Team 1: MSG-088

Data Farming in Support of NATO

Team 1 Members

In conjunction with IDFW 25, MSG-088 members met in Istanbul for meeting 9 of the task group. At this meeting the group completed drafts of all of the main chapters of the report to include chapters on each of the six domains of data farming and two case studies. The group completed work at meeting 10 in Paris in January 2013. The following article is the overview of the report of the task group.

Introduction

Data Farming, introduced by Horne (1997), is a process that has been developed in order to support decision-makers in answering questions that are not addressed by traditional modelling and simulation processes (Horne and Meyer, 2004). Data farming uses rapid prototyping, simulation modelling, experimental design, high performance computing, and analysis and visualisation to examine questions of interest with large possibility spaces. Using these five domains within a sixth, a collaborative framework, this methodology allows for the examination of whole landscapes of potential outcomes and provides the capability of executing enough experiments so that outliers might be captured and examined for insights. An international community has been conducting common activities for over a decade now around data farming ideas. Workshops have taken place approximately twice a year since 1999 in order to exchange knowledge in the area of data farming and apply data farming to military questions.

In 2010 the NATO Research and Technology Organization (RTO) started the Modelling and Simulation Group MSG-088 to evaluate and further develop the Data Farming methodology to be used for decision support within the NATO. The MSG-088 Report summarizes the work of this task group and the first six chapters summarize the six realms of data farming, representing the work of the corresponding subgroups of the MSG-088. Also as part of the "Program of Work" of MSG-088 (Horne, 2010), proof-of-concept explorations regarding questions and models of interest to NATO nations were conducted, with the objective of illustrating the power of data farming for decision support. In order to realize this MSG-088 objective, the task group set up two case studies. The first is related to "Humanitarian Assistance and Disaster Relief (HA/DR)", whereas the second case study involves the topic "Force Protection". These case studies are documented in Chapters 7 and 8 of the MSG-088 report.

The Development of Data Farming

The discovery of surprises and potential options are made possible by data farming. But many disciplines are behind these discoveries and their use in the overall data

