2010-03

Team 7: Data Farming to Support Model Validation of the BTRA-BC Battle Engine (BBE)

Blais, Curtis; Stork, Kirk; Upton, Steve; Eaton, Jeffrey; Hoffman, Matthew; Rollins, Steven

https://hdl.handle.net/10945/35683

Team 7: Data Farming to Support Model Validation of the BTRA-BC Battle Engine (BBE)

TEAM 7 MEMBERS

Blais, Curtis
Stork, Kirk
Upton, Steve
Naval Postgraduate School, US

Eaton, Jeffrey
Hoffman, Matthew
Concurrent Technologies Corporation, US

Rollins, Steven
USMC Logistics Command, US

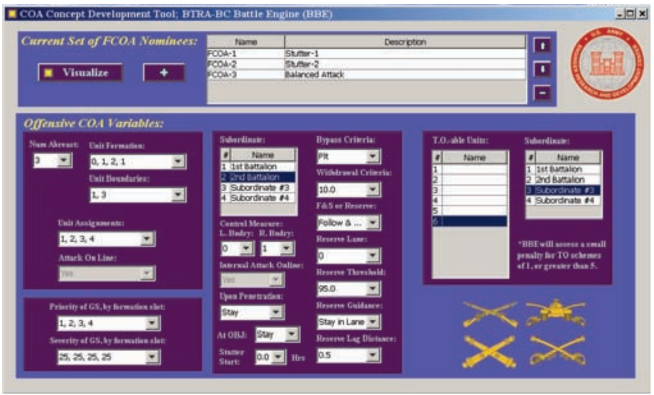


Figure 1: BBE Friendly COA definition window

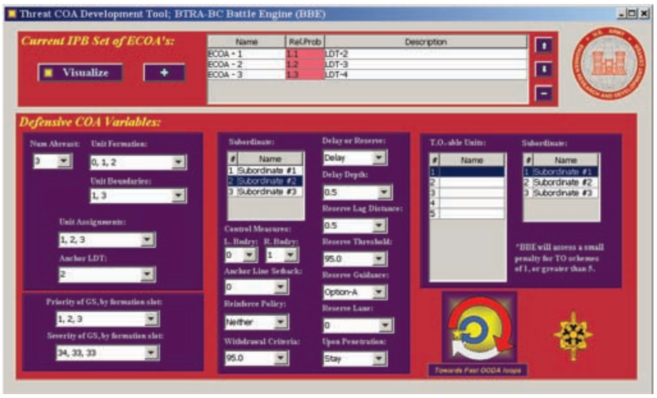


Figure 2: BBE Enemy COA definition window

INTRODUCTION

The Battlespace Terrain Reasoning and Awareness Battle Command (BTRA-BC) Battle Engine (BBE) [1] is a software tool designed to assist commanders and staffs in developing and analyzing Friendly Courses of Action (FCOAs) in the context of mid-to-high intensity combat operations.[1] It is designed to automate a number of sub-tasks of the Military Decision Making Process (MDMP) [2] that previously have been the exclusive domain of the human planner. Using BBE, commanders and staffs can quickly generate and evaluate an unprecedented number of FCOAs. BBE is intended to increase the speed of tactical decision making without sacrificing the quality of those previously manually-developed alternatives.

A major subcomponent of the MDMP is the Intelligence Preparation of the Battlefield (IPB) process [3], culminating in development of Enemy Courses of Action (ECOAs). This process mirrors FCOA generation, but is focused on identification and evaluation of potential enemy activities. A simplified set of procedures and analysis tools within BBE can also be used for generating ECOAs.

Gaming, in the most basic sense, is an attempt by one player to devise and implement a strategy to defeat an opponent or overcome a set of circumstances. In its most basic form, a game requires game pieces, a game environment or game board, and game rules. BBE game pieces represent military units, both enemy and friendly, and interactions between these pieces represent the fire and maneuver of tactical combat operations. The BBE game board represents an abstraction of the battlespace that preserves those tactical aspects of terrain that influence tactical operations. The game board greatly resembles the traditional Modified Combined Obstacle Overlay (MCOO) produced during the IPB process. The BBE reference model governs how game pieces interact with each other and with the game board; it represents the "rules" of a BBE game run.

