Data Farming: A Meta-technique for Research in the 21st Century

Dr. Alfred G. Brandstein
Dr. Gary E. Horne

Introduction: Returning to our Roots

In the summer of 1996, the Defense Science Board (DSB) recommended several fundamental changes to the analytic culture of the Department of Defense (see figure 1). These recommendations provided additional catalysis to the process that had been evolving in the Marine Corps' analytic community. Stimuli for the Marine Corps' process came from the realization that traditional simulations were incapable of representing many of the attributes of the world required to help explore contemporary military issues together with the observation that advances in computer science and calculational capabilities could, if applied properly, provide data and insight for more appropriate abstractions which would be useful in addressing many military questions.

What *is*	**What *should be***
Closed - Bureaucratic review - Accredited analysis	Open - Peer review - Competitive analysis
Model orientation - Mechanical - Data poor - Rigid approvals - Stable algorithms	Subject matter orientation - Meaningful - Data rich - Learning and adaptation - Unstable phenomena
Suppresses uncertainty	Illuminates uncertainty
Suppresses risk	Illuminates risk
Cold War orientation - Few, accredited scenarios - Point threat estimates	Oriented to now and the future - Broad range of scenarios - Robustness to threat variations

Figure 1. DSB recommended changes to military analytic culture.

The Marine Corps' process, which reflects and amplifies the DSB approach, has a basis which is conservative, even reactionary. It consists of regressing to the World War II roots of operations research. Namely, it is the use of multi-disciplinary teams and the scientific method. As symbolized by our troika of troikas in figure 2, this seminal approach, which is question based, yields increasingly valid and accredited approximations to the answers to the selected questions.

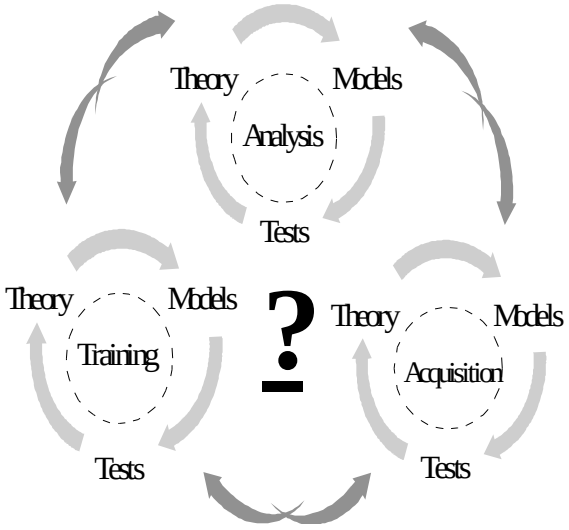


Figure 2. Troika of troikas.

The new component of our developing process is made possible by a convolution of advancements as this century closes.

- Advances in agent-based models which have the promise of capturing some of the adaptability and other key factors inherent in conflict.
- Advances in computing power which enable us to increase our volume of data.

- Advances in our ability to organize, analyze, and visualize scientific data.

We call this new component *data farming*.

Records of historical events are often rich in detail and fascinating to examine, but each is essentially only one data point in the landscape of possible outcomes. Large twentieth century computer models are used by the analytic community to run specific scenarios with many details. But they take many hours to set up for what again amounts to one data point on the landscape of possible outcomes. Thus, what if we want to take a look at questions of interest from the perspective of many data points? The meta-technique of data farming provides a framework to do this. But our motivation for using this technique stems not from any optimism generated by the many advances in science mentioned above. It's actually quite the opposite— the realization that the space of possibilities remains infinite in spite of any and all advances. Thus, we want to look at many data points in so far as they generate insight into the questions at hand.

The Technique: *Fertilizing, Cultivating, Planting, Harvesting,…*

Data farming can be thought of as nothing more than putting the advances mentioned earlier to work to automate the scientific method. Our technique depends on models that can be run multiple times to bring larger portions (as opposed to simply points) of the landscape to light. We will use in our examples data from agent-based models which we believe have the promise of eventually bringing the landscape to light by getting at some of the key elements of conflict. But the reader must realize that our data is illustrative only at this point---we consider our work in this area to be, at best, merely in the preliminary portion of the exploratory stage. We call the effectiveness accomplishing a mission the *mission fitness* and thus we call the results over a certain parameter space the *fitness landscape.*

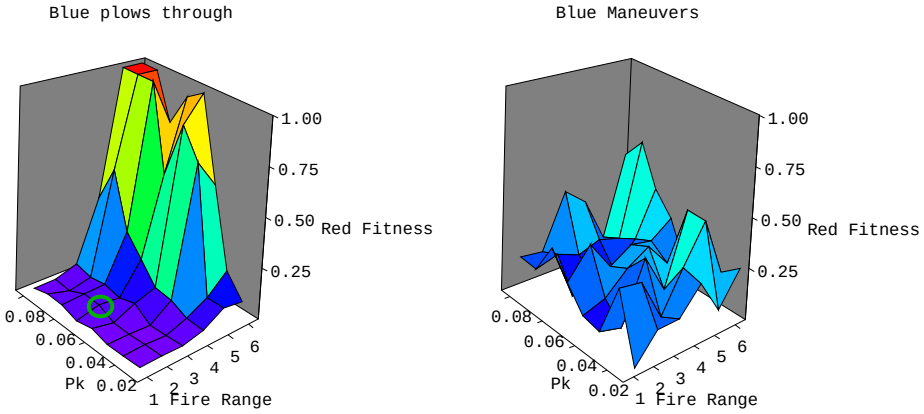
The technique goes something like this:

- *Fertilize* the minds of military professionals and other experts with ideas on how to capture the important aspects of conflict which we have not captured well in the past such as morale, leadership, timing, intuition, adaptability, etc.
- *Cultivate* ideas from these professionals concerning what might be important in a given situation.
- *Plant* these ideas in models to the degree made possible by the model in use and run the model over a landscape of possibilities for variables of interest.
- *Harvest* the data output from the model using developing techniques for understanding scientific data.

