Technical Evaluation Report

Prepared by Dr. Andreas Tolk
Virginia Modeling Analysis & Simulation Center (VMASC)
Old Dominion University
Norfolk, Virginia 23529
UNITED STATES

atolk@odu.edu

OVERVIEW

The NATO Modelling and Simulation Group (NMSG) Symposium (MSG-035) "Effectiveness of Modelling and Simulation – From Anecdotal to Substantive Evidence" was conducted in Warsaw, Poland from 13 to 14 October 2005. All sessions of the Conference were unclassified. The Conference audience of approximately persons included experts from NATO countries, Partners-for-Peace (PfP) nations, as well as invited nations.

Out of the more than 40 submitted abstracts, 17 Papers were selected for presentation. In addition, three invited papers and two keynote presentations were given. This technical evaluation report summarizes the core ideas and results presented in this wide variety of valuable contributions from NATO countries, PfP nations, and invited nations. Furthermore, the report provides an overview of the discussions conducted during the conference following each presentation and gives recommendations.

INTRODUCTION

The importance of Modelling and Simulation (M&S) in NATO and National activities has been well recognised for a number of years. On the technical side, advances in low cost, high power computers, graphics and telecommunications/networking advances became key enablers that provide opportunities to use distributed simulations, (including humans and hardware/live systems in the loop) in new and exciting ways. On the application side, developments such as summarized in the report RTO-MP-MSG-028 on last year's conference on "Modelling and Simulation to Address NATO's New and Existing Military Requirements" give examples for the great variety of requirements already dealt with in current M&S applications. This shows that M&S are indeed powerful tools assisting the search for improved operational effectiveness and yield value for money within NATO. However, many early M&S applications promised cost-effectiveness without delivering an objective proof in the end. In the light of ongoing force structure reductions and shrinking budgets, there is an increasing requirement to provide substantive evidence for the cost-effectiveness that M&S can provide.

The conference focused on papers that demonstrate and provide substantive evidence for the effectiveness of M&S in priority areas for NATO and the Nations. In addition, it addresses research, development and the application of models and simulation in the defence environment. The call for papers explicitly asked for papers focusing on how M&S can provide increased effectiveness to the NATO Response Force, Military Exercising, Training and Decision Support, Network Centric Warfare and Defence against Terrorism. In addition, the conference addressed the more general themes of research, development and M&S-related standards, such as the NATO Transformation process and implication for M&S, the NATO Joint Warfare Centre – current and future applications of M&S, the NATO Joint Analysis and Lessons Learned Centre – M&S Lessons Learned and Plans for Viking 05 Exercise. Based on their main topic, the selected papers were grouped into six categories coping with the methodology, case studies and assessment, evidence analysis, the way from anecdotal to substantive evidence and the results and way ahead.

This technical evaluation report does not intend to recite paper abstracts or give pure summaries. It wants to show the general developments and common ideas allowing the analyses of trends within NATO/PfP nations and other invited nations. To this end, discussion points, recommendations made during or after the presentation, and everything else of interest to the NATO M&S community is comprised with the section.

KEYNOTE PRESENTATION

As the academic keynote speaker Prof. Radosław Trębiński, Pro-Rector of the Military University of Technology (MUT), Technical, Educational and Scientific Centre of Polish Armed Forces, gave an overview of history and structure of the university. He presented the recent demographic data on students, educational tracks and ranking within the polish universities and relevant activities at MUT. Of particular interest for NATO MSG is that – although MTU focuses and excels in research on optoelectronics – there are R&D Laboratories for Combat Information System (CIS) Interoperability and Integration of Information Systems. MUT is very active in NATO activities and has representatives sitting in all RTO panels. The current research copes with a Polish Technological Platform of Security Systems and is conducted in a group with other national and international research organizations. The main objective is to establish a comprehensive and long-term strategy towards developing of main security systems technology involving both public and private sector representatives.

The second presentation introduced the NATO MSG programme Pathfinder, which is the flagship programme of NATO MSG initiated by the NATO M&S Master Plan. NATO's M&S Coordination Office (MSCO) Whitepaper summarizes the core idea as follows: *"The vision of the Pathfinder programme is to provide the technical capability for federations of national models and Decision Support Tools integrated and linked to NATO, and eventually national operational C3I systems, to exercise and train the Combined Joint Task Force (CJTF) Headquarter staff and Component Commands in the conduct of Crisis Response Operations (CRO)."* This vision shows the long-term goal and requires the orchestration of several contributing MSG activities; some of them will are topic of more detailed papers in this report. One of these activities is MSG-027 dealing with the definition of an "Integration Environment for the Multi-Purpose Application of Distributed Networked Simulations". This Integration Environment envisioned for the Pathfinder programme is a web-based facility that will leverage NATO and national M&S integration expertise. Intended users are national and NATO organisations responsible for the application, development and provision of simulation support to the allied forces. It will be usable in a distributed environment and capable of supporting collaborative High Level Architecture (HLA) federation development. It was pointed out that in order to migrate the different supporting technical activities into the overarching Pathfinder programme and to allow a gradual development and improvement, a rigorous standard management by NATO's MSG is required.

