17th ICCRTS

"Operationalizing C2 Agility"

Title of Paper

Validating Large Scale Simulations of Socio-Political Phenomena

Topic(s)

Primary
Topic 6: Modeling and Simulation
Alternate
Topic 3: Data Information and Knowledge

Authors

Grace I. Scarborough, Ph.D.	David T. Signori, Jr., Ph.D.	Szymon T. Letowski
Evidence Based Research 1595 Spring Hill Road Suite 250 Vienna, VA 22182	Evidence Based Research 1595 Spring Hill Road Suite 250 Vienna, VA 22182	Evidence Based Research 1595 Spring Hill Road Suite 250 Vienna, VA 22182

Point of Contact

Grace I. Scarborough, Ph.D.
703-287-0376
scarborough@ebrinc.com

Name of Organization

Evidence Based Research 1595 Spring Hill Road
Suite 250
Vienna, VA 22182
www.ebrinc.com

Validating Large Scale Simulations of Socio-Political Phenomena

Introduction

A significant issue of interest for users of a complex Human Social Cultural Behavioral (HSCB) simulation model is the identification of the active factors in the model for the scenarios under consideration. For models with a small number of factors, this can often be achieved by performing a large set of sensitivity analyses. The number of sensitivity analyses required rises very quickly, however, as the number of factors and interactions between factors increases, and the whole process can easily become a very involved and tedious task. Box and Meyer (1993) introduced a Bayesian method for identifying active factors using 2-level experimental designs that consolidates the process and considers interaction effects along with main effects. The method works by generating, for each of the factors, a (posterior) probability of that factor being active. The Box-Meyer method was developed for real-world physical experiments, in which the number of factors to be considered is generally limited and the number of active factors is assumed to be a small subset of all the factors. Its applicability to HSCB simulation models with numerous factors was not immediately certain. Webb and Bauer (1994) did an initial comparison of the Box-Meyer method to two other approaches for identifying active factors using a large-scale Air Force simulation (THUNDER) and found the Box-Meyer method to be superior, but their analysis was done on the basis of a subset of 13 factors. Additional work by Samset and Tyssedal (1998) looked at using the Box-Meyer method with models containing a large percentage of active factors, and they introduced an extension to the original method to correct for an underestimation of the active factor probabilities.

The goal of our research was to adapt the Box-Meyer method to work with larger models than previously considered and to develop a tool that would allow HSCB simulation model users to automatically identify the active factors for specified scenarios without needing to understand the details of the Box-Meyer method. Once the active factors have been identified and understood, the simulation behavior can be further and more deeply analyzed with other tools.

The Research Phase

The primary challenge in our research was scalability — dealing with models that have large numbers of factors (e.g., 30 or greater). When a model has this many factors, it is computationally infeasible to include them all in a single Box-Meyer analysis. This is because, internally, the Box-Meyer method utilizes all-subsets regression, which involves performing 2^N individual regressions. Thus, in order to use the Box-Meyer method, only subsets of all the factors can be analyzed at one time. However, simply partitioning the factors into reasonably sized, disjoint subsets and applying the method to each partition separately is not possible as the effects of any interactions between factors in different partitions will be lost.

We developed an approach based on splitting the factors into disjoint groups and performing individual Box-Meyer analyses on a set of overlapping "partitions." These partitions are defined as all the possible combinations of 2, 3, 4, etc. of the disjoint factor groups. The number of groups used in creating the partitions is determined by the order of factor interaction that the

Box-Meyer analysis will consider. This parameter of the Box-Meyer algorithm specifies the maximum number of factors that a term in the regression will be composed of. A 1st order (factor interaction) Box-Meyer analysis will only consider main effects (x_i), a 2nd order analysis will take interactions between pairs of factors into account ($x_i x_j$), a 3rd order analysis will look at 3rd order interactions ($x_i x_j x_k$), and so on. Clearly every combination of (the appropriate number of) factor groups must be analyzed if no potential interaction is to be missed. The number of analyses that must be completed can likewise become quite large, but an overall analysis of the model will nonetheless frequently remain feasible.

