allowed distance from reality". Conducting V&V shows, whether the acceptability criteria are met by the simulation model or not. Note that there is a great difference between the statements "as close to reality as possible" and "meets its requirements". Validity is always accredited *with respect to the requirements* which have to be met by the model to be appropriate for the intended use. However, talking to people involved in model use and model development, we noticed that it is often quite difficult to make up those requirements prior to model use accurately.

The *required reliability of the V&V results* indicates the amount of necessary V&V and is derived from the worst possible case of model malfunction or unsuitability. Vice versa the accredited degree of validity should indicate the effort spent on V&V of the simulation model to the model user. The introduction of risk classes or levels and the assignment of a required degree of validity to each risk level may enhance the determination of the necessary V&V activities (Mugridge 1999).

Due to, e.g., classification or simply failure in documentation, the *available information about the simulation model* itself may be insufficient. If for any reason this information can not be reconstructed, V&V will be less intense as with the information available. Type and amount of the available model documentation heavily influence the degree of understanding of the simulation model that the accreditation agent can achieve.

Keeping in mind that the accreditation agent should not be directly involved in model development (independent V&V, (Balci 1998)), information about the simulation model must be communicated to the accreditation agent. Several life cycles of simulation models exist which identify significant phases of model development and significant intermediate results associated to these phases. Those intermediate results reflect the development stages of the model and can serve as the communication base between model developer and accreditation agent.

To set an anchor for the following V&V process, a very abstract model development process is depicted in Figure 1. We are aware of the fact that this linear model development process is further away from reality as, e.g., an incremental model development process. As we want to introduce the intermediate results which are also gained during most of the more detailed model development processes, we keep the process itself as simple as possible yet (Lehmann 1999).

After identification of the main influences on V&V, indicators for the degree of validity are summarized below.

