The GM-VV Tailored for a Naval Ship-Handling Training Simulation

Jeroen Voogd
TNO - Defence, Security and Safety
P.O. Box 96864, 2509 JG
The Hague, The Netherlands
jeroen.voogd@tno.nl

Manfred Roza
NLR – Training and Simulation Systems - Application, Validation and Quality Assurance
Amsterdam, The Netherlands
manfred.roza@nlr.nl

Ad van Lier
Simulation Expertise Centre, Netherlands MoD
The Hague, The Netherlands
Ac.vanLier@mindef.nl

Keywords:
verification, validation, GM-VV, motion simulation

ABSTRACT: *The Generic Methodology for Verification and Validation (GM-VV) is a generic and comprehensive V&V methodology for acceptance of M&S assets. The GM-VV methodology is currently prepared for standardization within the Simulation Interoperability Standards Organization (SISO), and is at the same time under consideration by defense directorates of various nations to be incorporated as part of their M&S policies. The GM-VV provides a conceptual and implementation framework to efficiently develop argumentation to justify why M&S assets are acceptable or not acceptable for a specific intended use. This argumentation is intended to support stakeholders in their risk-analysis based decision-making process on the development, application and reuse of such M&S assets. GM-VV is a generically applicable methodology which means that it needs to be tailored to fit the specific V&V needs of an M&S organization, project or application domain. GM-VV offers as an inherent part of its framework a set of principles to guide the tailoring process.*

This paper illustrates how the GM-VV tailoring principles have been applied to create a specific V&V solution for a Navy Ship-Handling training simulation. This illustration is mainly based on a recent research project for the Royal Netherlands Navy while using a novel motion-based simulator. This project involved the V&V of developing a prototype simulator for the intended use within a heavy weather ship-handling training program for navy officers.

1 Introduction

Modeling and Simulation (M&S) has become an integral part of many training and education programs. For some domains (e.g. aerospace, defense) the risks of negative training effects are too high to not invest in a rigorous Verification and Validation (V&V) of the simulation based training assets. Experience shows, however, that V&V is often more of an afterthought than a built-in part of any M&S development and procurement policy. This is due to the fact that V&V for M&S is still a relatively new field of technology and practice, with many very divergent opinions. The choice which method for V&V works best in a given situation depends on the individual needs and constraints of an M&S organization, project, application domain or technology. Therefore, many different approaches to V&V exist that rely on a wide variety of different V&V terms, concepts, products, processes, tools or techniques. In many cases the resulting proliferation restricted or even worked against the transition of V&V assets and results from one M&S organization, project, technology or application domain to the other. This context was the key driver behind the development of the Generic Methodology for Verification and Validation (GM-VV).

The GM-VV development started in an international joint project, called REVVA, and is now continued within the NATO-MSG-073 task group. This cooperative effort of multiple nations (CAN, FRA, GER, NLD, SWE and TUR) aims at delivering a common framework for V&V of models, simulations and data, which will be shared between these nation's defense organizations. The GM-VV is currently prepared for standardization within the Simulation Interoperability Standards Organization (SISO), and is at the same time under consideration by various national defense directorates (DoD, MoD, etc) to be incorporated as a part of their M&S policies.

Section 2 of this paper gives a brief overview of the GM-VV common framework for V&V. Next, the paper introduces the context of the V&V study that has been conducted in the Netherlands using the GM-VV: the assessment of whether physical motion is important for simulation based training of ship handling in heavy weather situations. (Section 3). This paper exemplifies how GM-VV has been tailored and applied for this particular case (Section 4 and 5). Finally, in Section 6 the paper presents some results focusing on lessons-learned and recommended practices for using GM-VV.

The case as described in section 3 is actually not yet completely finished. Although the experiments were conducted in December 2010, not all M&S results have been processed yet. Hence the VV&A process was not completely finished when this paper was written. For the context of this paper this is not a handicap since the general approach to the VV&A work is the topic of the paper, rather than the overall experimental result.

2 GM-VV Overview

The GM-VV [1][7] provides a generic framework to efficiently develop an argumentation to justify why M&S assets (e.g. models, results) are believed to be acceptable, or not acceptable, for a specific intended use. This argumentation, in the form of an acceptance recommendation, is intended to be used by M&S stakeholders in their acceptance decision making process on such assets.

GM-VV attains its generic quality by means of a reference model approach, instead of trying to cover or merge all possible and existing V&V methods into a single one-size-fits-all V&V method implementation. This means that the GM-VV is not directly tied to any specific M&S application domain, standard, technology, organization or other distinctive M&S implementation details for V&V. The GM-VV seeks to provide common semantics and components for V&V that can be used unambiguously across and between different M&S organizations, projects, technology or application domains. Therefore, the GM-VV framework is an abstract framework that consists of two parts. The conceptual frame work provides unifying concepts to facilitate communication, common understanding and execution of V&V within an M&S context. The implementation framework translates these concepts into a set of generic building blocks for the development of consistent V&V method implementations supporting an individual M&S organization, project, and technology or application domain. GM-VV provides tailoring principles and guidance to develop and cost-efficiently apply such V&V method instantiations.