As a simple example, consider a model with 12 factors X_1, X_2,…X_{12}.[1] We are concerned about 2nd order interactions, and we want to use groups of 3 factors so that the partition size will be 6 factors. The factors are split into four groups of three factors each: $g_1 = \{ X_1, X_2, X_3\}$, …, $g_4 = \{ X_{10}, X_{11}, X_{12}\}$, so the 6 "partitions" needed to cover every pair-wise combination of factors consist of the following:

$$P1 = \{g_1 g_2\} = \{X_1, X_2, X_3, X_4, X_5, X_6\}$$
$$P2 = \{g_1 g_3\} = \{X_1, X_2, X_3, X_7, X_8, X_9\}$$
$$P3 = \{g_1 g_4\} = \{X_1, X_2, X_3, X_{10}, X_{11}, X_{12}\}$$
$$P4 = \{g_2 g_3\} = \{X_4, X_5, X_6, X_7, X_8, X_9\}$$
$$P5 = \{g_2 g_4\} = \{X_4, X_5, X_6, X_{10}, X_{11}, X_{12}\}$$
$$P6 = \{g_3 g_4\} = \{X_7, X_8, X_9, X_{10}, X_{11}, X_{12}\}$$

The number of factors also impacts the feasibility of using the Box-Meyer method in another way: the size of the experimental design matrix. Specifically, the number of rows in the matrix, which determines the number of model executions, is again 2^N for a full (2-level) factorial design. So, even though the number of regressions was addressed by our partitioning method, it may still not be feasible in practice to perform the total number of model runs required without reducing the size of the design matrices (for each of the individual partition analyses).

Though the Box-Meyer method is a general algorithm that can be used with any experimental design, in our application we consider only factorial designs. This is so that we can easily control the level of confounding in the design matrix, which is given by the resolution of the factorial design. A set of confounded terms in a design is associated with a single effect, and it is impossible to determine (from the underlying data) which of the terms in each set are in fact responsible for the effect. To prevent any confounding between terms of interaction order less than or equal to the Box-Meyer analysis order, the design used must have a resolution greater than twice the analysis order.

The number of rows in the design matrix can be reduced by using a fractional factorial designs. As the fractionality of the design increases, the number of rows in the design matrix decreases but so does the resolution of the design. Thus, the maximal reduction in the number of runs required, without introducing confounding, can be achieved by using the most fractional design with the minimum required resolution.

---

[1] Of course, since the total number of factors is small, they could all be analyzed at once. This example is only meant to illustrate the general idea.

The impact of group size and design fractionality on the results of the Box-Meyer analysis using our partitioning approach was investigated. We applied the Box-Meyer algorithm to sets of randomly generated models described by linear equations with a (normal) noise component. The generated models contained 20-50 factors, had maximum factor interaction orders from 1 (x_i) to 3 ($x_i x_j x_k$), and used 5 different noise levels.[2]

With the analysis of these models without partitioning there are three potential sources of error. First, if the maximum order of factor interactions in the model is greater than the order of analysis, important higher order interactions may be missed, which may then cause active factors to be missed (false negatives). Second, if the resolution of the experimental design is insufficient to prevent confounding, the effects of some important terms may be assigned to other terms, causing both false negatives and false positives. Third, the presence of the noise component can produce false negatives, if the noise component masks the effect of a factor, and more rarely false positives.

In our investigation, we only considered analyses in which the order of the analysis was equal to or greater than the maximum order of factor interactions in the model. Clearly, for a simulation model the maximum order of important factor interactions will most likely not be known. As part of our tool development, we developed indicators to help users gauge whether the interaction order used in an analysis was sufficient. Also, we only considered experimental designs of sufficient resolution to prevent confounding. Thus any errors will be due to the interactions between the partitioning parameters (group size and design fractionality) and the noise level.

