Verification, Validation and Uncertainty Quantification Methods and Techniques

(An Overview and their Application within the GM-VV Technical Framework)

Dr. Manfred Roza
National Aerospace Laboratory NLR
Q-tility V&V Expertise Center
Anthony Fokkerweg 2, Amsterdam
NETHERLANDS

Manfred.roza@nlr.nl or Manfred.roza@q-tility.nl

ABSTRACT

The Generic Methodology for Verification and Validation (GM-VV) is a generic and comprehensive methodology for structuring, organizing and managing the verification and validation (V&V) of M&S assets. The GM-VV is a new recommended practice within the Simulation Interoperability Standards Organization (SISO), which is the result of a joint development effort with NATO. The GM-VV provides a technical framework to efficiently develop arguments to justify why M&S assets are acceptable or unacceptable for a specific intended use. This argumentation supports M&S stakeholders in their acceptance decision-making process regarding the development, application and reuse of such M&S assets. The GM-VV technical framework assures that during the execution of the V&V work the decisions, actions, information and evidence underlying such acceptance arguments will be traceable, reproducible, transparent and documented. To collect appropriate evidence for such acceptance arguments, evidence solutions in the form of V&V methods and techniques must be selected and applied, along with a quantification of (residual) uncertainty.

This paper describes how V&V methods, techniques and uncertainty quantification fits in the GM-VV technical framework. Next it provides possible taxonomies of V&V and uncertainty quantification techniques and methods. A characterization approach is presented identifying and specifying the information relevant for selecting such appropriate methods and techniques throughout a V&V project. Within the GM-VV framework this provides a base line for risk-based tailoring approaches. This approach is translated into a V&V and uncertainty quantification catalogue that is embedded into a V&V enterprise repository and can be reused and refined over multiple V&V projects. An overview of tools is provided that implement these V&V and uncertainty quantification methods and techniques, and support the effective and efficient implementation of the GM-VV and V&V in general.

1.0 INTRODUCTION

The GM-VV attains its generic quality from a three-part technical framework [1],[2],[3]: the conceptual, implementation and tailoring framework. The implementation framework provides the implementation building blocks to set up en execute V&V projects and underlying technical activities. These technical activities can be summarized by the following simplified V&V life-cycle model steps [10],[11], which directly relate to a goal-claim network implementation of the GM-VV argumentation structure approach (Figure 1). The V&V life-cycle model steps are:

1) **V&V context analysis:** Acquiring all relevant and necessary information for the V&V of the M&S system of interest (M&S SoI).

2) **Acceptability criteria specification:** Development of acceptability criteria to be met by the M&S SoI to fit the intended use.

3) **V&V experimental frame design:** Specification and selection of appropriate evidence solutions to demonstrate that the M&S SoI meets the acceptability criteria.

4) **V&V experimental frame implementation and execution:** The implementation and execution of the selected evidence solutions

5) **V&V results analysis and integration:** Analysis of the V&V results produced by the executed evidence solutions, and integration into claims on the acceptability criteria compliancy by the M&S SoI.

6) **Acceptance case development:** Development of an acceptance case supported with all prior arguments and items of evidence on how well the M&S SoI fits its intended purpose.
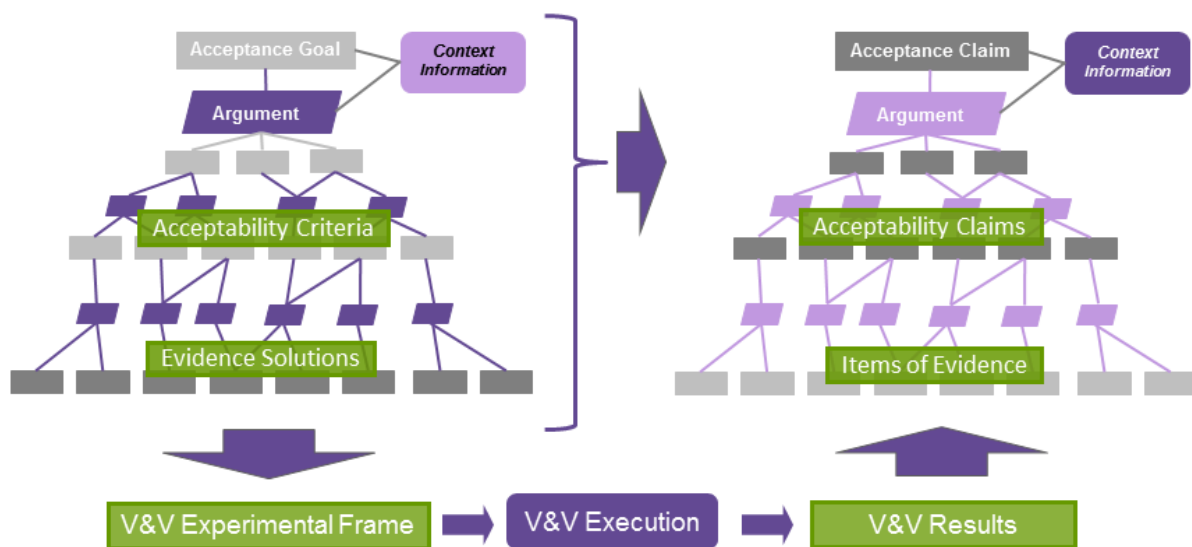

Figure 1: GM-VV V&V Argumentation Structure: Q-tility Goal-Claim Network.

This paper primarily focusses on the identification and selection, by means of sound argumentation, of evidence solutions during the V&V experimental frame design step. Evidence solutions in this regard comprise the set of V&V techniques and methods to acquire the right evidence to demonstrate that the acceptability criteria for relevant utility, fidelity and correctness properties of the M&S SoI are met. GM-VV also identifies V&V quality criteria that have to be met or specified to be able to make well-informed decision making during the V&V process, i.e. risk-based tailoring, and to assess the residual risks after the completion of the V&V effort [1],[3],[4],[6]. These V&V quality criteria primarily comprise identification of risk-factors and selection of uncertainty quantification methods and techniques for both the M&S SoI and the V&V process itself. Both the selection of V&V and uncertainty quantification techniques and methods should be substantiated with proper arguments and backing evidence inside the right side of a V&V goal-claim network (Figure 1).

This paper provides an overview of possible V&V (Chapter 2) and uncertainty quantification (Chapter 3) techniques and methods. Possible concepts to document, store and retrieve these techniques and methods by means of catalogue concept a GM-VV enterprise memory are discussed (Chapter 4). The paper ends in Chapter 5 with a practical approach to implement this concept as part of V&V experimental frame design step of the V&V project life-cycle.