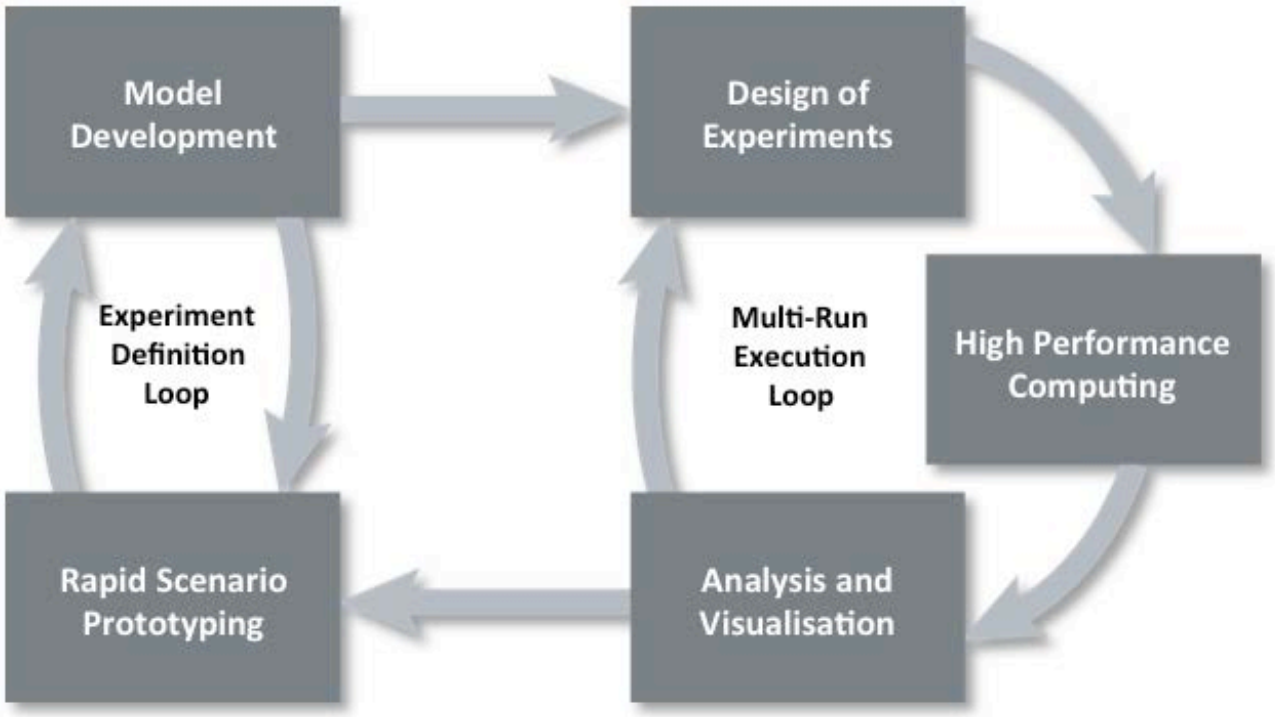

Figure 1. Data Farming "Loop of Loops"

farming process evolved over a period of time. In this section we summarize the account of this development as presented in Horne and Meyer, 2010.

Six realms or domains were incorporated into the data farming methodology from 1997 to 2002. They are each presented in a chapter of the MSG-088 report in this order: Rapid Scenario Prototyping, Model Development, Design of Experiments, High Performance Computing, Analysis and Visualisation, and Collaborative Processes. But we will present them here in this section in the rough chronological order with which they were incorporated into the data farming process.

Initial data farming efforts in the 1997-98 time frame relied upon the combination of two domains. The first was model development. These models, often called distillations at the time, may not have a great deal of verisimilitude but could be focused to specifically address the questions at hand. (Horne, 1999). The second was high performance computing as analysts in Quantico gained access to the resources of the Maui High Performance Computing Center. This capability, using high performance computing to execute models many times over varied initial conditions, allowed for improved understanding of the possible outliers, trends, and the distribution of results.

The models need not be agent-based models, but because of the ease with which they can be prototyped, agent-based models were used during this beginning time period. The huge volume of output from the simulations made possible by the high performance computing resulted in a need to develop visualisation tools and methods commensurate with this tremendous amount of data. Thus, visualisation of simulation data and rapid prototyping of scenarios became important to data farming efforts in the 1999-2000 time frame.

The simulations that defence analysts use are often large and complex. An evaluation of complete landscapes is extremely time consuming, sometimes not even possible. Also, even the smaller more abstract agent-based distillations referred to above can have many parameters that are potentially significant and that could take on many values. Even with high performance computing and the small models used in data farming, gridded designs, where every value is simulated, are unwieldy. Thus, using efficient experimental designs is essential. The Naval Postgraduate School in Monterey, California joined Project Albert researchers in the early 2000s with their expertise in this area.

Finally, collaborative processes help to integrate the other five domains of data farming through interdisciplinary work in creating models and data farming infrastructure and during the iterative process of prototyping scenarios and examining output from model runs. Collaboration also takes place between people from different organizations and nations sharing information and perspectives at various points in approaching common questions. With the addition of design of experiments and collaborative processes in 2001-2002 to data farming efforts, much attention then focused on applications.

Since the incorporation of the above six domains into the data farming process, many articles have captured the fundamental elements of data farming (e.g., Horne and Meyer, 2005). But the key tenet in the data farming process has been the focus on the questions and since 2002 many application efforts have been documented. For example, at the

Naval Postgraduate School many theses have been completed which have used data farming. And over the past decade, over a hundred international work teams have formed around questions at International Data Farming Workshops. These work teams fall into areas, or themes, which include: Joint and Combined Operations (e.g., C4ISR Operations, Network Centric Warfare, Networked Fires, and Future Combat Missions), Urban Operations, Combat Support (e.g., UAV Operations, Robotics, Logistics, and Combat ID), Peace Support Operations, the Global War on Terrorism, Homeland Defence, Disaster Relief, and others.

