Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

USING RESPONSE SURFACE METHODOLOGY TO LINK FORCE STRUCTURE BUDGETS TO CAMPAIGN OBJECTIVES

James B. Grier
T. Glenn Bailey
Jack A. Jackson

Department of Operational Sciences
Air Force Institute of Technology
2950 P Street, Bldg. 640 (AFIT/ENS)
Wright-Patterson AFB, OH 45433-7765, U.S.A.

ABSTRACT

Effective and timely execution of the Department of Defense's (DoD) Program Objective Memorandum (POM) and Joint Warfare Capability Assessment (JWCA) process requires objective measures of how Air Force programs support the Defense Planning Guidance (DPG) and the Chairman's Program Assessment (CPA). Using Factor Analysis (FA) and Response Surface Methodology (RSM), this paper presents a modeling approach that provides metrics which link expenditures to campaign level measures of outcome. Specifically, various alternative force structures are evaluated with regard to their combat capability as measured in terms of theater level campaign objectives (CO).

1 INTRODUCTION

Effective and timely execution of the Department of Defense's (DoD) Program Objective Memorandum (POM) and Joint Warfare Capability Assessment (JWCA) process requires objective measures of how Air Force programs support the Defense Planning Guidance (DPG) and the Chairman's Program Assessment (CPA). Since the Air Staff makes many force structure budgeting decisions in relative isolation from each major functional area, it is difficult to develop a comprehensive assessment of the total effect on the Air Force's ability to meet theater level campaign objectives (CO). Compounding the difficulty of this problem is the budgetary process itself. The biennial cycle seeks to reconcile the number of competing agendas within the Air Force, DoD, and Congress, and deal with the simultaneous processing of different fiscal year budgets at various stages of their respective planning and programming process. All of these items contribute to the need for a better way to link dollars spent to campaign level measures of outcome.

Our approach to capturing this linkage is a unique combination of Factor Analysis (FA) and Response Surface Methodology (RSM) as applied to a theater-level combat model. We present our research in the following manner. Section 2 describes our overall methodology, Section 3 presents the results for an unclassified notional Southwest Asia (SWA) scenario, and Section 4 concludes with suggestions for further research.

2 METHODOLOGY

Figure 1 presents our overall methodology. The first row of boxes essentially describe the standard steps for constructing an experimental design. Once the experimental design is completed, FA is employed to derive combat-related indices for use as the dependent responses in stepwise linear regression. The resulting response surfaces can then be used as the objective functions within appropriate linear programs, where their respective constraints link budget expenditures to combat capability as represented by the combat index of interest. The following subsections present this approach in further detail.

2.1 Experimental Design

THUNDER, a theater-level warfare simulation, is the basis for the experiment. A two-sided, stochastic computer simulation of conventional air, land, and naval warfare used for force structure evaluation, wargaming, and senior staff training, THUNDER is a SIMSCRIPT II.5 model used by a large number of U.S. and allied defense organizations and contractors.

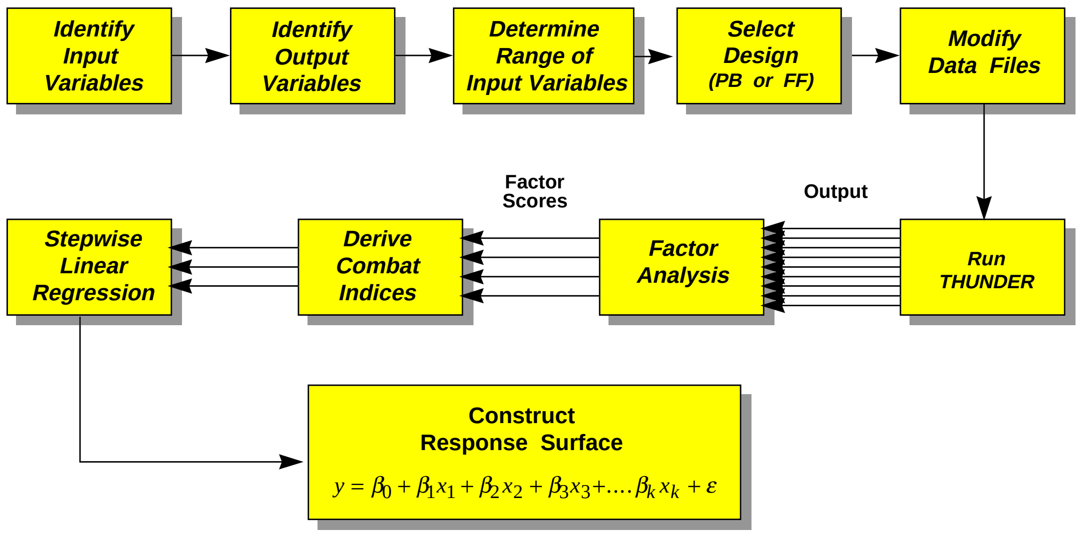

Figure 1: Methodology

It is capable of simulating 22 different air missions and generating Air Tasking Orders (ATOs) based on theater-level apportionment and targeting priorities (Air Force Studies and Analysis Agency 1995).