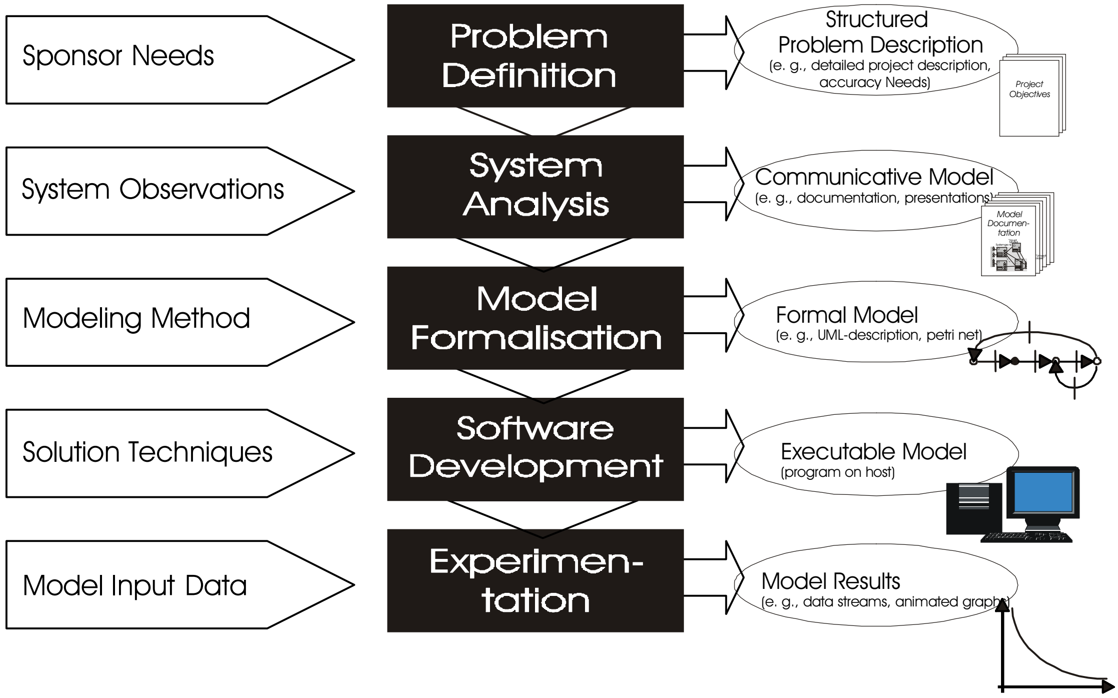

Figure 1: Intermediate Results gained during a Simple Model Development Process

2.2 Indicators for the Reached Degree of Validity

The achievable degree of validity mainly depends on the topics listed in Section 2.1. Both achievable depth and achievable width of V&V depend on the knowledge about – again – both real system and simulation model which determine the maximum achievable intensity. The minimum intensity is given by the desired validity which is derived from the risk incident to model use. Thus, those factors influence the *intensity* with which V&V are executed. For precise V&V intensity, we differentiate between the *depth* and the *width* of V&V.

With respect to intensity "width" describes the variety of questions asked during V&V concerning the previously introduced intermediate results (e.g., documentation, model concept, or implementation). Indicators for the width of the applied V&V are:

- the type of intermediate results taken under consideration and associated documentation evaluated;
- the variety of test data and test data sources (from both simulation model and reality);
- the domains where subject matter expert (SME) came from;
- the variety of techniques applied.

For example, to increase width, the structured problem description is validated, SME validates the conceptual model, the modeler verifies the formal model, while statistical techniques are used to compare model results to test data from the real system.

The "depth" describes the accuracy of examining different simulation model contents. Indicators for the depth of the applied V&V are:

- the concentration on one specific intermediate result of the simulation model and the level of detail of the documentation evaluated;
- the density of test data (from both simulation model and reality);
- skill of SME consulted on a specific topic;
- the techniques applied.

For example, to increase depth, further assumptions made by the modeler are checked by using deeper scientific knowledge.

The degree of validity depends on what was checked (width) and how intensively it was checked (depth). The less is known about either the real system or the simulation model, the lower is the maximum achievable intensity. However, the determination of "intensity" and the derivation of a degree of validity still depends on V&V expert opinion.

3 PROPOSED V&V PROCESS

3.1 Phases and Subphases

Usually each V&V phase parallels a model development phase (Department of the Army, Headquarters 1999a; Department of the Army, Headquarters 1999b). The intermediate result gained during each model development phase contains significant information about the simulation model and serves as one of the inputs to the associated V&V phase. The contained information about the model is subject to the V&V activities belonging to the associated V&V phase. Often in representations of the V&V process there are techniques assigned to the phases.

In Section 2.1 relevant bullets have been identified which should be reflected in the V&V process model. To allow a more precise assignment of relevant information or activities, the V&V phases have been split into subphases depicted in Figure 2.

