A Clinical Paper On Efficient Search Strategies in High-Dimensional Complex Models

Major Thomas M. Cioppa, United States Army, Doctoral Candidate, United States Naval Postgraduate School, Department of Operations Research, Monterey, CA 93940

Dr. Thomas W. Lucas, Associate Professor, United States Naval Postgraduate School, Department of Operations Research, Monterey, CA 93940

Simulation models have become increasingly complex, often with a large number of factors (variables) requiring examination. A comprehensive exploration of these factors may not be feasible even as processing speed increases or with parallel computing. For example, a model with 26 factors, each at two levels, with only one replication would require over 67 million observations in a traditional full factorial design (2^{26}). This could be considered a "small" problem since many computer simulations have thousands of factors. If one is willing to assume negligible interactions, then a fractional-factorial design is satisfactory, but still the required number of observations may be quite large. These "traditional" experimental designs and associated theory have been well researched (Box, Hunter, and Hunter (1978) and Hicks (1993)). Experimental designs relating specifically to computer simulations and models with a large number of factors were studied by Jacoby and Harrison (1962), Hunter and Naylor (1970), Kleijnen (1975), Biles (1979), Welch et al (1992), and Bates et al (1996).

As the number of factors available for experimentation has increased, methodologies were developed to reduce the number of observations required to identify significant main effects. Earlier work by Dorfman (1943) with group screening was extended by Watson (1961) and Li (1962). Subsequent screening techniques to identify the most significant factors, denoted as factor screening, were developed by Ott and Wehrfritz (1972), Montgomery (1979), Smith and Mauro (1982), Schruben (1986), and Mauro (1986). More recent factor screening methodologies include sequential bifurcation by Bettonvil and Kleijnen (1996), Latin supercube sampling by Owen (1998), and supersaturated designs by Yamada and Lin (1999). The primary focus of these methodologies is efficiently (defined as reducing the number of observations) identifying the critical factors (defined as contributing the most to the outcome measure). The critical assumption in these methodologies is that interactions are negligible. The objective is to efficiently identify the main effects, and then, if necessary, perform additional experimentation for refinement.

In certain areas, for example command and control in military conflict, interactions are prevalent and cannot be assumed negligible, even from the initial onset of the experimentation. The presence of non-monotonicity and chaos in combat models, defined as battles on the edge, is an obscuring factor in the search (Dewar, Gillogly, and Juncosa (1991)). Finding general patterns of behavior is complicated by the difficulties associated with exploring high-dimensional non-linear model surfaces. Existing search methodologies often restrict analysts to comprehensively exploring a tiny hyperplane of model space, finding extreme points, or varying many factors simultaneously by confounding main effects with interactions. The result of the aforementioned approaches may fail by either identifying a local extrema as a global extrema or failing to identify critical interactions.

Hencke (1998) developed an agent-based combat simulation to analyze information and coordination. His model succeeded in "providing a simulation where the agent's actions are reasonably well behaved, that is reminiscent of well-trained forces." His work is an excellent representation of a high dimensional complex model where interactions are significant and will provide a basis for the proposed experimentation. Assume an imminent conflict will occur between two opposing forces, denoted as Blue and Red. Table 1 shows the factors, each a variable capable of multiple levels, which can be adjusted and will serve as an excellent source for experimentation. Archimedes, an agent-based simulation being developed under the Marine Corps Combat Development Command's Project Albert, contains a much larger factor set than Hencke's model and may serve as an excellent model to further explore a proposed search strategy.

