Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

SENSITIVITY ANALYSIS OF A LARGE SCALE TRANSPORTATION SIMULATION USING DESIGN OF EXPERIMENTS AND FACTOR ANALYSIS

Glenn G. Rousseau

Studies and Analysis Flight
Headquarters Air Mobility Command
402 Scott Drive, Unit 3L3
Scott Air Force Base, Illinois 62225, USA

Kenneth W. Bauer, Jr.

Department of Operational Sciences
Graduate School of Engineering
Air Force Institute of Technology, 2950 P Street
Wright-Patterson Air Force Base, Ohio 45433-7765, USA

ABSTRACT

This paper describes how design of experiments and factor analysis were used to conduct sensitivity analysis on multivariate output from a large scale transportation simulation model. Specifically, this research focused on the sensitivity of airlift system performance to changes or errors in a list of transportation requirements. The general approach included perturbing a time-phased list of transportation requirements according to an experimental design and using a simulation model to estimate the airlift system performance response. We used factor analysis to reduce the dimensionality of the multivariate output data and to generate sensitivity plots, which proved to be valuable graphical tools for sensitivity analysis. Additionally, we identified how factor analysis can be used as a verification and validation tool for large stochastic simulation models.

1 INTRODUCTION

The Airlift Flow Module (AFM) is a large-scale transportation simulation model used primarily for estimating strategic military airlift system performance. Some of the basic airlift system features modeled in AFM include: aircraft, aircraft crews, air bases, routes, aerial refueling, ground refueling, materiel handling equipment, cargo, and passengers. The performance of an airlift system, as a whole, is the result of a complex interaction of all these features. This paper describes how AFM performance sensitivities provide insight for improving airlift system performance.

This paper is organized as follows. In Section 2, we provide background information on the AFM simulation, the modeled scenarios, and airlift system performance measures. In Section 3, we discuss how we perturbed four characteristics of time-phased force deployment data (TPFDD) according to a 2^{4-1} fractional factorial experimental design. In Section 4, we present a quick review of factor analysis and then discuss specific results from an airlift scenario example. In Section 5, we highlight implications pertinent to verification and validation. Finally, in Section 6, we recapitulate our main findings. This paper is based on Rousseau (1996).

2 BACKGROUND

2.1 The AFM Simulation

The AFM simulation contains over 50,000 lines of FORTRAN code and models to the individual aircraft and pallet level of detail. AFM was selected for this research because of its flexibility in representing airlift system activities and its ability to estimate many detailed aspects of airlift system performance. Currently, this flexibility and detail come at the costs of long setup and run times.

Being demand driven, AFM needs an airlift requirement for cargo or personnel in order to plan a mission. The AFM mission planning algorithm is deterministic, starting at the top of the TPFDD and reading sequentially to plan a mission for the first available and compatible load with the next available aircraft. During AFM mission execution, the aircraft's simulated flight times and ground times are drawn from random distributions defined by the user.

2.2 The Modeled Scenarios

Two different airlift scenarios were modeled and they will be generically referred to as the "small" and the "large" scenarios, respectively. Only pertinent results from the small scenario are presented in this paper. The small scenario models a minor conflict in the Caribbean. It requires 75,854 tons of cargo and 139,480 passengers to be delivered over a fairly simple air base network using 138 total aircraft. The large scenario models a major conflict in Southwest Asia. It requires 402,796 tons of cargo and 498,369 passengers to be delivered over a complex air base network using 448 aircraft.

2.3 Multivariate Output Data

Airlift system performance was broken down into three major categories: aircraft performance, throughput, and timeliness. For each run, the AFM simulation was stopped at day 30 to record the following performance measures for later analysis:

Aircraft Performance Measures (by aircraft type)
- utilization rate
- ground time per cycle
- average payload
- average million-ton-miles per day

Throughput Measures
- outsize tons delivered
- total cargo tons delivered
- total passengers delivered

Timeliness Measures
- percent of shipments delivered ontime
- average number of days late

Some of these variables are further explained for those not familiar with airlift. The aircraft performance measures represent per aircraft averages across all aircraft of a homogeneous type (e.g., C-141, C-5, B-747, etc.) modeled in the simulation. Utilization rate is a gross estimate of aircraft productivity and is a number between 0.0 and 24.0, representing the average number of hours per day an aircraft of that type flew (Kowalsky 1977). Conversely, ground time per cycle is a gross estimate of non-productivity, representing the average number of hours each aircraft of that type was on the ground during an average mission cycle. A mission cycle starts when an aircraft takes off to start a mission and ends when the aircraft lands at the recovery base after mission completion (Kowalsky 1977). The average million-ton-miles per day measure is used as gross estimator of the productive workload accomplished by each aircraft. Most of the large tanks, missile batteries, and other firepower that need to be airlifted are classified as outsize cargo. Therefore, total outsize tons delivered is an indirect measure of airlift effectiveness with respect to the capability to conduct combat operations.