Objectives

Team 7 sought to achieve the following objectives:

- Use design of experiments and data farming to support BBE validation
- Key Question: What are the factors having the greatest effects on BBE-computed outcomes and the scoring of FCOAs against specified ECOAs?

---

[1] Our thanks to the developers of the BBE tool, Mr. Jerry Schlabach and Mr. Eric Nielson, from the U.S. Army Geospatial Center, for support in learning the operation of the tool, creation of a version of the software for use in conducting runs of the software separate from the Graphical User Interface, and for the description of the software used in this paper's Introduction.

Model Validation

The team members were drawn to Team 7 by an intense interest in the use of data farming as a tool in model validation. On the one hand, Mr. Blais, Mr. Stork, and Mr. Upton are members of an NPS team that have been funded by the Army Geospatial Center to perform various validation studies on the BBE tool to develop evidence that can be used by accreditation authorities to determine if the tool is fit for its intended purpose. Mr. Eaton, Mr. Hoffman, and Mr. Rollins came from a background of common work on validation of models for the USMC Logistics Command (LOGCOM). Discussions in early working sessions of the team dealt with the concept and practice of model validation (in some cases, in contrast to the concept and practice of model verification). In [4], Dr. Petty describes two principal comparisons as the focus of validation activities: (1) comparison of the real world to the conceptual model; and (2) comparison of results from the executable model to the real world. This description is refined in [5] to (1) comparison of the referent(s) (i.e., what is known about the real world relative to the intended use of the model) to the conceptual model; and (2) comparison of the results from the executable model to the referent(s). In light of these considerations, the team wanted to generate evidence that would support a decision that the model is (or is not) useful for its intended purpose through the following actions:

- Investigate computational behavior (sensitivities) of the model for expected and anomalous outcomes
- Confirm expectations of the BBE conceptual model
- Provide a "conversation-starter" between the software developer and the validation team to advance common understanding of the intended model behavior.

Our purpose was definitely not to conduct verification studies with the model; that is, we were not investigating the correctness of the implementation of the software logic with respect to the conceptual model. On the other hand, it was recognized that any "disconnect" found by data farming, in light of expectations raised by knowledge of the conceptual model for the tool, either could be indicative of an issue in the implementation itself (a possible verification finding) or indicative of an issue in the conceptual model (a possible validation finding). In either case, the objective of the study with respect to model validation was not to judge whether the outcomes were right or wrong (against some criteria), or good or bad (based on some valuation), but to produce evidence that could be used by others in position to make such assessments with respect to the intended use of the tool.

Data Farming Approach

All but one of the members of Team 7 were "IDFW Rookies;" the one exception being Steve Upton from the NPS SEED Center. This meant that the team was particularly motivated to internalize the guidance and best practices of data farming as described in one of the plenary sessions and demonstrated throughout the working sessions by Steve. Perhaps the following summary will be useful to readers of this article, and potential attendees of future IDFWs:

- Determine an initial set of factors to explore. Because the foundation of the BBE processing logic is the underlying representation of important features of the terrain (maneuver network), we decided to begin by examining the sensitivities of model outcomes to