2.1 GM-VV Conceptual Framework

The basic premise of the GM-VV is that models and simulations are always developed and employed to fulfill the specific needs of their stakeholders (e.g. users/sponsors, trainers, analysts, decision makers). The GM-VV assumes that VV&A always takes place within such a larger context and uses a four-world view of M&S based problem solving to structure this context (Figure 1) [1]. These four worlds cover the whole life-cycle from Real World need to operational usage. GM-VV defines a VV&A world in parallel. Within this world the VV&A efforts take place. Depending on the VV&A requirements, the VV&A effort could span the whole or specific M&S life-cycle phase of the four worlds or could focus on one specific or multiple M&S products.

The objective of the VV&A world is to convincingly show that an M&S asset will satisfy its intended use inside the four world context. This objective is articulated as a set of acceptability criteria for the asset. For these criteria evidence must be collected to demonstrate their satisfaction by the M&S asset. The GM-VV identifies three classes of M&S properties for which acceptability criteria can be defined:

- Utility of the M&S asset (e.g. value, cost, risk)
- Validity of the modeled or simulated real-world
- Correctness of the M&S asset implementation

Based on how well the M&S asset satisfies these defined acceptability criteria, a recommendation can be made regarding the acceptability of the asset for its intended use. However, to make an acceptance decision one also needs to know the convincing force of this acceptance recommendation. For this purpose, the GM-VV identifies quality properties for the acceptability criteria; the process of developing and demonstrating criteria to be met satisfactorily (e.g. rigor, evidential value, uncertainties).

Developing an acceptance recommendation may involve the identification and definition of many interdependent acceptability criteria, for which many different items of evidence must be collected and assessed to make acceptability claims based on their satisfaction. Such items of evidence will vary in convincing force and some may even contradict other evidence. This is influenced by the design and implementation of the experimental frame used to collect this evidence. The GM-VV provides a VV&A Goal-Claim network approach to perform this effort in a structured manner and assure that the used reasoning in here is transparent, traceable and reproducible, see Figure 2. The VV&A Goal-Claim network as such encapsulates, manages and consolidates all underlying evidence and argumentation necessary for developing an appropriate and defensible acceptance recommendation.