2.0   V&V METHODS AND TECHNIQUES TAXONOMIES

Within the M&S community a wide variety of V&V methods and techniques have been identified and are used. Several M&S proponents have developed taxonomies for V&V techniques of which the work of Sargent and Balci can be considered to be seminal for all other efforts [14],[15]; such as the commonly used DoD VV&A Recommended Practices Guide for V&V techniques [16]. The most up to date taxonomy that is currently available is found in the work of Petty [17]. Though he strongly builds upon the work of Balci, he succeeds to extend this taxonomy more widely and with more practical details.

All these taxonomies have one thing in common and that is that they solely focus on assessing the M&S SoI level of fidelity, against a specific validation referent. These V&V methods and techniques are usually divided into the following four categories:

1)   **Informal:** V&V techniques that are usually executed and interpreted by humans. Typically these require few resources and can be executed in a short time. The convincing force, however, depends on the trust in the humans doing the work and the process they use.

2)   **Formal:** V&V techniques are based on mathematical proofs of correctness of model implementations. The application of formal methods, however, is often limited due to large resource costs even for relatively small M&S systems and their use. If applicable, the convincing -force of these V&V results are usually very strong.

3)   **Static:** V&V techniques can be applied early in the development process because no execution is required. It is typically applied in the concept phase and parts of the development phase. Typically specialized tools are used to do automated checks. The required resources are normally limited. It is required to have access to documentation and half-products. The strength of the convincing force is dependent on the rigor of the tests.

4)   **Dynamic:** V&V techniques require execution of the M&S System in part or as a whole. The dynamic properties of the M&S System are studied and checked. Typically specialized tools are used to do automated measurements and checks. The required resources are normally limited. Dynamic V&V techniques may require access to parts of the M&S System that are usually not available. The strength of the convincing force is dependent on the rigor of the M&S System check.

Within these categories about eighty V&V methods and techniques in general are identified (Table 1). For the complete description the interested reader is referred to the public available sources from Mikel Petty and others [14],[15],[16],[17].

Table 1 with V&V methods and techniques however do not cover all aspects of assessing M&S SoI correctness and utility properties as identified by GM-VV [1],[3]. First of all GM-VV identifies V&V as an extension of standard software and systems engineering V&V. In this regard M&S SoI correctness also covers standard software and systems V&V testing practices, for which many sources of methods and techniques are available. For assessing these aspects of M&S SoI the reader is referred to the work that is made available through the International Software Testing Qualifications Board (www.istqb.org) or through organizations such as RBCS (www.rbcs-us.com).

Table 1: V&V Method and Techniques.

Informal	Formal	Static	Dynamic
• Audit	• Induction	• Cause-Effect Graphing	• Acceptance Testing
• Desk checking	• Inductive Assertions	• Control Analysis	• Alpha Testing
• Documentation Checking	• Inference	• Calling Structure Analysis	• Assertion Checking
• Face validation	• Logical Deduction	• Concurrent Process Analysis	• Beta Testing
• Inspections	• Lambda Calculus	• Control Flow Analysis	• Bottom-Up Testing
• Reviews	• Predicate Calculus	• State Transition Analysis	• Comparison Testing
• Turing test	• Predicate Transformation	• Data Analysis	• Compliance Testing
• Walkthroughs	• Proof of Correctness	• Data Dependency Analysis	• Authorization Testing
		• Data Flow Analysis	• Performance Testing
		• Fault/Failure Analysis	• Security Testing
		• Interface Analysis	• Standards Testing
		• Model Interface Analysis	• Debugging
		• User Interface Analysis	• Execution Testing
		• Semantic Analysis	• Execution Monitoring
		• Structural Analysis	• Execution Profiling
		• Symbolic Evaluation	• Execution Tracing
		• Syntax Analysis	• Fault/Failure Insertion Testing
		• Traceability Assessment	• Field Testing
			• Functional (Black-Box)Testing
			• Graphical Comparisons
			• Interface Testing
			• Data Interface Testing
			• Model Interface Testing
			• User Interface Testing
			• Object-Flow Testing
			• Partition Testing
			• Predictive Validation
			• Product Testing
			• Regression Testing
			• Sensitivity Analysis
			• Special Input Testing
			• Boundary Value Testing
			• Equivalence Partitioning Testing
			• Extreme Input Testing
			• Invalid Input Testing
			• Real-Time Input Testing
			• Self-Driven Input Testing
			• Stress Testing
			• Trace-Driven Input Testing
			• Statistical Techniques
			• Structural (White-Box) Testing
			• Branch Testing
			• Condition Testing
			• Data Flow Testing
			• Loop Testing
			• Path Testing
			• Statement Testing
			• Submodel/Module Testing
			• Symbolic Debugging
			• Top-Down Testing
			• Visualization/Animation

GM-VV also identifies M&S utility properties that need to be assessed with V&V methods and techniques. These properties relate to assessing (non)functional requirements of an M&S SoI and the added value in the relation to the intended purpose. There are yet no real fixed taxonomies available since this a new "all inclusive" perspective on V&V of M&S SoI. One of the lessons-learned from the Q-tility V&V studies is that such V&V methods and techniques usually center around aspects from the real-world and problem world of the GM-VV four world view [10],[11],[13]:

- **Measures of Performance M&S SoI:** are specific M&S SoI performance metrics to be accomplished to be useful in the real-world or problem domain application, such as how fast the M&S system can be configured or initialized.

- **Measures of Effectiveness M&S SoI:** are specific metrics that are used to measure M&S SoI results in the overall real-world mission or problem world of assigned tasks, for example the accomplished transfer of training of an M&S based training or training cost-reduction.

3.0   UNCERTAINTY QUANTIFICATION METHODS AND TECHNIQUES TAXONOMIES

Modeling and simulation attempts to bring together what is known about the real-world in terms of certainty and what is uncertain in the application of the M&S system of interest. Training, science and engineering application of M&S have a strongly tended to emphasize what we know, or think we know, instead of what is uncertain [19]. Essentially all models are wrong because they are always an abstraction and approximation of the real-world [12]. Despite that all models and simulations are wrong some are useful [18]. However, to determine whether these models and simulations are useful and appropriately benefit from the results they produce requires a correct understanding of the level of M&S fidelity and application context of those results [20]. M&S communities often assess models and simulation results without fully considering uncertainties that might impact the level of M&S fidelity and application context of those results[1]. This creates a risk for inappropriate utilization of and overreliance on such models and simulations. Literature shows it is impossible to apply V&V methods and techniques to such an extent that 100% certainty is achieved regarding the M&S system correctness, fidelity and utility [7],[9],[12],[19].

To aid a responsible development and utilization of M&S systems and their results, not only V&V techniques as discussed in the previous chapter should be applied. Also the M&S uncertainties should be identified and, where possible, be quantified in an integral manner. GM-VV facilitates this through what is called V&V quality criteria that should be specified and assessed by means of the argumentation structure approach [1],[3]. These V&V quality criteria are the basis of the GM-VV risk-based tailoring approach to provide the M&S system developer or user with a quantified, where possible, indication of risks associated with using M&S results and an indication of confidence (trust) that the assessment of M&S system adequacy and reliability for its intended use warrants (Chapter 5).

