Proceedings of the 1985 Winter Simulation Conference
D. Gantz, G. Blais, S. Solomon (eds.)

# SIMULATION MODEL DECOMPOSITION BY FACTOR ANALYSIS

Kenneth W. Bauer
Bipin Kochar
Joseph J. Talavage
School of Industrial Engineering
Purdue University
West Lafayette, Indiana 47907.

## ABSTRACT

This paper offers a solution to the simulation model decomposition problem discussed briefly in Overstreet and Nance [1] and elaborated in detail in Overstreet [2]. The solution scheme involves the use of principal components analysis. We offer an example of the technique on a simple directed graph and then demonstrate the method on a small model given in Overstreet [2].

## INTRODUCTION

In moving toward a model development environment for discrete event simulation, Overstreet and Nance [1] have advanced a model specification language to systematically bridge the gap between a conceptual model and an executable representation of that model. In doing so, they introduce a formalism based on a condition specification. They demonstrate contributions of their model specification language to simulation theory and point out the utility of the formalism in an implementation enviroment.

One important area of application is the decomposition of models. Simply put, given a model specification, how to systematically decompose this representation into possible submodels? If a large model specification can be decomposed into submodels, this would facilitate model implementation because the task could be divided between programming groups. Overstreet [2] puts forth a framework for model decomposition, using a Cluster Interaction Graph. This directed graph represents a condition specification by nodes in a graph. The nodes represent "action clusters" constructed with the primitives of the specification language. These action clusters are linked through attributes. Attributes characterize objects in the model, and a change of an attribute in one cluster can trigger the actions of another cluster. Say for example, in a model specification of a machine shop, a machine fails in an action cluster. The failure is reflected as an attribute change and this in turn triggers the activities of another cluster. In this fashion, Overstreet links model actions to model attributes. He reasons that a link exists between action clusters when an output attribute for one cluster serves as an input/control attribute for another cluster. Further, action clusters with a high degree of interaction ought to be in the same component of the model. Now, within the Cluster Interaction Graph this "flow" of information between the nodes is represented by arcs. Overstreet suggests that the graph can be decomposed into two minimally interactive submodels by partitioning the graph into a pair of nonempty subgraphs with a minimal number of arcs connecting them. This problem has a complexity of $2^N$ (where N is the number of nodes). Another unfortunate feature of this cutset approach is that it offers no stopping rule.

We propose a solution to this problem which is not of high order complexity and has a built-in stopping rule. The Cluster Interaction Graph is reformulated into an association matrix and a factor analytic method is applied to the matrix. This technique is similar to that found in Garrison and Marble [3]; their research dealt with transportation networks. Another application of this technique is found in McClain [4].

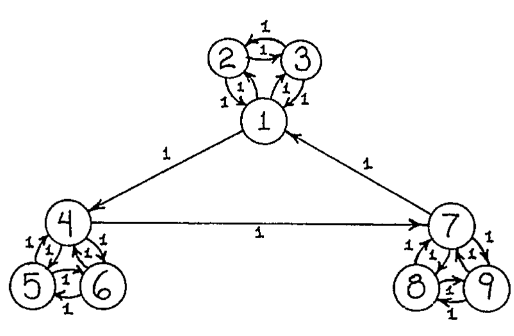
The association matrix entries represent the strength of relationships between the nodes. It is based on the number and direction of attributes they share. The eigenvalues and associated eigenvectors are extracted from the association matrix and then rotated to a solution best exhibiting the "simple structure" (in the sense of Thurstone (indirectly referenced through McNichols [5]) the model may exhibit.

We show an example of the technique on a simple directed graph and then demonstrate the method on an example from Overstreet [2].

## A SIMPLE EXAMPLE

Consider the following simple directed graph.

Figure 1. Simple Graph

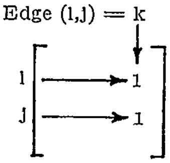


Assume that the numbers assigned to the arcs represent the amount and direction of information that is shared between nodes. Visually assessing this graph, one sees that there are basically three subgraphs. Nodes 1,2, and 3 form a logical subgraph, as do nodes 3,4, and 5, as well as nodes 6,7, and 8.

We may construct an edge-incidence matrix E as follows. If node i communicates with node j, then E can be constructed by

Edge (i,j) = k

Node



Next construct the matrix $A = E\,W\,W^T\,E^T$, where W is a weighting matrix. Various weighting schemes are possible. For instance, it might be reasonable to weight the bidirectional arcs more heavily than unidirectional arcs.

