HOW SIMPLE IS SIMPLE ENOUGH? MILITARY MODELING CASE STUDIES

T.E. MEYER,* M. KOEHLER, P. BARRY, and B. TIVNAN,
The MITRE Corporation, McLean, VA

ABSTRACT

All models are abstractions of the real world. Determining the appropriate level of abstraction is a balancing of the complexity of the system being modeled, the available data resolution provided by data sources and subject matter experts, the needs of decision makers, and the limitations of the computational and developmental resources. Results from algorithmically linear, physical, closed-system simulations can often be improved by using higher-resolution inputs and by modeling lower-order phenomena. It is not as obvious; however, that ever-increasing resolution will necessarily improve the results from modeling complex systems. Two military course-of-action (COA) development case studies are examined to determine what level of model resolution is sufficient to provide significant insight into COA development. We examine the appropriate level of fidelity for modeling force structures and behaviors as well as the appropriate level of detail for modeling the terrain and physical environment. Methods for evaluating and comparing the results of varying model resolutions are presented.

Keywords: Model fidelity, distillation, abstraction, brigade

INTRODUCTION

Because of their nature, it is very difficult to build models of combat that are able to provide complete answers to a military decision maker's questions. Most combat situations are open systems whose initial conditions are poorly known and for which the motivation of opposing forces is unknown. Further, because the (1) motivation of allies is not well known, (2) communication and interactions among entities are extremely complex, and (3) number of possible courses of action available to the participants could be huge, closed-form modeling is problematic. Typically, the military modeler's response to this problem has been to build increasingly complex models to represent as significant a portion of the combat situation as possible. As a result, these models have tended to be extremely expensive and to require an extensive amount of training, time, and human resources to set up and run. Given the uncertainties of the inputs, the results of even the most competently validated models are suspect at best. Several military organizations are addressing these issues by implementing models focused on specific questions and by establishing processes for executing the models a sufficient number of times to see the potential range of outcomes.

Both the Army G8 Laboratory and U.S. Marine Corps (USMC) Warfighting Laboratory are investigating the use of agent-based models (Axtell 2000) and high-performance computing

---

* *Corresponding author address*: Theodore E. Meyer, The MITRE Corporation, 7515 Colshire Dr., Mail Stop H305, McLean, VA 22102; email: tedmeyer@mitre.org.

(HPC) to provide the analytic capability to examine alternate tactical courses of action at the brigade level. Scenarios to address specific questions are developed relatively quickly in the agent-based modeling environment and tested in real time. The scenarios become the "sandbox" for testing alternate courses of action (COAs) in a fast-turnaround process. As is always the case, however, modelers, subject-matter experts, and analysts frequently identify aspects of the model or scenario data that can be improved. Because results from algorithmically linear, physical, closed-system simulations can often be improved by using higher-resolution inputs and by modeling lower-order phenomena, it has been assumed that similar improvements will enhance the verisimilitude of question-focused agent-based models. The purpose of this paper is to document some preliminary examinations of the effects of input data precision on brigade-level scenarios implemented in agent-based models.

ABSTRACTING BRIGADE-LEVEL COMBAT

Distillations

Everything should be made as simple as possible, but not simpler.

Any intelligent fool can make things bigger, more complex….

It takes a touch of genius and a lot of courage to move in the opposite direction.

~ Albert Einstein

Data farming (Brandstein and Horne 1998) is a process developed by Project Albert, a USMC program aimed at supporting military decision makers. It is the process of running models many times in an HPC environment and varying the initial conditions in order to find outliers, examine the potential range of outcomes, and test the model across its parameter space. Models used within the paradigm of data farming are referred to as "distillations" (i.e., abstractions). It is recognized that all models are distillations of the real world. It is only by the judicious implementation of specific aspects of a system that we can produce models that are helpful. The quotes by Einstein above capture the intent of modeling within the realm of data farming. Distillations should be complex enough to address the question, but no more complex than that. Distillations should be:

- Intuitive — the users must be able to understand the parameters and rules that define the model and how they relate to the system being modeled;
- Transparent — the users must be able to understand how the behaviors in the model emerged from a set of parameters and rules; and
- Transportable — the model must be portable to a data farming environment (Horne and Meyer 2004).

