Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

MARINE CORPS APPLICATONS OF DATA FARMING

Adam J. Forsyth
Gary E. Horne
Stephen C. Upton

Referentia Systems Incorporated
Airport Trade Center
550 Paiea St., Suite 236
Honolulu, HI 96819, U.S.A.

ABSTRACT

Project Albert is a modeling and simulation initiative of the United States Marine Corps that combines the rapid prototyping of agent-based distillations with the exploratory power of data farming to rapidly generate insight into military questions (Fry 2002). Data farming focuses on the complete landscape of possible system responses, rather than attempting to pinpoint an answer. This "big picture" solution landscape is an invaluable aid to the decision maker in light of the complex nature of the modern battlespace. And while there is no such thing as an optimal decision in a system where the enemy has a vote, data farming allows the decision maker to more fully understand the landscape of possibilities and thereby make an informed decision. The goal of data farming is that decision makers will no longer be surprised by surprise. This paper outlines some data farming explorations conducted over the past few years.

1 PROJECT ALBERT

In 1998 a new program called Project Albert was initiated at the Marine Corps Combat Development Command, with the vision of addressing the needs of military decision makers that were not supported by traditional operations research methods. Specifically, Project Albert addresses the areas of highly non-linear systems where small changes as the scenario unfolds can lead to dramatically different outcomes, intangibles such has bravery, trust and cohesion, and co-evolution – the fact that blue and red strategies do not evolve independently. Co-evolution is often summed up in the phrase "I think that he thinks that I think …."

Project Albert focuses on rapidly generating insight into complex problems. Models are built directly on the basis of subject matter expert (SME) input– a modeler and SME sit together to build the model. The models used are usually simple, transparent, abstract models referred to as "distillations." The rapid prototyping nature of the Project Albert tools means that models can be constructed in hours or days, and insight can be given to decision makers within days or weeks, rather than the months or years that traditional modeling and simulation often entails. The methodology used to analyze these models is referred to as Data Farming (see, for example, Horne 1997, Brandstein 1998, Horne 1999, Horne 2001) which will be described in the following section.

2 DATA FARMING

In essence, data farming allows the modeler to rapidly explore a large parameter space, with sufficient depth to provide valid statistical results. This massive data exploration capacity allows the modeler to rapidly focus in on those areas of the parameter landscape that represent potential problems or opportunities to exploit. Due to the large numbers of replications possible with such a technique, not only does the user get an appreciation for the mean and standard deviation of possible outcomes, but those "one-in-a-million" possibilities that reside in the tails of the distribution can also be discovered. Traditional analysis often "throws out" these anomalies as they do not fit the pattern. Project Albert cherishes these outliers, as they may represent the true threats or opportunities. It is this appreciation of the entire range of possibilities, rather than simply the most likely outcomes, that makes data farming such a powerful tool.

For example consider a force on force scenario that is run 1000 times with different random seeds. Let us assume that red wins 999 times and blue only once. Often we can learn much more from the one case where blue was successful against what are obviously overwhelming odds that we do from the remaining 999 iterations put together. If we can learn the conditions required for success from this one scenario, then in the field we can begin to shift the odds from one in a thousand to closer to one in one.

3 PROJECT ALBERT: THE EARLY YEARS

Project Albert began as, and still is, a question-based multi-disciplinary scientific method of inquiry. In the early years the project was focused on understanding the mesh of the nonlinear sciences and complex adaptive systems with the study of warfare. Notional scenarios were created to begin to examine questions and spur the development of the data farming infrastructure.

One such effort from approximately 6 years ago was the creation of scenarios designed to notionally compare attrition and maneuver warfare. The scenario set was named AMY (attrition-maneuver yardstick) and the model used was the only model in the Project Albert tool-kit at that time, ISAAC (Irreducible Semi-Autonomous Adaptive Combat). We will depict two AMY scenarios here where a blue force of 25 agent in the upper right opposed by a red force of 150 (the blue agents each have a fighting capability greater than the individual red agents) is trying to reach a goal in the lower left.

The first figure depicts a scenario where the blue agents have been given a propensity to move away from the red agents. The result is the appearance of a "maneuver" type behavior. The second figure depicts a scenario where the blue agents have some propensity to go after the red agents. The result is a more "attrition" type behavior where the blue agents head for the goal. Comparing these two scenarios show that the maneuver behavior results in fewer casualties while the attrition behavior allows the blue force to reach the goal faster.

A great amount of Data Farming was performed on these notional AMY scenarios, giving a good feel for the tradeoffs in time to goal and casualties due to different force capabilities for both blue and red. This work was part of the early developmental work, but as Project Albert has matured the ability to go beyond the notional and take on real questions has increased. Below we describe some recent data farming explorations for the United States Marine Corps (USMC).

