Proceedings of the 2004 Winter Simulation Conference
R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

AN EFFICIENT SCREENING METHODOLOGY FOR A PRIORI ASSESSED NON-INFLUENTIAL FACTORS

Thomas M. Cioppa

US Army Training and Doctrine Command
PO Box 8695
Monterey, CA 93943, U.S.A.

ABSTRACT

The choice of which factors to choose to vary in a simulation to effect a change in the measure of interest is difficult. Many factors are *a priori* judged not to effect the measure of interest. There is often no subsequent analysis of whether this judgment is valid or not. The proposed methodology outlines a sequence of limited runs to assess the accuracy of the *a priori* belief of non-influential factors. These runs can either identify if an influential factor(s) has been omitted or confirm *a priori* beliefs. Executing these sequences of runs before focusing on the suspected influential factors can contribute to subsequent analysis by confirming that suspected non-influential factors do not significantly impact the measure of interest. An example of the methodology using an agent based model is presented.

1 INTRODUCTION

Computers have permitted a more complex environment to conduct experiments. A computer simulation is a computer model that attempts to imitate the actual, real-world problem or scenario. Hunter and Naylor (1970), Ignall (1972), and Montgomery (1979) address experimental designs for computer simulations by using factorial, fractional factorial, rotatable, and response surface designs. Still, the maximum number of factors examined in any of these designs is seven; a small number of factors when compared to the number of factors examined in many simulation models today (Law and Kelton 2000).

The Pareto principle is a common assumption made in experimental designs for computer simulations. This implies that although there are a plethora of input factors, only a small number of these factors will have a significant effect on the output response. This is known as factor sparsity. A preliminary investigation of a computer simulation may involve group screening for these significant factors as a precursor for more detailed experimentation. Group screening originated with Dorfman (1943) who demonstrated the decrease in the number of experiments required in laboratory testing by placing individual samples into groups and then testing each group, as if it were an individual sample, for presence of the desired characteristic. His goal was to decrease the total number of tests required to identify those World War II drafted males who had syphilis.

Group screening is a common technique used in simulations with many factors. The goal is to identify the critical, important, or influential factors in a limited number of runs and then conduct further experimentation on these critical factors. Common assumptions and limitations of group screening designs (e.g., Watson 1961, Montgomery 1979) include:

- only a small number of the total possible factors are important,
- each factor is restricted to two levels of testing,
- no interactions between factors are present,
- main effect directions are known,
- all factors have the same probability of being important, and
- observation errors are independently normal with constant known variance.

Dorfman's group screening has been extended by numerous researchers. Watson's (1962) group screening designs extended Dorfman's work by grouping the k factors into g groups and then using a two-stage procedure. Li (1962) and Patel (1962) extended and generalized the procedure for more than two stages. Mauro (1984) extends Watson's work by not assuming the direction of possible effects. Bettonvil (1995) proposes a modification of Jacoby and Harrison's (1962) sequential bifurcation, a method of group screening.

This paper extends group screening by using a novel procedure that capitalizes on expert opinion or prior knowledge and then uses specifically tailored simulation runs to confirm perceived non-influential factors. In this paper, the term non-influential implies that the factor, regardless of its setting, does not cause a significant change in the output measure(s) of interest. Instead of verifying important factors, this method is used to verify the assumed non-influential factors. The remainder of the paper is organized as follows. Section 2 provides details of the methodology, section 3 presents justification for the procedure, section 4 illustrates the use of the methodology, and section 5 concludes with suggestions for subsequent research.

2 METHODOLOGY DESCRIPTION

This section describes, in detail, the sequential methodology used to assess the accuracy of the *a priori* knowledge of the non-influential factors. In analyses, people who implement a model, those who analyze the model output, and/or area subject matter experts with real-world experience develop a rigorous insight concerning the effect of different factors or parameters on the measures of interest. It is not uncommon that people have over a decade's worth of experience with a particular model. Furthermore, querying those with actual operational experience is important to enhance the model's accuracy in portraying the simulated environment. Not taking advantage of these resources would not be wise, but concurrently, we must ensure that we temper this judgment with actual model runs to verify the intuition.

If a specific set of factors has been identified as the focal point of investigation, then the following described method does not require implementation. This situation may occur when the experiment is scoped or tailored to investigate particular factors. Alternatively, if there is no predetermined factor list identified for experimentation and doubt exists as whether to classify a factor as important or unimportant, then this method will be beneficial.