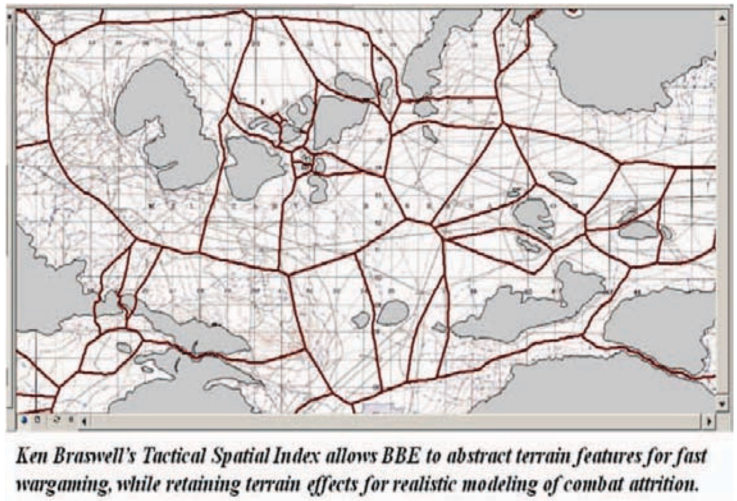


Figure 3. Terrain Analysis Mobility Network

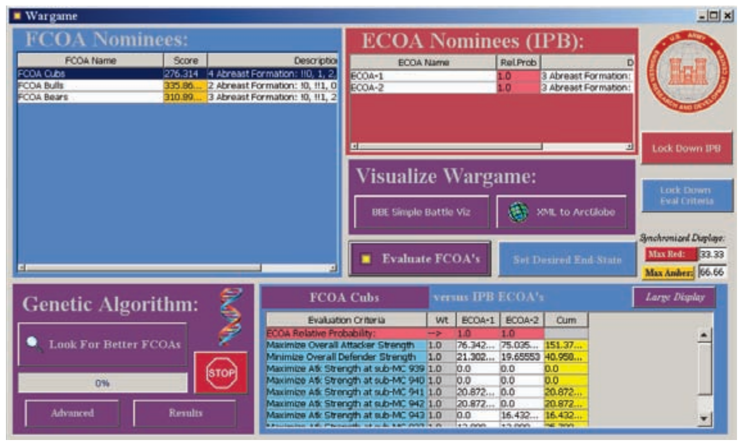
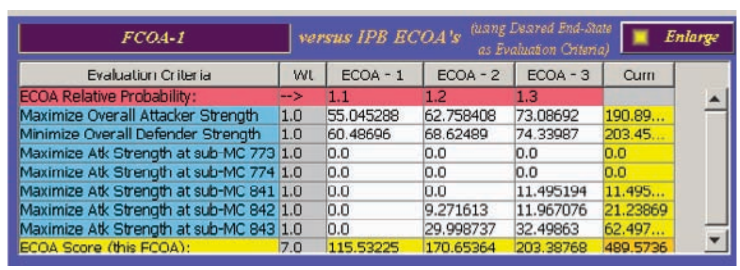


Figure 4. Wargaming FCOAs against ECOAs in BBE



Evaluation Criteria	Wt	ECOA - 1	ECOA - 2	ECOA - 3	Cum
ECOA Relative Probability:	-->	1.1	1.2	1.3	
Maximize Overall Attacker Strength	1.0	55.045288	62.758408	73.08692	190.89...
Minimize Overall Defender Strength	1.0	60.48696	68.62489	74.33987	203.45...
Maximize Atk Strength at sub-MC 773	1.0	0.0	0.0	0.0	0.0
Maximize Atk Strength at sub-MC 774	1.0	0.0	0.0	0.0	0.0
Maximize Atk Strength at sub-MC 841	1.0	0.0	0.0	11.495194	11.495...
Maximize Atk Strength at sub-MC 842	1.0	0.0	9.271613	11.967076	21.23869
Maximize Atk Strength at sub-MC 843	1.0	0.0	29.998737	32.49863	62.497...
ECOA Score (this FCOA):	7.0	115.53225	170.65364	203.38768	489.5736

Figure 5. Friendly COA Evaluation Scores in BBE

settings of the five terrain modifiers implemented in the model.

- Use the SEED Center's Nearly Orthogonal Latin Hypercube spreadsheet tool [6] to generate design points for the five terrain modifiers characterizing the maneuver network. We generated 513 design points across a range of values from 0.1 to 2.0 for each modifier.
- Use the SEED Center's OldMcData tool [7] to generate XML (Extensible Markup Language) excursion files from a base BBE scenario definition file (also in XML).
- Execute the model once (since the model computation is deterministic) for each of the excursion files.
- Use a post-processing tool (customized script developed by Steve Upton) to gather the output data into a single file to load into JMP (http://www.jmp.com) for statistical analysis.
- Perform quantitative and qualitative analysis on the results.
- From examination of various views and analyses in JMP, form new hypotheses possibly identifying other variables of interest. Then iterate the process.

RESULTS AND ANALYSIS

In the span of the workshop's four days, Team 7 was able to iterate over the above process three times, generating approximately 2000 individual outcomes from the BBE model. All of these runs were conducted in the context of a single demonstration scenario provided by the developers of the model. It should be noted that the model runs very quickly; although we used a small cluster for our runs, we could have easily performed all the runs on a laptop. Due to space limitations, we will only discuss our first two iterations in this paper.