THUNDER's long running time for a single replication presents a textbook example of the need to minimize the number of runs. Therefore, our approach to developing the experimental design follows the standard practice of first running preliminary screening experiments with follow-up fractional-factorial central composite (CCD) designs (see Naylor 1969, Law and Kelton 1991, Box and Draper 1987, or Myers and Montgomery 1995). Our choice for a preliminary screening design is Plackett-Burman (PB), and in this regard we follow the earlier efforts of Webb and Bauer (1994) in applying PB designs using THUNDER.

Table 1 lists the variables used in the SWA scenario. The lower bound numbers are (i) aircraft inventories at THUNDER's default levels, and (ii) munitions at an estimated 80% expenditure rate. (Prepositioned munitions and intra-theater re-supply features are disabled, and all munitions are in place at the commencement of hostilities.) The number of squadrons is representative of those that deployed and fought in the Persian Gulf War, where with few exceptions a typical squadron is assigned 24 primary aircraft. Regarding the upper bound numbers, we assume the greatest single increase of authorized aircraft approved by Congress in a given year would be 50%; therefore, we base our upper range on the number of additional weapons and aircraft that such an increase would allow. Center point values are available for each variable as well. THUNDER produces 34 output metrics for this SWA scenario, such as estimates on the number of friendly and enemy aircraft, tanks, personnel carriers, infantrymen, and defense infrastructure sites destroyed.

2.2 Factor Analysis

The underlying idea of Factor Analysis (FA) is to simplify the relationship among a set of observed variables by explaining it in terms of fewer, conceptually meaningful, independent factors (see Kleinbaum and Kupper 1978, or Dillion and Goldstein 1984). One method of accomplishing this is the varimax rotation, where the variation of squared factor loadings within

Table 1: Input Variable Levels

Aircraft	Lower	Center	Upper	Munitions	Lower	Center	Upper
F-15C	120	150	180	AIM-120	4,250	5,312	6,375
F-15E	48	60	72	AIM-9	6,600	7,500	8,400
F-16	228	282	336	20 MM	3,250	4,062	4,875
A-10	144	180	216	MK-82	87,275	109,094	130,912
F-111	96	120	144	AGM-65	8,707	10,883	13,060
EF-111	18	21	24	ARM-88	567	708	850
F-4G	72	90	108	B-Delay	375	468	562
F-117	12	15	18	B-Lethal	1,644	2,060	2,476
Tomahawk	120	150	180	CBU-87	1,300	1,631	1,962
JSTAR	6	9	12	CBU-97	23,895	28,373	32,852
AWACS	12	15	18	LGB	2,930	3,632	4,335
				GPS LGB	60	75	90

a factor is maximized. By comparing these reduced factors with other known features of the system being modeled, one will hopefully develop a better understanding of its performance.

Within the present context, the 34 separate responses from THUNDER's output can be converted to matrix form. Using FA, several varimax rotations can then be accomplished on one set of observations to determine which number of factors best represent the Campaign Objectives (CO). The Air and Space Power Validation Group (1995), HQ USAF, has identified the Operational Objectives (OO), and Operational Tasks (OT) that THUNDER is capable of measuring with regards to the COs. These include:

- CO-1: Halt Invading Armies.
- CO-2: Marshall and Sustain In-Theater Assets.
- CO-3: Evict halted Armies from Friendly Territory.
- CO-4: Gain and Maintain Air Superiority.
- CO-5: Gain and Maintain Sea Control.
- CO-6: Gain and Maintain Space Control.
- CO-7: Gain and Maintain Information Dominance.
- CO-8: Deny Possession and Use of Weapons of Mass Destruction (WMD).
- CO-9: Suppress National Capacity to Wage War.

Historically, the specific metrics from THUNDER have not been translated from simulation model output to the COs used in the budget process. Therefore, we introduce FA as a method for mathematically mapping the simulation output to a set of combat indices that capture these objectives. (As illustrated in Figure 2, multiplying the second set of averaged observations by the factor matrix produces a set of indices for each CO.) Once this mapping is established, a second set of observations can then utilize this rotation to construct a response surface of the reduced set of factors - i.e., the COs.