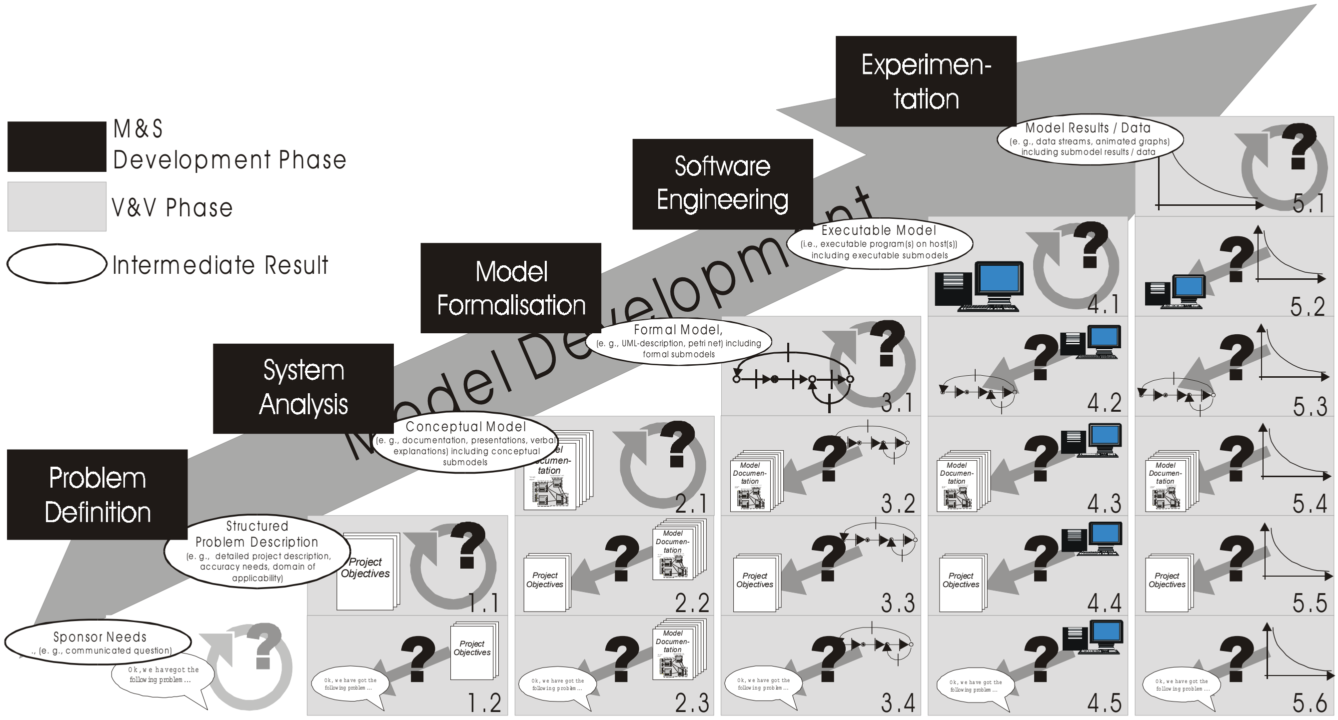

Figure 2: V&V Phases and Subphases Assigned to Each Intermediate Result

In each of the five V&V phases (the columns in Figure 2; phase 1: Specification validation; phase 2: conceptual model V&V; phase 3: formal model V&V; phase 4: executable model V&V; phase 5: model results V&V) one intermediate results is examined. To each subphase a single intermediate result (subphases x.1) or a pair of intermediate results (subphases x.2 or greater) is assigned. The phases contain information about what to do with the intermediate results and a separate examination aim for each V&V subphase. In each subphase the fields shown in Table 1 are addressed.

3.1.1 Intermediate Result Subject to V&V

This field states which intermediate result is subject to V&V. Referring to Figure 2, this is for, e.g., all subphases of phase 3 the “formal model”.

3.1.2 V&V Examination Aims

Specific examination aims on which the V&V subphase focuses. One problem with V&V is the extend and complexity of most simulation models to be examined. The question “Is the simulation model appropriate for the intended purpose?” is split up into many subquestions. Thus, the field “V&V examination aim” shall guide the accreditation agent through the process.

The examination aim of, e.g., subphase 3.2 is to show, whether the conceptual model was properly transformed into a formal model or not. The check of the formal correctness (which is independent from the contents of the conceptual model) is the examination aim of V&V subphase 3.1.

3.1.3 Required Intermediate Result for Comparison

If the examination aim is checking the consistency with a previously achieved intermediate result, this field states, to which intermediate result the comparison is done (as in, e.g., subphase 3.2: formal model versus conceptual model). Again, this supports focusing comparison activities by clear identification of the required comparison information.