We do not want to call the actions just described "steps" because they are all intertwined into the inquiry process of the scientific method that allows us to grow in our understanding. But, just as you do not grow crops or raise livestock in a vacuum, the growth resulting from data farming has a larger purpose. The reason for data farming is to feed our desire for answers to questions. We can grow an overwhelming amount of data, so we continually re-focus on the question at hand and grow data which promises to add to our understanding.

As an illustrative example we consider the question of maneuver vs. attrition. To get the data we use the ISAAC model. The scenario is Red defending versus Blue attacking where the mission fitness for Red is measured by keeping Blue from the area Red is defending and by the number of Blues killed. Figures 3 and 4 show fitness landscapes for Red over the parameters of Red's fire range and probability of hit (here denoted by p_k) for each shot fired. The figure on the left shows the case of Blue heading straight for the objective and engaging in essentially attrition warfare. On the right, Blue engages in more of a maneuver type of behavior on its way to

the objective. Notice that Blue does quite well in the attrition cases except in the patch of the landscape where Red's p_k and fire range appear to synergize to spell Blue's demise. In the maneuver case, Blue does not do as well as the attrition case at most points on the landscape, except in the aforementioned zone where Blue does much better.



Figures 3 and 4. Attrition vs. Maneuver (Transform images courtesy Fortner Software).

Thus, **IF** (a very big if) this had some connection to reality, the lesson for Blue (if it could know Red's capabilities) might be for Blue to simply plow through Red with its razor sharp effectiveness, unless Red had the higher p_k and fire ranges that resulted in the high Red fitness.

But what about Red's options? Red, knowing what his fire range and p_k are, might want to know what to do to improve his seemingly miserable prospects at that point. To illustrate, we grow some more data by changing Red's sensor range and combat aggressiveness. We show in Figure 5 the data we grew in taking a

closer look at a point which might be what Red estimates his fire range and pk actually to be (we use the point circled in Figure 3 to illustrate). The illustrative results show that Red's prospects are still dim. But if forced to fight in this situation, the data suggest that Red might want his forces to become very aggressive (the lower the aggressiveness parameter the more aggressive) and give his forces a medium sensor range.

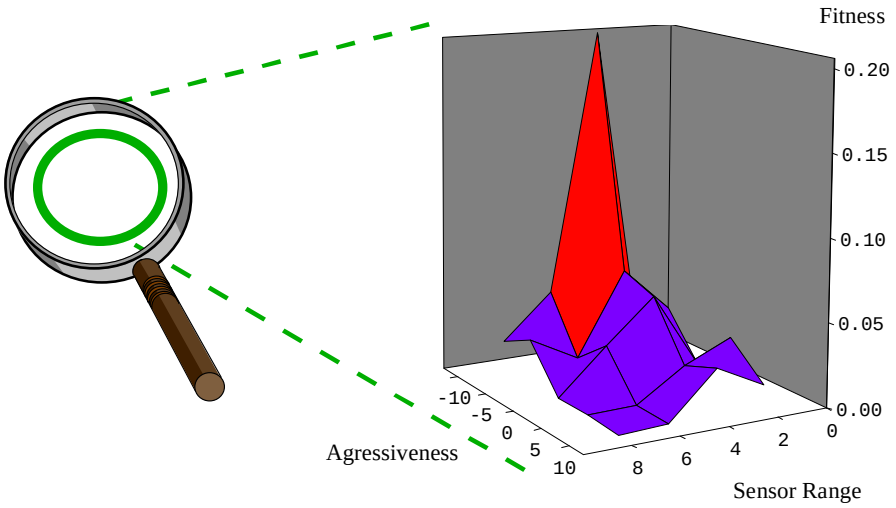


Figure 5: Taking a closer look at a point in the landscape (Transform image courtesy Fortner Software).

Future Directions: Crop Rotation and JWARS

Crop rotation is the metaphor we use to describe a process to farm landscapes in more than one way. This process is one of using a series of models to abstract insights from more realistic models as

well as to demonstrate that insights derived from abstract models have applicability in the real world. The process is in reality a meta-troika in that it enables the extension of regions derived from field experiments on the one hand as well as contributing to the design of these real world trials on the other. We plan to move in this direction in the near future by using a SWARM simulation variation at the level just after ISAAC and simulations with 100,000 entities using massively parallel machines at the level just before field experiments. The goal is to develop algorithms that can then be incorporated into models to be used to assist Department of Defense decisionmakers in the 21st century, in particular the JWARS model.

JWARS is a closed form analytic model being developed for use in addressing joint questions. Its applications include the evaluation of alternative courses of action and resource allocation issues. A concerted research program into exploring these non-linear topics should enable JWARS to have increasing returns in addressing the issues for which it is designed. These research efforts should include developing

- a maneuver warfare framework,
- improved, verifiable (dynamic) algorithms,
- methods for satisficing, and
- dynamic methods for trading off between mission areas.

The structure of JWARS allows for the inclusion of non-linear results and methods that will help JWARS better represent the real world. Thus, we must quickly and relentlessly engage in farming the data across multiple landscapes in pursuit of the nonlinear insights needed to fully exploit this opportunity. But once we have a model that satisfactorily represents warfare, the real challenge will confront us—how do we use the model and interpret the results to answer the questions at the core of military analysis?