INVITED PRESENTATIONS

The symposium was launched with three invited presentations. The first two are coping with results from two relevant technical activities currently being conducted under the leadership of NATO MSG: the technical activity NMSG-031 dealing with the cost effectiveness of M&S and NMSG-026 evaluating the feasibility of M&S Processes in the Proposal of Early Warning Systems for Defence against Terrorism (DAT) and Crisis Management. The third invited presentation represents the customer of M&S by giving a user specific overview on the Viking 2005 exercise planning process.

The first presentation described the progress to date on MSG-031 "The Cost Effectiveness of Modelling and Simulation" and presented the results so far. The activity started with the observation that many programmes currently make a specific business case for any M&S outlay and there is a need for generic business case guidance. One reason for this is lack of evidence to sustain the case for M&S use. Therefore,

MSG-031 has set out to break down some of the barriers at present in the way of the extended use of M&S by identifying the areas of which are not cost effective to suggest further research and providing evidence of the areas that are cost effective. The objective is to provide a framework to collect evidence associated with the cost effective application of M&S based upon the NATO procurement process. (This focus on NATO's procurement process distinguishes this activity from the general evaluations on M&S economics, which are described in paper 20 of this report.) A first necessary step is mapping the NATO process to the national procurement process and to agree on form of metrics and aggregation of M&S related costs to facilitate comparison and share of data. To this end, MSG-031 started to agree on a common taxonomy for M&S classifications, defence applications for M&S and M&S categories. In order be able to collect relevant data, these taxonomy needs to made public and adopted by various users. Unfortunately, the taxonomy is partly MSG-031 specific and the evaluator assumes that major adjustments and alignments will be necessary before this will be the case. Furthermore, it is well known that it is possible to determine cost effectiveness by weighing the costs associated with M&S against the benefit accrued, but determining the contributing factors to cost is not trivial, as costs may have already been invested (sunk cost), or costs incurred no matter whether M&S is employed or whether a more traditional approach is used, or costs did not incurred because M&S is used, or costs only incurred because M&S is employed. Unfortunately, the study so far only states identified problem areas and the need for metrics. As such, the paper gives a valuable problem description and points to new research, but lacks reusable and applicable solutions so far, which is not – given the fact that it is an interim description of an ongoing study – surprising. Nonetheless, the presenters introduced a first prototype for a web based Evidence Tool enabling the collecting and querying of data, which is definitely a necessary step into the right direction assuming that the alignment of the underlying taxonomy with other technical activities is feasible. Candidates for data exchange are the Pathfinder Web Portal of MSG-027, the Simulation Resource Library of MSG-012 and the Framework for Reusability of MSG-042.

The second presentation describes a concept of modelling and simulation process in the prototype of "Early Warning Systems (EWS) for Defence Against Terrorism" on behalf of MSG-026. Again, this is an ongoing study and only preliminary results are presented. Nonetheless, the ideas mainly driven by the polish lead nation are based on significant contributions but MUT students very mature, but they need to be generally agreed to be all team members before completion. It is agreed that in the taxonomy of security services EWS can be considered as a part of Crisis Management Systems and understood as information systems that process any information from any sources about escalatory developments, be they slow and gradual or quick and sudden, far enough in advance in order for a national government or an international or regional organization to react timely and effectively, if possible, still leaving them time to employ preventive measures. As in the previous paper, this taxonomy must be shared and aligned between the users of the study. Furthermore, MSG-026 evaluates a simulation based diagnostic support tool with its associated algorithms that realises and supports gathering the information relevant to terrorism threat estimation and intelligence data analysis, analysing and simulating the information in order to forecast terrorism threat over long periods of time, the stability of the threat factors and the signalling when a break-through of a pre-determined threshold is detected and visualising the output for potential users. The underlying formal model of the EWS is described in sufficient detail in the paper. It combines a mathematical model of potential terrorist crises, an expert method of terrorist threat assessment being able to learn and pattern recognition algorithms as a tool of automatic evaluation of terrorist threats. The evaluator believes that the resulting model can benefit from considering work on web based data mining to collect evidence of terrorist threats currently conducted at George Mason University, Fairfax, Virginia, and Old Dominion University, Norfolk, Virginia. The success of this sophisticated and mathematically well-funded initiative will depend on the alignment of data and its meaning in form of a general accepted ontology, at least for NATO's MSG community.