Table 1: The Checklist associated to each V&V Subphase

Subphase Chart		
No	**Field Name**	**Content (exemplary)**
1	Phase ID	*3.1*
2	intermediate result subject to V&V	*formal model*
3	V&V examination aims	*checking syntactical correctness*
4	required intermediate result for comparison	*none*
5	required real world representation for comparison (availability, source)	*modeling theory*
6	required knowledge of the real world (application domain), modeling and simulation, or Software/Hardware design	*model description formalisms*
7	V&V techniques	*syntax checking*
8	V&V tools	*<appropriate modeling tool>*
9	documentation templates for V&V report	*<pointer to template>*
10	V&V subphase skipping requirements and consequences	*This phase should never be skipped. Following examinations of the formal model may be disturbed by syntactical errors*
11	impact of failure indicator	*low impact; repeat model formalization phase*

3.1.4 Required Real World Representation for Comparison

If the examination aim of the subphase is checking the contents of an intermediate result, this field should hold information about availability and general sources of relevant real world information. This may vary from fundamental scientific references to special databases.

3.1.5 Required Knowledge

Description of the knowledge domain a SME must master to be well suited for the application of subjective test methods.

3.1.6 V&V Techniques

A selection of techniques appropriate to examine the intermediate result according to the examination aim (Department of Defense 1996).

3.1.7 V&V Tools

A selection of tools to support the selection and application of appropriate V&V techniques. This field is only useful, if it is up-to-date (Pace 1999).

3.1.8 Documentation Templates for V&V Report

For accreditation, the V&V results must be evaluated. Standardized templates for V&V documentation are helpful to do the evaluation efficiently (Sullivan 1999).

3.1.9 V&V Subphase Skipping Requirements/Consequences

If the examination aim was already achieved during a previous V&V subphase (decision of the V&V agent), this subphase may be skipped to reduce the V&V costs. This is further discussed in Section 3.2.

3.1.10 Impact of Failure Indicator

Consequences for the model development, if a non-appropriateness or malfunction is detected. If, for example, a test fails in subphase 3.2, then the formal model either is subject to a syntax error which was not detected in subphase 3.1, or the consistency between the conceptual model and the formal model is violated. This means that the phase "model formalization" from the model development process has to be (at least partially) repeated.

Most of the field contents of subphases in the lower left corner of the triangle (Figure 2) depend on the application domain of the model. The further the model development advances, the more influences from modeling and implementation are reflected in the fields.

Both V&V and accreditation can benefit from this subphase approach.

3.2 V&V Planning

Often the same examination aim can be achieved by evaluation of different information sources. Thus, there is redundancy of examination aims all over the V&V subphases. For example, the question, whether the simulation model fulfills the requirements documented in the structured problem description, can be examined by checking the formal model against the structured problem description or checking model results against the structured problem description. The desired degree of validity determines, how often a specific examination aim is subjected to V&V activities which rely on different data/information.

With the required degree of validity, the model information available, and the knowledge about the real system available in mind, the accreditation agent decides, where the foci of the V&V activities will be set. To do so, the accreditation agent may use the triangle to assign an "amount of required V&V" to each subphase. To keep V&V effective, the V&V agent makes the *deliberate* decision which examination aims are less important to achieve than other examination aims. The agent must be aware of the possible consequences of assigning low V&V intensity to the examination aim of this subphase. In the following, this process is referred to as "weighting". During V&V planning the agent weights the examination aims of the subphases and "tailors" a list of the most important examination aims.

A V&V subphase and all associated activity may be lightly weighted or even skipped, if the common examination aim is reached during another V&V subphase much easier and the desired degree of validity allows it. Skipping of subphases accelerates V&V, but the loss of redundancy reduces the final trustworthiness granted to the simulation model by application of V&V. It is a challenge to find the best ratio between necessary V&V and the impact of non-appropriateness of model results. Again, expert opinion is required to decide about this, but risk classification and V&V levels support the decision.

Another reason for assigning only little weight to a subphase is that there is just no or to less comparison information available that is relevant in the specific subphase. For example, usually in battle field simulation it is not sufficient for V&V to compare the output of the simulation model to well documented real scenarios, because the number of those scenarios is very limited. Thus, the main V&V focus on the model concept. Contrarily, the V&V of technical simulation models developed in analogy to the real system (e.g., for rapid prototyping) may focus on the comparison between model output and numerous performance data of the real system that were measured, e.g., in the lab.