Uncertainty Quantification (UQ) is gaining more and more interest within the M&S community due to increasing reliance on M&S results in today's society. Available UQ literature focusses on techniques and methods for determining simulation uncertainties and their propagation through the model or simulation. The majority of the research work is done by the mathematical and computing communities [19]. This chapter provides an introduction to the types and sources of M&S uncertainty, along with the most important UQ methods and techniques without going in the details of their often complex mathematical constructs.

3.1   Types of Uncertainties

Many categorizations of uncertainty have been developed tend to mix the nature or essence of a type of uncertainty with how or where it might occur in model and simulation life-cycle or in the (sub)model chain

---

[1] In GM-VV this application context is considered in the V&V context analysis of the real-world and problem world [4].

[18],[19]. For example parametric uncertainty, structural uncertainty, geometrical, data uncertainty and boundary/initial condition uncertainty [20]. However, a sound taxonomy would only categorize uncertainty types according to their fundamental essence, and then discuss how that essence could be embodied in different aspects of an M&S system. The risk assessment community, primarily the nuclear reactor safety community, has developed the most workable and effective categorization of uncertainties [19]. This taxonomy is depicted in Figure 2:

1) **Aleatory uncertainty:** *uncertainty due to inherent randomness*. Aleatory uncertainty is also referred to as stochastic uncertainty, variability, inherent uncertainty, uncertainty due to chance, and Type A uncertainty. The fundamental nature of aleatory uncertainty is randomness, e.g., from a stochastic process. The mathematical representation, or characterization, most commonly used for aleatory uncertainty is a probability distribution. Aleatory uncertainty can be embodied in two ways in M&S system: in the underlying model form itself and in parameters of this model.

2) **Epistemic uncertainty:** *uncertainty due to lack of knowledge*. Epistemic uncertainty is also referred to as reducible uncertainty, knowledge uncertainty, and subjective uncertainty. The fundamental source of epistemic uncertainty is incomplete information or incomplete knowledge of any type that is related to the simuland or het M&S system. Epistemic uncertainty is a property of the modeler or observer, whereas aleatory uncertainty is a property of the simuland being modeled or observed. The lack of knowledge can be related to modeling issues for the simuland, computational issues of the M&S system, or experimental data needed for validation (See Section 3.2). Epistemic uncertainty can be divided into two types: recognized uncertainty to mean any epistemic uncertainty that has been recognized in some way (e.g. explicit assumptions or approximations) and blind uncertainty to mean any epistemic uncertainty that has not been recognized in some way (e.g. human errors, blunders, or mistakes in judgment and interpretation).
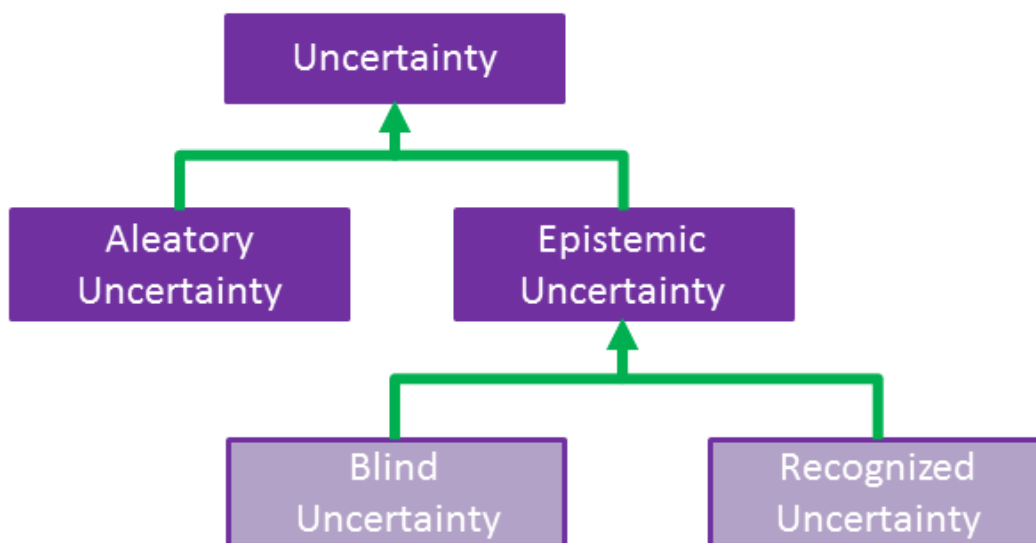

Figure 2: M&S Uncertainty Taxonomy.

The above categorization showed to be applicable and effective within GM-VV case-studies by Q-tility [12],[11].

3.2   Sources of Uncertainties

There are many sources for uncertainty in the whole M&S system life-cycle. The most comprehensive paradigm to characterize these sources is developed by Pace [18] and provided a good basis to identify

sources of uncertainty within the Q-tility GM-VV case studies [10],[11],[13]. This paradigm consists of the following seven major source of M&S uncertainty that should be considered in a V&V effort:

1) **M&S system application domain.** Uncertainty in the simulation application domain comes from a variety of sources. Firstly, there is the variability of the simuland. Such variability is an aleatory uncertainty. A second aspect of application domain uncertainty comes from lack of understanding the principles that govern how things of the simuland are and how they behave and interact with other things represented of in the M&S system. A third source of application domain uncertainty comes from inability to completely know the situation of things for which the governing principles are fully understood. It should be recognized that uncertainty in the simulation application domain establishes general limits on the level of fidelity that should be expected from a M&S system.

2) **M&S system intended purpose.** Uncertainty in M&S objectives, intended use statements, and requirements can be avoided, but usually is present because of vague, incomplete, inconsistent and contradictory material in the documents that establish an M&S's purpose. It is impossible to generalize about simulation purpose uncertainties, but there have been many instances of simulations used outside their appropriate realms because of uncertainty in intended purpose. A typical and often seen example is the lack of good Training Needs Analysis (TNA) prior to acquiring a training simulation device resulting in inadequate or even negative training outcomes.

3) **M&S system concept to design.** In the translation of the M&S system objectives into a design, often through a conceptual model, many explicit and implicit decisions are made about simulation fidelity in terms of the completeness of factors considered, level of resolution and accuracy of algorithms. Consequently, recognized and blind uncertainties can enter in through: (a) Assumptions, choices about simulation resolution, algorithms (including approximations), aggregation, errors and variability of what is represented in the simulation (b)) Computational considerations may dictate selection of algorithms that are recognized as approximations of the desired calculations, (c) Aggregation, many M&S systems employ aggregation in their level of simuland representations, (d) Numerical computational errors and rounding, (e) modelling of the simuland variability.