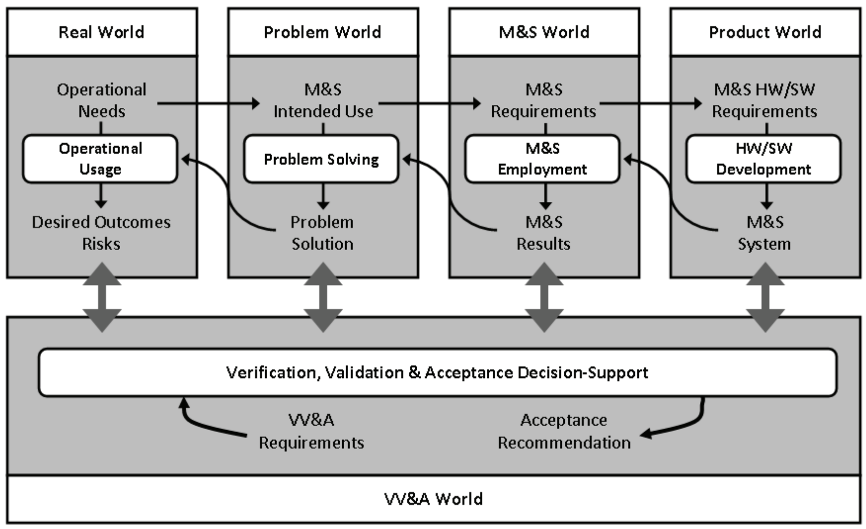


Figure 1 Four Worlds of M&S Based Problem Solving

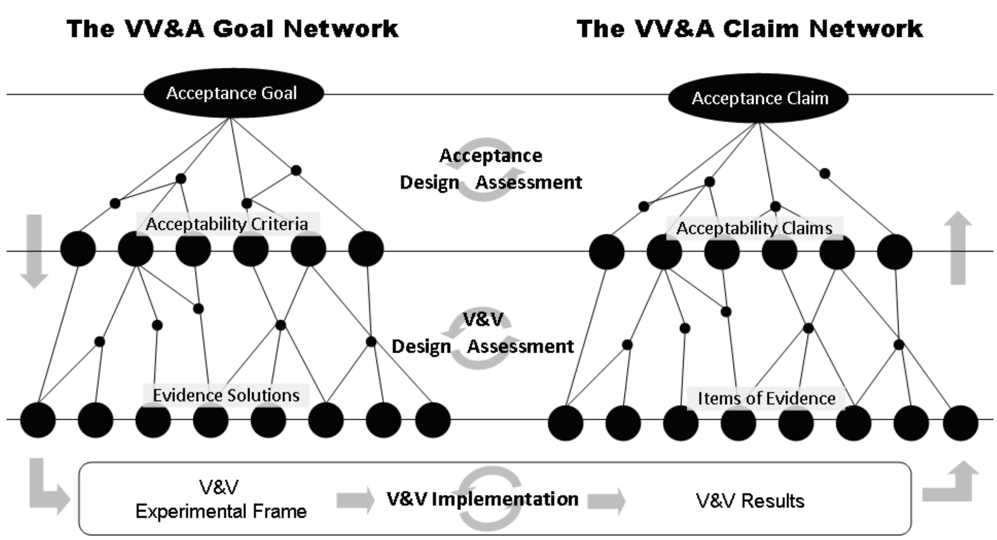


Figure 2 VV&A Goal – Claim Network Structure

To facilitate the efficiency and quality the VV&A efforts mentioned before, these efforts should be executed in an organized way in the VV&A world. Therefore, the GM-VV defines a managed project, the VV&A project, to develop and deliver an acceptance recommendation for an M&S asset. In support of a VV&A project, the GM-VV defines an organizational structure, the VV&A enterprise, which establishes, directs and enables the execution of VV&A projects. More importantly it retains information from past and current efforts to support the cost-effective execution of future VV&A work. Such information could be for example M&S technology or application domain specific recommended practices, acceptability criteria, VV&A Goal-Claim Network design patterns, V&V methods, tools and techniques.

2.2 GM-VV Implementation Framework

The GM-VV implementation framework translates the GM-VV basic concepts into a set of generic VV&A building blocks or components. These components are classified in three interrelated dimensions: product, process and organization. These GM-VV components are intended to be used and combined to implement tailored VV&A solutions that fit the needs of any particular M&S organization, application, and technology or problem domain. These components are classified in the following three interrelated dimensions.

<u>GM-VV product dimension</u> includes VV&A products that may be developed and used throughout a VV&A effort. These products are grouped into project management, technical and support products. Project management products can be used for the establishment and management of a VV&A project. Technical products can be used for the development of evidence and argumentation in support of the acceptance decision. Support products can be used for the information, knowledge and configuration management of the previous two groups of products. Since all these products are abstract information products, they can have multiple instances, representational and documentation formats.

<u>GM-VV process dimension</u> includes processes related to the life-cycle of VV&A products. The lifecycle processes deliver the GM-VV product dimension products. The GM-VV life-cycle processes are grouped into project management, technical and support processes. Project management processes can be used to manage VV&A projects. Technical processes can be used to develop acceptability criteria, evidence and argumentation to support acceptance recommendations. Support processes can be used to establish the organizational environment in which the project management and technical processes should be conducted. The processes can be carried out recursively, concurrently and iteratively within and between organizations or projects.

<u>GM-VV organization dimension</u> includes the components that facilitate the organization of VV&A enterprises and VV&A projects, which are specified in terms of roles played either by people or by organizations. These roles could be played either by separate organizations or people.

2.3 Tailoring principles

The GM-VV is intended as a generic, high-level implementation framework for VV&A, which should be tailored or "customized" for each individual M&S organization, project or application domain. The basic premise of the GM-VV tailoring concept is that the GM-VV should first be cast into a tangible VV&A method fit for an organization or application domain, and secondly this instance should be optimized for a VV&A project. The objective of this tailoring is to adapt the GM-VV products, processes and organization, to satisfy the specific requirements and constraints in the environment in which GM-VV is applied. The GM-VV tailoring process applies four basic tailoring approaches:

- Extension: adding elements not specified in the GM-V&V (e.g. additional products.)
- Reduction: cutting out GM-VV elements (e.g. activities and tasks.)
- Specialization: adaptation of GM-VV elements (e.g. using domain specific V&V methods.)
- Balancing: adaptation to find optimum cost-benefit-ratio (e.g. M&S use-risk and project resources.)

The result of a successful implementation of the GM-VV tailoring process is a modified or new VV&A method instance according to the GM-VV. This consists of concrete VV&A organization, products and processes, which should achieve the VV&A purposes of an M&S organization, project, technology or application domain.

3 Case: Heavy Weather Ship Handling

In order to be able to operate effectively and safely the Royal Netherlands Navy needs well trained personnel and appropriate doctrine. Currently no specific training for Heavy Weather Ship Handling (HWSH) is available. Learning to handle ships in heavy weather is learned on the job in real world situations under guidance of experienced officers. The navy has a lot of practical experience in these real world situations, but training and doctrine evaluation in a land based simulation would make training safer, more cost effective and save time. Currently the navy uses a Full Mission Bridge Simulator (FMBS) in educational programs, however that simulator is fixed based. In a fixed based simulator many of the necessary procedures can be practiced, but the question remains to what extent one can perform the same procedures during real world heavy weather situations?

The Netherlands Defence Materiel Organisation (DMO) is responsible for all materiel within the defense organization: from procurement and maintenance to disposal. The DMO questions whether or not the FMBS is sufficient in case the navy wants to offer HWSH as part of their educational program and as a doctrine identification and evaluation aid. Part of that question deals with the issue of whether a motion base is beneficial or even necessary for a HWSH simulator? To answer that specific question the DMO tasked TNO, the Netherlands Organization for Applied Physics research [6], to conduct a scientific experiment in a controlled environment. TNO is an independent research organization and a strategic partner of the Netherlands Ministry of Defence.

TNO designed an experiment to determine training effects due to motion simulation via an in-simulator comparison approach. In this experimental design two groups of test subjects were used, one that is trained with motion and one without motion. All test subjects followed the same test sequence consisting of a habituation period followed by a pre-test, the training and a post-test. Both groups did the habituation, testing and training in the same simulator, the latter either with or without motion simulation. The scenario consisted of a number of tasks that also under heavy weather conditions needed to be performed (following a ship, changing course, making a 180 turn.) During the experiment subjective, subject matter expert (SME), and objective measurements were taken to assess the test subject task performance.