Additionally, treating all false negatives equally is not ideal. Missing a very active factor is certainly worse than missing an only slightly active factor. In order to measure the level of "activeness" of a factor, we developed the "absolute contribution" metric, which is the average absolute impact that changing a factor's value from minimum to maximum (or vice versa) has on the model output, with the averaging being done over the full factorial design. The absolute contribution of the j-th factor is then given by:

$$ac_j = 1/2^{n-1} \sum_{i=1}^{2^{n-1}} \left| f_j(x_j^{\max}, \langle x \rangle_i^j) - f_j(x_j^{\min}, \langle x \rangle_i^j) \right|$$

$\langle x \rangle^j$ ranges over all 2^{n-1} combinations of min and max values of the other factors

$$f_j(x_j, \langle x_1, \dots x_{j-1}, x_{j+1}, \dots x_n \rangle) = f(x_1, \dots x_{j-1}, x_j, x_{j+1} \dots x_n)$$

To absolute contribution is normalized by diving by the sum of absolute contributions to get the absolute contribution ratio (*acr)*:

$$acr_j = \frac{ac_j}{\sum_k ac_k}$$

[2] The noise levels corresponded to the percent (5%, 10%, 15%, 20%, 25%) of the model range that was used as the standard deviation of the noise component.

On the basis of testing and observation, basic thresholds were selected for strong, medium, and weak active factors. Inactive factors were introduced into the models by specifying that the models contained more factors than actually appeared in the underlying linear equation.

Table 1 below shows the percent of strong and medium active factors missed and the percent of inactive factors misidentified as active. The threshold for weak active factors was set such that they might be better termed "very weak," and whether they are identified or not is not particularly significant. Most are, in fact, already masked by the noise component at Noise Level 1, so (very) weak false negatives are not included in the following tables.

Factors	Group Size	Noise Level: 1			Noise Level: 3		
		Strong FN	Medium FN	FP	Strong FN	Medium FN	FP
20	3	0.0%	16.7%	2.1%	8.1%	61.4%	3.4%
	4	0.0%	0.0%	0.3%	0.0%	30.1%	2.8%
	5	0.0%	0.0%	0.0%	0.0%	3.9%	0.8%
30	3	0.0%	12.9%	3.9%	10.9%	54.9%	6.3%
	4	0.0%	2.7%	2.4%	0.0%	32.0%	4.3%
	5	0.0%	0.0%	1.4%	0.0%	6.1%	1.0%
40	3	0.0%	7.5%	7.4%	11.7%	42.3%	11.8%
	4	0.0%	0.4%	4.4%	0.0%	17.4%	6.3%
	5	0.0%	0.0%	1.8%	0.0%	3.4%	3.3%
50	3	0.0%	6.4%	11.6%	3.6%	45.4%	12.5%
	4	0.0%	0.4%	8.6%	0.0%	16.6%	7.5%
	5	0.0%	0.0%	3.6%	0.0%	1.2%	3.6%

Table 1: Rates of false negatives (FN) and positives (FP) for models with 2^{nd} order interactions [2^{nd} order analysis with full factorial design and partitioning]

The results in Table 1 show the expected trend of increasing noise producing significantly more false negatives and slightly more false positives. In terms of group size, larger groups provide better results, which is likewise to be expected as the effects of noise are mitigated by larger design matrices (those with more rows). The main finding is that the group size should be kept above 3 to prevent strong false negatives and reduce medium false negatives and false positives.

It is also clear that for a fixed group size the number of false positives increases along with the number of factors in the model. This stems from the fact that as the number of factors goes up, there are more groups of a given size and so each factor appears in more partitions. The more partitions an inactive factor appears in, the greater the chance that in some partition the noise will be unevenly enough distributed between the model runs in which the factor is set at its minimum and maximum values to cause it to be falsely identified as active. The associated decrease in medium false negatives with increasing model size is related to the same phenomenon.

If these error rates are compared to the rates of strong false negatives and positives for models of 6, 8, and 10 variables analyzed directly and not as part of a partitioned analysis of larger models, they are noticeably higher. In particular, there are no false positives in the stand-alone analyses. The relevant data are given in Table 2 below.