Finally, the user perspective as customer of M&S was given in the overview of the Viking 2005 planning. Sweden conducts this exercise, supported by USJFCOM as agreed on bilateral basis, although NATO experience, expertise and additional manpower support is highly desired and needed. In recent years, there

was a change of focus of Viking exercises from the strategic to the tactical planning level, resulting in increased interactions between military and civilian organizations. Viking is recognized as a unique exercise by NATO and will be observed and evaluated. The demand for a distinct control organisation in a distributed exercise was recognized and a sophisticated White Cell support is in place, which is described in detail in the presentation. Security is handled using commercial solutions of the Internet and the Partnership-for-Peace (PfP) Network. Of particular interest is that the PfP-Network will not be taken down after the exercise but will remain in place to become a backbone for future work, similar to the idea of Joint National Training Capability (JNTC) networks of the training directorate or Distributed Continuous Experimentation Environment (DCEE) of the experimentation directorate of US Joint Forces Command. The presentation shows that the international M&S applications for distributed exercise support have been significantly improved in the recent years. More examples underlining this result will be given in following sections, in particular when presenting the M&S infrastructures supporting other distributed exercises, such as distributed mission operations (DMO) of the Air Force or the new US Army M&S system OneSAF Objective System (OOS).

THE METHODOLOGY

The paper "Role of Effects-based Metrics in Advancing R&D Agility through Modelling and Simulation Based Exercises" gives a high level review of the current M&S value paradigm based on evaluations conducted by the Defence R&D Canada. M&S value is currently more or less focused on traditional measures of cost-effectiveness associated with the consequence of earlier decisions on follow-on stages in the lifecycle, likely in form of cost avoidance. While this is appropriate for procurement applications, the application domains of M&S are much broader, including training and doctrine evaluation for the transformation of the armed forces. The terms of central interest are effects and capabilities. The operational community defines capability in form of requirements to be able produce a desired effect; the system of systems provides the means to do this. Consequently, the metrics used to measure M&S value must be rooted in the operational ideas of effects and capabilities as well. This enables to rapidly react in response to evolving operational mission requirements through an M&S-based framework. The paper describes the resulting framework and its key performance parameters within a metrics in the well-known categories of people, processes and material/equipment. By providing a consistent strategic and diagnostic measurement approach, based on "effects", both end-state and incremental attainment can be evaluated. The core idea is to bring M&S as a tool into the operational community to avoid the perception that "M&S is done by others."

The follow-on paper on the "Modelling and Simulation and Capability Engineering Process (CEP)" supports these findings and describes the process itself in more detail. CEP promotes SBA concepts exploiting M&S across all acquisition phases in form of spiral developments. This results in the positive effect that models and simulations created in the early phases of acquisition can naturally evolve in a consistent way throughout the CEP while embedding the metrics proposed in the first paper of this section. In summary, CEP aims at efficiently integrating M&S/system engineering processes to optimize how capabilities are delivered better, faster and cheaper. The metrics are used to deliver the evidence that capability requirements actually are met and how they values are in comparison with alternative solutions. Beside allowing to optimise M&S values, this approach also allows to efficiently implement the notion of collaboration and sharing across the life cycle, because the process in connection with the metrics allows to define which and how M&S activity value documentation must transit from one phase to an other.

The final methodology paper on "Case-Based C2 Modelling and Effective Development, Implementation and Experimentation for Simulation Based Operational Training Support System" evaluates the possibility for implementing decision automats on the tactical level. The system was prototypically implemented by MUT and is currently installed in Warsaw, Poland, to be used in selected applications by experts of the Polish Army. The paper gives a sufficiently detailed description of the concept for modelling and effective

development, implementation and experimentation on the tactical level. Core idea is that the automat executes two processes: the decision planning process – in three stages – and generating commands for the simulated entities. The decision planning process starts with the identification of a decision situation, continues with the generation of decision variants in form of alternative courses of action and ends with evaluating the alternative and choosing the best one under the given constraints. The paper exemplifies the underlying algorithms, which are rooted in graph theoretic approaches. A comparison with the Joint Theatre Level Simulation (JTLS) is planed, but the underlying Multi-Resolution Modelling (MRM) problems are very challenging. The approach has clearly the potential to add realism to decision automats, e.g., in training applications. However, it should be pointed out that the approach is not of descriptive, but of reproductive nature, which means that the behaviour of a unit looks the same from the outside, but the internal algorithms proposed in this paper work on mathematical optimisation problems that are not used by real commanders in their "real-world" decision process. An alternative approach to this approach using artificial intelligence methods is presented in the result section.