4) **M&S system implementation.** Implementation uncertainty comes in various forms. These include software bugs, federation communication and control factors that potentially can cause uncertainty in federate and federation performance and results, computational processes such as table lookup errors or grid selection. M&S systems that involve special facilities, hardware and systems (e.g. live or virtual simulations, visual systems or motion system) must consider uncertainties related to those factors as well as the factors mentioned previously. This also includes the maturity level of the M&S system technology applied, the newer a technology the more uncertainties may be implied due to more complex models.

5) **M&S system inputs and use.** Input uncertainty comes from three sources. One kind is concerned with poor-quality data and parameter values that are to be used in the M&S system: the GIGO problem. A second kind is concerned with errors in entering the information of configuring the M&S system. Many large and complex simulations require thousands of parameter values for a run. The third kind of input uncertainty comes from the scenarios that are used in simulation runs. A proper set of scenarios must be used including proper initial and boundary conditions. Uncertainty related to use of an M&S system is sometimes called ''user effect'' or ''user effects.'' This uncertainty is derived from variation in how the M&S system user chooses to interpret the problem to be run and sets the simulation up to run the problem, monitors and controls or influences the simulation execution by additional inputs and events.

6) **V&V referent data.** The V&V referent is the standard to which simulation results are compared. The validation referent is subject to all the uncertainty of the application domain knowledge, since the V&V referent can be considered a subset of application domain knowledge. The subset of application domain knowledge used as V&V referent may consist of test results that are not known to the simulation user prior to running the simulation. The subset of the application domain

knowledge used as the V&V referent could also be test results known to the user prior to the simulation runs. Many simulations contain parameters that can be adjusted to improve coherence between simulation results and the standard with which simulation results are compared. Often this data is used for the M&S system calibration. Consequently, few consider comparison of simulation results with test values known to the simulation user prior to running the simulation as a reliable guide to the estimation of simulation fidelity or predictive accuracy.

7) **M&S results interpretation.** Uncertainty enters the interpretation of M&S results mainly through the people performing the interpretation. Often SMEs are involved in M&S results interpretation. This introduces many of the uncertainties associated with SME use. Such uncertainties can arise from the SME perspective in regard to M&S system's intended use, SME bias or slant on the subject and differences among SMEs if more than one SME is involved.

3.3 Uncertainty Methods and Techniques

There are no reliable methods or techniques for estimating or bounding the magnitude of blind uncertainties, their impact on a model, its simulation, or on the simuland response [19]. As a result, the primary approach for dealing with blind uncertainties is to try to identify them through such techniques as: (a) redundant procedures and protocols for operations or analyses, (b) various software and hardware testing procedures, (c) use of different experimental facilities, (d) use of a variety of expert opinions, and (e) use of broader sampling procedures to try to detect a blind uncertainty.

For aleatory and recognized epistemic uncertainties of M&S systems there are various kinds of quantification methods and techniques available. However, there is yet no real taxonomy available of possible UQ techniques and method to can be applied to M&S systems in general. The UQ techniques and methods available are tightly coupled to specific M&S systems, applications and problem domains. Usually these UQ techniques and methods can be classified in the following broad categories as discussed in the next subsections.

3.3.1 Sensitivity Analysis

Sensitivity analysis methods and techniques are one of the most popular UQ approaches and study how the uncertainty in the M&S system can be apportioned to different sources of uncertainty in its input variables and model parameters. Sensitivity analysis also addresses how M&S results depend on assumptions, or mathematical models, for the M&S intended use. Sensitivity analysis informs the user about what factors are the most important or most contributing to the uncertainty of the M&S results.

In general all sensitivity methods and techniques adhere to the following outline:

1) Quantify the uncertainty in each input (e.g. ranges, probability distributions).

2) Identify the M&S results to be analyzed. This M&S property of interest should ideally have a direct relation to the real world problem tackled by the M&S system.

3) Run the M&S system a number of times using some design of experiment dictated by the method of choice and the input uncertainty.

4) Using the resulting simulation outputs calculate the sensitivity M&S property of interest due to uncertainties in the input.

There are a large number of methods and techniques to perform a sensitivity analysis [19],[20]. These can be divided into local and global sensitivity analysis. Local sensitivity is mostly based on a Taylor series approximation to the model under consideration. They are useful in exposing the effects of small perturbations from the base-case values at which the Taylor series is developed and subsequent uncertainty analyses based on variance propagation are straightforward. Global sensitivity analysis uses a variety of

methods to decompose the total variance of a model output into contributions of the input parameters [20]. Sensitivity indices are computed as the ratios of a partial variance contributed by a parameter of interest over the total variance of the output. Most of the methods examine the full range of each input parameter, the effects of parameter interactions can be determined, and the analysis does not depend on linearity assumptions.

3.3.2   Forward and Backward Uncertainty Propagation

Uncertainty quantification focusses on propagation of input/parametric and model-form uncertainties. Of both sources, thee input/parametric methods and techniques are the most widely developed. By propagated it is meant that the sources of uncertainty, wherever they originate, are mapped (most often model mathematically) to the uncertainties in the M&S results. Uncertainty quantification has gained popularity as methods and techniques to assess the effect of variability and lack of knowledge on the output of a simulation model [19],[20]. This problem is referred to as a forward uncertainty propagation problem. Existing forward uncertainty propagation approaches include probabilistic approaches and non-probabilistic approaches. Probabilistic approaches are the most rigorous and often used, and include among others [23]:

- **Simulation-based** methods like well know Monte Carlo simulations based on repeated random sampling. These methods use very many simulations runs order to obtain the distribution of an unknown probabilistic entity.

- **Local expansion-based** methods like Taylor series expansions and perturbation methods that are based on local linearization of the simuland behaviour.

- **Functional expansion-based** methods such as Neumann expansion, orthogonal or Karhunen-Loeve expansions, with polynomial chaos expansion and wavelet expansions as special cases.

For non-probabilistic approaches, interval analysis, Fuzzy theory, possibility theory and evidence theory are among the most widely used [24]. An example of evidence theory that is referred in literature the Dempster-Shafer theory [25]. As shown is this publication, Dempster-Shafer theory allows for the allocation of a probability mass to sets or intervals. Moreover, Dempster-Shafer theory does not require an assumption regarding the probability of the individual constituents of the set or interval. This is a potentially valuable tool for the evaluation of risk and reliability in engineering applications of M&S systems when it is not possible to obtain a precise measurement from experiments, or when knowledge is obtained from expert elicitation. An important aspect of this theory is the combination of evidence obtained from multiple sources and the modeling of conflict between them.