Factors	Group Size	Noise Level: 1			Noise Level: 3		
		Strong FN	Medium FN	FP	Strong FN	Medium FN	FP
6	**6**	0.0%	8.0%	0.0%	0.0%	72.0%	0.0%
8	**8**	0.0%	0.0%	0.0%	0.0%	11.9%	0.0%
10	**10**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 2: Rates of false negatives (FN) and positives (FP) for models with 2^{nd} order interactions [2^{nd} order analysis with full factorial design without partitioning]

The reason for this difference is that the standard deviation of the noise for a given noise level is calculated from the overall model output range. As the output range within a single partition can be considerably less, the level of noise with respect to the partition range can therefore be much higher than the nominal noise level. A higher noise level, in turn, can mask more strongly active factors, causing strong false negatives, and amplify the effect of any unevenness in the noise distribution, making false positives more likely. If false positives are suspected, it is possible to perform a further analysis on just the active factors initially identified to verify whether they are all in fact active.

Table 3 show the impact of design fractionality on error rates with group size set to 5. The most fractional design considered was the 2^10-3 design, as the resolution of the 2^10-4 design is 4 (IV), which is not sufficient to prevent confounding between 2^{nd} order factor interactions. The trends with respect to noise and number of factors are the same as discussed above. The main result with respect to fractionality is that with higher noise, the most fractional design of sufficient resolution may have too few rows to mitigate the effects of the noise.

Design	Factors	Noise Level: 1			Noise Level: 3		
		Strong FN	Medium FN	FP	Strong FN	Medium FN	FP
2^10 (Full)	**20**	0.0%	0.0%	0.5%	0.0%	4.1%	0.0%
	30	0.0%	0.0%	1.0%	0.0%	6.4%	1.0%
	40	0.0%	0.0%	1.0%	0.0%	3.1%	2.9%
	50	0.0%	0.0%	2.3%	0.0%	3.0%	4.4%
2^10-1	**20**	0.0%	0.5%	0.5%	0.0%	13.1%	0.2%
	30	0.0%	0.0%	1.6%	0.0%	15.0%	3.5%
	40	0.0%	0.0%	1.8%	0.0%	8.7%	7.1%
	50	0.0%	0.0%	3.6%	0.0%	8.6%	7.5%
2^10-2	**20**	0.0%	1.8%	0.5%	0.0%	38.6%	2.1%
	30	0.0%	1.5%	1.6%	0.0%	29.2%	3.3%
	40	0.0%	0.7%	4.2%	0.0%	19.2%	5.9%

	50	0.0%	0.0%	4.9%	0.0%	21.5%	8.8%
2^10-3	20	0.0%	11.8%	0.5%	2.0%	55.5%	0.7%
	30	0.0%	6.4%	1.2%	3.6%	51.5%	3.9%
	40	0.0%	4.2%	2.9%	2.8%	42.3%	5.3%
	50	0.0%	2.2%	4.1%	1.3%	38.3%	5.9%

Table 4 presents the stand-alone results for 10-factor models using fractional designs.

Design	Factors	Noise Level: 1			Noise Level: 3		
		Strong FN	Medium FN	FP	Strong FN	Medium FN	FP
2^10	10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2^10-1	10	0.0%	0.0%	0.0%	0.0%	5.9%	0.0%
2^10-2	10	0.0%	0.0%	0.0%	0.0%	20.9%	0.0%
2^10-3	10	0.0%	2.1%	0.0%	0.0%	43.3%	0.0%

The research indicates that the appropriate use of the Box-Meyer algorithm together with our partitioning approach allows models with 50 variables to be analyzed with no strong false negatives and negligible false positives.

The Tool Development Phase

The algorithms and application methods resulting from this research were instantiated in a software prototype tool, named DIET (the Driver Identification and Exploration Tool), that is intended to interface with the simulation being analyzed and automate the key steps in the end-to-end analytic process. The execution of this process is facilitated by a flexible user-friendly interface for inputs and analysis outputs in a manner that does not require detailed understanding of the underlying algorithms and statistical framework embedded in the tool. The top level architecture is described in Figure 1.