Although any model could be data farmed, distillations are intended to be a bottom-up reduction to the essence of a question. Typically, distillations are expected to be developed quickly — potentially in a matter of a few days or hours. Current distillation development applications use abstraction judiciously, thus representing a number of phenomena in a few paradigms. For example, various types of interchanges (such as food, resources, and positive or negative messages or propaganda) may be abstracted and proxied by weapons exchanges in some distilation modeling environments. Location or proximity in a model can be abstracted to represent relative aspects of other relational parameters. Modeled obstacles can represent walls, floors, borders, or sociological or psychological obstructions in non-geoterrain or combat interchanges. In short, using creative abstraction can keep the models computationally simple, allowing for a large number of model executions in a relatively short period of time. This guideline encourages distillation modelers to innovate and use imagination to define abstractions.

Military decision makers often must address questions with answers that are dependent on intangibles; examples are “How will the morale of my men affect this battle?” “How tired are they?” and “What does the enemy know about my positions?” These types of questions rarely have precisely defined initial conditions or a complete set of algorithms that describe the system being considered. As implied above, we have been using data farming with a wide variety of possible variable combinations to provide insight into these complex questions. Looking at the distribution of results over a large number of runs can provide insight that can be used to address these complex questions. The accomplishment of this data farming relies on two basic ideas:

1. Use HPC to execute models many times over varied initial conditions to gain an understanding of the possible outliers, trends, and distribution of results and

2. Develop models called distillations that are focused to specifically address the question.

Model Scenarios: IMEF and Army G8

Figure 1 is a screen shot of the Map Aware Non-uniform Automata (MANA) modeling environment (Roger et al. 2002) implementing a scenario near an airfield in Camp Pendleton, California. This model is an illustrative example of the type of models that are currently of interest to the USMC First Marine Expeditionary Force (IMEF) and the Army G8. The models often represent hundreds of units including tanks, unmanned aerial vehicles (UAVs), ground troops, artillery, aircraft, and unmanned vehicles, and all of the associated weapons, armor, communication, and command structure.

The scenario in Figure 1 takes place over a 12-kilometer grid and incorporates ground-based units such as infantry, field artillery, and armored vehicles from two opposing forces: red and blue. The pictured scenario includes more than 200 agents. The blue and red dots are agents that represent various entities within the blue and red forces. The blue and red lines represent the entities firing on the opposing force. The entities are displayed on a standard topographic map. Not visible but an essential part of the scenario are the topographic data that the agents within the scenario sense and respond to, both from a visibility and a mobility perspective. This scenario was used as a demonstration of MANA’s ability to handle scenarios of this scale when we began to design a larger scenario that covered a different, wider (35-kilometer-square) area; incorporated air support; and included more than 400 agent types. The analysis described in the subsequent sections used this more complicated scenario.