In contrast to the forward approach there exist also uncertainty methods and techniques where one starts from observed simuland performance (i.e. output properties of interest) and one infers the input quantities (i.e. backward uncertainty propagation). These methods are less popular and well developed for M&S uncertainty quantification, and are usually rooted in the Bayesian framework. Bayesian inference derives the posterior probability as a consequence of two antecedents, a prior probability and a "likelihood function" derived from a probability model for the data observed. Bayesian inference computes the posterior probability according to Bayes' rule [26]. The Bayesian framework is a well-studied and successful framework for inductive reasoning, which includes hypothesis testing and confirmation, parameter estimation, sequence prediction, classification, and regression [27]. As such Bayesian methods are also possible formal alternatives, hence more complex, for the more graphical V&V argumentation networks as suggested by GM-VV [1].

4.0   CHARACTERIZATION OF EVIDENCE SOLUTIONS AND CATALOGUE CONCEPT

The collective set of evidence solutions in the form of V&V and uncertainty quantification methods and techniques discussed in the previous section cover more than hundred different types of solutions that can be

selected and applied within a V&V project. Most of these methods and techniques have multiple derived evidence solutions for specific M&S technologies or applications, and new methods and techniques are continuously research and developed. Hence, the set of evidence solutions is large and still expanding, which makes it difficult to retrieve and select appropriate V&V and uncertainty quantification methods and techniques during a V&V project. Therefore, several attempts have been made to develop characterization schemes in support of retrieving and selecting such methods and techniques [7],[8],[10]. These schemes are than used to build a catalogue of reusable V&V and uncertainty quantification methods and techniques for multiple V&V projects. GM-VV supports this approach by means of its concept of V&V enterprise memory and associated enterprise level processes [2],[1].
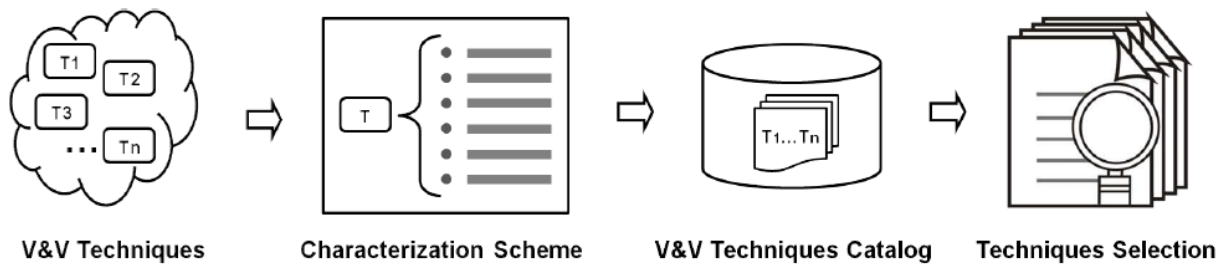

**Figure 3: V&V and UQ technique and method characterization
and application approach (Source: Wang [8]).**

GM-VV shows that V&V is a systematic effort that can be applied continuously throughout the whole M&S system life-cycle or at one or more specific moments and phase in this life-cycle. This requires that V&V and UQ techniques and methods can be selected in a systematic manner with the right information to justify the risk-based decision whether or not to apply a method or technique, its evidence provided and costs of its application matches the given V&V context (e.g. M&S use risk, available project budget, timeframe, technology and life-cycle phase). This justification must be embedded in the GM-VV argumentation structure (Figure 1). Based on this consideration, the attributes defined in a V&V and UQ method or technique characterization scheme should address at least the following three information categories according to GM-VV [3]:

1) Applicability M&S domain and technology, M&S or V&V life-cycle phase/activity

2) Expected evidential strength or convincing force,

3) Expected costs.

Wang provides a very detailed and field tested V&V technique and method characterisation schema instance [8]. This schema does however not address the second category as indicated by the GM-VV. The self-explanatory Table 2 shows the attributes addressed by Wang's schema.

Table 2: Wang's V&V Technique and method characterization schema.

Category	Attribute	Description
Applicability	V&V activity	*V&V stages in which the V&V technique can be applied, such as V&V of M&S requirements specification, V&V of conceptual model, V&V of executable model or V&V of data.*
	Object	*Artifacts that the V&V technique is able to examine, including work products, their documentation, and any other documents created in the course of a simulation project*
	Deficiency type	*Types of model deficiencies that the technique helps to detect.*
	Development paradigm	*Model development paradigm to which the technique is linked, e.g., component-based, object-oriented and agile development*
	Modelling formalism	*Modeling formalism to which the technique is linked, e.g., Petri nets, process algebra, Discrete Event System Specification (DEVS)*
	Simulation type	*Simulation types to which the technique is linked, e.g., stochastic, deterministic, discrete or continuous event simulation.*
	Simulation language	*Simulation language to which the technique is linked, including general-purpose programming languages, general purpose simulation language, and special purpose simulation packages.*
	Observable system	*Whether or not an observable system is required for applying the technique?*
	Model execution	*Whether model execution is required, when applying the technique.*
	Modality	*Is the technique subjective or objective?*
	Dependency	*Relationships of the technique with others.*
Cost	Data quality level	*Requirements for identifying, preparing, and applying test data, e.g., low, medium, high.*
	Formality level	*Level of using formalized structure and process, formal logic, and mathematics in the technique, e.g., low, medium, high.*
	Comprehensibility	*Effort required for understanding the technique, e.g., low, medium, high*
	Human resource	*Effort and time exposure required for applying the technique e.g., low, medium, high.*
	Type of application	*How is the technique applied, teamwork or single-handed?*
	Participant	*Role(s) involved in application of the technique.*
	Knowledge	*Knowledge required for applying the technique.*
	Experience	*Experience required for applying the technique.*

The Risk Based Methodology for Verification, Validation, and Accreditation (VV&A) developed by Johns Hopkins Applied Physics Lab provides a similar template as Wang [7]. However, it adds the following additional useful attributes that make the retrieval and selection of V&V and UQ methods and techniques a bit easier:

- **Keywords:** This field contains some words that might be used to search for information on the method or technique

- **Application process:** This field provides an overview of the steps to accomplish the method technique in sufficient detail to determine if the technique meets the constraints of the VV&A effort in terms of human skills and levels, schedule, budget, and risk

- **Method or technique assessment:** Assessments by SME on how and what evidential strength from the method or technique can be expected, for which circumstances or M&S life-cycle phases.

- **Input and output:** defines the format of the input and output of the information needed and produced by the method and technique.

- **Maturity level:** this attribute gives an indication how mature the method and technique is from a theoretical basis (e.g. solid, questionable or unknown), and the maturity of applying it by the own organization or others (e.g. extensive, limited, unknown) and guidance available (e.g. extensive, limited or unknown). This field gives a good indication how well developed the method and techniques are, and how wide it has been applied. Mature methods and techniques usually result in more reliable outcomes and less uncertainty.