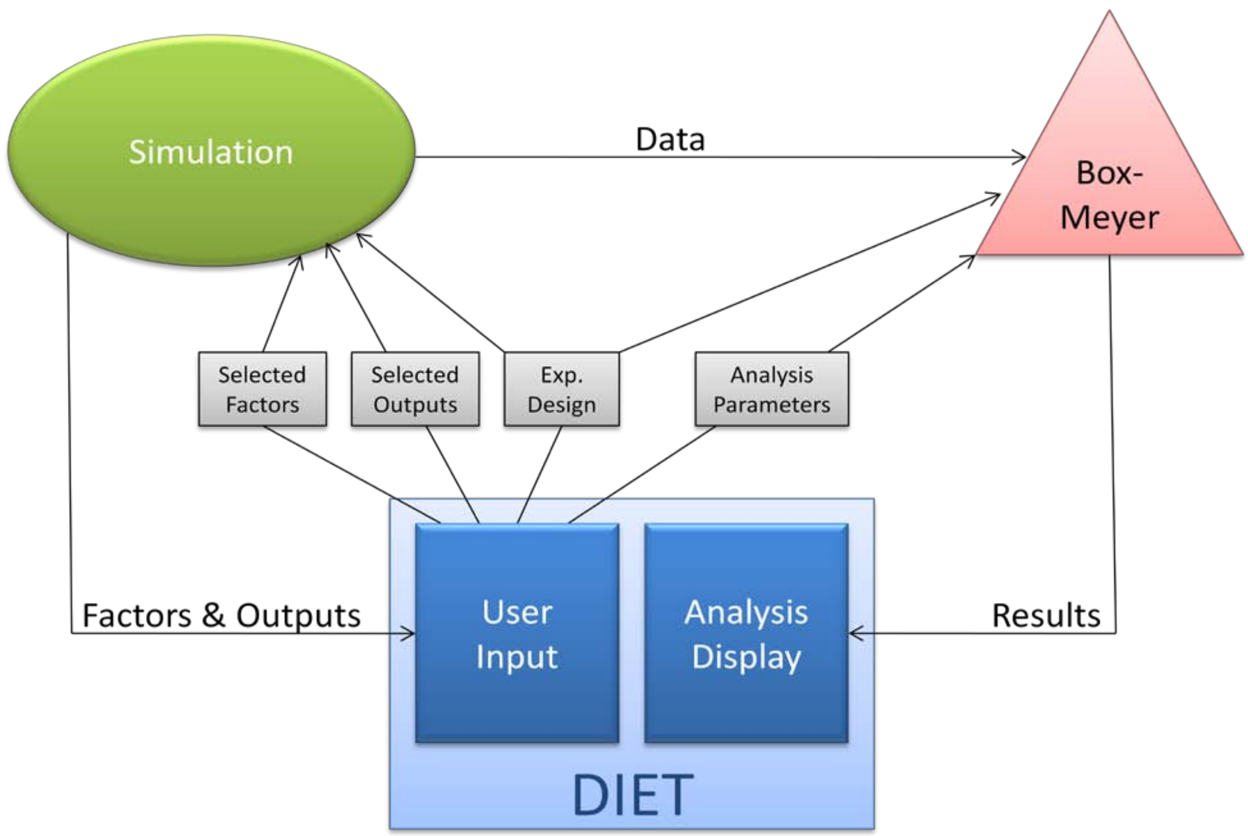

Figure 1: Top Level Architecture for DIET

Users select the simulation factors and outputs to analyze and choose the experimental design based on the level of factor interaction desired and the amount of time available for doing the simulation runs. The basic result of the analysis is a ranking of the active factors in accordance with their level of activeness. This is calculated by applying the absolute contribution metric described above to the regression model generated by the set of active factors. In cases where many active factors are identified, the tool does not automatically set an "activeness threshold" to limit the set of active factors to a most active subset, but rather provides a graphical view of the distribution of the factor activeness levels so that the users themselves can visually determine the optimal cut-off point for their purposes.