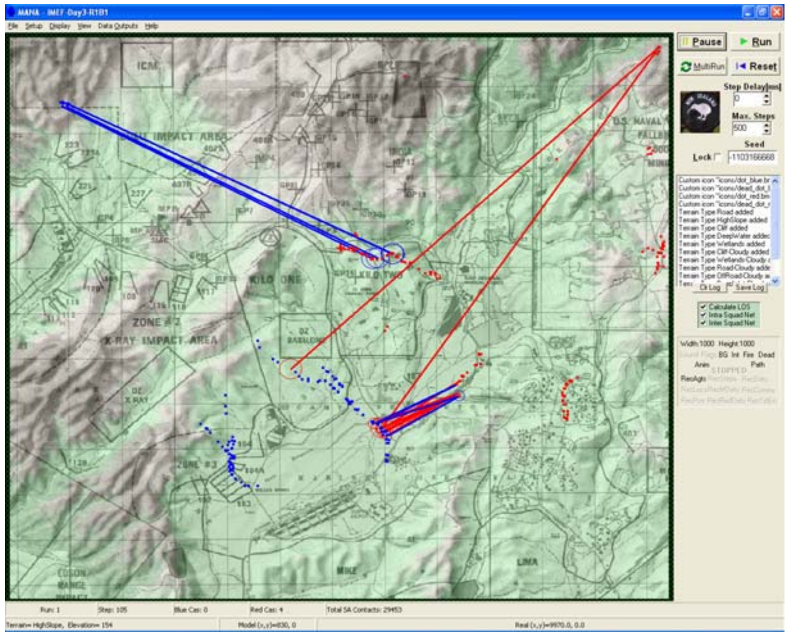

FIGURE 1 MANA modeling environment

Abstracted Model Parameters

Models can become more complex by adding additional detail to the algorithms, rules, and data inherent in them. One of the easiest ways to "improve" a model is to use readily available, higher-resolution data sources with input having more "real-world" accuracy. The question this paper examines is whether this added detail is useful in providing better and more insightful results.

For this study, two input data sources are examined at several levels of abstraction. The MANA simulation environment uses an elevation grid to determine line-of-sight. Weapons are modeled by defining a probability of kill (Pk) profile that determines effectiveness at various ranges. Although implementing "improvements" to these input data is straightforward, it was unclear whether there would be any significant improvement in the quality of the results produced. Therefore, as part of this study, the resolution of the digital terrain elevation data (DTED) was varied, as was the Pk. The specifics of this process are discussed below.

In the scenario, four red force 155-mm Howitzer artillery units play a vitally important role in the outcome. The units are well-protected and, depending on the specifics of the Howitzers' implementation in the model, can be devastating to the blue forces. Various public data sources can be used to acquire information about the effectiveness of Howitzer shells. These sources are questionable, though, and may not be based on anything but an untrained observer's comments. Classified sources can provide more "accurate" data, potentially creating a more valid representation of the weapon in the model.

However, before spending time and resources implementing a classified version of the model in order to execute runs using validated Pk profiles, it is appropriate to test the model by using various unclassified Pks to determine whether variations in the weapon's Pk profiles have a significant effect. Because of the significance of these weapons in the outcome of the battle, it was anticipated that variations in the Pk would be likely to have a significant impact.

Three variations of the Howitzer's Pk profiles were examined for this study, as shown in Figure 2. The base case uses a nominal profile acquired from unclassified sources. Pk A (on the left in Figure 2) represents a hypothetical "real" Pk profile with "fine" adjustments to the profile. The right side of Figure 2 represents Pk B, a "simple" version of the profile.

The model currently uses DTED Level 0 at a 1-kilometer resolution. Given the vagaries of sensor limitations, tree lines, unit positioning, and other terrain features, the importance of implementing high-resolution elevation data was an open question. For the purposes of this study, two abstractions of the scenario's elevation were tested: the DTED Level 0 data and a flat surface. With DTED Level 0 data, the line-of-sight is affected by elevation obstacles. With the flat surface implemented, only an agent's sensor range and the terrain type affect how far the agent can see.

Data Farming and Results

In order to examine the impact of the changes to the Howitzer Pk and elevation, the model was data farmed. The model was run 500 times — 100 times for each of these five variations:

1. Base Case — Base-case Howitzer profile and DTED Level 0;

2. Control — Same Howitzer Pk profile and elevation as in the Base Case, with only a change in the random seed;

3. Howitzer Pk A — The "refined" Howitzer blast radius Pk profile and DTED Level 0;

4. Howitzer Pk B — A simplified Howitzer blast radius Pk profile and DTED Level 0; and

5. Simple Elevation — Flat surface (no variation in surface height) and same Howitzer Pk profile as in the Base Case.

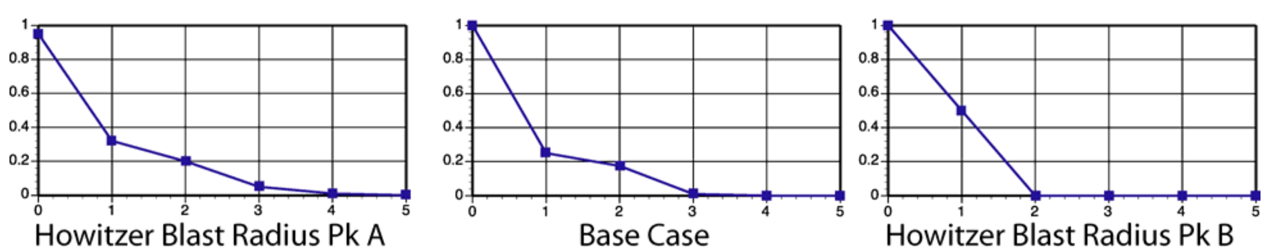

FIGURE 2 Probability of kill blast radius profiles (X axis, one unit is 40 m)

Data farming is a statistical sampling technique. Each execution of the model with a different random seed, referred to as a “replicate,” is a sample of the almost infinite population of potential model outcomes for a particular set of parameter inputs. Varying input parameters produces “excursions” that expand the population and provide analysts with a comparative stratification of the outcomes. Data farming addresses the issue that a model may produce a range of results depending on random or designed variations in inputs. Only by sampling the input data space can the full scope of possible outcomes be understood. Methods using formal experimental design to effectively explore this excursion/replicate space have been developed (Lucas et al. 2002).

Each of the five variations enumerated can be considered an excursion. For each excursion, 100 replicates were run in order to acquire enough data to be able to statistically compare distributions.

For each of the 500 model runs, a set of more than 800 end-of-run measurements of effectiveness (MOEs) were generated by the model. These MOEs include red and blue force casualties for each agent group. Other outputs from the model can include data on agent position, casualty location, enemy contact and detection, and communication for every time-step. For the purposes of this study, the end-of-run MOE data were reduced to total blue casualties in order to examine the impact of small variations on the large-scale results.

Figure 3 represents the MOE results from these runs. Each plot represents a histogram of the distribution of the number of blue casualties over 100 runs of the model. The Base Case and Control represent the model scenario with no changes to the original input parameters. The difference between the Base Case and Control is a change in the set of random