Now A is symmmetric and positive semi-definite, hence it can be converted to an association (pseudo-correlation) matrix by multiplying it by the matrix D, where

$$D = \begin{bmatrix} \ddots & & \\ & 1/\sqrt{a_{ii}} & \\ & & \ddots \end{bmatrix}$$

Now we have $C = D^T A D$ . The eigenvalues and associated eigenvectors are extracted from this matrix. The eigenvalues can be examined to see if a potential reduction in dimensionality is feasible. Based on this examination certain factors (scaled and normalized eigenvectors) are retained and these are rotated to simple structure.

This procedure was accomplished for the graph in Figure 1. Principal components analysis was performed using SPSS. Principal components analysis is equivalent to the eigenvalue analysis described above.

Table 1 details the factors and associated eigenvalues as extracted from the pseudo-correlation matrix prepared for the graph presented in Figure 1.

Table 1. Extracted Factors

| Factor | Eigenvalue | Pct. of Var | Cum. Pct |
|---|---|---|---|
| 1 | 2.00000 | 22.2 | 22.2 |
| 2 | 1.83333 | 20.4 | 42.6 |
| 3 | 1.83333 | 20.4 | 63.0 |
| 4 | .83333 | 9.3 | 72.2 |
| 5 | .50000 | 5.6 | 77.8 |
| 6 | .50000 | 5.6 | 83.3 |
| 7 | .50000 | 5.6 | 88.9 |
| 8 | .50000 | 5.6 | 94.4 |
| 9 | .50000 | 5.6 | 100.0 |

Notice that 63% of the variance in this graph is explained by 3 factors. Also notice that the fourth eigenvalue is less than 1. Retaining only those factors whose eigenvalues are at least 1 is a popular retention criteria attributed to Kaiser [6]. Some empirical evidence for the use of this criteria is offered by Bauer [7]. Table 2 is a table of initial factor loadings. This table relates the nodes to the factors, the higher the loading the greater the linear correlation to the factor.

Table 2. Initial Factor Loadings

| Node | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Node1 | .53452 | -.51044 | .21215 |
| Node2 | .43644 | -.62516 | .25983 |
| Node3 | .43644 | -.62516 | .25983 |
| Node4 | .53452 | .43895 | .33598 |
| Node5 | .43644 | .53760 | .41149 |
| Node6 | .43644 | .53760 | .41149 |
| Node7 | .53452 | .07149 | -.54813 |
| Node8 | .43644 | .08765 | -.67132 |
| Node9 | .43644 | .08765 | -.67132 |

Since the first extracted factor tends to be a general factor (see Nie [8]), relating all the variables to one another, not much is gleaned from this initial factor loadings matrix. Application of varimax rotation to this matrix yields much more interesting results. Table 3 is the new rotated factor matrix.

Table 3. Rotated Factor Matrix

| Node | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Node1 | .08294 | .08294 | .75994 |
| Node2 | -.02441 | -.02441 | .80475 |
| Node3 | -.02441 | -.02441 | .80475 |
| Node4 | .75994 | .08294 | .08294 |
| Node5 | .80475 | -.02441 | -.02441 |
| Node6 | .80475 | -.02441 | -.02441 |
| Node7 | .08294 | .75994 | .08294 |
| Node8 | -.02441 | .80475 | -.02441 |
| Node9 | -.02441 | .80475 | -.02441 |

Before addressing this matrix we refer back to Table 1. Note that by Kaiser's criterion we retained 3 factors. Now, looking at Table 3, we note that 3 nodes load heavily on each factor (circled). Hence, we summarize by saying that there seems to be 3 principal subgraphs here (3 retained factors) and each is composed of the 3 nodes indicated.

## DECOMPOSITION OF A CONDITION SPECIFICATION

To demonstrate the method's potential in model decomposition, we apply it to a Cluster Interaction Graph of a harbor simulation model offered by Overstreet [2]. The harbor model has ships arriving at a harbor. These ships wait in an area just outside the harbor until a tug is available to move them to their assigned berth. Once the ships are unloaded and a tug is available, the ships are moved out of the harbor. Overstreet presents a condition specification

Table 4. Cluster Interaction Graph (From Overstreet [2])

| Clusters | init | arrival | enter | unload | endunl | deberth | endeb | mttp | taap | mtto | taao | term |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| initialization | | 1 | 3 | | | 4 | | 5 | | 4 | | |
| arrival | | 1 | 1 | | | 1 | | 1 | | 1 | | |
| enter | | | 2 | 1 | | 2 | | 3 | | 3 | | |
| unload | | | | | 1 | 1 | | 1 | | 1 | | |
| end unload | | | | | | 1 | | 1 | | | | |
| deberth | | | 1 | | | 3 | 1 | 3 | | 2 | | |
| end deberth | | | 1 | | | | | 1 | | 1 | | 1 |
| move tug to pier | | | 1 | | | | | 1 | 1 | 1 | | |
| tug arr at pier | | | | | | 1 | | 1 | | 1 | | |
| move tug to ocean | | | | | | 1 | | 1 | | 1 | 1 | |
| tug arr at ocean | | | 1 | | | | | 1 | | 1 | | |
| termination | | | | | | | | | | | | |

for the harbor model based on 12 action clusters. He gives self-explanatory names to each cluster (node) and presents an arc-incidence matrix which details the shared information between the nodes. Table 4 is from Overstreet. We converted this table to an association matrix and factored it using SPSS. The final rotated factor loading matrix is Table 5.