By "scenario" we mean a single order of battle, along with a fixed set of three enemy courses of action, and a fixed set of evaluation criteria all taking place in a single physical setting. Initially, the scenario included two human-crafted FCOAs. The FCOAs are identified as "Devin Hester" and "Jay Cutler," or by their abbreviated names ("Devin" or "Hester" for the first, and "Cutler" for the second). In the final iteration of our process we also examined several machine-generated FCOAs. With respect only to the FCOAs and ECOAs, our experimental design is full-factorial.

Initial Line of Inquiry

The variables included in a course of action are many, and the dimensionality is not constant. That is, there are COA choices that add to the number of COA variables in a hierarchical way. Because our team was generally new to Data Farming, we chose to avoid that complexity and focus on a key premise of the model; i.e., that terrain properties are a key contributor to combat outcomes.

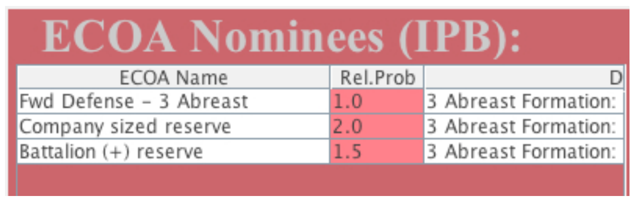
The model abstracts terrain into Maneuver Corridors (MC), which represent possible paths over which military units can travel. The MCs have a set of properties that are used by the model to compute rates of advance, and to modify the outcomes of combat activity that takes place within them. The modifiers available for each MC are:

- Road Speed
- Attack Maneuver
- Defense Maneuver
- Attack Fire Support
- Defense Fire Support

The Maneuver modifiers are applied to maneuver units such as tanks and infantry, and the Fire Support modifiers are applied to fire support units, such as artillery.

Our first iteration, then, farmed over these MC multipliers as a way to explore model behavior over various assumptions about the impact of terrain on model outcomes. As mentioned previously, we applied the SEED Center's NOLH DOE tool to explore values between 0.1 and 2.0 on each MC multiplier. This range covers terrain input values that might never occur in practice; however, we wanted initially to explore a full range of possibilities, with the expectation that future iterations will narrow our focus based both on our findings and the advice of subject matter experts.

The outcome of each model run is a score for each FCOA/ECOA pair. These pair-wise scores are also aggregated in a user-defined way that reflects the IPB estimates of the likelihood of encountering each ECOA (i.e., a weighted average). For our runs, we weighted the ECOAs as provided for in our example scenario, so the "overall scores" are based on the weightings in Figure 6.


ECOA Nominees (IPB):

ECOA Name	Rel.Prob	D
Fwd Defense - 3 Abreast	1.0	3 Abreast Formation:
Company sized reserve	2.0	3 Abreast Formation:
Battalion (+) reserve	1.5	3 Abreast Formation:

Figure 6. ECOA Weightings

The first thing we noticed from these runs was the relative unimportance of the Fire Support modifiers. In Figure 7, we show the per-ECOA scores across all the MC Multipliers, where you can see that Road and Maneuver multipliers are the only ones with visible effect. We took particular interest in a second feature, that there seem to be values of the Road Speed modifier that cancel out all other multipliers and clip the score to 250.

Because most of the team was also new to using the JMP software, we devoted a substantial amount of time to exploring the data with this new tool. As we thought about the use of this model as a decision aid, it occurred to us that the score might be of less interest than the relative ranking of FCOAs. With some JMP magic by Steve Upton, we produced Figure 8, which shows a clear break in preference from the Devin FCOA to the Cutler FCOA when the Road Speed multiplier gets below about 0.7.