CASE STUDIES AND ASSESSMENT

With exception of the last of the four presentations, the talks coped with mature and relevant applications of M&S applications within NATO. They are all potential candidates for assessments and in particular the presentations on Distributed Mission Operations (DMO) have been used to collect data to move from anecdotal to substantive evidence. The first two papers of the section on "Analysing the Evidence" will give details on the process and produces forms and queries.

The first case study presented gives a high-level view on DMO of the US Air Force and its relations with the Joint National Training Capability. Many of these developments are driven by operational experiences made in the recent wars in Afghanistan and Iraq. The enumeration of driving factors for effective and efficient training is a guideline on what to measure to proof substantive evidence, namely High Fidelity Virtual Battlespace, Full Sensor & C2 Representation, Composite Force Operations, Full Combat Mission Rehearsal, Sensor-to-Shooter Interactions and Joint/Coalition Interoperability. The US Air Force Aerospace Expeditionary Forces Training (AEF) conducts experiments in conjunction with the US Joint Forces Joint National training Capability (JNTC) to evaluate what is possible today. The JNTC underpins a global, information age joint training and advances the transformation of training to include enabling Multinational, interagency and intergovernmental network-centric operations. A connection to NATO's Joint Warfighting Centre (JWC) seams to be obvious, but is not yet established, the security challenge being one of the driving problem factors: Multi-Level System Security linking distributed assets of varying classification must be solved and Network Information Assurance, including but not limited to the common use of Intrusion Detection and Prevention Devices, must be integrated. First international systems participating in exercises have been hosted, but there is much room for improvement.

However, the potential of this training transformation can be summarized in the quotation of a national guard F-16 pilot after he went through the DMO training before participation in the air attacks during operation "Iraqi Freedom". After flying the real mission, he said: *"I was 'downtown' at the start of tonight's activities and got to launch the first HARM of Operation Iraqi Freedom. The location of my flight and the tactics employed were EXACTLY like we were practicing in the F-16 Mission Training Centre at Shaw before we left. Talk about Mission Rehearsal!"*

The second presentation and paper gives additional information about DMO on the technical level, in particular how it is embedded with Distributed Mission Operation Centre (DMOC). The Virtual Flag Exercise is part of the description. One main focus is Command and Control. The time of training pilots in a plan to shot-down enemies or to drop bombs are over and are replaced in coordinated and orchestrated events combined into a distributed missions, which requires Command and Control to become the "glue" between the otherwise singular events. DMOC is an umbrella term for six interrelated applications –

Training, Full Spectrum Mission Rehearsal, Testing and Evaluation, Range Integration, Experimentation and Decision Support – and supports currently elements from the engagement level of war up to a combat operations cell at the operational level of war by merging live, virtual, and constructive elements. The presentation stressed that there are two cardinal errors which should by avoided: First, just because it is technically feasible to merge live, virtual and constructive doesn't mean it should be done; each merger must enhance combat capability for all involved. Second, using one group of soldiers as a training aid for another group of soldiers is problematic, as complex support operations in DMO require experts to gain the full support, not laymen; the practice of using soldiers as training aids must be rethought. In summary, he Joint Synthetic Battlespace exists today. There are still challenges and open questions, such as rapid scenario and data base generation, multi-level security and cohesive brief/debrief capability, but progress has been made.

The third paper deals with the role of BMC3I modelling within the NATO Active Layered Theatre Ballistic Missile Defence (ALTBMD) programme. Modelling and Simulation has played a pivotal role in advancing the NATO ALTBMD Programme, which is one of the largest and technically most complex in the history of the NATO Alliance. A suite of models from NATO nations has been used within the programme to provide the analytical foundations from which the system requirements were derived. It should be pointed out that the detailed description of the Extended Air Defence Simulation (EADSIM) does not imply that EADSIM is an exclusive tool of choice for ALTBMD. The programme actively supports MSG activities like MSG-039 and a federated approach utilizing multinational resources is the goal of the approach, but currently EADSIM support most requirements in a sufficient way to progress within ALTBMD without having to wait for other technical activities. Nonetheless, ALTBMD expressed explicit interest in MSG-027 and MSG-042 and is a potential contributor to the knowledge base of NATO concerning reusability and integration of multinational resources. NATO's Theatre Missile Defence (TMD) Programme reached a key milestone in Alliance efforts to field an Active Layered Theatre Ballistic Missile Defence (ALTBMD) capability by 2010. In March 2005, the North Atlantic Council approved the Charter for the ALTBMD Programme Management Organisation (PMO). Side by side with NATO's JWC, the ALTBMD programme is among the most important candidates of NATO programmes being potential candidates to proof the substantive evidence of M&S and to implement the processes requested to set up a database with relevant data.