As the simulator being used is a generic R&D device known as DESDEMONA, it needed to be configured for the specific task at hand. For the development of the M&S system TNO subcontracted two partner organizations: MARIN and Desdemona Ltd. MARIN is the Dutch Maritime Research Institute which provided an extensively verified ship motion dynamics model including wave dynamics, called FREDYN [2] suitable for simulating extreme motions in the nonlinear time domain. They also delivered a maritime simulation environment with controls, displays, visuals, etc. to be directly coupled to the Desdemona motion simulator. Desdemona Ltd. is the company which exploits the advanced motion simulator system of the same name: a six degrees of freedom motion simulator, disorientation trainer and research lab, all in one [5].

Based on the experimental design a Conceptual Model (CM) for the M&S environment was constructed in cooperation with all parties, together with DMO and navy officers with relevant experience. The CM encompasses the ship, its environment, and tasks to be executed by the ships crew. The minehunter was chosen for this experiment because of the behaviour of the ship: heavy weather has a large impact on the selected ship's type motion and handling characteristics. Another more practical reason was that a suitable ship motion dynamics model was available. Desdemona, through its design, however has a significant limitation; only one person can be seated. As a result the choice was made to use the officer of the watch as the test subject, and to place him in the position of the helmsman. The simulated sea state during the experiment is set between 4 to 6 in a blue water environment.

The M&S system implementation consisted of the Desdemona motion simulator fitted with a simple mock up of a ships bridge with a field of view of about 180 degrees. Due to the given platform physical limitations, only essential ship controls and navigational instruments were made available. This was considered as sufficient as the experimental frame focused on the influence of motion on the primary control task training. The ship's motion was calculated by the FREDYN model. Prior to the experiment, the M&S system was evaluated and its motion cueing algorithms were fine-tuned by navy officers who have served many years on the minehunter being simulated.

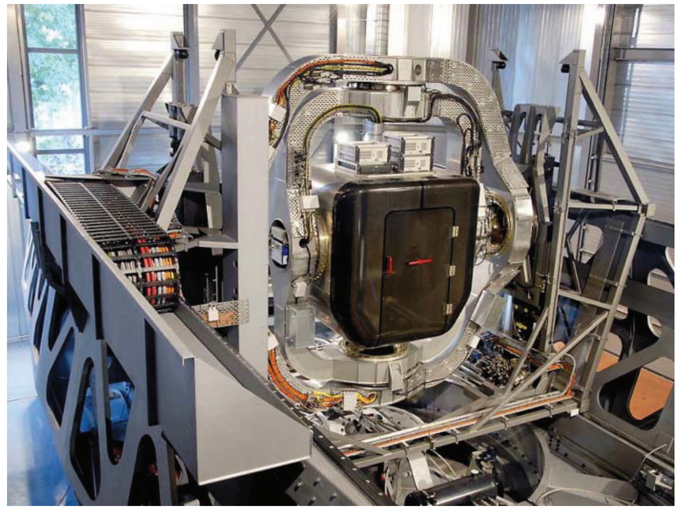
 

Figure 3: Desdemona motion simulator (right) with test subject (left)

4 Case instantiation and tailoring

The GM-VV instantiation started with the identification of the VV&A User/Sponsor role for the HWSH experiment. In practice this role proved to be two separate roles. The VV&A User was a person whose objective was to obtain a decisive acceptance recommendation for the training of HWSH and to identify and evaluate doctrines, either with or without motion. The VV&A Sponsor was a person whose interest was to show to the defense organization the importance and benefits of V&V. Since they both work for the DMO and try to obtain synergetic effects, they are in this paper treated as one role. The GM-VV four world view of M&S based problem solving was then used as the start point of the GM-VV tailoring process to create a V&V method that fitted the needs of this VV&A User/Sponsor (Figure 1).

The HWSH experiment focuses on determining whether physical motion is important in simulation based training and doctrine evaluation in the Problem World. Therefore, in Figure 1 the arrow from 'M&S Results' to 'Problem Solving' must be interpreted as an advice to the Royal Netherlands Navy whether the simulation should include physical motion in case the navy decides to use simulation to support HWSH training or doctrine evaluation. The V&V is intended to make sure that the advice is sound and thus focuses on the 'M&S Results' and not - as is normally the case - on the arrow from 'Problem Solution' to 'Operational Usage'.

For a decent V&V recommendation, the resources spend on this V&V effort must be in balance with the real world risk of adhering to the advice. For the HWSH experiment context, an erroneous advice claiming that motion *is* necessary, would result in the acquisition of too expensive simulations. An erroneous advice claiming that motion *is not* necessary, would result in personnel less well trained than achievable and missing or insufficient doctrines that might compromise the safety of ship operations and its crew. Based on these considerations a budget for V&V was allocated by the V&V User/Sponsors.

In order to build an acceptance recommendation for the M&S Results, the V&V activities also focused on the "M&S Employment" including the experiment's execution, which measurements were performed, and how they were performed. Following the arrows further back in Figure 1 acceptability criteria were formulated for the M&S System. We could have followed the arrows back to "M&S Requirements", "HW/SW Development" and even further back. It was however decided to start with the combined examination of "M&S Results", "M&S Employment" and "M&S System", keeping in mind that if insufficient concluding evidence could be found for these arrows, these can be traced back to perform further V&V.