Expert opinion is queried using some collection and analysis methodology (e.g., survey or Delphi technique). The goal is to produce a categorization of the factors into three stochastically ordered groups (e.g., Agresti 1990). These groups consist of the categories of most likely to be important [ξ], unknown [ψ], and not likely to be important [τ]. It is recommended that the unknown category contain as few factors as possible. This does not seem a prohibitive restriction since the expert opinion and prior knowledge should reduce the uncertainty of a factor's effect. Additionally, we assume that the direction of the effect of a factor is known (e.g., Bettonvil 1995). The methodology is applicable to both deterministic and stochastic simulations, but stochastic simulations will require replications.

With the factors grouped into one of three stochastically ordered groups, a sequence of runs described in the following section is conducted. Note that "+" refers to the maximum setting (favorable) for that factor and "-" refers to the minimum value (unfavorable) for that factor. Thus, in a similar fashion as done with group screening, if the category of [ξ] is assigned a "+" setting, then all factors within that category have a "+" setting.

2.1 Sequence I Simulation Runs

Sequence I consists of two runs and establishes a baseline to better determine factor effects. The first run sets all factors at their maximum setting and thus should yield the best possible outcome for the desired measure of interest. The second run sets all factors at their minimum setting and thus should yield the worst possible outcome for the desired measure of interest. Table 1 summarizes these two runs where the observation corresponds to the measure of interest.

Table 1: Sequence I Simulation Runs

ξ	ψ	τ	Observation
+	+	+	γ_1
-	-	-	γ_2

From these two runs, we establish the following equation.

$$\gamma_1 - \gamma_2 = \delta_1 . \qquad (1)$$

In theory, there should be a sufficiently large δ_1, but this may not be evident (e.g., if the error term has a complicated variance structure). Thus, it would be unwise to conclude with finality that the factors have minimal impact upon the measure of interest if the difference between the two observations is small. Note that when the terms of large and small corresponding to the δ's is made, this suggests that the user is the one to define the comparative magnitude. After only two runs, no defining characteristics of the relationship can be made, but future runs will yield additional information permitting more detailed explanation.

2.2 Sequence II Simulation Runs

Sequence II consists of two runs whose purpose is to verify those factors designated as not important. We employ the following rationale. Since [τ] contains perceived non-influential factors, there should not be a significant δ found between γ_1 and a modified run corresponding to [τ] being assigned a "-" setting. Similarly, there should not be a significant δ between γ_2 and a modified run corresponding to [τ] being assigned a "+" setting. Table 2 summarizes these two runs.

Table 2: Sequence II Simulation Runs

ξ	ψ	τ	Observation
+	+	-	γ_3
-	-	+	γ_4

From these additional two runs, we now establish the following two equations.

$$\gamma_1 - \gamma_3 = \delta_2 \quad (2)$$

$$\gamma_2 - \gamma_4 = \delta_3 . \quad (3)$$

At this point, four runs have been conducted (Tables 1 and 2) so it is possible to gain some insight into the measure of interest's variance and the accuracy of the expert opinion. If the [τ] category truly contains non-influential factors, then δ_2 and δ_3 should be small. Thus, there are four possible scenarios which may exist and are arranged in order from most desirable to least desirable where desirability is measured in terms of additional runs required to verify non-influential factors.

- Scenario 1: If sequence I has a large δ_1, paired with small δ_2 and δ_3, then this suggests that the [τ] category does contain non-influential factors. This can be considered the best situation. Simulation runs focusing on the important factors can commence.
- Scenario 2: If sequence I has a small δ_1, and δ_2 and δ_3 are small, then this suggests that the measure of interest may not be substantially influenced by any of the factors. Simulation runs focusing on the important factors can commence, if desired.
- Scenario 3: If sequence I has a large δ_1, and one or both of δ_2 and δ_3 are large, then this suggests that the [τ] category may contain an influential factor(s) not found with expert opinion or a complicated variance exists for the measure of interest. Sequence III (described below) runs should be conducted.
- Scenario 4: If sequence I has a small δ_1, and one or both of δ_2 and δ_3 are large, this suggests that the expert opinion or previous knowledge may have incorrectly classified factors (in both the [ξ] and [τ] categories) or a complicated variance for the measure of interest exists. This may be considered the worst situation, but in practice, a gross factor misclassification of this nature should occur very rarely. If a complicated variance is the root cause, it will become evident as the procedure ensues. Although sequence III may be conducted, it is recommended instead to conduct a thorough reexamination of the categories and re-executing sequences I and II.

Note that we do not separately examine the [ξ] category, but assume that influential factors are contained in this category. Otherwise, we would have encountered scenario 2 and possibly terminated the simulation experiment or encountered scenario 4 (a rare event) which requires further investigation and refinement of the categories. Furthermore, if non-influential factors are incorrectly contained in the [ξ] category, these will be identified during subsequent simulation runs.