Two indicators of potential error were also developed to help users determine the adequacy of the performed analysis. Both indicators are based on the fit of the individual partition regression models to the associated simulation output. The first considers the level of factor interaction and the second looks at the size of the factor domain (the min-max value ranges of the factors). The factor interaction level indicator is intended to help the user decide whether another analysis with a higher level of factor interaction is warranted. A low value means there is a lack of fit for partitions with high variation in output and indicates that important higher order factor interactions are being missed. The factor domain indicator is intended to help the user decide whether the min-max ranges of the factors should be further constrained for another analysis. A low value means that the simulation is behaving differently over the individual partitions and indicates that the assumption of linear-like behavior for the model over the whole factor domain is not correct. As it is based on the use of a 2-level (min/max) experimental design, the Box-

Meyer method can only identify "linearly" active factors and so is intended to be used with factor domains over which the simulation is approximately linear.

The tool also provides detailed information for the individual Box-Meyer analyses performed for each of the partitions. This can help not only with the identification of problem partitions or partitions for further analysis but also with gaining a better understanding the behavior of the simulation more generally. It likewise includes the capability for comparing results over multiple simulation outputs as well as comparing results from different analyses in which the factor domains and analysis parameters (such as level of factor interaction) have been varied.

Use Case (SENTURION)

The tool and methodology were tested on the Senturion™ simulation for forecasting how the policy positions of key individuals and groups evolve over the course of a negotiation. Senturion combines recent theoretical advances in game theory with data drawn from Subject Matter Experts and advanced computational methods to generate a round-by-round simulation of stakeholder position changes as the stakeholders interact over time. The negotiation used for the test was a US-PRC military diplomacy issue that had been previously modeled for use with Senturion for PACOM. It involved 43 actors (24 Chinese and 19 US), giving a total of 129 factors as each actor is described by 3 parameters (position, influence, and salience). Position is the initial position of an actor on the issue under consideration, influence quantifies how influential the actor is with respect to the issue, and salience represents the importance of the issue to the actor. (All three parameters are specified on a scale from 0-100.) Our intention was to use all 43 starting positions as factors in the analysis, but due to limited access to the simulation, the analysis was restricted to just the 24 Chinese positions.

There are also 43 outputs that could potentially be analyzed: each of the actors' final positions. Since in this case, the final position of the Chinese president Hu Jintao was taken to determine the outcome of the negotiations, his final position was selected as one of the outputs to be analyzed. The other output was the final position of the actor whose position (on average) changed the most over all of the simulation runs. This turned out to be Henry Kissinger.

The results for both outputs are shown below in Figures 2 and 3. The fundamental difference between the analysis as done for Jintao's and Kissinger's final positions is that in the case of the Chinese president, his starting position is one of the factors included in the analysis. Not surprisingly then, Jintao's starting position is the most important factor in determining his final position. This is particularly true in our analysis because the difference between his minimum and maximum starting positions is quite large (min = 0, max = 50)[3]. The presence of his starting position in the set of analysis factors is also the reason that the results for Jintao's final position are better in terms of both the error indicators (level of factor interaction and size of factor domain).

---

[3] The minimum and maximum values for each actor's starting position were established by a subject matter expert and provided to us.