3 EXPERIMENTAL DESIGN

Four general TPFDD characteristics were selected and perturbed according to a 2^{4-1} fractional factorial design to identify their effects on airlift system performance. The four characteristics were: location, timeline, total demand, and cargo flavor. A short definition of these characteristics and brief descriptions of the perturbation strategies are presented in the following paragraphs.

The location characteristic refers to the onload and offload locations, or in airlift terms, the aerial ports of embarkation (APOE) and debarkation (APOD). To perturb this location characteristic, new TPFDD files were generated that were identical to the original TPFDD except for the specified APOEs and APODs. For each line of the original TPFDD, the specified onload and offload locations were replaced by randomly selected locations, then output to the new TPFDD file. The random selection process was constrained to keep the new APOE-APOD distance within 10 percent of the old distance. The intent of this constraint was to keep the airlift workload, estimated in ton-miles, fairly constant across all generated TPFDD files.

The timeline characteristic refers to available-to-load (ALD) and required delivery dates (RDD) for each requirement line in the TPFDD. To perturb this timeline characteristic, new TPFDD files were generated that were identical to the original TPFDD except for the specified ALDs and RDDs. For each line of the original TPFDD, the specified ALD and RDD were replaced by randomly selected dates, then output to the new TPFDD file. The random selection was accomplished by choosing -1, 0, or 1, with equal probabilities, and adding it to both the ALD and RDD. Since the AFM mission planning algorithm reads a TPFDD sequentially, from the beginning, each time it plans a mission, there is an implicit assumption of priority for each line of the TPFDD based on its order in the file during execution. In reality, changing the RDD usually implies a change in airlift priority. To effect the same conceptual change in priority during AFM execution, the lines in the newly generated TPFDD were sorted appropriately.

The total demand characteristic refers to the total amount of cargo and passengers needing airlift. The cargo portion is broken down further into outsize, oversize, and bulk in AFM to determine compatibility with the different aircraft types. To perturb this total demand characteristic, new TPFDD files were generated that were identical to the original TPFDD except for the specified cargo and passenger requirement amounts. For each line of the original TPFDD, the specified demand requirement values were increased by 10 percent, then output to the new TPFDD file.

Cargo flavor is not really an airlift term, but seemed to be an appropriate simple label for this characteristic. For this paper, the cargo flavor characteristic refers to the relative amounts of cargo within each requirement. To perturb this cargo flavor characteristic, new TPFDD files were generated that were identical to the original TPFDD except for the outsize and oversize cargo requirement amounts. For each line of the original TPFDD, the outsize cargo requirement was doubled and an equal tonnage was subtracted from the oversize cargo requirement, then output to the new TPFDD file.

An orthogonal 2^{4-1} fractional factorial design formed the basis for the experimental design shown in Table 1, allowing for unbiased estimation of the effects of the four perturbation schemes and all possible 2-way interactions (Box and Draper 1987). Two more design points were added to directly observe the effects of the two random perturbation schemes. Ten replications of each design point were then fed into the AFM simulation and the desired output variables from all 100 runs were recorded in a flat file for later analysis.

Table 1: Experimental Design for Evaluating Four TPFDD Perturbation Effects

Design Point	*Location*	*Timeline*	*Total Demand*	*Cargo Flavor*
0	0	0	0	0
1	0	0	1	1
2	0	1	0	1
3	0	1	1	0
4	1	0	0	1
5	1	0	1	0
6	1	1	0	0
7	1	1	1	1
8	0	1	0	0
9	1	0	0	0

4 FACTOR ANALYSIS

Factor analysis was used to interpret the effects of the four TPFDD perturbations on the selected AFM output. Factor loadings matrices provided pertinent information on the relationships between the variables as they responded to the TPFDD perturbations. In addition, visual pictures of output sensitivity to the perturbations were found by plotting two factors against each other. These "sensitivity plots" proved to be useful tools for sensitivity analysis.