The final paper of the case studies and assessment section had a slightly different direction. It presented another relevant NATO MSG activity on the Effectiveness by Reusability, conducted by MSG-042. Although the activity is still ongoing, there are some first findings, in particular that a common technical architecture alone is not enough to ensure interoperable solutions, but additionally guidelines and specifications and organizational considerations are needed. The recommendation is therefore to work on a framework for interoperable M&S solutions, taking into account the needs of C2 as well, the Framework for Simulation Resources Reusability (FSRR). The framework distinguishes three categories: organizational considerations, Guidelines and Specifications and Common Technical Architecture elements. The methodology being currently seen to have the most promising potential are ontologies. However, in order to be relevant, a common way to make use of and apply ontologies for interoperability, but many workshops are currently coping with such options. Furthermore, the findings are supported by research of leading M&S organizations as well. During the SISO/SCS Panel Discussion on "Priorities for M&S Standards;" conducted during the IEEE Spring Simulation Interoperability Workshop in Orlando, Florida, March 2003, Professor B. Zeigler explicitly stated in his presentation that *standardization must be aimed at the modelling level* to ensure interoperability between systems, i.e., the standardized level must be higher than the programming level standards currently applied. Work conducted at the Virginia Modeling Analysis and Simulation Center, Norfolk, Virginia, supports that for "meaningful interoperability," the sharing of standardized data via standardized protocols, such as the Distributed Interactive Simulation (IEEE1278) protocol or the High Level Architecture (IEEE1516) standard, is necessary, but not sufficient. The coordination of the underlying conceptual models and the harmonization of the operational ideas simulated, are the real crux to create interoperable solutions. Instead of only

standardizing the information exchange requirements, the underlying modelled cause-effect-chains must also be coordinated. Ontologies are seen to be an applicable solution.

ANALYSING THE EVIDENCE

The first two papers of the section on Analysing the Evidence comprise a state-of-the-art example on how to use objective and subjective measures and means to move towards formal proof of substantive evidence. The first paper gives a First Wave overview plus focus on Dutch participation as a case study. The exercise First Wave was evaluated under the aspect of effective and efficient Mission Training through Distributed Simulation (MTDS), which required assessing training effectiveness for all players at all locations. The resulting data collection requirements are divided into subjective data requiring consistent surveys and protocol at all sites, and objective data requiring establishing exercise protocols and recording network data. The core idea of the first paper is to produce forms that can be reused for future evaluations (subjective and objective). A practical problem is always that surveys are often seen to be "in the way of the real work", so they must be integrated as smoothly as possible using proctors and expert-tailored questions. Also, improvements in automated data collections are needed. This goes well with the recommendation of "NATO Code of Best Practice for C2 Assessment", written by the expert in the NATO SAS-026 effort. The Code of Best Practice recommends establishing an automated data collection facility for operational research studies using operational C2 systems. Web services and applets are technical solutions helping to make the interview process as smooth as possible, but experiences show that without proctors the data collection will not happen. It was evaluated to have a tremendous positive effect that Fist Wave was build on an operational scenario, so that operators found themselves in a used-to environment. Furthermore, the exercise made the diversity in national training obvious and showed potential to avoid misunderstandings in operational domain, clearly an evident contribution. On the technical side it showed that a stable network is necessary for planning, briefing, and debriefing. Distributed Training is much more than plugging the simulators together. Generally, there was a positive feedback between DMO and First Wave.

The second paper presented the methods and results of evaluating the impacts of mission training via distributed simulation on live exercise performance using the US/UK "Red Skies" event followed by the "Red Flag" event. This paper is a direct follow-on to the first paper of this section, as the same technologies presented before could be reused and refined. Red Skies is a simulation driven event, while Red Flag is part of Joint Red Flag and a live exercise. Within the conducted study, the same crews that participated in Red Skies stayed together a couple of weeks later in the US/UK exercise "Joint Red Flag". In this joint exercise, much more data was collected before, during, and after the exercise, often reusing forms and interviews from First Wave evaluations. This made the comparison of simulation (Red Sky) and real flight (Red Flag) possible. The results of this comparison are in the paper and show a significant overlap of objectives in both events. Red Skies supported training in more than 70% of aircrew collective competencies and comparable training to Red Flag for 75% of the Squadron Training Objectives. However, it is worth mentioning that the training is normally focused in command and control interoperability in the joint coalition force, not the technical skill of pilots, and that orchestration of the forces becomes focal point. These elements were not evaluated in sufficient detail in the study described in the paper. Furthermore, control groups allowing statically significant tests were not established due to the constraints of the events. A potential future research should be a comparison between operations on the joint level against operations of individual national play boxes with additional orchestration. In summary, this paper deals with very valuable and significant first steps to seriously analyse the evidence on M&S applications.

The final paper of this session had a different focus. The work on "Sustainability Simulations for Fighter Aircraft in Peace and at War" shows the gap in readiness and is designed to make pilots and air force personnel aware of connections and cybernetics of the problem of deployment, readiness, and training.