- **References:** provides literature references upon which the method en techniques description is based, and with additional information how to apply the method or technique.

Based on templates as outlined V&V and UQ method and techniques catalogues can be developed by making an inventory of applicable V&V and UQ methods and techniques for a specific M&S application domain, technology, life-cycle model and other M&S organisation specific needs. Such a technique catalogue, implemented in a V&V enterprise repository, is not a static objects since M&S technology and objective change overtime, new V&V and UQ methods and techniques are developed, and existing are refined with information from applying them. Therefore, the catalogue should be updated on a regular basis with lessons-learned and other helpful information from V&V/M&S projects executed and from other public sources in the M&S community. This process requires an V&V knowledge management process to be implemented in the M&S or V&V organization using this catalogue. Such processes are specified at the GM-VV enterprise level [2] [5].

5.0   PRACTICAL EVIDENCE SOLUTIONS SELECTION APPROACH: RISK-BASED STRATEGIES

Risk-based strategies are needed during all V&V activities since it is practically impossible to verify or validate an entire M&S system in any project. Exhaustive verification and validation (i.e., 100% coverage of all aspects) of an M&S system exists only in theory [12] [19]; requiring infinite time and V&V resources. In practice, there is always a limited time and budget available for a V&V project. Moreover, there is always the pressure on the M&S system development to provide the needed capabilities (i.e., functionality) on time and usually more capabilities (i.e., nice-to-have features). In practice this means that the original time and budget allocated for performing V&V is often reduced by such M&S system development requests and constraints. This requires continuously balancing the time schedule, budget and resources available for V&V against what should and could be verified or validated throughout the life-time of a V&V project (Figure 4). The general rule is that the more risks and lower risk acceptance, the more rigorous V&V must be executed, hence more *advanced V&V and UQ methods and techniques* must be applied. However, these are usually also more costly, hence require more investments.
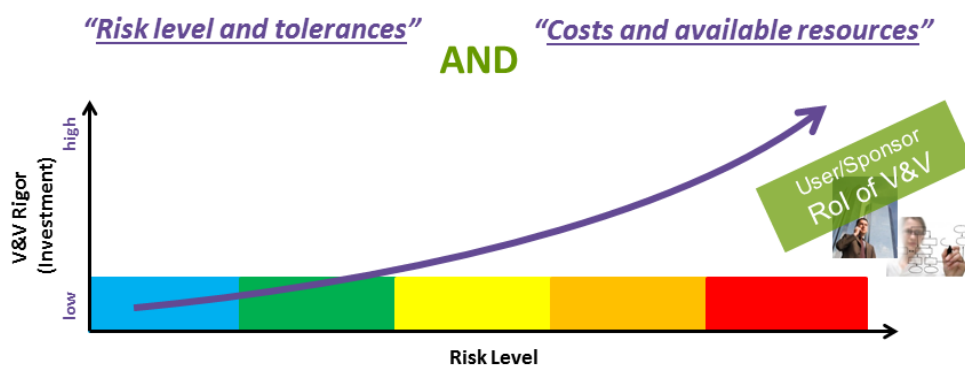

Figure 4: Risk-based balancing of V&V rigor versus costs.

Risk-based strategies are practical means of balancing, also referred by GM-VV as tailoring [4]. Risk-based V&V centres the verification and validation around the M&S use risks. M&S use risks are the risks directly related to usage of the M&S system and what the impact could be if the M&S system isn't (completely) fit for the intended use. Risk-based V&V identifies and analyses the M&S use risks, and aims at addressing these risks by guiding the technical V&V activities towards the level of risk of each identified risk item. A risk-based approach responds to these M&S use risks as follows:

- **Target technical V&V activities:** allocating V&V effort and *selecting V&V and UQ techniques* based on the level of risk of each identified risk item; matching the rigor and extensiveness of V&V techniques to the level of risks.

- **Sequencing of technical V&V activities:** prioritizing the risk items, starting with verifying and validating the most important M&S use risk items first and work down to the less important ones.

- **Reduction of technical V&V activities:** if the initial time, budget and resources are limited or are reduced throughout the life-time of the V&V project, *V&V and UQ methods and techniques* can be reduced in reverse-risk priority order, starting with least important ones.

- **Reporting of technical V&V results:** reporting V&V results in terms of residual M&S use risks (e.g., *V&V and UQ techniques and methods* executed, not executed, executed with limitations or omitted).

There exist not yet very many risk-based V&V strategies to support the selection of V&V and UQ methods and techniques during a V&V project. Wang does provide a high level approach, which is generally applicable but stays at an abstract conceptual level [8]. The M&S Use Risk Methodology (MURM) for V&V is directly applicable but a very computational formal and complex approach [7]. Within Q-tility a more pragmatic and semi-formal approach is used, which is rooted in the risk-based software testing approach, called PRISMA [9] [10].