Why Data Farming?

The data farming methodology applies a simulation-based, holistic and iterative approach to analyze complex systems. In general, the challenge of all simulation systems is the fact, that running one simulation only provides one singular result regarding just the one given situation and circumstances. In this case, no conclusions as to different circumstances - including (identification of) best / worst case scenarios - can finally be drawn. A wider description of the underlying system would be most valuable to obtain a deeper insight. And awareness of this fact gave rise to the establishment of data farming, a simulation based analysis process that enables quantitative analysis of complex questions, obtaining robust results, the comparing of results, and "What-if?" analyses.

The nucleus of Data Farming builds on myriad simulation runs, conducted on high-performance supercomputers, with numerous input parameters varied along a deliberately defined plan, measuring the output and finally examining the mutual interrelationships. Within this activity, Data Farming enables the ability to check assumptions, to gain new insights into relevant relationships and, last but not least, to obtain more robust statements on opportunities and risks of specific mission situations. Briefly, to obtain a more detailed insight into the properties of the examined complex system. This goal is achieved through a deliberate alternation of parameter values of decided input parameters, assuming them to be crucial as regards the measures of effectiveness. Data generated in this way can be of a different nature. Depending on its extent, the analysis can be exploratory or descriptive.

Data Farming is an Iterative Team Process

Data farming is an iterative team process (Horne and Meyer, 2004), a set of embedded loops that incorporate the six realms of data farming. We could list collaboration first as it underlies the entire data farming process, although in the MSG-088 report we will describe it in chapter six as it pulls together the other five domains which are described in more detail in Chapters 1 through 5. And we will present these five in the basic flow of the iterative loop of loops as depicted in 1, which is the following order: Rapid Scenario Prototyping, Model Development, Design of Experiments, High Performance Computing, and Analysis and Visualisation.

Data farming should be regarded as question-centric. It engages an iterative process that scientifically and systematically refines an operational question from its initial raw version (commonly colloquially formulated) into a corresponding answer (at best in a most suitable jargon). The data farming process enables a refinement of questions as well as obtaining answers and insight into the questions.

The first realm "Rapid Prototyping" emphasizes the importance of scenarios as a crucial qualification to answer the initial questions. A rapidly generated scenario accelerates and drives the scenario discussion and its correct implementation into a specific simulation model. The resulting scenario should not only include the definition of the measures of effectiveness and the input parameters including corresponding value ranges varied through the data farming experiment.

The second realm is "Model Development" where a model needs to be developed in order to simulate the required scenario on the required level of detail with the given set of input parameters and measures of effectiveness. Ensuring that the scenario is representative is crucial to the final acceptance of all examination results. This realm combines with the Rapid Scenario Prototyping to make up the “experiment definition loop” in Figure 1.

The next three realms make up the “multi-run execution loop” in the figure. The third realm "Design of Experiments" comprises the statistical experiment planning. Design of Experiments can cut down the sampling requirements by orders of magnitude, yet make it possible and practical to develop a better understanding of a complex simulation model. As stated in Sanchez (2006) a well-designed experiment allows the analyst to examine many more factors than would otherwise be possible, while providing insights that cannot be gleaned from trial-and-error approaches or by sampling factors one at a time.

The fourth realm "High Performance Computing" copes with the techniques to efficiently perform thousands of simulation runs on high performance computer clusters thus providing reasonable runtimes even for encompassing experiments. High performance computing allows for the multiplicity of the numerous individual simulation runs that is both a necessity and a major advantage of data farming.

The fifth realm is "Analysis and Visualisation" that involves techniques and tools for data processing of large datasets resulting from the data farming experiment. The concluding statistical analyses examine the simulation output data upon anomalies, outliers, unexpected developments or simply the underlying interdependencies as described throughout the MSG-088 report. The analysis and visualisation of results allows for the support of decision-making through answering the what-if questions.