4.1 Tailoring of the Organization Dimension

The basis for the V&V organization instantiation and tailoring is the fact that the HWSH experiment team had no a-priori interest in one particular experimental outcome. Therefore, it was decided by the VV&A User/Sponsor that independent V&V was not necessary. This meant that experimental team members could be part of the VV&A project team. For this HWSH experiment all VV&A project level roles were instantiated to create the VV&A team. The VV&A project team was managed by one of the authors who did not participate in the activities of the experiment team. The one role at enterprise level, however, was not instantiated because no V&V enterprise exists yet. This is a form of tailoring by reduction.

Tailoring of the organization was mainly applied to the subject matter experts (SME) employed by the V&V Leader in the construction of the VV&A Goal-Claim Network and the execution of the V&V experimental frame. The role as an V&V implementer could be assigned to SME's of all parties involved, depending on the acceptance criterion at hand. Due to the wide spectrum of available SME's, resource limitations and the level of risk, no external V&V implementers were involved. This is a form of tailoring by balancing.

4.2 Tailoring of the Product Dimension

The VV&A User/Sponsor focus was on an overall acceptance recommendation, including evidence for a number of specific questions where he expected to be faced with task critical conditions. For efficiency reasons the VV&A User/Sponsor and V&V project manager made a collective decision to document the results in one single document containing most of the technical products defined by the GM-VV. This is a form of tailoring by specialization. An overview of these resulting products is presented in section 8.

The VV&A Goal-Claim Network has been constructed using the ASCE tool [3]. The complete VV&A Goal-Claim Network has been build iteratively and used extensively in discussions with SME and other stakeholders. For the final report however only an overview of the most important findings and those findings requested by the VV&A User/Sponsor from this network will end-up in the recommendation report to be delivered.

4.3 Tailoring of the Process Dimension

The most important input for tailoring the processes was the requirement that the VV&A had to be executed concurrently with both the development and running of the experiment. This is a form of tailoring by specialization. The mutual benefit of this concurrent VV&A was that key SMEs were continuously available and several large meetings with all SMEs could be used in the construction of the VV&A Goal-Claim Network. On the other hand the experiment team could look at the criteria that were being set and make sure their M&S system and experiment matched them.

The experiment itself was carried out in two consecutive weeks. In the first week the whole M&S system and experimental set-up was build and fine-tuned with the aid of experienced navy officers. Some of these officers had a large body of experience with the minehunter being simulated. In the second week the experiment was executed. The V&V required some of the measurements to be taken during the experiment, but also measurements to be taken before or after the actual experiment. For example one acceptance criterion is that experienced minehunter SMEs judge the simulator to be close enough to reality within the context of the experiment. The VV&A team was able to assess this aspect in the first week. Another set of acceptance criteria deals with processing of the experimental measurements. These need to be consistent before any meaningful conclusion can be drawn. This set of criteria can only be assessed by the VV&A team if the experiment's measurements have been processed.

The GM-VV draft Implementation Guide [4] defines many process activities and tasks. It is beyond the scope of this paper to present in detail how they were tailored. For this reason only high level examples of how the GM-VV tailoring principles were employed are given in the following paragraphs;

Process Tailoring by Reduction

For the processes some reduction tailoring was applied resulting in not all processes, activities or tasks listed in the GM-VV being executed. For example, since no enterprise organization was involved all processes, activities and tasks related to enterprise management were omitted.

Process Tailoring by Specialization

As stated above, the argumentation network has been constructed partly by having direct face-to-face meetings with SMEs, often also indirectly during larger project meetings where it was particularly important for the VV&A team to listen to the navy experts. In other words, many process activities and tasks have been executed in a way that is typical for doing concurrent VV&A in a setting found for experiments involving human test subjects in order to determine learning effects.

Overall VV&A work has been executed with varying levels of formality allowed by the required level of rigor and the relationship with the VV&A User/Sponsor. As an example the production of VV&A project reports with status and issues was not executed by formal written documents such as a V&V plan or report; instead continuous informal updates were given because of the frequent contact with the V&V User/Sponsor.

Process Tailoring by Balancing

During design of the VV&A Goal-Claim Network a decision needs to be made for each identified goal whether or not to continue decomposition of the goal in smaller sub-goals. Factors included in such a decision are:

- Availability of test methods to obtain evidence (e.g. performing measurements, literature study, SME opinion)
- Costs (budget, time, needed expertise, etc.) to execute the available test methods
- Expected convincing force of the obtained evidence from the test methods
- The availability of ways to decompose the goal into sub-goals with an estimation of the above stated factors applied to the sub-goals.

For the HWSH case the overall optimization by balancing resulted in a thorough study on the M&S system, the experimental set-up and the way the experiments were performed. The cost/benefit balancing was such that no V&V-initiated human factors studies were performed and no additional validation of e.g. FREDYN was executed.

5 VV&A Case-Study Results

The case as described in section 3 is actually not yet completely finished. Although the experiments were conducted in December 2010, not all M&S results have been processed yet. Hence the VV&A process was not completely finished when this paper was written. The VV&A Goal-Claim network, and ultimately the acceptance recommendation, still requires missing evidence on the experimental data processing and M&S results before it can be completed. For the context of this paper this is not a handicap since the general approach to the VV&A work is the topic of the paper, rather than the overall experimental result.