We were initially somewhat surprised that changes to the Road Speed multiplier would cause such a sharp delineation of the recommended COA. However, upon reflection, this result is not so surprising. The model is telling us that not only does the Devin COA capitalize on high mobility terrain,

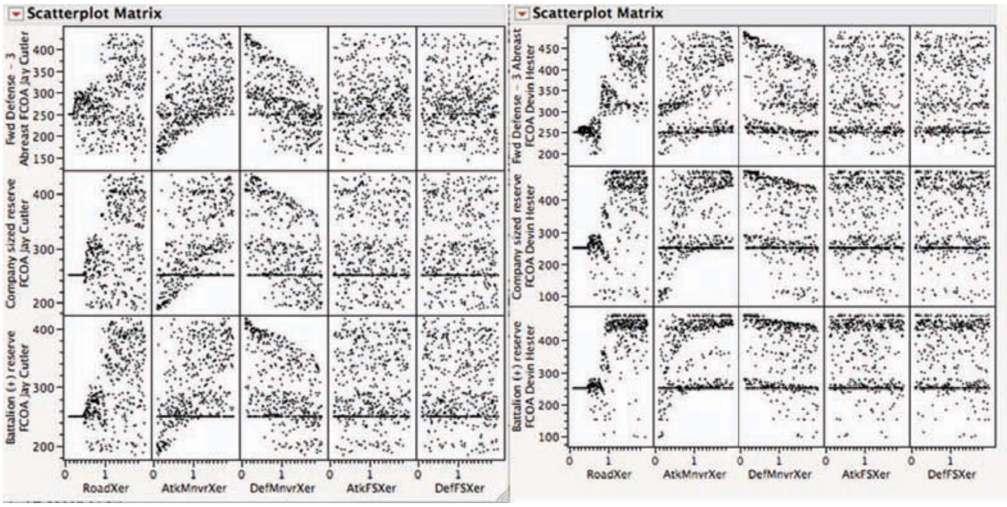


Figure 7. Scatterplots of FCOA Scores versus Terrain Factors

but that it may be a very bad choice when mobility is constrained. This conclusion of the model can now be subjected to expert criticism; for example, all things considered, for this specific scenario, do experienced tacticians agree with this conclusion?

As newcomers to Data Farming, we found this outcome to be encouraging with respect to using data farming as a tool to support questions of model validation.

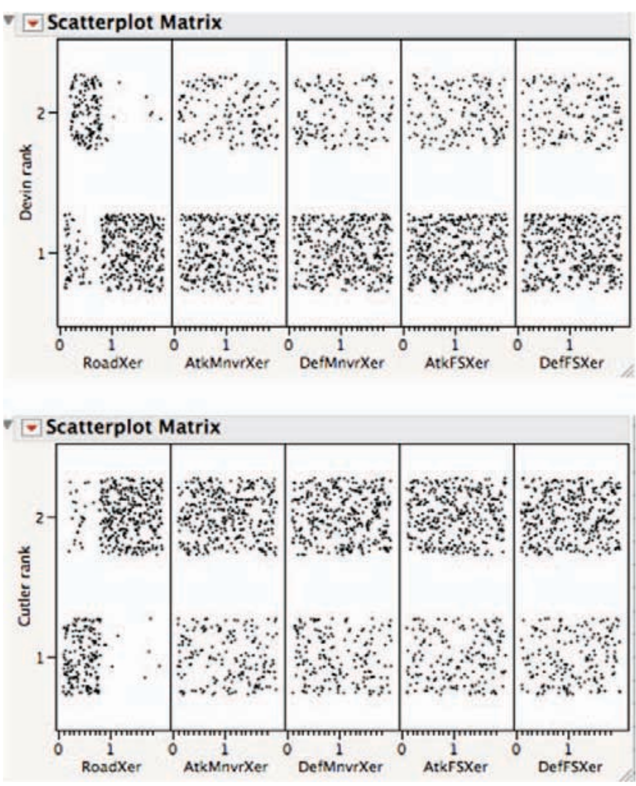


Figure 8. Scatterplots of FCOA Rank versus Terrain Factors

New Questions for Iteration Two

The BBE model gives the planner a choice for the time resolution of the combat model. Time slices available are 6, 12, 18, 24 and 30 minutes. The importance of the road speed multiplier in the first iteration caused us to wonder how the time slice selected would impact the results. We formulated two new research questions.

a. Is the outcome of the model (i.e., rank and/or scoring of a COA) dependent on the time slice selected?

b. Are the conclusions about why one COA is better than another consistent across all the time slices? (i.e., is the conclusion about Devin's superiority in high road speed multiplier consistent as we change the time slice?)

To perform this test, we examined the outcomes across the 5 time slice options provided in the model.