After the multi-run execution loop is performed, perhaps many times, the process may return to the experiment definition loop where scenarios and models are adjusted as informed by the work done and discoveries made during the process to that point. At this point questions may be revisited and refined as well as the parameters and scenarios.

The scenarios continue to be defined in close collaboration with subject matter experts, but this collaboration represents just one aspect of the sixth realm of collaborative processes. Collaboration as defined in data

farming ties together effective partnerships and ways of integrating the efforts of modelers, analysts, subject matter experts, decision makers, and all those working on the questions at hand throughout the other 5 realms.

Recommendations and Summary

The objective of MSG-088 was to document and assess the data farming capabilities that could contribute to the development of improved decision support to NATO forces. Proof-of-concept explorations in the form of two case studies involving questions and models of interest to NATO nations were also undertaken. The first 6 chapters of the final report of MSG-088 document the six realms of data farming and provide an assessment of each realm. In chapters 7 and 8 we illustrate the use of data farming through the lens of two case studies that answer illustrative questions in the areas of both humanitarian assistance/disaster relief and force protection. The results of both the assessment and case study explorations indicate the potential high value of data farming to NATO decision-makers and answering their questions.

Harnessing the power of data farming to apply it to our questions is essential to providing support not currently available to NATO decision-makers. This support is critically needed in answering questions inherent in the scenarios we expect to confront in the future. Thus we recommend implementing data farming methods as codified in the MSG-088 report in NATO modelling and simulation contexts and we recommend undertaking specific efforts to apply data farming to NATO questions. Some possible areas of application are force protection, humanitarian assistance/disaster relief, future resources/combined resource initiatives, cyber security, chemical/biological/radiological/nuclear, non-lethal weapons, critical infrastructure protection, and joint sea basing.

In summary, the data farming process can be viewed as the arrows in Figure 1. Data farming is a method that can be viewed as the six realms coming together in an iterative loop of loops. As we illustrated in the case study explorations in the MSG-088 report, the essence of data farming is that it is first and foremost a question-based approach. The basic question that is asked over and over again in different forms in different contexts is: What if?

REFERENCES

[1] Horne, G. (1997). Data Farming: A Meta-Technique for Research in the 21st Century, Naval War College. Newport, RI.

[2] Horne, G. (1999). Maneuver Warfare Distillations: Essence Not Verisimilitude, in: A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, (eds.), Proceedings of the 1999 Winter Simulation Conference, pp. 1147–1151. Phoenix, AZ.

[3] Horne, G. and Meyer, T. (2004). Data Farming: Discovering Surprise, in: R. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, (eds.), Proceedings of the 2004 Winter Simulation Conference, pp. 171–180. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

[4] Horne, G. and Meyer, T. (2005). Data farming: Discovering Surprise, in: M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, (eds.), Proceedings of the 2005 Winter Simulation Conference, Orlando, Florida.

[5] Horne, G. (2010). MSG-088 Data Farming in Support of NATO, MSG-088 Program of Work, NATO Research and Technology Office (RTO).

[6] Horne, G. and Meyer, T. (2010). Data Farming and Defense Applications, in: MODSIM 2010 World Conference and Expo, Hampton Roads Convention Center, Hampton, VA, USA.

[7] Kallfass, D. and Schlaak, T. (2012). NATO MSG-088 Case study results to demonstrate the benefit of using Data Farming for military decision support, in Proceedings of the 2012 Winter Simulation Conference, Berlin, Germany.

[8] Sanchez, S. M. (2006). Work smarter, not harder: guidelines for designing simulation experiments, in: L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, (eds.), Proceedings of the 2006 Winter Simulation Conference, pp. 47–57. Piscataway, NJ: Institute of Electrical and Electronic Engineers, Inc.