5.1 VV&A Goal-Claim Network

The VV&A Goal-Claim Network starts with the acceptance goal. This goal is to be understood within a certain explicitly given context. This top goal should encompass all of the relevant VV&A needs of the VV&A User/Sponsor. For the HWSH case the following context was specified:

- The navy wants to make sailing in high sea states sufficiently safe, given the operational tasks
- Improving safety, for the experiment at hand, focuses on training and doctrines
- Simulation can contribute to training of sailing in high sea states as well as identification and evaluation of relevant doctrines
- It is not clear whether or not physical motion plays a significant role in heavy weather simulation
- An experiment is performed to establish whether physical motion is significant

A goal consists of a number of information items: the part of the VV&A system of interest the goal is referring to, observables, the criticality which is indirectly derived from the M&S use risks, and most important, the proposition which states what needs to be shown of the observables of the system of interest under consideration. The proposition of the acceptance goal was formulated as: *"The results of the experiment are useful in the determination of the significance of physical motion in simulation of high sea states for training and doctrine evaluation"*. This proposition ultimately needs to be demonstrated with evidence. As can be expected, however, this top goal is still rather vague and high level. Therefore, no test method to obtain evidence is available. This problem is tackled by defining sub-goals via a decomposition strategy. For a strategy it must be shown that the decomposition is justified, including that it completely covers the parent goal. Also an inference rule is to be specified indicating how the satisfaction of the sub-goals determines the satisfaction of the parent goal. This is especially important if the satisfaction of one of the sub-goals already implies that the parent goal is also satisfied, or conversely if the rejection of one sub-goal is sufficient for the parent goal to be rejected.

Via a number of decompositions, goals are defined with propositions on a number of topics:

- the increase of performance,
- finding a significant difference between novice and experienced officers,
- and the ability for test subjects to better sense the boundary between safe and risky ship handling.

The above mentioned goals are all utility goals, see section 0. These goals are subsequently mapped to acceptance criteria on two main groups of goals: the experiment must be executed correctly and the experiment must deliver valid results.

Acceptance Criteria for Experimental Correctness

The assessment of the experimental correctness consists of the evaluation of acceptance criteria on:

- The relevance of the experiment for the test-subjects. If the test-subjects have no professional interest in HWSH they are unsuitable for the experiment. The Acceptance Criterion (AC) specified that the past or current occupation of all test subjects must be relevant for HWSH.
- The use of an experiment group and a control group. The experiment uses in-simulator comparison, and the AC therefore states that one group in the experiment undergoes the "intervention" phase and one group is used as "control". In this case all test subjects undergo the same phases (habituation, pre-test, training, post-test) but there is a difference in the training: one group is trained with motion switched on (intervention) and the other group was trained without motion (control).
- The number of test subjects used. In order for the results to be significant a sufficient number of test subjects must participate. Since two groups are used, both must have sufficient test subjects.
- The presence and order of phases in the experiment. This AC states that the defined phases must all be present and they must be executed in the correct order.
- The time allotted to the experimental phases. This AC states that for all phases sufficient time is available. For example at the end of habituation the test subject must be comfortable with the simulator.

Acceptance Criteria for Experimental Validity

The assessment of experimental validity consists of the evaluation of acceptance criteria on three main subjects, each with a number of sub-criteria:

- The simulator is realistic enough
    - The navy ship handling SME must judge the simulator as sufficiently realistic
    - The test subjects must judge the simulator as sufficiently realistic
    - A detailed examination of the simulator finds it to be sufficiently realistic

- The operational tasks the test subjects must perform are sufficiently realistic
    - The procedures the test subjects must perform are relevant (representative)
    - The setting in which the tasks are performed are relevant

- The human factor is taken into account
    - Performance is measured correctly
    - Workload is measured correctly
    - Situational awareness is measured correctly
    - Well being is measured correctly
    - Personal factors are measured correctly
    - All human factors (the previous bullets) are correctly taken into account during data processing

The sub criteria on human factors each have smaller criteria stating how these must be measured during the experiment, at which time during the experiment and on the consistency of the results. Note: for VV&A it does not matter what the measurements result in, just that they are consistent such that the results of the experiment are consistent.

The goal stating that a detailed examination of the simulator finds it to be sufficiently realistic, is further developed via decomposition. The result is presented in Figure 4. In this figure the rectangles are goals, the parallelograms the strategies with the description of how goals are decomposed into smaller more detailed goals. The highest goal in this figure is not the overall acceptance goal, it is only a part of the complete VV&A Goal-Claim Network. The bottom three nodes of each branch provide the V&V solutions: the method to be used for evidence collection (parallelogram), the precise description of how the method must be executed: who does what, when, and with what equipment, using which reference data, etc. (circle), and the V&V Results (square).

A CM was developed for the experiment and thoroughly checked by all involved parties. The strategy with which the evaluation of the realism of the simulator is decomposed is based on the Conceptual Model (CM).

1. the realism of the implementation of all elements of the CM that have been implemented in the simulator
2. justification of not implementing some elements of the CM
3. justification of all elements not in the CM but which are found in the simulator
4. realism of combinations of elements (both in and not in the CM).