Gap may be closed by simulation. The presented software can be used to investigate the change in production and readiness on the airbase by changing factors such as maintenance, pilot availability, aircraft reliability, aircraft demand can help the user to simulate the impact a deployment will have on the airbase, to identify problems that may occur in situations with a change in demand for aircraft and to test whether the proposed changes will have the desired effect or not, including the effects of unscheduled maintenance. Although not in the mainstream of the other papers of the programme, the results are relevant in particular in the lights of constraints in the resources and the necessity to be aware of the effect of deployment of systems in real world operations and the effect on live training with these systems.

COMBINING ANECDOTAL TO SUBSTANTIVE

The three papers presented in this section were very different from their applications and the background of the contributors. The all deal with different aspects of the necessity to come of with unbiased objective processes to proof substantive evidence.

The first paper started with the questions: How good can PC based flight simulators prepare you for a real flight? The underlying thesis and antithesis are that (a) to date, manual flying skills supposedly can only be acquired in high-fidelity simulators or the real aircraft versus that (b) it is now possible to do with PC's that in the past could only be done with expensive systems. Studies on landing, basic manoeuvres and instrument handling already have been evaluated with positive transfer from PC based simulation to actual flight, but so far the evaluation was based on instructors opinion. The study evaluation presented in the paper is based on objective measures in the aircrafts. To conduct the study, participants were divided into three groups: Group A was trained in addition to the live education in the airplane using a standard Personal Computer (PC); Group B used an enhanced PC for the training (with additional features, such as instructors voice, realistic throttle, etc.) and Group C was the control group. The selection criteria for participants excluded experienced users of flight software or people with real flight experiences. The experiment led to very interested results: While the instructors perceived a difference in the performance, the objective measures did not. This led to the conclusion that no skill transfer from PC-based simulations was objectively measured, but the measurable skills are significantly determined by other factors such as flight time in the real aircraft, pilot ability score of the student and presence of a particular flight instructor. Nonetheless, positive effects of the PC based training were shorter briefing times (students already know the theory better), remedial instructions and an increased procedural efficiency (less non-specific flying activities). Another significant result of this study is the necessity to base the evaluation of substantive evidence as much as possible on objective, automatically collected data, as the biases of Subject Matter Experts (SME) can otherwise lead to ambiguous results.

The second paper highlights the fact that when evaluating the evidence of Synthetic Environment (SE) this is often done by comparing simulation experiments with field trials. However, the shortfalls of field trails are seldom taken into account as well as the short falls of SE. The paper gives a couple of examples and how this "unfair comparison" can lead to wrong conclusions. The presented recommendation is to accept field trials and SE experiments as complementing technologies and not as competing alternatives. SE are generally cheaper (which does not mean that they are cheap) but costs may be shared for field trials when they may piggyback on exercises training specific and relevant capabilities. In summary, there are many overlapping or hybrid technologies, where only a mix of field trials and SE generates a sufficient database to give answers. We need guidelines and best practices to understand limitations and select appropriate techniques. It is correct to be cautious using simulations, but we need equal caution for field trials and be unbiased in the selection of appropriate methods.

The last paper in this section presents SimEC3, a Simulation Based Acquisition Tool for the France's cooperative Fighting System. This paper gives an overview on current French efforts and could have easily been placed in the case studies and assessment section. EC3 stands for "Engin de Cohérence du

Combat de Contact", which is a light armoured vehicle replacing the current AMX 10RC. The goal was to design a multi-role platform with some innovative capabilities in the field of Network Centric Warfare concepts using M&S as the supporting technology, leading to the SimEC3 idea. A Thales and GIAT consortium, led by Thales High Tech Services for the French MO¨D Procurement Agency (DGA), develops the SimEC3 programme. The technical aspects are that SimEC3 is developed according to the Federation Development and Execution Plan (FEDEP, IEEE1516.3) to support interoperability between models and potential growth through the use of HLA. A generic model library comprising components that can be combined to generate more complex objects such as platform, unit or system supports composability. Models and complete system qualified are using a dedicated V&V process defined with DGA experts. To support standardizable evaluation the assessment is based on Multi-Criteria Decision Making (MCDM) algorithm. Although the French system has no C2 systems integrated yet, this is an example for a very mature simulation and integration of OR devices.

THE RESULTS AND WAY FORWARD – TO DATE!

For the last session of the symposium, presentations and papers exemplifying the current state-of-the-art with potential to show a way forward were selected. While most of the papers could easily have been selected for this session, the presented candidates all show at least one aspect of particular interest in the area of substantive evidence.