In the Q-tility risk-based approach the V&V goal-claim network (Figure 1) and SME are used to identify the M&S system of interest (SoI) properties along with associated acceptability criteria for the intended use. Next these M&S SoI properties are scored by various SME, in a similar way as in PRISMA, for their likelihood of occurrence (i.e. technical risks) and impact (i.e. business risks) on the intended uses of the M&S system. For each M&S SoI, all averaged scores are computed and weighted to determine a position in a five scale risk-level matrix (Figure 5). In this way quick overview of the most critical M&S properties and associated acceptability can be identified. Depending on the M&S life-cycle phase a relative priority order can be constructed of the acceptability criteria. Where in case of V&V during M&S development phases, the acceptability with the highest technical risks are given priority, and are addressed first (Figure 6). In case of a post-hoc V&V during M&S system acceptance or usage phase, the acceptability with the highest business risks are given priority, and are addressed first.
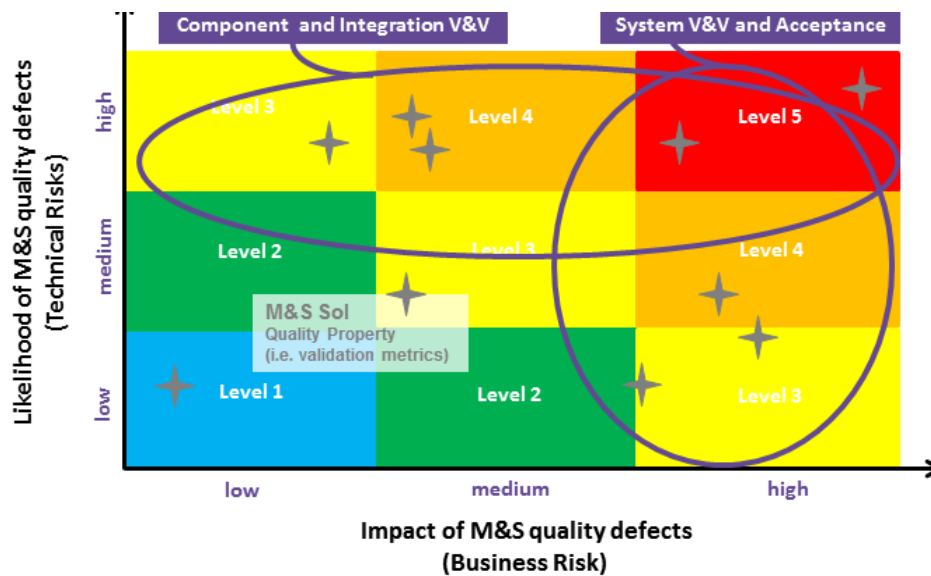
Figure 5: M&S Use-Risk Identification and Prioritization Schema (Source: Q-tility [10] [11]).
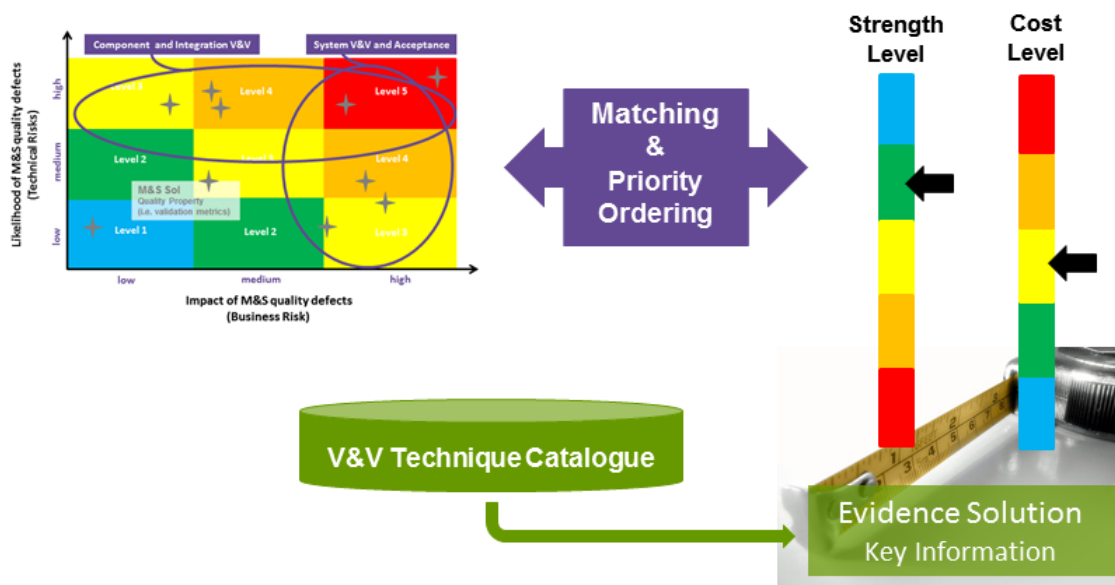

**Figure 6: M&S Use-Risk Based Selection of V&V and UQ
Methods and Techniques (Source: Q-tility [10] [11]).**

This priority order is used to make a first selection of the most appropriate V&V and UQ methods and techniques from the Q-tility V&V and UQ catalogue for each acceptability criteria. For this the V&V and UQ method and technique expected evidential strength and cost level are the key aspects for the initial selection (Figure 6). Highest priority acceptability criteria are given, relatively, to the most rigorous and cost effective V&V and UQ methods and techniques to acquire the evidence to demonstrate that the M&S system meets the acceptability criteria. This is done by help of SME and all decisions are underpinned with argumentation through expanding the V&V goal-claim network with these evidence solutions (Figure 1). Possible alternative evidence solutions are also identified, if available. Next practical feasibility and applicability (Table 2) aspects are determined and assessed for each initially selected V&V and UQ methods and techniques. Those not feasible or applicable are eliminated. Finally, since usually the total costs exceeds

the available V&V project resource an overall trade-off analysis is done over all feasible V&V and UQ methods and techniques to find the optimum balance of evidential strength and expenditure of available resources. Usually, several iterations have to be performed weighing also the alternatives. In this process the contribution and final selections is negotiated and consolidated with the V&V user/sponsor [5] [11]. Again SME input in this regards is vital and all decisions are underpinned with argumentation within the V&V goal-claim network (Figure 1).

6.0   V&V AND UNCERTAINTY QUANTIFICATION TOOLS

In this chapter a high level overview of possible useful tools to support the GM-VV implementation and the application of the V&V and uncertainty quantification method and techniques discussed in this paper. The list doesn't attempt to be exhaustive and complete, but provides a basis to search for V&V tools and select those that fit the specific needs of a V&V project or organization.

6.1   Process Support Tools

The V&V process support doesn't require complex or expensive tools. Q-tility V&V projects showed that general of the shelf office tools like word, spreadsheets and database programs are very useful if one uses them in combination with GM-VV based information artefact templates [2],[6],[11],[14]. Furthermore, Mindjet (www.mindjet.com) or any other mind-map software application is a useful tool to visualize SME and AOCS stakeholder discussion results during the V&V project. If more formal V&V process modelling and guidance is required to document and apply a GM-VV based V&V life-cycle model in an V&V enterprise setting like Q-tility [10], a tool like Eclipse Process Framework proves to be effective and efficient (https://www.eclipse.org/epf). Another tool to document V&V process information is the US-Navy Validation Documentation Tool (VDT) (http://nmso.navy.mil).

6.2   Argumentation Tools

Argumentation tools come in different flavours and levels of formality. The following tools have proven to be useful in the context of GM-VV applications conducted by Q-tility and other NMSG073 nations:

1)   **yEd** is a powerful freeware desktop application that can be used to quickly and effectively generate high-quality diagrams including argumentation network implementations [11],[13],[6]. It is possible to develop your own template structures and therefore often used in the research community (http://www.yworks.com/en/products_yed_about.html).

2)   **ASCE** is a commercial tool that implements Goal Structure Notation (GSN) and claim arguments and evidence (CAE) notations, which are often used in safety case community. With minor user-made modifications it fits the GM-VV (http://www.adelard.com/asce/choosing-asce/index.html).

More on argumentation tools can be found in the LS123 paper of Polacsek [21].

6.3   V&V Tools

A list of public and commercial tools that implement the majority of V&V methods and techniques discussed in Chapter 2 can be found in the Tools section of the US DoD VV&A Recommended Practice Guide [22]. This document is an excellent start point to discover more details about available V&V tools.