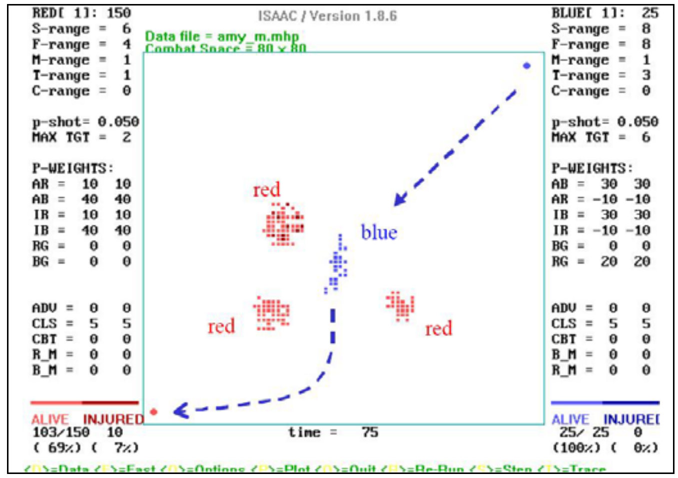

Figure 1: AMY_M Snapshot taken at time step 75

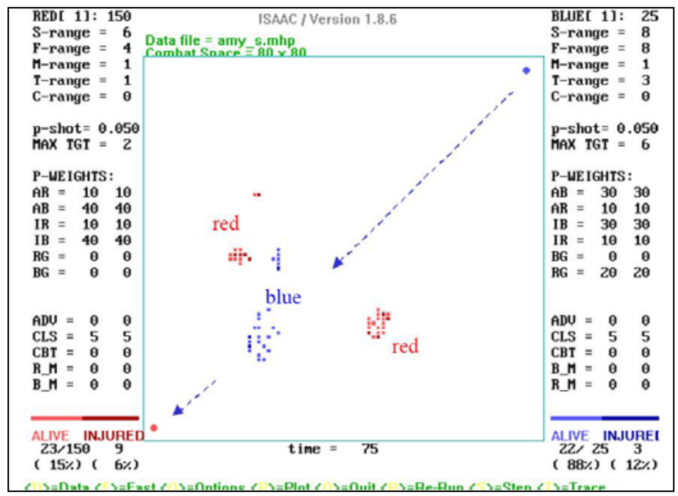

Figure 1: AMY_S Snapshot taken at time step 75

4 COURSE OF ACTION DEVELOPMENT AND ANALYSIS

In May of 2005, a rapid Course of Action (COA) development and analysis demonstration was held at the 10th Project Albert International Workshop in Stockholm Sweden. Two teams, led by personnel from the 1st Marine Expeditionary Force (IMEF), conducted a competition to investigate and assess the utility of data farming to assist COA development and analysis.

This competition brought together international groups of military and simulation experts who used Project Albert tools, specifically the Map Aware Non-uniform Automata (MANA) which is developed by the New Zealand Defence Technology Agency (DTA), to rapidly data farming a set of COAs.

The scenario chosen for the COA development and analysis required both blue and red forces to try to secure an airport. Both forces represented reinforced infantry battalions consisting of infantry, mechanized and reconnaissance units, and indirect fire

The initial force layouts and avenues of approach are shown in Figure 3.

A team was considered to have successfully completed its mission if it had significant forces, defined by a force ratio of at least 2:1, at the airfield and had less than 70% casualties overall.

Teams would use MANA to develop COAs during the day. At night local time, the multiple blue and red COAs were run on the supercomputers at the Maui High Performance Computing Center (MHPCC). The results were

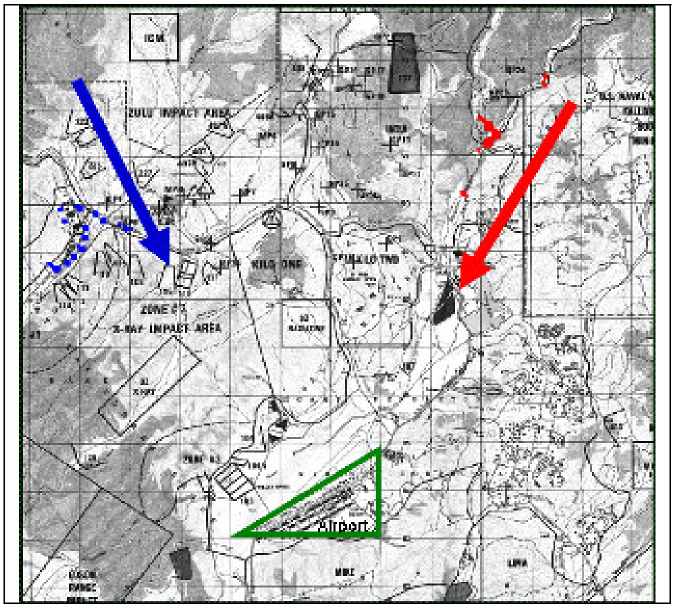

Figure 3: The initial force layout and avenues of approach for both blue and red. The objective is the airport contained within the green triangle.