Factor analysis is a multivariate data reduction technique that identifies common factors underlying a set of observed variables. The formal factor analysis model describes each original observable variable in terms of a linear sum of unobservable common factors and a single latent unique factor, as shown by Equation (1) (Dillon and Goldstein 1984).

$$\mathbf{X} = \Lambda\mathbf{f} + \mathbf{e} \tag{1}$$

where

$\mathbf{X} \equiv$ p-dimensional vector of observed responses
$\mathbf{f} \equiv$ q-dimensional vector of unobservable common factors ($q << p$)
$\Lambda \equiv p \times q$ matrix of weights or factor loadings
$\mathbf{e} \equiv$ p-dimensional vector of unobservable unique factors

Assuming the common and unique factors are uncorrelated with themselves and each other, the covariance matrix of the response vector **X**, denoted by Σ_{xx}, can be expressed as shown in Equation (2) (Dillon and Goldstein 1984). In simplified terms, the object of factor analysis is to replace many variables, **X**, with much fewer variables, or factors, **f**, preserving as much information as possible during the transformation to the new variables (factors). The preserved information is represented in the common variance, $\Lambda\Lambda'$.

$$\Sigma_{xx} = \Lambda\Lambda' + \Psi \tag{2}$$

where

$\Sigma_{xx} \equiv p \times p$ variance-covariance matrix
$\Lambda \equiv p \times q$ matrix of weights or factor loadings
$\Lambda\Lambda' \equiv$ common part of variance-covariance matrix
$\Psi \equiv p \times p$ diagonal matrix containing p unique variances

Once a solution to Equation (2) is found, Λ represents a matrix of possible correlation coefficients between the new factors and the original variables. The patterns of loadings contained in Λ identify which variables each factor represents and are subject to interpretation by the analyst. This interpretation can usually be enhanced by a rotational transformation of the loadings matrix using matrix multiplication. The rotated matrix of factor loadings represents an alternative interpretation of the data and allows the analyst a degree of flexibility (Dillon and Goldstein 1984). The number of such rotational transformations, orthogonal or oblique, is infinite, leading to criticism that any solution chosen is, mathematically speaking, arbitrary (Basilevsky 1994). However, selection of certain criterion functions can define unique rotations and confront this criticism with more generally accepted statistical practice (Basilevsky 1994).

A commonly used rotation criterion results in a unique rotation called the varimax rotation. The varimax rotation seeks to maximize the variation of the squared factor loadings within each factor, thereby forcing the loading coefficients to either really high or low values. With not many of the loadings falling into the middle "gray area" it is usually easier to find an interpretation from the loadings matrix structure.

Table 2 shows the results of a varimax rotation and highlights the significant factor loadings from the factor analysis of the small scenario's data. (All loadings were not shown to simplify this discussion.) These loadings were deemed to be significant if they were the largest (in magnitude) value in any particular row of the table. Selecting the significant loadings in this manner identified which factors correlated most closely with the original variables. Notice that one variable, C-17 MTMs/acft/day, has two entries in this table because the loadings were fairly equal. All this means is the information contained in C-17 MTMs/acft/day was split between factors 1 and 4, which are orthogonal dimensions in the new factor space.

Table 2: Small Scenario Factor Loadings Matrix

Factor	1	2	3	4
C-141 MTMs/acft/day	-.91			
C-141 use rate	-.89			
C-5 MTMs/acft/day	.89			
C-141 average payload	-.88			
C-5 use rate	.85			
B-747P use rate	.84			
KC-10 use rate	.83			
KC-10 MTMs/acft/day	.77			
C-17 ground cycle	.76			
C-141 ground cycle	.76			
Total Passengers	.75			
C-5 average payload	.73			
C-17 MTMs/acft/day	.63			.59
KC-10 average payload	-.59			
Total outsize tons		.93		
C-17 use rate		.84		
C-17 average payload		-.80		
% Shipments ontime			-.81	
Average days late			.78	
Total cargo tons				.94
Eigenvalues	9.46	2.98	2.39	1.71
% of Total Variance	47%	15%	12%	9%
Cumulative %	47%	62%	74%	83%

The dimensionality of the original data was reduced from 20 to 4, reflecting the simplicity of the small scenario. The 4 new dimensions account for 83 percent of the information. For comparison, though not listed in this article, the dimensionality of the more complex large scenario data was reduced from 29 to 8 factors, accounting for 80 percent of the variance/covariance information. For a mild increase in the number of original variables, the number of factors doubled, indicating the impact of increasing the complexity of the model greatly increases the dimensionality and complexity of the output.