Table 5. Harbor Model: Rotated Factor Matrix

| Node | Factor 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| taao | -.088 | -.0241 | .868 | .011 | -.008 |
| arriv | -.014 | .046 | .001 | -.005 | .926 |
| enter | .042 | -.068 | .228 | .593 | .314 |
| unload | .705 | -.001 | .077 | .061 | .019 |
| endunld | .791 | -.044 | .047 | -.082 | -.058 |
| deber | .412 | .250 | -.148 | .397 | .211 |
| enddeb | -.045 | .074 | -.026 | .815 | -.215 |
| mttp | -.018 | .643 | .226 | .197 | .106 |
| taap | .009 | .868 | -.042 | -.106 | -.051 |
| mtto | .272 | .226 | .611 | .084 | .031 |

Referring to Table 5, first notice that 5 factors are retained. Since we are factoring a 10 variable set, we see that we may have trouble finding separable submodels since the overall model is so interrelated. However, a couple of observations are warranted. Notice that Unload and Endunld load heavily on the first factor and that MTTP and TAAP score heavily on the second factor. This suggests that "pier activities" and "harbor activities" might represent reasonable submodels. Note also that TAAO and MTTO load heavily on the third factor, perhaps indicating a submodel of "ocean activities". It is interesting to note that the arrival of ships to the system loads on the fifth, independent factor.

## SUMMARY

The use of factor analytic techniques, in particular the method of principal components, offers a low complexity solution to model decomposition problem. Cluster Interaction Graphs can be converted into association matrices and these matrices can be factored using standard factor analysis routines such as those found in SPSS.

## REFERENCES


[1] Overstreet, C. M. and Nance R. E. " A Specification Language to Assist in Analysis of Discrete Event Simulation Models," *Communications of the ACM* **2**, February 1985, 190-201.

[2] Overstreet, C. M. " Model Specification and Analysis for Discrete Event Simulation", Unpublished Doctoral Dissertation, Virginia Polytechnic Institute and State University, 1982.

[3] Garrison, W. L. and Marble D. F. *The Structure of Transportation Networks,* TCREC Technical Report 62-11, Transportation Center, Northwestern University, Evanston, Illinois, May 1962.

[4] McClain, Denny R. " A Systems approach to the Aeromedical Aircraft Routing Problem Using A Computer Based Model," Unpublished Doctoral Dissertation, University of California, Berkeley, December 1984.

[5] McNichols, Charles W. " An Introduction to Applied Multivariate Data Analysis ( Course Notes )," Air Force Institute of Technology, WPAFB, Ohio, 1980.

[6] Kaiser, H. F. " The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement,* **20**, 1960, 141-151.

[7] Bauer, Kenneth W. " A Monte Carlo Study of Dimensionality Assessment and Factor Interpretation in Principal Components Analysis," Unpublished Master's Thesis, Air Force Institute of Technology, 1981.

[8] Nie, N. H. et. al. *Statistical Package for the Social Sciences,* 2nd Edition, New York: McGraw Hill, 1975.

KENNETH W. BAUER is a Captain in the United States Air Force and currently a Ph.D. student in Industrial Engineering at Purdue University.

Grissom Hall, Room 374
Purdue University
West Lafayette, Indiana 47907
(317) 494-1797

BIPIN KOCHAR holds a M.S. in Industrial Engineering from Purdue University and is currently employed by Pritsker and Associates.

Grissom Hall, Room 374
Purdue University
West Lafayette, Indiana 47907
(317) 494-1797

JOSEPH J. TALAVAGE is a professor of Industrial Engineering at Purdue University. He holds a Ph.D. (Systems Engineering) from Case Institute of Technology, 1968. In addition to teaching simulation courses at the graduate level, he is engaged in research on simulation methodology, including the development of a manufacturing decision support system, and the use of simulation to design improved hierarchical control systems for steel production. He has been a consultant to numerous companies and govermental agencies, and is the prime developer of the microNET simulation language.

Room 1318C
Michael Golden Laboratories
Purdue University
West Lafayette, Indiana 47907
(317) 494-5412