returned to the groups each morning for analysis. Based on these results, new COAs would be designed. The week culminated with a "final battle" between the "best" red and blue COAs.

The current planning cycle at the MEF level can be summarized as follows:

- After evaluation of mission and commander's intent, two or three COAs are developed
- The friendly COAs are fought against possible enemy COAs. This is usually restricted to the Most Dangerous and Most Likely enemy COAs
- The COAs are then ranked and presented to the commander

At the MEF level, this is usually completed over a two to three day period. Using the Project Albert methodologies however, 12 COAs were developed by each team over the course of 2 standard military days.

The current process is limited by both staff and time restrictions. The limitations of the standard planning methodology are:

- Only a small number of friendly COAs can be developed
- The number of enemy COAs that can be developed is limited. Again, this is usually restricted to the Most Dangerous and Most Likely enemy COAs. One other problem is that the Most Dangerous and Most Likely enemy COAs are usually chosen *a priori i.e.* before any wargaming of the plans takes place
- Time limitations restrict the number of critical details that can be examined for their effect on the COAs
- The impact of terrain and mobility is not always analyzed to the detail it should be.

The advantages of Project Albert's rapid COA development and analysis include:

- More potential COAs can be tested quickly
- There is less cost to testing ideas
- The COAs can be run under various conditions such as weather
- Synchronization schemes and task organizations can be tested
- Multiple enemy COAs can be developed and analyzed to test which are the Most Dangerous and Most Likely
- As these COAs are run many times for each set of conditions, the decision maker is given an understanding of all of the possible outcomes of the scenario, not just the most likely outcome
- Data farming analysis gives the decision maker an understanding of which factors have the greatest influence on the outcome. For instance, if a particular COA is successful 90% of the time, but the top 5 factors that determine the outcome are factors that red influences, this may not be the best COA to consider
- By examining the effect of blue COAs on red and red COAs on blue the courses of action co-evolved.

The process should be looked upon as a planning support tool, not a decision support tool. The goal is to explore the possibilities of a particular COA, not to predict a specific outcome.

5 STABILITY AND SUPPORT OPERATIONS

In January 2004 IMEF held a workshop with Project Albert to explore ways the methodology of data farming could be used to assist with their missions around the world. One of the projects that developed from that workshop was to examine how agent-based modeling and data farming could be used to examine the effect of coalition actions on stability and support operations (SASO). Full details of this project can be found in Koehler 2004.

5.1 General Model Description

The model was constructed using NetLogo (Wilensky 1999). A screen shot from the model is shown in Figure 4.

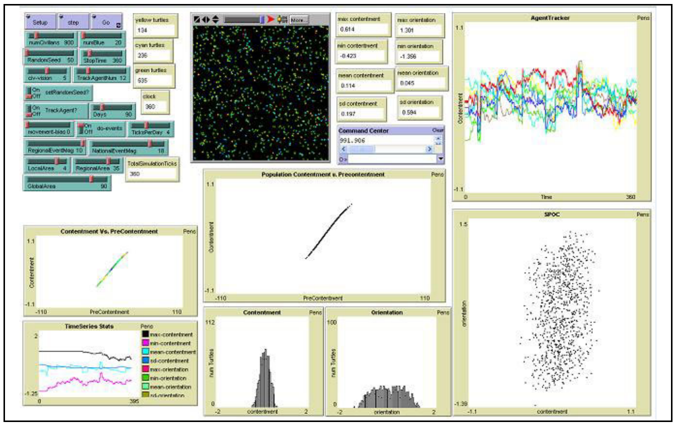

Figure 4: SASO Netlogo Model Screen Shot

This model simulates the interaction of coalition forces with a nominal civilian population. Specific actions of the coalition such as building a school or distributing food are not modeled directly, rather we aggregate all of these behaviors into a generic "good" action. Similarly, actions of those trying to undermine the coalition efforts such as throwing a rock through a window or detonating a roadside bomb are aggregated in to a generic "bad" action. Good events are defined as those that increase the standard of living for the population. Conversely, bad events are those which decrease standard of living. We further subdivide good and bad actions into three categories each: those that are significant on a local level such as (i.e. distributing food), those that are significant on a regional level (i.e. building a school) and those that are significant at a national level (such as instigating free elections). The frequency and range and magnitude of influence of these events are adjustable parameters within the model. This is in essence an "effects-based" model – we model the effect of the actions, not the actions themselves.