Table 2 further indicates, in general, a positive correlation between factor 1 and positive widebody aircraft (B-747, C-5, C-17 and KC-10) performance. Passenger throughput is also positively correlated with factor 1, however, C-141 performance is negatively correlated. One possible valid interpretation of the factor 1 loadings indicate that the changes made to the TPFDD caused an increase in widebody aircraft performance at the expense of C-141 performance. The dimension of throughput most highly affected by this change in aircraft performance was passenger throughput. Factor 1 could be labeled as a widebody and passenger **index**, which is the label chosen for the rest of this discussion. It could also be labeled as a widebody and C-141 passenger performance **contrast**. In either case, positive values for this factor indicate positive widebody and passenger throughput performance, and negative values indicate positive C-141 performance.

A similar interpretation of factor 2 loadings identifies a positive relationship between total outsize tons throughput and C-17 usage. In other words, variance in the C-17 use rate accounted for the majority of the variance in total outsize tons delivered. The negative correlation with C-17 payload amounts reflects the less dense nature of outsize cargo (in tons/unit volume). Factor 2 is most definitely an index reflecting total outsize tons throughput. Similarly, factor 3 is some sort of lateness index, and factor 4 is a total cargo tons throughput index.

Most statistical packages that can solve Equation (2) can also solve Equation (1) for **f**, the factor scores. In terms of the small scenario, for each of the 100 runs conducted for the experimental design, the 20 output variable values were transformed into 4 factor variable scores. These factor scores were used to graphically represent the AFM simulation's output for sensitivity analysis.

The sensitivity plot shown in Figure 1 is a result of plotting the factor 2 scores against the factor 1 scores for all 100 runs from the small scenario. The plotted points were coded to identify the experimental design points listed in Table 1. The axes were labeled as indices to remind the reader of the general information represented by each factor. The cluster of points representing design point 0 provides a focus for interpreting this plot because the TPFDD was not perturbed for these 10 runs. The only difference among these 10 runs was the random seed selected at the start of the simulation. The size (area) of the reference cluster in Figure 1 approximately represents the variance, or random noise, expected from the stochastic nature of the AFM simulation. Design point clusters larger (covering more area) than the reference cluster show an effect on variance that is attributable to the design point. Clusters for design points 4, 5, 6, 7, and 9 were larger, indicating that perturbing the location characteristic greatly increased the variance of some airlift system performance measures.

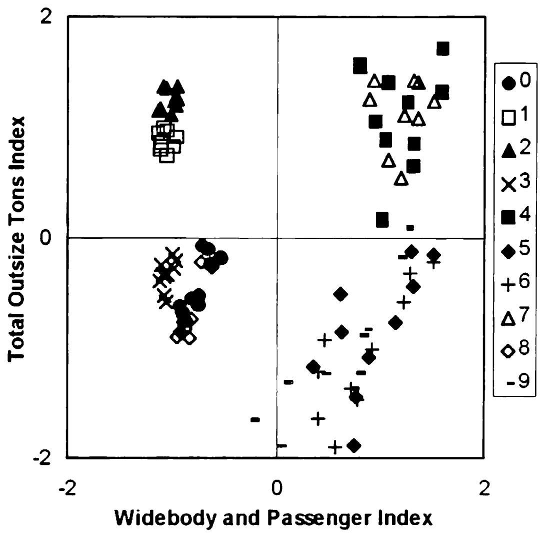

Figure 1: Sensitivity Plot of All Runs

Another observation from this sensitivity plot (Figure 1) has to do with the shape of the design point clusters. The cluster for design point 0 seems to be broken into at least two distinct groups. Furthermore, the clusters for design points 6, 8, and 9 also seem to exhibit the same phenomenon. The investigation into this phenomenon led to a discovery about the instability of one of the algorithms in the AFM simulation. The overall effect of this instability was not considered significant by the users of the AFM simulation, but the discovery did identify one potential model improvement area.

The plot shown in Figure 2 is a result of plotting the centroid of each design point cluster. Displacements between the plotted centroids indicate a change in airlift system performance, as measured by the AFM simulation. Once again, the reference is design point 0. Noticeable displacements to total outsize tons throughput are identified by design points 1, 2, 4, and 7. These 4 design points reflect the increase in outsize throughput resulting from perturbing the cargo flavor characteristic, where the demand for outsize cargo was increased. Similarly, the centroids for design points 4, 5, 6, 7, and 9 are displaced from the reference point, reflecting an increase in widebody aircraft performance and an increase in passenger throughput performance.