2.3 Response Surfaces

We employ RSM to map a response surface – in this case, an aggregate combat capability index called Total Combat Index – over a particular region of interest; e.g., alternative force structures. RSM methods are widely examined within the experimental design literature (see Box and Draper 1987, Myers and Montgomery 1995). Furthermore, our approach conforms to the use of metamodels to study the behavior of computer simulation, particularly parametric polynomial response surface approximations (Kleijnen 1987, Barton 1994). Once the combat indices are derived as the dependent response in separate stepwise linear regressions, each of the resulting response surfaces is a candidate objective function in a linear program, where the constraints link budget dollars to the combat capability of alternative force structures in terms of campaign objectives.

3 SWA SCENARIO RESULTS

The scenario used to develop this metamodel was based on a Major Regional Conflict (MRC) in SWA. The unclassified database represents "real world" information with respect to the number and type of aircraft and munitions modeled; however, the target array is very limited in scope. Three major experi-

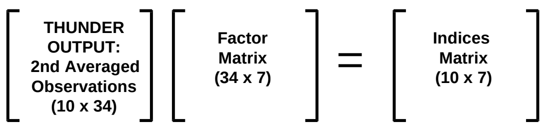

Figure 2: Application of Factor Analysis

mental designs were run. First, our initial run was a Resolution III PB, where 30 replications were accomplished at each design point (with two additional center point runs providing measurement of pure error). After running the PB design, we found the input variables A-10, F-15E, MK-82, and AGM-65 were the only statistically significant input factors.

This led to the second design based on the PB results, where thirty replications were performed at each design point of a 2_{IV}^{4-1} design. As in the initial PB design, for each design point the average number of each output variable was calculated as the response, and two center point runs were added to more accurately measure pure error (thus a total of 10 design points). The third design simply repeated the second design using a different random number seed value, thus providing an independent estimate of the responses at each design point. In both designs, the same range of values were used for the four variables.

FA on the second design found that five of seven factors could be clearly defined in terms of the COs:

- CO-1: Halt Invading Armies (Halt).
- CO-3: Evict Invading Armies (Evict).
- CO-4: Gain and Maintain Air Superiority (Air Superiority).
- CO-7: Gain and Maintain Information Dominance (C^3).
- CO-9: Suppress National Capacity to Wage War (Interdiction).

For example, Table 2 shows how the five output metrics (out of THUNDER's total of 34) associated with stopping the enemy's ground advance load significantly on one of the factors produced by the varimax rotation. Since these set of loadings clearly represent CO-1 (Halt Invading Forces), this factor is identified with that campaign objective. Similarly strong correlations occurred for the remaining factors and their respective COs.

Table 2: Significant Factor Loadings for CO-1 (Enemy Assets Destroyed)

THUNDER Output Metric	Loading
Tanks	.96
Armored Personnel Carriers	.96
Infantry Killed	.95
Artillery	.42
Air Defense TELs	.64

Once the results of FA are obtained from the second experimental design, a reduced set of combat indices are derived (Figure 2). We further define the Total Combat Index (TCI) as an equally weighted linear combination of the five individual indices. Stepwise linear regression is then conducted for each of the individual indices as the dependent variable using data from the third experimental design, thus providing independence between the factor analysis and response surface estimation. A summary of the results and associated response surface is listed in Table 3.

While the adjusted R^2 value for Evict is disappointing, the remaining values explain anywhere from 50% (Interdiction) to 94% (C^3) of the variation within THUNDER. Examining the magnitudes of the parameter estimates, as well as the sums-of-squares, clearly shows the A-10 as the dominant input variable of the 23 considered. The fractional factorial design identifies some synergistic effects between the A-10 and AGM-65 maverick missile, but no other second-order effects are statistically significant.

Dominance of the A-10 can be explained in part by the notional scenario. Greater than 75% of the targets in the data base are located in the first 40 miles of enemy territory, and consisted mostly of tanks, armored personnel carriers, mobile SAM, and bridges. This, along with the nature of the terrain in SWA, enhances the dominance of an A-10 aircraft well suited for the desert environment. Similarly, the number of strategic targets loaded in the data base is limited, down playing the importance for precision guided weapons delivery, and thus helping the AGM-65 maverick missile. Finally, the absence of significant air-to-air input variables in the indices is consistent with the Persian Gulf War, where a hostile air-to-air threat did not exist after the first 24 hours of hostilities.