With the subphases and the fields listed in 3.1 the V&V triangle supports V&V planning. Excerpts of the V&V triangle already were successfully used in an experimental V&V study. The Amt für Studien und Übungen der Bundeswehr (AStudÜbBw - Department for Studies and Exercises of the German Federal Armed Forces) / Operations Research did exemplary V&V on the battlefield simulation model SiRa-Brigade, and used the explicitly stated examination aims of the triangle's subphases to identify required comparison material.

3.3 V&V Result Evaluation

The association of relevant V&V information, including specific information about the model (intermediate results), model information for comparison, real world data for comparison, and V&V techniques to V&V subphases structures test results.

With respect to intensity, the effort spent on reaching the examination aim of a single subphase is referred to as "depth". The selection of subphases (and associated examination aims) is referred to as "width". Thus it becomes evident to the accreditation instance, how intensively V&V was done, i.e., how much effort was spent to, e.g., validate the conceptual model against the sponsor needs, or check the executable model against its formal specification. Much more important, gaps (i.e., subphases examined with a low intensity) in the chain of V&V activities can be discovered and judged more easily.

4 REPRESENTATION OF V&V RESULTS

Doing V&V is a matter of increasing trust in a specific simulation model. To do so, the V&V result must state, how intensively the overall appropriateness of the simulation model was examined. The user needs one overall indicator to judge, whether the validation effort applied to the simulation model outmatches the risk incident to the model use. In section 2 we motivated the

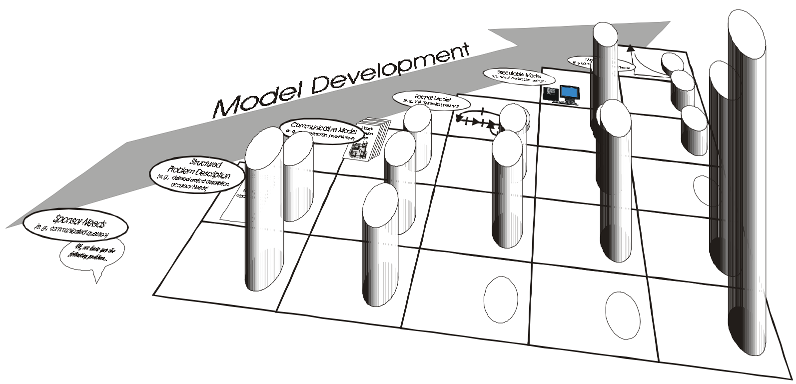

Figure 3: Visualization of "Depth" (column height) and "Width" (column distribution) of V&V Intensity

introduction of a "degree of validity" and addressed the need for a measure of validity like a degree of validity.

But there are different persons confronted with the "validity" of a simulation model. Information which may be interesting to one person could confuse the other. Different persons confronted with the simulation model require different depths of understanding of the V&V activities and follow different ways of gaining trust into the model. For example, the accreditation agent who does the first accreditation of the simulation model for the given intended purpose needs to know every single raw test result to grand a certain degree or level of validity. However, after, e.g., re-implementation of an already validated conceptual model, only those V&V activities have to be repeated which are associated to implementation. Thus, details of the previous V&V phases just slow down accreditation. Section 3 states that V&V results should be structured and assigned to different development stages (which are reflected by the intermediate results) of the simulation model.

It is possible, to arrange the three representations of V&V information (raw test results, structured test results, and overall result) hierarchically. This is depicted in Figure 4. Although there are neither algorithms for calculating intensity "depth" from the selection and extent of applied test techniques, nor algorithms for calculating an overall degree of validity from intensity "depth" and "width", this structure supports V&V expert decision.

The V&V result presented in the top layer informs the model user about the overall effort of V&V. Thus the model user gets an impression of the trustworthiness of the simulation model, but does not learn about foci of V&V activities. Showing further details, the medium layer helps organizing and interpreting raw test results. The fine decomposition of the V&V process in phases and subphases supports the accreditation agent, to document the V&V results, by arranging the raw test results of the bottom layer into groups.