Figure 9 shows the score distribution (aggregated scores, recall Figure 6) for two time slice selections. Clearly Devin does better in the upper frame, which is a time slice value of 12 minutes, than it does in the lower frame, where the time slice is 6 minutes.

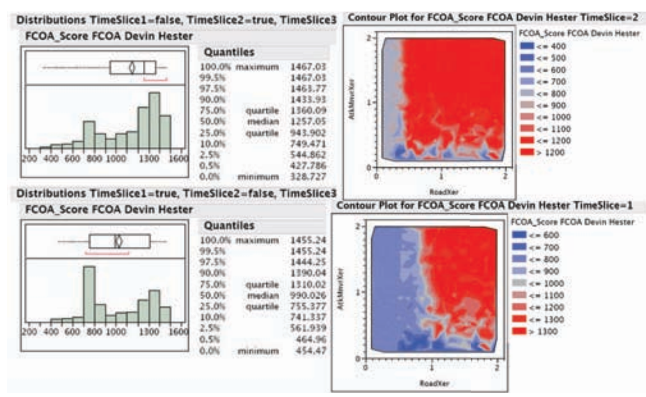


Figure 9. Distribution of FCOA Scores by Timeslice Value

We again return to the purpose of the model as a decision aid and investigate the relative rankings. Figure 10 confirms that Hester (i.e., Devin) does relatively better with a larger timeslice, winning 443 times in a 12 minute time slice and only 339 in a 6 minute time slice.

In retrospect, we would not use the aggregated scores for this analysis. As we are attempting to validate the underlying combat resolution mechanism, dealing with the probabilities of encountering any particular ECOA only serves to cloud the results. If we can make sound statements about the model's recommendations for each FCOA/ECOA pair, then we have done our job. The validity of weighting by relative probability of occurrence is a separate matter.

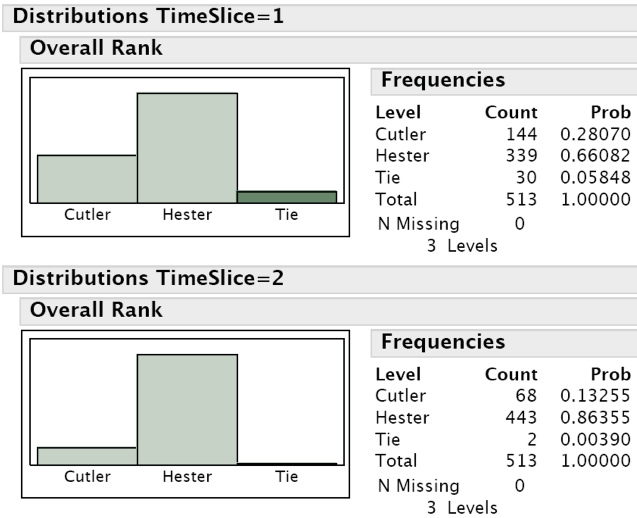


Figure 10. Distribution of FCOA Ranks by Timeslice Value

Perhaps of even more interest is the second question. This interest is because our intuition would suggest that the implementation detail of selecting a time slice should not cause the model to give different conclusions about how the world works.

T

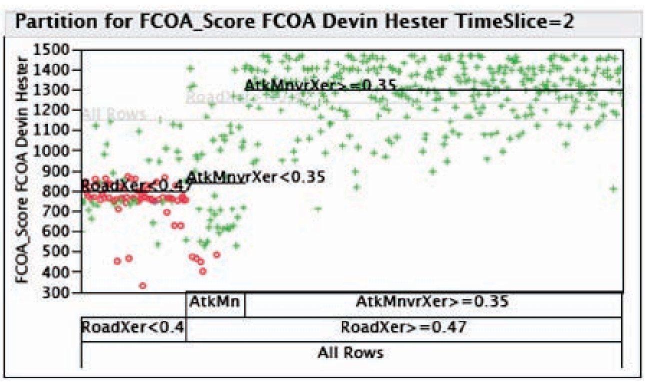


Figure 11. Regression Tree for Selected FCOA and Timeslice 2 (12 minutes)