The first paper of this session dealt with the potential to combine different areas and application domains of M&S into a cohesive support using the German experiment VIntEL ("Verteilte Integrierte Erprobungs-landschaft"). The VIntEL Experiment took place in Fall 2004 and is an example on coupling a master constructive simulation – the Simulation System Papst – with several live and virtual components. In addition to combining Live-Virtual- Constructive elements, CAD/CAM data were used for visualization of future platforms. Objective of the experiment was to demonstrate the feasibility to test systems being still under development in a simulated but realistic environment in collaboration with live systems, simulators and other systems still in the design phase. In summary, the first VIntEL experiment could successfully demonstrate that it is possible to build a federation for live-virtual-constructive integration to support development and testing of new capabilities in connection with existing and planed system and doctrines, hence supporting transformation in an effective way.

The second paper describes the current French developments to support the Rapid Reaction Corps (RRC-FR) under French Command. The French Research & Technology programme CALIPSO ("Concepts avancés d'interopérabilité pour la connexion de la simulation avec les SIO") is currently conducted to enable to evaluate experiment architectures for M&S-C2 system coupling, new models and provide code of best practices and recommendations for future use. The paper gives a sufficiently detailed overview of the RRC-FR to show what user elements have to be supported. As an operational NATO Headquarter, the RRC-FR has to follow the Operational Planning Process (OPP), which is described in the GOP (Guidelines for Operational Planning) and that is also described in the paper. As technical enablers, the Command and Control Information Exchange Data Model (C2IEDM) and the related Data Replication Mechanism (DRM) and Message Replication Mechanism (MRM) – as they are defined in NATO Multilateral Interoperability Programme (MIP) – in connection with web services have been identified. The work is therefore closely related with the current activities in the Coalition Battle Management Language (C-BML) activities of NATO's MSG as well as SISO study groups. The requirements for the system are described in detail in the paper and reflect current research. One of the central requirements for simulation systems to be applicable as decision support systems within headquarters is that the application cannot be man-intensive but nonetheless must be realistic. Therefore, the use of intelligent software agents to generate situation adequate orders for the simulated entities is mandatory. The approach chosen by the CALIPSO developers is different from the one presented in the last paper of the methodology section: two different types of agents are used to produce a realistic behaviour including an explanation component.

The approach was successfully presented during the recent Fall Simulation Interoperability Workshop in Orlando, Florida, and paper 05F-SIW-032 published in the IEEE Computer Science proceedings gives more details. In summary, CALIPSO addresses planning for Corps, RRC-FR and Divisions for Article V and Non-Article V operations, M&S and C2 systems interoperability and high-aggregated command agents based on current Artificial Intelligence technology. As such, CALIPSO provides a technology platform to collect lessons learned on C-BML and the French contribution to the upcoming NATO MSG-048 activity on NATO C-BML. It is worth mentioning that CALIPSO is of higher resolution and more flexible in the behaviour and complements the NATO simulation system GAMMA, which is designed to support OPP.

The third paper describes current developments for OneSAF Objective System (OOS), the main US simulation system for the Army. OOS replaces older systems, fosters interoperability, and supports the future force. The source code can be exchanged within US DoD to support joint activities. This is possible as well as with allies to support combined activities, if bi-national or multi-national exchange agreements are in place. OOS supports multiple and variable resolution from platform down to component level. OOS comprises many tools, for example a composer for entities (movement component, communication component, weapon components, etc), units (out of entities), and behaviour of simulated entities (using primitives and other behaviours). It is taken also into account that the role of the environment changes dramatically, in particular in the urban environment. Questions like “Can a helicopter land on a roof of a given building?” or “Do you hear boot-steps on the floor in the level below in a building in which terrorist hold captives?” are relevant to the user and respective capabilities must be reflected in new systems. Additionally, the environment becomes agile and changeable; terrain is no longer a constant play box, but it has its own attributes and state changes and reacts with weather and weapons. Also, alliances and perceptions must change over the time (friendly can change to neutral or even hostile). The focus shifts towards non-military entities as far as they effect the operation. It is of interest that Validation and Verification is currently not a formal process. However, the use of OOD is truly joint: the Air Force Integrated Environment uses OOS, USMC uses OOS for their training and JFCOM uses OOS in various exercises and experiments. The architecture of OOS allows to support DIS 1278.1 and HLA 1.3 NG (HLA Pro) with FOM Mapping tools (allowing, e.g., RPR-FOM use). In summary, OOS represents a simulation develop supporting the latest user requirements.