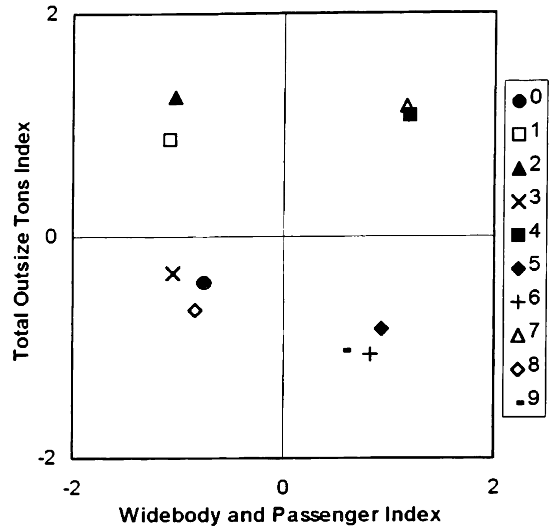

Figure 2: Sensitivity Plot of Design Point Centroids

5 VERIFICATION AND VALIDATION

The following definitions of verification and validation from Pritsker (1986) are presented to support the following discussion. Verification is the process of establishing that the computer program executes as intended. Validation is the process of establishing that a desired accuracy or correspondence exists between the simulation model and the real system.

A slightly different look at the factor analysis model in Equation (1) leads to implications supporting stochastic model verification and possibly validation. The view assumed so far sees a factor analysis model which attempts to account for variation in multivariate data by identifying a small number of common factors which encompass as much of the original variation as possible. A slightly different view is analogous to putting on factor analysis glasses to look inside the multivariate data and see what factors, or processes, are causing the data to change. If the latter view is accepted, factor analysis can be seen as a tool to identify the source of variation in multivariate data.

By changing the random seed from run to run in a complex, stochastic simulation model, the simulation's internal stochastic processes are the known source of variation in output data. If we then put on factor analysis glasses to look at the simulation's multivariate output data, we should be able to see the simulation's stochastic processes as the source of output data variation. If factor analysis results adequately reflect the simulation's stochastic processes then the program is working as intended and we have a method of verification. The choice of output variables selected determines the level of verification, allowing the analyst to tailor this methodology to focus on whatever process or combination of processes needs verification.

Using the small scenario example, Table 3 shows the loadings matrix for the output from 50 runs, where the only difference between the runs was the random seed. All the relationships between variables that loaded on the same factor were explainable through knowledge of AFM's internal processes and knowledge of how those process are supposed to work. With the aggregate variables selected to measure airlift system performance, an aggregate, surface-level verification of AFM was the result. From the information contained in Table 3, the AFM simulation's processes seem to work as designed.

Table 3: Factor Loadings Matrix for 50 Seed Runs

Factor	1	2	3	4	5
Total cargo delivered	0.97				
C-17 MTMs/acft/day	0.95				
Passengers delivered	0.91				
C-141 use rate	0.87				
C-141 MTMs/acft/day	0.86				
C-17 use rate	0.85				
C-17 ground cycle	-0.74				
C-17 avg payload	0.71				
KC-10 use rate		0.92			
KC-10 MTMs/acft/day		0.85			
KC-10 avg payload		-0.61			
C-5 MTMs/acft/day			0.88		
C-5 use rate			0.82		
C-141 avg payload			-0.68		
Average days late				0.83	
% Shipments ontime				-0.72	
B-747P use rate				-0.56	
C-5 avg payload					0.75
Outsize delivered					0.56
C-141 ground cycle					0.41
Eigenvalues	6.24	2.40	2.34	1.96	1.44
% of Total Variance	31%	12%	12%	10%	7%
Cumulative %	31%	43%	55%	65%	72%

Furthermore, the small scenario was designed to favor the smaller C-17 and C-141 aircraft, which is reflected in the factor 1 loadings. As can be seen in Table 3, passenger and cargo throughput are positively correlated with the C-17 and C-141 aircraft performance variables. Since the AFM formulation of the small scenario appears to execute similar to expected performance of the real world airlift system, there appears to be some potential for validation with this method, as well.