Though coalition actions are highly abstracted, the population model is relatively rich. Civilians have four canonical types : married female, married male, unmarried female and unmarried male. Each civilian type is further divided into religious groups, clans, economic class and social influence. A civilian's membership of these groups determines the interactions that civilian can undertake with other members of the population. As we simulate a three month period only, we make the assumption that there is no transition between groups.

Each civilian agent also has 3 internal parameters, Contentment, Orientation and Predilection.

5.1.1 Contentment

Contentment is just what is sounds like – how content the agents are with life. This is their perceived quality of life. Within this model, Contentment is modeled like a well-buffered solution. If an individual is particularly content, it will take a large number of small-scale bad events to make the happy individual unhappy. Similarly for an unhappy individual. However, in the region between happy and unhappy there is a relatively sharp tipping point where a few small-scale events can change dramatically change an agent's contentment. For the technical details of these interactions, please see Koehler 2004.

5.1.2 Orientation and Predilection

Orientation is a measure of how strongly the agent supports or opposes the coalition. Predilection is an agent's tendency to be swayed towards the coalition values as its quality of life improves. Essentially these two parameters control how an agent interprets events. For instance, if an agent is strongly opposed to the coalition, "good" events – those that raise the standard of living – will actually make the agent unhappy. Again, for technical details of these interactions, please see Koehler 2004.

5.2 Next Steps

The models that have been developed have produced "reasonable" answers that have passed informal face validity testing with subject matter experts. We are now looking to formally tune the model with respect to real world data to explore whether the models may have utility in a decision support environment. Once this has occurred, we can then use data farming to explore the effects of possible coalition courses of action. Furthermore, we wish to explore accepted theories of social interaction to further improve the model.

6 CONCLUSION

The Project Albert philosophy of Data Farming – simple distillation models harnessing the massive computing power of the Maui High Performance Computing Center - is beginning to influence decision makers within US military organizations. The research has evolved over the past decade to the point where we have moved our focus from research to application, as evidenced by the case studies presented here.

REFERENCES

Brandstein, A. and Horne G. 1998. Data farming: A meta-technique for research in the 21st century. *Maneuver Warfare Science 1998*. USMC Publication.

Fry, A. and Forsyth, A. 2002. Project Albert: Pursuing the leading edge of military thinking and technology development. *Maneuver Warfare Science 2002*. USMC Publication.

Horne, G. E. 1997. *Data Farming*, Briefing presented to the Strategic Studies Group at the Naval War College, Newport, Rhode Island.

Horne, G. E. 1999. Maneuver warfare distillations: Essence not verisimilitude. *Proceedings of the 1999 Winter Simulation Conference*.

Horne, G. E. 2001. Beyond point estimates: Operational synthesis and data farming. *Maneuver Warfare Science 2001*. USMC Publication.

Koehler, M. T. K., Barry, P. S., Widdowson, B. L. and Forsyth, A. J. 2004. Case study: Using agents to model stability and support operations. *Proceedings of the Agent 2004 Conference.*

Wilensky, U. 1999. NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. [online]. Available via http://ccl.northwestern.edu/netlogo/ [accessed August 1st 2005]

AUTHOR BIOGRAPHIES

ADAM J. FORSYTH is the Chief Scientist for Referentia Systems Inc., in Honoulu, Hawaii. He has a Ph.D in Theoretical Physics from Monash University, Australia. His research interests include the modeling and simulation of complex adaptive systems, particularly as it applies to the defense domain. His email address is adam@referentia.com.

GARY E. HORNE is the founding member of the Data Farming Team at Referentia Systems Inc. He holds a B.S in Mathematics and an M.Ed. in Education from the University of Maryland and a D.Sc. in Operations Research from the George Washington University. He is the originator of the Data Farming methodology and the Director of Project Albert. His email address is ghorne@referentia.com.

STEPHEN C. UPTON is a Senior Systems Engineer with Referentia Systems, Inc. He holds advanced degrees in Physics, Operations Research, and Systems Engineering. His research interests include artificial immune systems, evolutionary computation and agent-based simulation and their application in studying complex adaptive defense systems. He is a member of INFORMS Military Applications Society, the Military Operations Research Society, and the Association for Computing Machinery. He previously spent 24 years on active duty with the U.S. Marine Corps. His e-mail address is supton@referentia.com.