The first point (the evaluation of the realism of implemented CM elements) contains goals, possibly sub-goals and acceptability criteria related to: FREDYN for the ship simulation model that is controlled by the test subject; Wave and wind model; Visualization of the 3D world; FREDYN used for simulating ship traffic (in this case the target ship that needs to be followed in the tracking task); Desdemona, the motion simulator; Sound, including slamming sounds; Controls: steering and speed; Position of test subject in Desdemona; Instruments, e.g. speed indicator; 3D visual model of minehunter; Instructor console and instructors; Scenario Implementation.

The second point (CM elements not implemented in the simulation) deals with all CM parts that have been mentioned by experienced navy officers, but which have been left out of the simulation. The missing parts were:

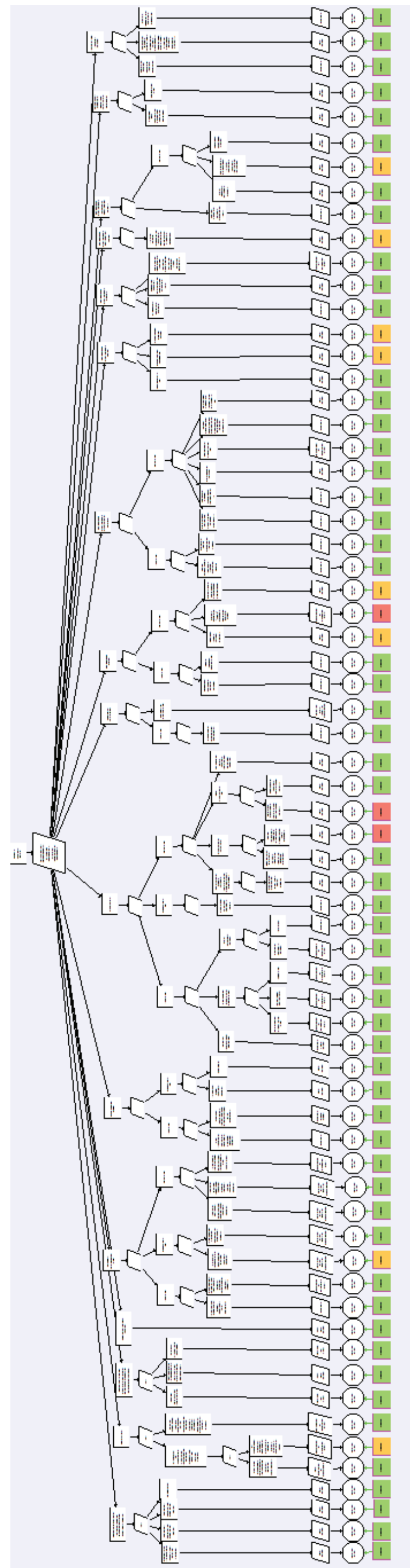
- the water model did not simulate current and swell,
- much less sensors and controls than on a real minehunter were presented and those that were are not exact replicas,
- there is no helmsman with which officers of the watch normally interact on a ships bridge.

The third point (elements present in the simulation but not in the CM) evaluates that there is no negative effect induced from:

- the sound of the motion simulator
- the position of the officer of the watch is more or less that of an helmsman including controls etc.
- an extra task was presented to the test subject in order to measure the workload: every few seconds a pedal had to be pressed when a small light went on.

The fourth point (combinations of elements) evaluates the combined effect of more than one element. One important overall criterion is the real time response: there should be an unnoticeably small delay from steering to visualization, sound and motion. Another criterion is the synchronization between the sounds, visuals and motion. If these are not synchronized they can lead for instance to simulator sickness.

Figure 4 (next column): Part of the VV&A Goal-Claim Network that deals with the evaluation of the realism of the simulation (about half of the complete VV&A Goal-Claim Network). It is deliberately made unreadable in this paper for reasons of confidentiality but gives an impression of the complexity and scale of the VV&A Goal-Claim Network for the HWSH experiment.



5.2 V&V Experimental Frame

The V&V experimental frame is defined by the last nodes on each branch of the VV&A Goal-Claim Network (Figure 4). The square at the very end of each branch contains the V&V Results. It is a summation of all results from executing the defined test methods. For the HWSH case the experimental frame specified a number of different test methods for obtaining V&V Results:

- Inspection is suitable in case an unambiguous AC (e.g. that some instrument must be present) that is easy to check. The cost of inspection is usually low and the residual uncertainty zero (it either *is* or *is not* present).
- Measurement is suitable for well defined AC but whose observable is difficult to see with the naked eye and thus requires a measurement. An example is the criterion that visuals, motion and sound must be synchronized within a specified fraction of a second. The measurement should not result in a too high degree of uncertainty and not cost too much. The residual uncertainty/cost ratio must be in balance with e.g. the maximum allowed residual uncertainty and available budget.
- Reuse of existing V&V results or other historical evidence for parts of simulation. The ship motion dynamics model FREDYN has already been extensively validated for a number of ship types. If the current use is close to the uses described in existing validation or test reports an argument may be constructed that this part is also appropriate for the current use. The residual uncertainty depends on the uncertainty in the judgment in the validation or test report and the uncertainty with which the claim can be made on the similarity in use.
- SME opinion can be used when correctness, validity or utility is difficult to measure rigorously and objectively. For example the feel of the motion of the simulated ship in the given environmental conditions is difficult to judge objectively. A quick and relatively cheap strategy is to use SME opinion. Some inherent uncertainty can be mitigated by using several independent experts and combine their results in some way.