5 CONCLUSIONS AND FUTURE WORK

Three significant layers for V&V results representation are identified, integrating raw test results, subphases with separate examination aim which are examined with a particular intensity, and an overall degree of validity.

The phases and subphases of the V&V process provide guidance for V&V planning. Each single phase owns a specific examination aim and describes, how this aim can be achieved. Furthermore skipping requirements are given which allow to "tailor" a set of V&V activities from the activities given all over the V&V process.

Our future research will focus on filling the fields of the checklist of each V&V subphase with more detailed information. Although the idea of V&V process decomposition into subphases and associated activities works fine for the V&V of a simple crossroad model and supports the domain experts of AStudÜbBw / BerOR in organizing exemplary V&V of a battlefield simulator at brigade level, it still needs to prove suitability in further applications. Following questions need to be answered:

- How can we support the determination of "V&V intensity" by expert opinion to make it less subjective?
- How is the intensity of subphase V&V activities "aggregated" to an overall V&V intensity?
- Can rules for decomposition of overall validation criteria to the subphases examination aims be made up?

The V&V triangle provides guidance for doing V&V and evaluation of V&V results which still strongly depends on V&V expert opinion. By answering the above questions we try to reduce this dependency.

ACKNOWLEDGMENTS

The research leading to the development of this paper has been sponsored by the Bundesamt für Wehrtechnik und Beschaffung (Federal Office of Defense Technology and Procurement) and the Amt für Studien und Übungen der Bundeswehr (Department for Studies and Exercises of the German Federal Armed Forces). The research is directed by Prof. Dr. Axel Lehmann at the University of the German Federal Armed Forces, Munich.

REFERENCES

Balci, O. 1998. Verification, validation and testing. *The Handbook of Simulation*, Chapter 10, 335–393, John Wiley & Sons, Inc.

Department of Defense. 1996. Department of Defense Verification, Validation and Accreditation (VV&A) Recommended Practices Guide. Defense Modeling and Simulation Office, Alexandria, VA. (Co-authored by: O. Balci, P.A. Glasgow, P. Muessing, E.H. Page, J. Sikora, S. Solick, and S. Youngblood)

Department of the Army, Headquarters. 1999a. VV&A of army models and simulations. Pamphlet 5-11, Department of the Army, Headquarters, Washington, DC.

Department of the Army, Headquarters. 1999b. Test and evaluation: M&S VV&A methodology. TECOM Pamphlet 73-4, U. S. Army Test and Evaluation Command, Aberdeen Proving Ground, Maryland.

Flight Mechanics Panel Working Group WG-12 on Validation of Missile Simulation. 1985. Final Report (excerpts): AGARD Advisory Report No. 206.

Lehmann, A., J. Lüthi, C. Berchtold, D. Brade, and A. Köster. 1999. Published in German: Zukunftsfelder der Modellbildung und Simulation. Zwischenbericht,