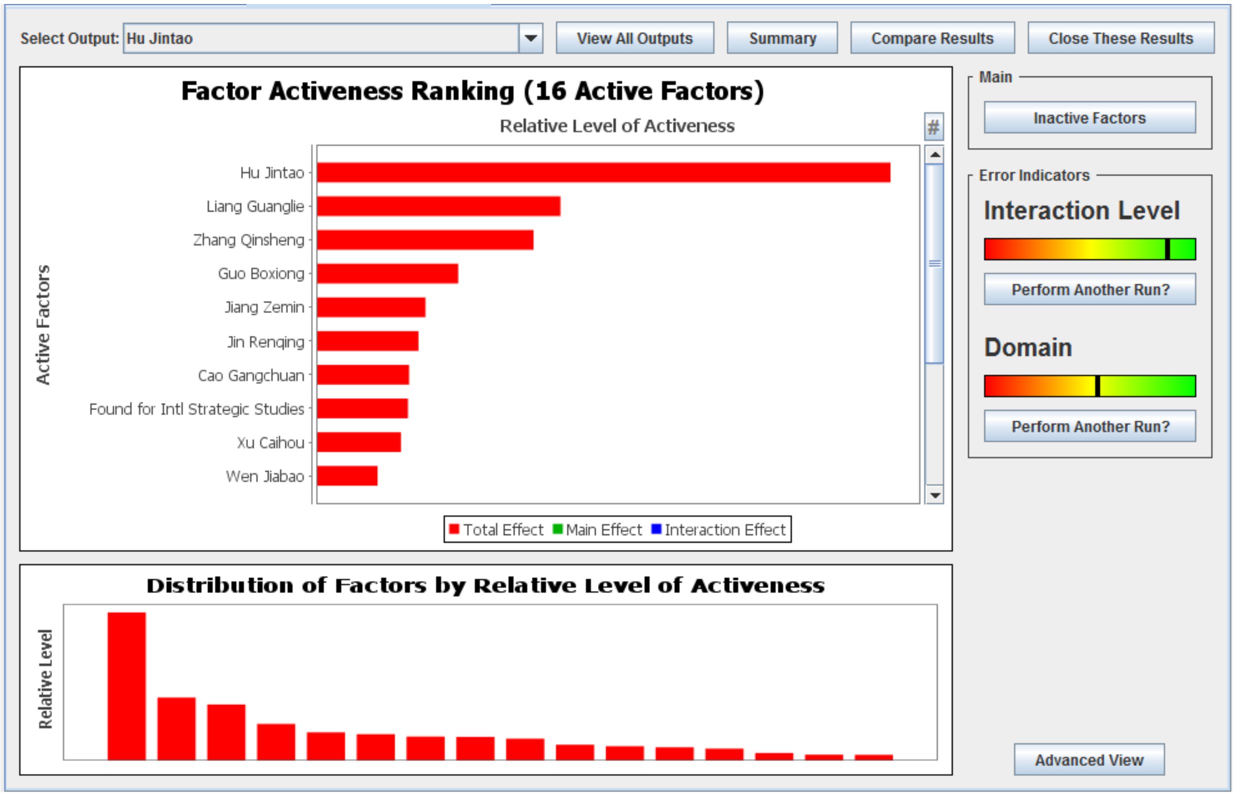

Figure 2: Results for Hu Jintao's Final Position

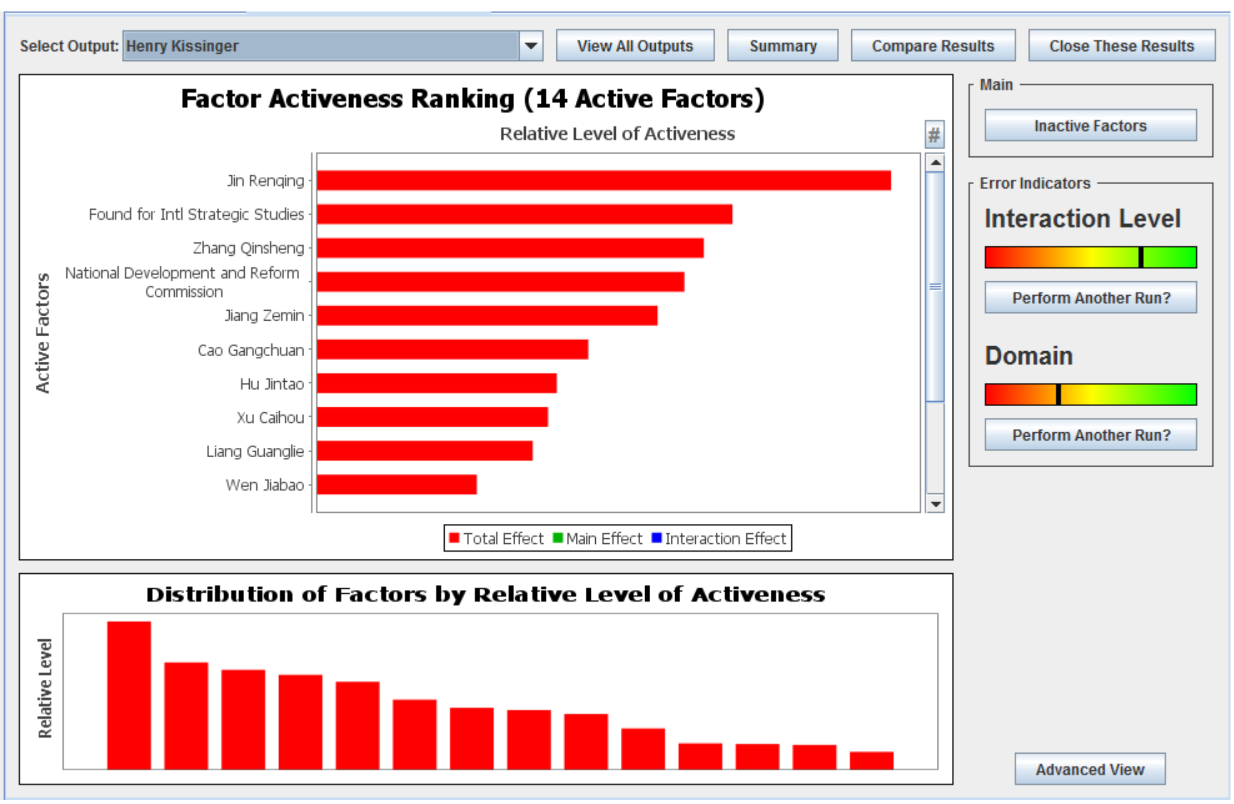

Figure 3: Results for Henry Kissinger's Final Position

When we look at the values of the other two actor parameters (influence and salience) for the actors whose starting positions are important in the analysis of Jintao's final position, the results seem quite reasonable. The "active" actors all have relatively higher influence and salience values. In the analysis of Kissinger's final position, however, the 1^{st} and 4^{th} most active factors are the two factors with the lowest influence (1.5) and salience (5) values. Though the analysis of Kissinger's final position in particular may not be significant in the scope of the specific issue being considered, the result demonstrates the potential our approach has for identifying "unexpected" or "surprising" active factors.

Summary

The Box-Meyer methodology can be successfully augmented by our partitioning approach to identify active factors for defined applications of large-scale simulation models with numerous factors. The research has produced a prototype of a tool that promises to be useful in assessing the credibility of simulations or in characterizing their behavior when applied to specific situations. A limited application of the tool to analyze the Senturion simulation using a specific US-China negotiation problem demonstrated that the tool can produce surprising insights into the behavior of the simulation being analyzed. These results highlight the potential utility of the tool in identifying the most important factors for user-defined scenarios. Our approach will be most useful for users who need to analyze a well-defined set of problems that can be understood by constraining the simulation's behavior to some finite local domain. Fortunately, many problems can be solved by bounding them in this way, such that a local solution based on a small piece of the response surface can answer the analytical questions at hand. The tool can also be used to explore the limits of the applicability of such solutions.

References

Box, G.E.P. and R.D. Meyer, "Finding the Active Factors in Fractionated Screening Experiments," Journal of Quality Technology, 25(2), April 1993.

Samset, O. and J. Tyssedal, "Study of the Box-Meyer Method for Finding Active Factors in Screening Experiments," NTNU Preprint Series: Statistics, 4, 1998.

Webb, T.S. and K.W. Bauer, Jr., "Comparison of Analysis Strategies for Screening Designs in Large-Scale Computer Simulation Models," Proceedings of the 1994 Winter Simulation Conference, 1994.