The last nodes specifying the V&V experimental frame also contain a discussion on how the obtained V&V Results must be interpreted in order to judge if they can be used as items of evidence in the V&V Claim Network. The following color coding is used for the status of obtained V&V Results nodes: white: no V&V Results have been obtained yet. In our case only results of the data processing from the huge, still to be analyzed, collected data set are missing (not in Figure 4); green: V&V Results are present and usable as evidence, and it show that the AC to which it belongs is met; orange: V&V Results are present but some issues prevent it from being usable as evidence, it is not yet determined whether the AC is met or not. Additional V&V Results are needed; red: the V&V Results are usable as evidence and show that there is something wrong. The AC is not met. The impact of this failure is to be evaluated in the VV&A claim network.

5.3 V&V Claim network

Since not all V&V Results are available, the VV&A Claim Network has not yet been fully constructed. For the evaluation of the realism of the simulator, however, some preliminary results are already available. One particular example concerns the controls available to the test-subject. In a real minehunter a small steering wheel is used in much the same way as in a car (turning clockwise makes the ship go to starboard side). In the simulator mock-up, however, a handle was available. It had been implemented as pushing the handle to the left steers the ship to port side, pushing the handle to the left, however, is a clockwise motion with the hand. This caused some initial confusion in several test subjects. One might say that the V&V Results indicate that the AC on the controls now failed. However, the V&V Results on the experimental correctness, and more specifically whether the habituation period in which test subjects can familiarize with the simulator, indicate that the habituation period was sufficiently long in order for the test subjects not be hindered by the unintuitive steering control direction. Therefore overall one piece of pro and one piece of counter evidence is found. The overall conclusion is that the steering problem introduces no problem for the experimental outcome.

As apparent in Figure 4 evidence has been found that indicate that some AC are not met. The severity of that failure and how it aggregates up the VV&A claim network needs to be determined.

6 Discussion and conclusions

Based on the case-study, as executed so far, we feel that the GM-VV contains all necessary high level ingredients for a rigorous approach to VV&A. This abstract methodology, however, needs to instantiated and subsequently the instantiated VV&A method needs to be

tailored in order to fit to the needs of the VV&A project at hand.
Tailoring has been applied in several ways: during instantiation elements were added or removed from the default GM-VV. During the execution of the processes specialization has been applied. One of the main technical products, the VV&A Goal-Claim Network, is build with a continuous tailoring by balancing. Defining the V&V Experimental Frame also required extensive balancing. The GM-VV tailoring principles worked well and resulted in a practical application of the abstractly defined GM-VV.

Some parts of the results of the case-study may be re-usable for other VV&A projects dealing with VV&A of training simulation or experimentation. Examples are the specialization of the processes and parts of the VV&A Goal-Claim Network. Besides finishing the case-study, all reusable information and lessons-learned from this study will be compiled into a recommended practice guide.

7 References

[1] "GM-VV Introduction and Overview", SISO-GUIDE-00X.2-201x-DRAFT-V1.10.5, April 2011

[2] FREDYN, http://www.marin.nl/web/show/id=70052, accessed 24 April 2011

[3] ASCE, http://www.adelard.com/web/hnav/ASCE/, accessed 24 April 2011

[4] "GM-VV Implementation Guide", SISO-GUIDE-00X.3-201X-DRAFT-V1.6.1, January 2011

[5] Desdemona, http://www.desdemona.eu, accessed May 2011

[6] TNO, http://www.tno.nl, accessed May 2011

[7] SISO GM-VV Product Development Group, http://www.sisostds.org/StandardsActivities/DevelopmentGroups/GMVVPDGGenericMethodologyforVVAintheM.aspx, accessed May 2011

Acknowledgements

The authors wish to thank all people from TNO, MARIN and Desdemona involved in the experiment and who contributed to the V&V work. Special thanks to KLTZ den Arend, KLTZ Lambregts, LTZ 2 Beemsterboer, LTZ 3 Kauwoh and cadet Zwakman.

Author Biographies

JEROEN VOOGD is a member of the scientific staff at the Modeling, Simulation and Gaming Department of TNO Defense, Security and Safety. He holds a Ph.D. in Computational Physics from the University of Amsterdam in the field of modeling and simulation biophysical systems on parallel distributed computation platforms. His current research interests include decision support with simulation, validity and agent technology.

MANFRED ROZA is Sr. R&D manager at the Training, Simulation and Operator Performance department of the NLR. He holds a M.Sc. and Ph.D. in Aerospace M&S Technology from the Delft University of Technology. He specializes in requirements, quality assurance, verification and validation of models and simulation, in particular for simulation-based training systems. He was previously a member of the scientific staff at TNO, during which time he participated in the presented project.

AD VAN LIER is a member of the staff at the Simulation Expertise Centre of the Royal Netherlands Armed Forces. He holds an M.Sc. in Information Technology (University of Liverpool). He is tasked with standardization, interoperability, VV&A and re-use of M&S. Areas of personal interest focus on intelligent training within M&S systems.