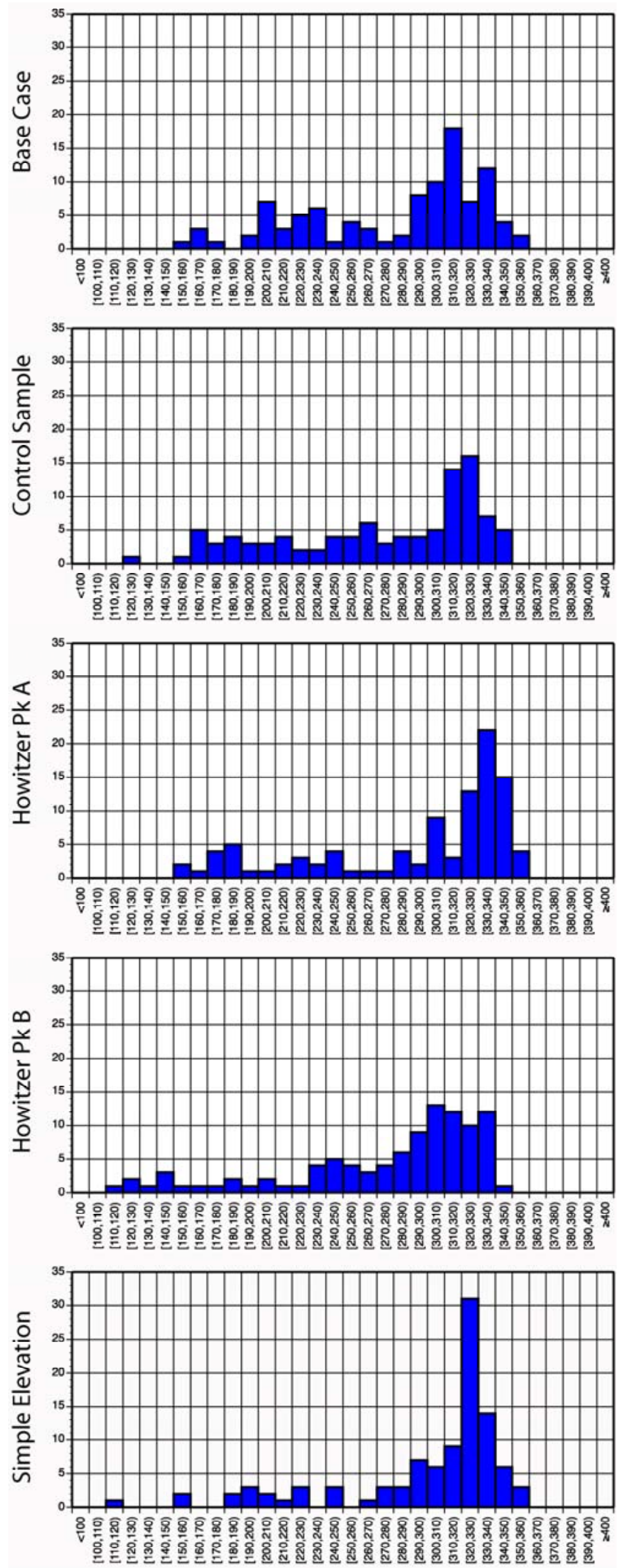

FIGURE 3 Blue casualty distributions

seeds used to provide small changes to initial starting positions and to adjudicate engagements. The authors' assumption was that the Control and Base Case should have MOE distributions that are statistically the same. The MOE distributions that resulted from the variations of Howitzer (Base case versus Howitzer Pk A versus Howitzer Pk B and Elevation (Base Case versus Simple Elevation) were examined in conjunction with the Control to discern whether these variations have any statistical impact on the model results.

Figures 4 and 5 represent quantile plots of the MOE distributions. Figure 4 shows the excursion distributions plotted against normal distributions. These plots indicate that the profiles of all of the distributions were very similar, but that there are visible, if not statistically significant, differences in the distributions.

The QQ plot on the left of Figure 4 shows the quantiles of Base Case and Control plotted against a normal quantile. The plot on the right includes the quantiles for the Base Case and the excursions. It is evident that the Control distribution is a reasonable match to the Base Case distribution, but that the excursions show some deviations from the Base Case. Figure 5 shows a QQ plot of the Control and excursions versus the Base Case. This plot indicates that the Control and Howitzer Pk B distributions adhere more closely to the Base Case and that the other two excursions have similarities in their profiles.

Table 1 provides the results of a statistical analysis of the distributions. Displayed are the Mann-Whitney nonparametric, independent, two-group "statistical comparisons." Mann-Whitney does not assume normal distributions and provides an indicator of the similarity of distributions. T-tests (which do assume normal distributions) gave similar results. In the table, significance values of less than 0.025 indicate that there is a difference in the distributions. It is interesting to note that the distributions fall into two groups (differentiated by coloring in the table): (1) Base Case, Control, and Howitzer Pk B group and (2) Howitzer Pk A and Simple Elevation group. This is also indicated (although not as obviously) in Figure 5.