6.4   Uncertainty Quantification Tools

An internet survey shows various kinds of tools that implement uncertainty quantification methods and techniques as presented in Section 3.3. The following list of tools seems to be useful in the context of GM-VV applications:

1) **DAKOTA** is a large toolset with a large number of optimization and uncertainty quantification tools. There are many different approaches for parameter screening, design of experiments for computer models, response surface design (surrogate models), and reliability analysis. DAKOTA also has a variety of optimization techniques and an optimization under uncertainty component (http://dakota.sandia.gov/index.html).

2) **GPMSA** is a toolbox for emulation (response surface), calibration (parameter estimation), prediction, validation, and sensitivity analysis, developed by the Statistical Science group at LANL (http://www.stat.lanl.gov/staff/DHigdon/Shortcourse.html).

3) **GLUE** is a tool that was developed to perform a Generalized Likelihood Uncertainty Estimation on hydrological model (http://www.paramo.be/R/GLUE.html).

4) **NLFIT** is a tool that implements Bayesian nonlinear regression techniques for uncertainty quantification (http://www.eng.newcastle.edu.au/~cegak).

5) **PEST** is an open-source, public-domain software suite that allows model-independent parameter estimation and parameter/predictive-uncertainty analysis (http://www.pesthomepage.org/Downloads.php).

6) **PSUADE** is a software tool that is used to study the relationships between the inputs and outputs of general simulation models for the purpose of performing uncertainty and sensitivity analyses on simulation models. PSUADE is targeted for simulation models that are expensive to evaluate, such as large scale multi-physics models. The software has enriched sets of sampling and analysis tools. (https://computation.llnl.gov/casc/uncertainty_quantification)

7) **SIMLAB** is a tool that offers a programming and development environment aimed to facilitate the integration of sensitivity analysis features into user's modeling software (http://simlab.jrc.ec.europa.eu).

8) **UCODE** is a tool developed to perform inverse modeling, but also includes sensitivity analysis; data needs assessment; calibration; prediction; and uncertainty analysis (http://igwmc.mines.edu/freeware/ucode/?CMSPAGE=igwmc/freeware/ucode).

7.0 CONCLUSION

This paper presented an overview of the various verification, validation and uncertainty quantification methods, techniques and tools that are available in literature and on the internet. The paper showed how these fit in and can be used in the GM-VV implementation framework architecture and underlying risk-based V&V tailoring strategies. The work presented here is based on original papers of the authors and others, and summarizes those based on the authors experience in performing V&V projects. The interested reader is referred to these documents as listed in the reference section.

8.0 ACKNOWLEDGEMENT

The authors would like to thank Zhongshi Wang of the Bundeswehr University Munich and Dr. Mikel Petty of the University of Alabama, Center for Modeling, Simulation, and Analysis, Dr. Dale Pace of the Johns Hopkins Applied Physics Laboratory for their contribution to this paper in the form of making their work available to use and the rest of the NATO M&S community.

9.0 REFERENCES

[1] SISO, "*GM-VV Vol. 1: Introduction and Overview*", SISO-GUIDE-001.1-2012, October 2012.

[2] SISO, "*GM-VV Vol. 2: Implementation Guide*", SISO-GUIDE-001.2-2013, June 2013.

[3] SISO, "*GM-VV Vol. 3: Reference Manual*", SISO-REF-039.3-2013, October 2013.

[4] Roza M, GM-VV Part 1: Introduction and Concepts, NATO STO LS123, April 2013.

[5] Roza M, GM-VV Part 2: Processes, Information Artefacts and Roles, NATO STO LS123, April 2013.

[6] Voogd J, GM-VV Part 3: Implementation Guidance for M&S Projects and Organizations, NATO STO LS123, April 2013.

[7] Risk-Based Methodology for Verification, Validation and Accreditation (VV&A): The M&S Use Risk Methodology (MURM), Johns Hopkins University Applied Physics Laboratory Report, NSAD-R-2011-011, April 2011.

[8] Wang, Z., Selecting Verification and Validation Techniques for Simulation Projects: A Planning and Tailoring Strategy. Proceedings of the 2013 Winter Simulation Conference. Download from: www.wintersim.org.

[9] Van Veenendaal E, Practical Risk-Based Testing: The PRISMA Approach. UTN Publishers, 's-Hertogenbosch, The Netherlands, 2012.

[10] Q-tility, www.q-tility.nl.

[11] Roza M, A. Lemmers, F. van Heck, 'Verification and Validation of Distributed Air Operation Mission Training Environment', AIAA modelling and simulation technologies conference, paper AIAA 2013-5073, Boston, August 2013.

[12] Roza, Z.C., Simulation Fidelity Theory and Practice: A Unified Approach to Defining, Specifying and Measuring the Realism of Simulations, Delft University Press, Delft, January 2005.

[13] Buiel E., J. Voogd, Maintaining public order in a virtual world, NATO M&S Conference, 2013.

[14] Balci, "Verification, Validation, and Accreditation of Simulation Models", Proceedings of the 1997 Winter Simulation Conference, ed. S. Andradhttir, K. J. Healy, D. H. Withers, and B. L. Nelson.

[15] Sargent, "Verification and Validation of Simulation Models", Proceedings of the 2010 Winter Simulation Conference. Download from: www.wintersim.org.

[16] DoD, "V&V Techniques", RPG Reference Document, 2001, vva.msco.mil.

[17] Petty D., "Model Verification and Validation Methods", I/ITSEC tutorial 1306, 2013. Download from: http://www.uah.edu/cmsa/downloads.

[18] Pace. D.K., Comprehensive consideration of uncertainty in simulation use, Journal of Defense Modeling and Simulation, I-14, Society of Computer Simulation, 2012.

[19] Oberkampf W.L., C.J. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, 2010.

[20] Satelli A., When all Models Are Wrong, Issues in Science and Technology, pp.79-85, Winter 2014.

[21] Polascek T., V&V argumentation structures and their application, NATO STO LS123, April 2013.

[22] US DoD, DMSO, "Recommended Practice for VV&A", Build 3.0, 2006.

[23] S. H. Lee and W. Chen, A comparative study of uncertainty propagation methods for black-box-type problems, Structural and Multidisciplinary Optimization Volume 37, Number 3 (2009), 239-253, DOI: 10.1007/s00158-008-0234-7.

[24] Jaulin, L.; Kieffer, M.; Didrit, O.; Walter, E. (2001). Applied Interval Analysis. Berlin: Springer. ISBN 1-85233-219-0.

[25] Sentz K, Combination of Evidence in Dempster-Shafer Theory, SANDIA Labs Report, SAND 2002-0835, April 2002.

[26] The Problem of Old Evidence", in §5 of "On Universal Prediction and Bayesian Confirmation", M. Hutter – Theoretical Computer Science, 2007 – Elsevier.

[27] Dawid, A. P. (2001) "Bayes' Theorem and Weighing Evidence by Juries. University College London, 2011.