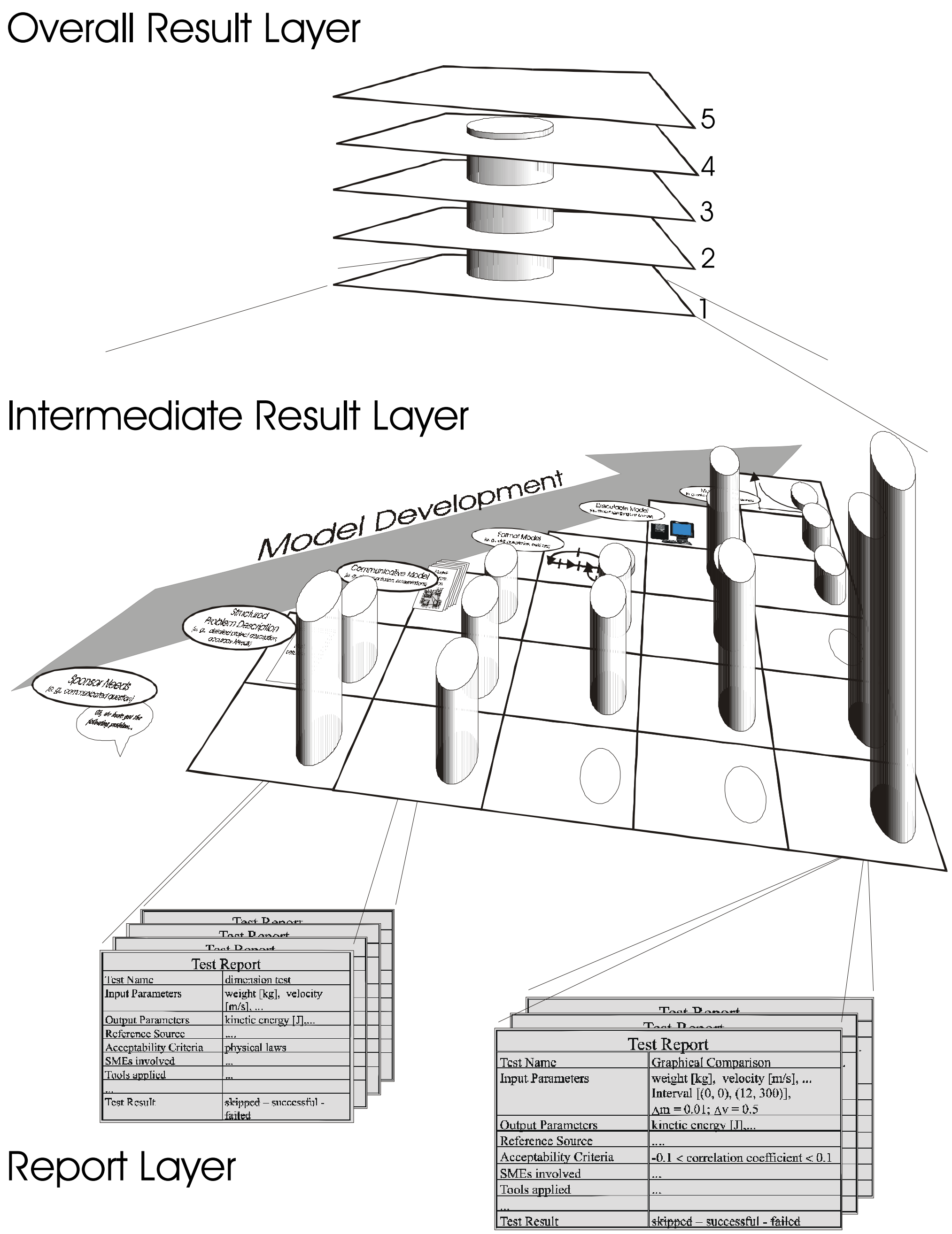

Figure 4: A Three Layer Representation of V&V Results. The Degree of Validity is exemplary ordered in Five Levels.

Institut für Technik Intelligenter Systeme e. V. an der Universität der Bundeswehr München.

Mugridge, C. 1999. Verification, validation and accreditation of models and simulations used for test and evaluation – a risk / benefit based approach. Internal Report, Technical Development Group, Defense Evaluation and Research Agency UK.

Pace, D.K. 1999. V&V technology. *Proceedings of the 1999 Summer Computer Simulation Conference*, 441 – 445.

Sullivan, C. and J. Chew. 1999. International Test Operating Procedure (ITOP) on verification, validation and accreditation of models and simulations, presented at the *Army Operations Research Symposium XXXVIII* (Fort Lee, Virginia).

AUTHOR BIOGRAPHY

DIRK BRADE is a Research Staff Member of Prof. Dr. Axel Lehmann at the University of the German Federal Armed Forces, Munich / Institut für Technik Intelligenter Systeme (Institute for Technology of Intelligent Systems). He successfully finished his studies of Computer Science at the Technical University of Darmstadt in 1998 and has been employed at the institute since then. His current research interests include component-based simulation development and VV&A. His email address is <brade@informatik.unibw-muenchen.de>.