Another tool to use during a validation effort would be a plot of the factor scores similar to what was done to create the sensitivity plots in the previous section. In this example, Figure 3 is the plot of the factor 2 scores against the factor 1 scores for all 50 random seed runs. The distinct grouping of the output into at least two groups led to a discovery of a weakness in one of the model's mission planning algorithms. In this case, some of the cargo was ignored by the AFM simulation if it was too small, or trivial, to justify dispatching an aircraft. The size of a trivial cargo load is aircraft dependent and therefore the smaller aircraft, with the smaller trivial load limits will pick up more cargo throughout the duration of the simulation scenario. In this case, some of the runs were running out of cargo to deliver before the simulation ended because cargo was being ignored. The source of this effect was a combined function of the size of each requirement line in the TPFDD and the trivial load limit algorithm. Graphical plots similar to that shown in Figure 3 should be considered an essential part of a verification effort using this factor analysis technique, along with a factor loadings matrix similar to that shown in Table 3.

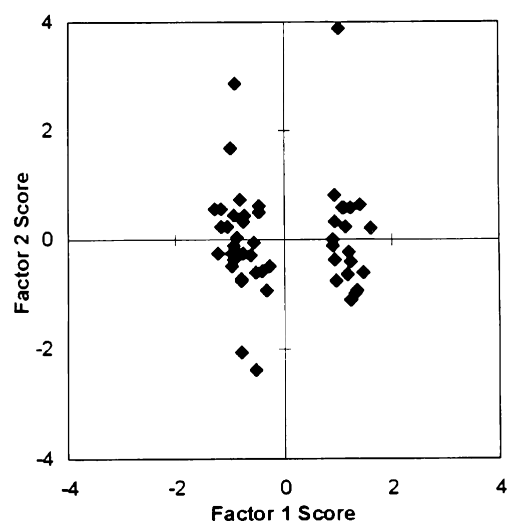

Figure 3: Plot of Factor Scores for 50 Seed Runs

6 SUMMARY AND CONCLUSIONS

Design of experiments and factor analysis techniques provide a powerful combination for analyzing large, stochastic simulations. This paper demonstrated an approach to conduct sensitivity analysis on the output of a large, stochastic airlift simulation model. We demonstrated how the factor loadings matrix can be used as a tool to interpret what happened in the simulation. In our example, the sensitivity plots created from the factor scores graphically depicted the relative magnitude and variance of the simulation's response to each experimental design point. We also successfully highlighted a new application area for factor analysis in stochastic simulation model verification and validation. The potential to tailor this verification and validation approach to any desired level of detail seems promising and worthy of further research.

REFERENCES

Basilevsky, A. 1994. *Statistical factor analysis and related methods: Theory and applications.* New York: Wiley.

Box, G. E. P., and N. R. Draper. 1987. *Empirical model-building and response surfaces.* New York: Wiley.

Dillon, W. R., and M. Goldstein. 1984. *Multivariate analysis.* New York: Wiley.

Kowalsky, T. 1977. Utilization rates and productivity. Study report, Studies and Analysis Flight, Air Mobility Command, Scott Air Force Base, Illinois.

Pritsker, A. A. B. 1986. *Introduction to simulation and SLAM II.* 3d ed. New York: Halstead.

Rousseau, G. G. 1996. Airlift system sensitivity to perturbed time-phase force deployment data. Master's thesis, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.

AUTHOR BIOGRAPHIES

GLENN G. ROUSSEAU, Captain, USAF, is an airlift pilot currently serving as a mobility analyst for Air Mobility Command at Scott Air Force Base, Illinois. He received a B.S. degree in engineering physics from the United States Air Force Academy in 1985, and he received an M.S. degree in operations analysis from the Air Force Institute of Technology in 1996. His research interests are in statistics and simulation modeling, primarily with application to transportation problems. He is a member of Tau Beta Pi and INFORMS.

KENNETH W. BAUER, JR., Lt Col, USAF, is an Associate Professor in the Department of Operational Sciences, School of Engineering, at the Air Force Institute of Technology (AFIT). He received a B.S. in Mathematics from Miami University (Ohio) in 1976, a M.E.A. from the University of Utah in 1980, a M.S. from AFIT in 1981, and a Ph.D. from Purdue University in 1987. His research interests are in the statistical aspects of neural networks and simulation modeling. His mailing address is: AFIT/ENS, WPAFB, OH, 45433-7765.