2.3 Sequence III Simulation Runs

If scenario 3 (or possibly 4) is found, sequence III is done in an attempt to determine which factors in the [τ] category are influential. The expert opinion or previous knowledge can be reexamined to determine if an oversight occurred. The [τ] category can then be revised and sequence II repeated with the goal of achieving scenario 1. The other recommended alternative is to conduct sequential bifurcation (Bettonvil, 1995) on the original [τ] category to identify the influential factors. With k representing the number of important factors in the [τ] category, m representing the total number of factors in the [τ] category, and ℓ an integer subject to the following conditions:

$$\begin{gathered} 0 \le \ell \le 2m \\ 2^{\ell-1} < k \le 2^{\ell}, \end{gathered} \quad (4)$$

then in the worst case, the number of required observations (Bettonvil, 1997) during sequence III is

$$1 + 2^{\ell} + k(m - \ell) . \quad (5)$$

2.4 Sequence IV Simulation Runs

Sequence IV focuses on the [ψ] category where the most uncertainty exists. Partition [ψ] into equal (or nearly) subgroups subject to the following constraint where n equals the total number of subgroups,

$$[\psi] = \sum_{i=1}^{n} [\psi_i] . \quad (6)$$

A resolution III experimental design is then conducted on the subgroups, explained later in this paper, but first we address the consequences of subgroup size and the impact of finding a subgroup containing an influential factor(s).

The first alternative is to transfer each factor in the subgroup to the [ξ] category. The second alternative is to conduct sequential bifurcation on the subgroup to isolate influential factors. Thus, choosing the size of the subgroups of [ψ] is an important consideration. If the subgroups are too large and the first alternative is chosen, then the [ξ] category may get unnecessarily large for the follow-on phases.

If the subgroups are large and the second alternative is chosen, then the sequential bifurcation will require additional runs to determine influential factors. Conversely, subgroups which are too small will yield excessive runs in the proposed resolution III design. Since the follow-on phases will require extensive runs, our recommendation is to use the second alternative and choose relatively small subgroups (no more than 10 factors in a subgroup).

Since the focus is to determine the influence of the [ψ] subgroups, we set the [ξ] category to the "-" level in order to magnify the effects of any influential factors in a [ψ] subgroup. From sequence 3, the [τ] category should contain the non-influential factors and therefore have no effect whether its factors are set at "+" or "-." Arbitrarily the [τ] category is set at "+." Thus, the [ξ] and [τ] categories are considered fixed, and a resolution III design is constructed for the [ψ] subgroups. Table 3 illustrates an example of this design where there are three [ψ] subgroups.

Table 3: Example Resolution III

ξ	τ	ψ_1	ψ_2	$\psi_3 = \psi_1\psi_2$	Observation
-	+	+	+	+	γ_5
-	+	+	-	-	γ_6
-	+	-	+	-	γ_7
-	+	-	-	+	γ_8

From the resolution III design and its associated analysis of variance, influential [ψ] subgroups can be identified. As previously noted, an entire subgroup may then be transferred to the [ξ] group or sequential bifurcation can be done on the influential [ψ] subgroup(s) to isolate the influential factors. The sequential bifurcation alternative is recommended to reduce the computational burden during subsequent computer experimentation.

3 METHODOLOGY JUSTIFICATION

Although the runs required to corroborate *a priori* non-influential factors appears to be excessive, this is a necessary step for two major reasons. The first reason has already been mentioned; namely, by conducting these simulation runs, fewer runs will be required during subsequent computer experimentation. The second reason is that if the overall process must be terminated early due to an accelerated decision requirement or some other unforeseen event, insights can still be acquired during any point of the sequence.

Theorem 1 is presented which details the worst-case performance for the number of runs required under scenarios 1, 2, and 3. Scenario 4 is omitted, although its worst-case performance could be found depending on the outcome of the re-execution of sequences I and II.

Theorem 1 *Under scenarios 1 or 2, in the worst case the number of observations is 4. Under scenario 3, in the worst case the number of observations is*

$$5 + 2^{\ell} + k(m-\ell) + \sum_{i=1}^{p} [1 + 2^{\ell_i} + k_i(m_i - \ell_i)], \quad (7)$$

where k, m, and ℓ are as previously defined, k_i represents the number of important factors in a particular [ψ] subgroup, m_i represents the total number of factors in the particular [ψ] subgroup, p represents the total number of [ψ] subgroups from the minimal resolution III design, and ℓ_i is an integer subject to the following conditions:

$$0 \le \ell_i \le m_i$$
$$2^{\ell_i - 1} < k_i \le 2^{\ell_i}. \quad (8)$$

The proof of Theorem 1 is straightforward. From sequences I and II under scenarios 1 and 2, the number of observations required is 4. Equation (7) follows from a combinatorial extension of Equation (5) from the minimal resolution III design. This concludes the proof.