Table 3: Polynomial Approximations of Indices (Coded Variables)

Index	Adj R^2	Polynomial Response Surface
Halt	.8655	26,422 + 2,248·X_{A-10} + 367·X_{AGM-65} + 676·X_{A-10}·X_{AGM-65}
Evict	.1697	13,226 + 604·X_{A-10}
Air Superiority	.7020	9,545 + 498·X_{A-10} + 356·X_{AGM-65}
C^3	.9488	-807 - 65·X_{A-10} + 25·X_{F-15E} - 13·X_{AGM-65} + 13·X_{MK-82} -16·X_{A-10}·X_{AGM-65}
Interdiction	.5012	6,340 + 367·X_{A-10}
TCI	.6674	54,725 + 3,652·X_{A-10}

4 CONCLUSIONS

In summary, this paper presents a three-prong modeling approach to provide metrics which link expenditures to campaign level measures of outcome. First, we introduce the multivariate method of Factor Analysis to bridge the gap between THUNDER output and Air Force campaign objectives through a simplified, balanced combination of indices to compute the overall combat index for each alternative force structure. Second, when integrated within a response surface framework, this approach provides a budgeting tool for linking combat indices to force structure changes. Third, we suggest incorporating the resultant first-order polynomial as the objective function in a linear programming problem. At a minimum, the feasible region would be constrained by the total available procurement dollars that fiscal year (FY); i.e.,

$$c_1x_1 + ... + c_{23}x_{23} \leq FYBudget$$

where c_i is the unit fly-away cost in current year dollars, and x_i is the number of units purchased that FY. (The coefficient of cost for each type aircraft and weapons system in the USAF inventory is published annually in AFI 65-503, Table 10-1, and 11-1, 1995.) Additional constraints, such as operational and maintenance costs for each weapons system, and the maximum number of aircraft that can be produced in a one year period of time, could also be included. Finally, future research should consider other methods of weighting the indices, and employing variance reduction techniques on THUNDER to refine the simulation output.

REFERENCES

Air Force Instruction (AFI) 65-503. HQ USAF, Pentagon ADM VA, 20330. 1995.

Air Force Studies and Analysis Agency. *THUNDER Analyst's Manual.* Version 6.3. Arlington VA: CACI Products Company. 1995.

Air and Space Power Validation Group (ASPVG). *ASPVG Checklist of Objectives for Model Evaluation: Major Regional Conflicts.* HQ USAF/XOM, Pentagon ADM VA, 20330. March 22, 1995.

Barton, Russell R. "Metamodeling: A State of the Art Review", in *Proceedings of the 1994 Winter Simulation Conference.* Ed. J.D. Tews, S. Manivannan, D.A. Sadowski, and A.F. Seila. Institute of Electronics Engineers, Washington D.C. 1994.

Box, George E.P. and Norman R. Draper. *Empirical Model-Building and Response Surfaces.* New York: John Wiley and Sons, 1987.

Dillion, William R. and Matthew Goldstein. *Multivariate Analysis: Methods and Applications.* New York: John Wiley and Sons, 1984

Kleijnen, Jack P.C. *Statistical Tools of Simulation Practitioners.* New York: Marcel Dekker, 1987.

Kleinbaum, David G. and Lawrence C. Kupper. *Applied Regression Analysis and other multivariate methods.* North Scituale: Duxbury Press, 1978.

Law, Averill M. and W. David Kelton. *Simulation Modeling and Analysis (Second Edition).* New York: McGraw-Hill, 1991.

Myers, Raymond H. and Douglas C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* New York: John Wiley and Sons, 1995.

Naylor, T.H. *The Design of Computer Simulation Experiments.* Durham, NC: Duke University Press, 1969.

Webb, Timothy S. and Kenneth W. Bauer, Jr. "Comparison of Analysis Strategies for Screening Designs in Large Scale Computer Simulations", in *Proceedings of the 1994 Winter Simulation Conference.* Ed. J.D. Tews, S. Manivannan, D.A. Sadowski, and A.F. Seila. Institute of Electronics Engineers, Washington D.C. 1994.

AUTHOR BIOGRAPHIES

JAMES B. GRIER, Major, USAF, received his Bachelor's in General Studies from the University of Maryland in 1983 and his M.S. in Operations Research from the Air Force Institute of Technology in 1997. He is a senior pilot currently assigned as Chief, F-16 Programming, Air Staff, the Pentagon, Washington D.C.

T. GLENN BAILEY, Lieutenant Colonel, USAF, is an Assistant Professor of Operations Research in the Department of Operational Sciences at the Air Force Institute of Technology, WPAFB, OH. A senior pilot and 1978 graduate of the U.S. Air Force Academy, Lt Col Bailey received his M.S. in Operations Research from AFIT in 1988, and his Ph.D. from the University of Texas at Austin in 1995. His research interests include simulation, response surface methods, and metaheuristics.

JACK A. JACKSON, Lieutenant Colonel, USAF, is the Director of the Center for Modeling, Simulation and Analysis at the Air Force Institute of Technology, Wright-Patterson AFB, Ohio. A 1976 graduate of the USAF Academy, he has flown fighter aircraft and served as an Air Staff tactical fighter analyst at the Pentagon. He received his M.S. from AFIT in 1989 and his doctorate from the Colorado School of Mines in 1994. A member of INFORMS, his current research interests include decision analysis, combat simulation and modeling, and applied statistics.