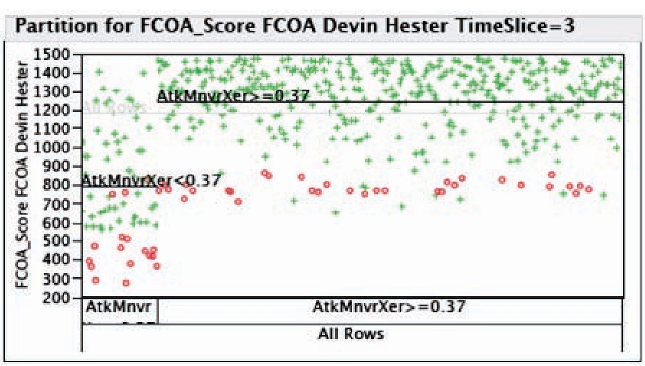


Figure 12. Regression Tree for Selected FCOA and Timeslice 3 (18 minutes)

The results actually show that the important factors in the model change as a result of changing the time slice selection. Figure 11 is a regression tree with time slice set to 12 minutes. This analysis applies equally to the runs performed in our first iteration, and shows the same result. The green FCOA (Devin Hester) performs best when the Road Speed multiplier is greater than 0.4. We remain suspicious that something else may be going on at these low values, since both COAs cluster tightly around the aggregated score of 800.

Figure 12 displays the same analysis for a time slice selection of 18 minutes. Here, though, the most important factor is the Attack Maneuver Multiplier.

Once again we have uncovered something significant for validation of the model. Both sets of conclusions about what factor is most important to the outcome in this scenario cannot be correct.

We are therefore led to ask, "Which time slice is the correct one?" (or are both of them wrong). For a decision aid, should the analyst even have access to implementation details that can have such an impact? At the very least, the data farming effort made it easy to discover something about the model that deserves more attention.

SUMMARY OF FINDINGS

We found that Data Farming is an efficient way to find out what the model thinks is important. This finding speaks directly to the process of validation. For example, for the BBE tool, we discovered that time slice selection affected the relative performance of the COAs, changing the distribution of scores, changing which COA is "best," and changing which terrain modifier was most influential in the outcome. Such information is useful for software developers and users of the tool to consider.

CONCLUSIONS

The main interest Team 7 had coming into the workshop was to gain insight into how data farming might contribute to the validation process for a model. We immediately experienced first-hand the power of space-filling experimental designs when our first iteration highlighted the importance of the Road Speed multiplier in our test scenario.

A second, and more direct, contribution to our validation effort emerged when the highly exploratory nature of the data farming process allowed us to investigate model time-slice selection. Our discovery that the model changes character based on the selected time-slice is a significant finding that will be of immediate concern to the software developers.

Other data farming possibilities:

- Find the "right" time-slice value (or, why provide the selection?)
- Farm over COA parameters
- Farm over the value systems, such as the Commander's evaluation criteria weightings
- Tune performance of the decision aid by farming over Genetic Algorithm parameters to find the most efficient settings.
- Investigate sensitivities and impact of the global force modifiers provided in the decision aid (posture, C2, morale, etc.)

REFERENCES

[1] Schlabach, J. (2008). "BBE Cognitive Amplification for Battle Staff Planning." ERDC Topographic Engineering Center. Battlespace Terrain Reasoning & Awareness Battle Command (BTRA BC) Joint-Geospatial Enterprise Service (J-GES). Vol 3:2, pp 5-10.

[2] U.S. Army Field Manual 101-5. Staff Organization and Operations.

[3] U.S. Army Field Manual 34-130. Intelligence Preparation of the Battlefield.

[4] Petty, Mikel D. (2009). "Verification and Validation." In Principles of Modeling and Simulation: A Multidisciplinary Approach. Edited by John A. Sokolowski and Catherine M. Banks. John Wiley & Sons, Inc. Hoboken, NJ.

[5] Appleget, J., Duong, D., Blais, C., Perkins, T., and Brown, R. F. (2010). Irregular Warfare (IW) Model Validation and Best Practices Guide. DRAFT. Naval Postgraduate School. March 2010.

[6] Sanchez, S. M. 2005. "Work Smarter, not harder: Guidelines for designing simulation experiments," Proceedings of the 2005 Winter Simulation Conference, 69-82. Software and references available at http://harvest.nps.edu

[7] Old McData and other software and references are available at http://harvest.nps.edu