Clearly we expect the actual performance of phase I to be superior to that prescribed under Theorem 1 since otherwise, the expert opinion or previous knowledge would be simply random selection.

4 IMPLEMENTATION EXAMPLE

The Map Aware Non-Uniform Automata (MANA) is an agent-based simulation that is being developed by the New Zealand's Defence Technology Agency (Lauren, Stephen, and Anderson 2001). MANA is a useful simulation for analysis of complex military problems (Cioppa and Brown 2002).

A peace enforcement scenario was developed that consisted of a friendly force that is subjected to a series of encounters with a hostile force and a non-hostile force that turns hostile after a specified time. There were 10 factors identified for the [ξ] category, three factors for the [ψ] category, and 11 factors for the [τ] category. The measure of interest was how many friendly forces were killed or wounded (a lower value is preferable). Since MANA is stochastic, 30 replications were done for each specified run to attain the average measure of interest.

The runs specified in Table 1 were executed. Equation (1) resulted in a value of 16. The runs specified in Table 2 were then executed. Equations (2) and (3) resulted in values of 7 and 5, respectively. These values were considered large indicating that a factor in the [τ] category may be influencing the measure of interest.

Sequence III runs were then executed by using sequential bifurcation and required five runs. The procedure identified a factor (movement speed) that influenced the number of friendly forces killed or wounded, although movement speed had not been identified *a priori* by expert judgment.

Although a total of nine runs were executed prior to detailed experimentation of the $[\xi]$ category factors, these runs identified an important factor, originally thought to be unimportant. Subsequent analysis from this detailed experimentation indicated that movement speed was the second most significant factor in number of friendly forces killed or wounded. Specifically, increasing movement speed resulted in a decrease of friendly forces killed or wounded.

5 CONCLUSIONS

Choosing which factors to focus upon within a simulation is difficult. Expert opinion, judgment, or experience with the simulation are used to determine which factors might influence the measure of interest. Often this judgment is not verified with actual simulation runs and can be inaccurate.

Although the majority of simulation and analysis effort should be upon the important factors, a subset of runs prior to this detailed experimentation should be used to verify *a priori* judged non-influential factors. The methodology presented in this paper provides an efficient approach for identifying potential important factors that were originally omitted from consideration.

Subsequent research will focus on two areas. The first area will be a comparison of this methodology between deterministic and stochastic simulations. The second area of research will examine strategies that decrease the likelihood of scenario 4 occurring.

ACKNOWLEDGMENTS

The author thanks Dr. Thomas W. Lucas and Dr. Susan M. Sanchez from the Department of Operations Research at the U.S. Naval Postgraduate School for their help and assistance.

REFERENCES

Agresti, A. 2002. *Categorical Data Analysis*. New York: Wiley and Sons.

Bettonvil, B. 1995. Factor Screening by Sequential Bifurcation. *Communications in Statistics: Simulation* 24 (1): 165-185.

Cioppa, T. and L. Brown. 2002. *Agent Based Models as an Exploratory Analysis Approach to Combat Simulations*. Monterey, California: US Army TRADOC Analysis Center.

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

Hunter, J. and T. Naylor. 1970. Experimental Designs for Computer Simulation Experiments. *Management Science* 16 (7): 422-434.

Ignall, E.J. 1972. On Experimental Designs For Computer Simulation Experiments. *Management Science* 18 (7): 384-388.

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

Lauren, M., R. Stephen, and M. Anderson. 2001. *Map Aware Non-Uniform Automata Version 1.0 Users Manual*. New Zealand: Defence Technology Agency.

Law, A.M. and W.D. Kelton. 2000. *Simulation Modeling and Analysis*. New York: McGraw-Hill.

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

Mauro, C.A. 1984. On the Performance of Two-Stage Group Screening Experiments. *Technometrics* 26 (3): 255-264.

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

Patel, M.S. 1962. Group-Screening With More Than Two Stages. *Technometrics* 4 (2): 209-217.

Watson, G.S. 1962. A Study of the Group Screening Method. *Technometrics* 3: 371-388.

AUTHOR BIOGRAPHY

THOMAS M. CIOPPA is a Lieutenant Colonel in the U.S. Army. He currently serves as the Director of the U.S. Army Training and Doctrine Command Analysis Center in Monterey, California. He has a Ph.D. in Operations Research from the U.S. Naval Postgraduate School. His email address is <`tom.cioppa@trac.nps.navy.mil`>.