Table 1: Factors for Experimentation

Red Forces	**Blue Forces**
Red agents in cell	Blue agents in cell
box center x	box center x
box center y	box center y
box size x	box size x
box size y	box size y
goal x	goal x
goal y	goal y
probability hit	probability hit
speed	Speed
sensor/shoot range	sensor/shoot range
charge ratio	charge ratio
runaway ratio	runaway ratio
maximum hits	maximum hits

The goal of this research is to determine a methodology and associated theory that has the ability to efficiently search across the breadth of factors to identify not only main effects, but also critical second-order (and perhaps third-order) interactions. A sequential design, where the number of observations may not be known in advance, which combines traditional full and fractional factorial designs, Latin supercube, and supersaturated designs together with group screening, prior information, and expert judgment may provide an effective approach to efficiently identify not only main effects, but also appreciable interactions. Furthermore, once an appropriate set of factors and their interactions have been identified, a perturbation approach to judge the stability of the proposed set will be integrated into the search methodology.

REFERENCES

Bates, R.A., Buck, R.J., Riccomagno, E., and Wynn, H.P., "Experimental Design and Observation for Large Systems," *Journal Royal Statistical Society* 58, No.1, pp. 77-94, 1996.

Bettonvil, B. and Kleijnen, J.P.C., "Searching for important factors in simulation models with many factors: sequential bifurcation," *European Journal of Operations Research* 96, pp. 180-194, 1996.

Biles, W.E., "Experimental Design in Computer Simulation," *1979 Winter Simulation Conference*, pp. 3-9, 1979.

Box, G.E.P., Hunter, W.G., and Hunter, J.S., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley and Sons, New York, 1978.

Dewar, J.A., Gillogly, J.J., and Juncosa, M.L., "Non-Monotonicity, Chaos, and Combat Models", Final Report, RAND Corporation, Santa Monica, CA, 1991.

Dorfman, R., "The detection of defective numbers of large populations," *The Annals of Mathematical Statistics* 14, pp. 436-440, 1943.

Hunter, J.S. and Naylor, T.H., "Experimental Designs for Computer Simulation Experiments," *Management Science*, Vol. 16, No. 7, pp. 422-434, 1970.

Jacoby, J.E. and Harrison, S., "Multi-variable experimentation and simulation models," *Naval Research Logistics Quarterly* 9, pp. 121-136, 1962.

Hencke, R.B., "An Agent-Based Approach to Analyzing Information and Coordination in Combat," Master's Thesis, United States Naval Postgraduate School, Monterey, CA, September 1998.

Hicks, C.R., *Fundamental Concepts in the Design of Experiments*, Saunders College Publishing, New York, 1993.

Kleijnen, J.P.C., "Screening designs for poly-factor experimentation," *Technometrics* 17, pp. 487-493, 1975.

Li, C., "A sequential method for screening experimental variables," *Journal of the American Statistical Association* 57, pp. 455-477, 1962.

Mauro, C.A., "Efficient Identification of Important Factors in Large Scale Simulations," *Proceedings 1986 Winter Simulation Conference*, pp. 296-305, 1986.

Montgomery, D.C., "Methods for Factor Screening in Computer Simulation Experiments," Final Report, Office of Naval Research, Contract N0014-78-C-0312, Georgia Institute of Technology, Atlanta, GA, 1979.

Ott, E.R. and Wehrfritz, F.W., "A special screening program for many treatments," *Statistica Neerlandica*, Vol. 26, pp. 165-170, 1972.

Owen, A.B., “Latin Supercube Sampling for Very High-Dimensional Simulations,” *ACM Transactions on Modeling and Computer Simulation*, Vol. 8, No. 1, pp. 71-102, 1998.

Schruben, L.W., “Simulation optimization using frequency domain methods,” *Proceedings 1986 Winter Simulation Conference*, pp. 396-369, 1986.

Smith, D.E. and Mauro, C.A., “Factor Screening in Computer Simulation,” *Simulation* 38, pp. 49-54, 1982.

Watson, G.S., “A Study of the Group Screening Method,” *Technometrics* 3, pp. 371-388, 1961.

Welch, W.J, Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., and Morris, M.D., “Screening, Predicting, and Computer Experiments,” *Technometrics*, Vol. 34, No. 1, pp. 15-25, 1992.

Yamada, S. and Lin, D.K.J., “Three-level supersaturated designs,” *Statistics and Probability Letters* 45, pp. 31-39, 1999.