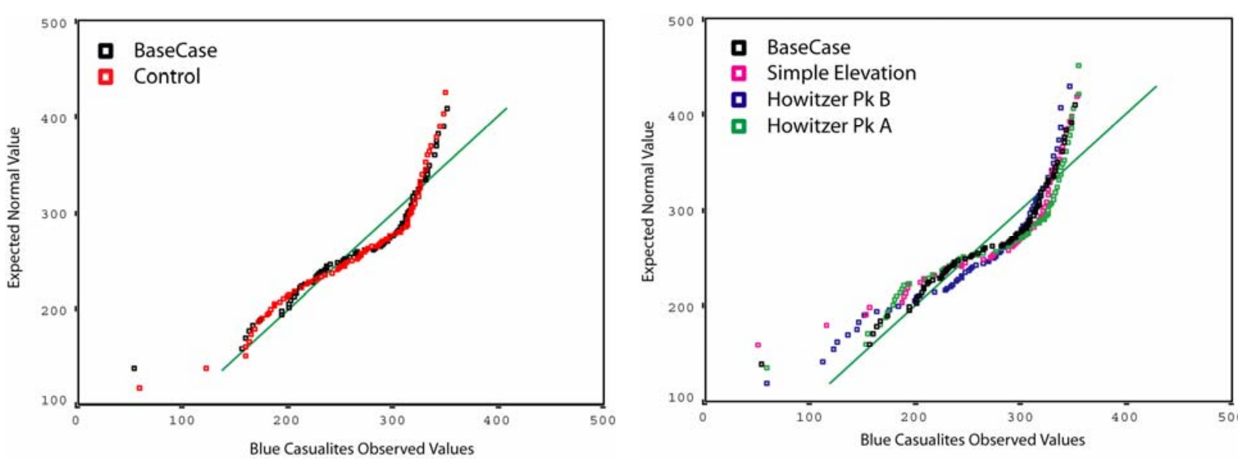

FIGURE 4 Normal QQ plot of Base Case and Control (left) and of Base Case and excursions (right)

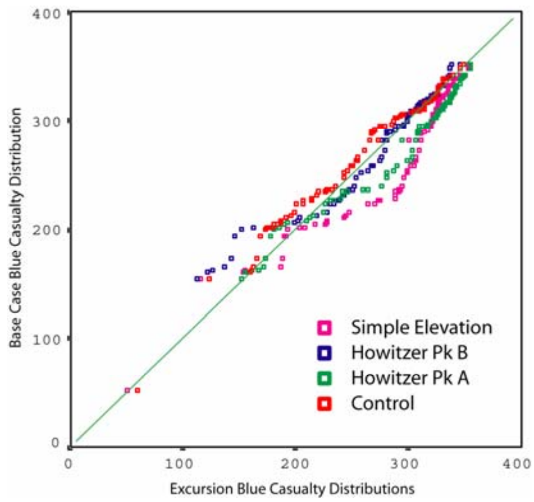

FIGURE 5 QQ plot of Base Case versus Control and excursions

TABLE 1 Statistical differences

Mann-Whitney Test	Base Case	Control	Howitz Pk A	Howitz Pk B	Simple Elevation
Base Case	-	0.396	0.005	0.495	0.002
Control	0.396	-	0	0.823	0
Howitz Pk A	0.005	0	-	0	0.489
Howitz Pk B	0.495	0.823	0	-	0
Simple Elevation	0.002	0	0.489	0	-

Why would the distribution fall into these groups and why is there a shared distribution between the Simple Elevation and the Howitzer Pk A group? After an examination of multiple model executions and their methods, the authors provide the following explanations.

Flattening elevation has a potentially large influence on both the red and blue forces' artillery effectiveness. Flattening the elevation provides all units with a 360° line of sight. Blue has significantly more artillery units that can take advantage of the flattened terrain. Figure 5 indicates that the Simple Elevation excursion is advantageous to the blue force.

At first glance, the profiles in Figure 2 would seem to show that the Pk B profile is a much larger variation from the Base Case profile than the Pk A profile. Pk A seems to consist of small adjustments to the Base Case profile. Pk B is a radical change in all but the zero radius point. Yet Pk A is statistically different than the Base Case, and Pk B is not. What is not obvious is that one "fine" adjustment to the Pk A is the adjustment of the zero radius point from a 100% kill probability to a 95% one. The Base Case and Pk B maintain a 100% probability at the zero radius point. The "minor" change to the zero radius point has more impact than a more radical change at higher radii. This change is also an advantage for the blue force and results in a distribution similar to that of Simple Elevation.