The final paper summarizes a great variety of valuable sources coping with the economics of simulation, which has been a study group effort of SISO. As many papers of this conference, the study group identified the need of metrics and information repositories with relevant information. To enable this, standards for measuring M&S costs, collecting data in the process – not after the fact – are required. The paper has a rich reference section pointing to sources of additional information, such as SISO’s Economics of Simulation Study Group, Society for Computer Simulation (SCS) Technical Chapter of the Economics of M&S or the current SimSummit activities. There are also obvious connections to the NATO MSG-031 activity on Cost Effectiveness of M&S. The paper gives many examples and is a very valuable resource for a “quick start” into the problem. However, there is much room for improvement. It is necessary to identify national and international points of contacts for effectiveness of M&S and to task and fund central Effectiveness of M&S study efforts. Only when we systematically preserve data already gathered on an advertised website and solicit new data submissions, we will be able to gather the necessary data to become statistically relevant. Therefore, we need to define necessary products and actions in form of a common roadmap. As many other organizations are interested in this work, NMSG must align and orchestrate their efforts with SISO, SCS and SimSummit. Objective of these collaborations must be to establish the standards to judge M&S Costs/Investments and to develop consensus metrics for Cost and Benefit.

SUMMARY AND RECOMMENDATIONS

This section shows the major trends as observed in the papers as well as in the discussions following the presentations and the sessions. Each paragraph summarizes a main trend and gives a recommendation on how NATO in general and the NMSG in particular should act regarding the results of the technical evaluation.

In general, the conference gave a good overview on the perceived evidence of M&S within NATO. All presentations on examples and applications showed firm technological trends and confirmed that the state of M&S within NATO is stable. However, although some first frameworks are prototypical established, there is still no common understanding or even a common methodology applied to measure the value of M&S using objective criteria. Psychology and Human Sciences contribute very relevant insights on how to structure interviews and – even more important – how to conduct interviews and collect data without "annoying" the users or unnecessarily disturb an exercise. Overall, the symposia confirmed that there is still a lot to be done and we are far from being able to measure substantive evidence in a generally accepted or even standardized manner.

It should always be kept in mind that cost effectiveness of M&S compared with alternatives is not all, cost-effectiveness compared with other M&S is equally important. While many practitioners know by experience what M&S applications are effective or more efficient given a special problem, there is no guide of best practice established to guide newcomers or decision makers in this area. The "NATO Code of Best Practice for C2 Assessment" is a valuable example on how to guide newcomers, analysts and decision makers when conducting operational research studies. A "NATO Code of Best Practise for M&S Selection and Value Estimation" – in particular when he takes the effect/capability based approach recommended during this symposium into account – could be a valuable effort within NATO's MSG.

In several discussion and some papers it was stressed that cost is not the only measure of merit for M&S values. Risk-Analyses and M&S should be included. Also, risk/benefit based assessment matrices, similar to those recommended for Verification and Validation, must be evaluated. Chris Mugridge recommended the risk/benefit based approach in "Verification, Validation and Accreditation of Models and Simulation used for Test And Evaluation – A Risk/Benefit Approach" (DERA UK, March 1999) summarized in Table 1. If the training application of M&S reduces, e.g., the risk from Catastrophic to Negligible in the domain of Personal Safety, than this is a measurable value.

Table 1: Risk/Benefit based Approach

	CATASTROPHIC	CRITICAL	MARGINAL	NEGLIGIBLE
Personal Safety	Death	Severe injury	Minor injury	Less than minor injury
Environmental Impact	Severe environmental damage	Major environmental damage	Minor environmental damage	Trivial environmental damage
Security Breach	Top Secret	Secret	Confidential	Restricted
Mission Impact	Mission loss	Severe mission degradation	Slight mission degradation	Mission delayed

Another aspect only captured only in the discussion and only slightly touched in some papers, such as the discussion of the Coalition Battle Management Language (C-BML), is that of the influence of new infrastructures such as the Global Information Grid (GIG). In the future, M&S applications are likely to share infrastructure components with other applications. If the M&S community will not be prepared to justify its use of these resources by substantive evidence of their services, they will not be able to participate. It is likely that the resource net bandwidth will be in the future as important as today only funding is.

Finally, NATO must agree on applicable and mandatory core standards to enable reuse of products developed in technical activities and contributing to an overarching programme, such as Pathfinder. It is not understandable that solutions, such as the Simulation Resource Library of MSG-012, cannot be reused in following or parallel activities, such as the Integration Framework of MSG-027 or the Reusability Framework of MSG-42. Furthermore, related efforts, such as the Evidence Tool proposed by MSG-031, must be integrated as well. While Internet and Web technologies like XML and web services are the technical backbone for the feasibility, a common taxonomy (shared class and relation hierarchy) or ontology (shared formalisation of a specifications) is required as well. The C-BML experiences show that national extensions are not only likely but also necessary.

In summary, NATO's MSG is not only on the right track, we are also moving in the right direction, but also we are far from being mature enough and research is necessary. However, the willingness of participating nations to commit to common findings and recommendations and to apply them in the national context will be the real milestone for MSG's technical activities.