Is Pk A or Pk B more “real?” In the real world, if a Howitzer shell hits within a zero radius of a unit’s position, it is a direct hit and that unit is killed. This scenario and model, however, employ a grid cell that is 40 meters in length. A zero radius implies a hit for anything within that cell. A 95% Pk at zero radius can be considered an adjustment to the real-world data to account for the fact that within the model abstraction, hits within 40 meters are not necessarily deadly. For example, if an artillery shell directly hits a tank, the tank will very likely suffer damage that will make it combat-ineffective. However, if the artillery shell hits 30 meters away, the tank will experience little damage. In MANA, in the current scenario, both of the above hits are considered the same — causing similar effects. Here is an example of an incompatibility between levels of abstraction or simplicity. The actual Pk data indicate that at zero radius, a target should suffer from more severe damage; however, the abstraction of the terrain generates an absurd result with this Pk table.

As one creates models and struggles with determining the proper level of simplicity or abstraction, one must be consistent. Although it is very tempting to abstract model features that are less well understood and to increase the “realism” of better-understood features, one may be tempting fate. Increasing the realism of one area of a model while abstracting other areas may bias your results in significant ways. Worse yet, these biases are introduced by parts of the model that do not, on their face, appear to be problematic (i.e., “These are the actual data, how could that be the problem?”).

CONCLUSION

While all models are abstractions, the abstract models discussed above are used to glean insight into complex phenomena for which there are no closed-form analytic solutions. Further, it is frequently difficult, if not impossible, to check the “correctness” of results of a distillation. Unlike the modeling of physical phenomena, which produces testable results and usually has an underlying mathematical model, the modeling of intangibles and complex phenomena has no “correct” solution. However, the very act of (1) describing the distillation at the appropriate level of abstraction for the question addressed, (2) explicating the associated assumptions, and then (3) exploring the parameter space has been shown to be useful in providing insights to decision makers.

This paper has demonstrated that when a distillation is being developed, it is important that the abstractions be consistent. “Validated” real-world inputs may be incompatible with other model abstractions, such as terrain resolution. The model described in this paper is based on a grid with cells that are 40 meters across. Therefore, it makes no sense and, more importantly, contributes little to the value of the results to have an “accurate” real-world Pk profile that would model the degradation in Pk by the actual distance in meters from the impact. For this distillation, the Pk profile needs to be abstracted — its resolution and probabilities refactored to the scale and functional requirements of the model — before being incorporated into the model. Understanding the overall implications of the abstractions within the distillations is necessary and critical for their appropriate use. As was demonstrated, increasing the fidelity on some parameters (e.g., Pk) with scant regard for the other parameters (e.g., terrain resolution) contributes little and frequently can give a false sense of precision.

We have only begun to examine the question “How simple is simple enough?” in this study, and only for the model scenarios that were specifically analyzed. For the scenarios modeled, only a small set of input parameters — weapon Pk and elevation — over a small range of variation was examined. It is evident from this study that the level of simplicity (abstraction) of the model and its input data can affect the outcomes in statistically significant ways, especially if they are inconsistently “simple.”

The lesson learned from this exercise is as follows: *Validated “real-world” data may not increase the accuracy of a distillation model and may have a negative impact if they are incompatible with the model’s abstractions.*

REFERENCES

Axtell, R., 2000, *Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences,* Working Paper No. 17, Center on Social and Economic Dynamics, Nov.

Brandstein, A., and G. Horne, 1998, “Data Farming: A Meta-Technique for Research in the 21st Century,” *Maneuver Warfare Science 1998,* U.S. Marine Corps Combat Development Command, Quantico, VA.

Horne, G., and T. Meyer, 2004, “Data Farming: Discovering Surprise,” in *Proceedings of the 2004 Winter Simulation Conference,* Denver, CO.

Lucas, T., S. Sanchez, L. Brown, and W. Vinyard, 2002, “Better Designs for High-dimensional Explorations of Distillations,” *Maneuver Warfare Science 2002,* U.S. Marine Corps Project Albert, Quantico, VA.

Stephen, R.T., M.A. Anderson, and M.K. Lauren, 2002, *MANA Map Aware Non-uniform Automata Version 2.0 User’s Manual*.