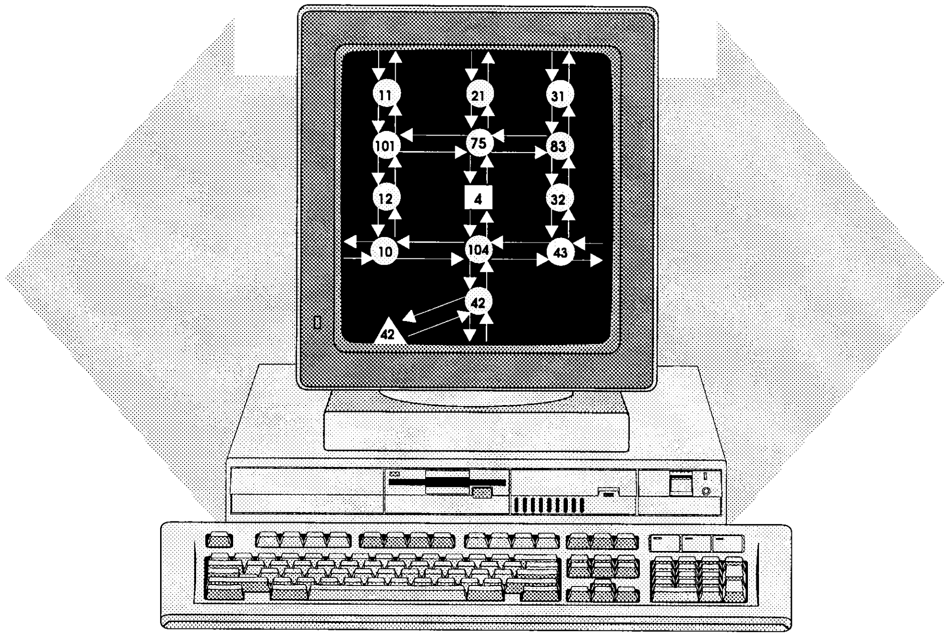
FINAL
CONTRACT REPORT

A METHOD TO ENHANCE THE PERFORMANCE OF SYNTHETIC ORIGIN-DESTINATION (O-D) TRIP TABLE ESTIMATION MODELS



R. SIVANANDAN
Assistant Director of Research
Center for Transportation Research
Virginia Polytechnic Institute & State University

DHRUV NANDA
Graduate Research Asssistant
Center for Transportation Research
Virginia Polytechnic Institute & State University

V·I·R·G·I·N·I·A
VTRC
TRANSPORTATION RESEARCH COUNCIL

VIRGINIA TRANSPORTATION RESEARCH COUNCIL

Standard Title Page - Report on Federally Funded Project

1. Report No. FHWA/VTRC 98-CR 23	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A Method to Enhance the Performance of Synthetic Origin-Destination (O-D) Trip Table Estimation Models		5. Report Date January 1998 6. Performing Organization Code
7. Author(s) R. Sivanandan and Dhruv Nanda		8. Performing Organization Report No. VTRC 98-CR 23
9. Performing Organization and Address Virginia Transportation Research Council 530 Edgemont Road Charlottesville, Virginia 22903		10. Work Unit No. (TRAIS) 11. Contract or Grant No. SPR 0577-030-940
12. Sponsoring Agencies' Name and Address Virginia Department of Transportation 1401 E. Broad Street Richmond, VA 23219	FHWA 1504 Santa Rosa Road Richmond, VA 23239	13. Type of Report and Period Covered Final Report, Feb. 1996-Dec. 1997 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The conventional methods of determining origin-destination (O-D) trip tables involve elaborate surveys, e.g., home interviews, that require considerable time, staff, and funds. To overcome this drawback, a number of theoretical models that synthesize O-D trip tables from link volume data have been developed. The focus of the research reported here was on two of these models – The Highway Emulator (THE) and the Linear Programming (LP) model. These models use target/seed tables for guiding the development of output trip tables. In earlier research conducted by the Virginia Tech Center for Transportation Research for the Virginia Transportation Research Council, it was determined that the performance of these models could potentially be enhanced by using a superior target/seed table. The research in this report uses readily available socioeconomic data and link volume information to develop a methodology for obtaining an enhanced target/seed table through application of the trip generation and trip distribution steps of the four-step planning process. The enhanced table was then used as the target/seed to THE and LP models, and their performance evaluated. In addition to measuring the closeness of the output tables to surveyed tables and their capability to replicate observed volumes, their improvements over the case when a structural table is used as target was also studied. Tests showed that the use of the enhanced target/seed table significantly improved the performance of the LP model. However, mixed trends were obtained for THE.

17 Key Words The Highway Emulator, O-D Models, Trip Tables, Synthetic Models, Link Volumes, Trip Distribution, Trip Assignment	18. Distribution Statement No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 40	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

FINAL REPORT

A METHOD TO ENHANCE THE PERFORMANCE OF SYNTHETIC ORIGIN-DESTINATION (O-D) TRIP TABLE ESTIMATION MODELS

R. Sivanandan
Asst. Director of Research & Asst. Professor
Dhruv Nanda
Graduate Research Assistant
Center for Transportation Research
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0536

Project Monitor
Eugene D. Arnold, Jr., Virginia Transportation Research Council

Contract Research Sponsored by the
Virginia Transportation Research Council

Virginia Transportation Research Council
(A Cooperative Organization Sponsored Jointly by the Virginia Department of Transportation
and the University of Virginia)

Charlottesville, Virginia

January 1998
VTRC 98-CR 23

NOTICE

The project that is the subject of this report was done under contract for the Virginia Department of Transportation, Virginia Transportation Research Council. The opinions and conclusions expressed or implied are those of the contractors, and, although they have been accepted as appropriate by the project monitor, they are not necessarily those of the Virginia Transportation Research Council or the Virginia Department of Transportation.

Each contract report is peer reviewed and accepted for publication by Research Council staff with expertise in related technical areas. Final editing and proofreading of the report are performed by the contractor.

ABSTRACT

The conventional methods of determining origin-destination (O-D) trip tables involve elaborate surveys, e.g., home interviews, that require considerable time, staff, and funds. To overcome this drawback, a number of theoretical models that synthesize O-D trip tables from link volume data have been developed. The focus of the research reported here was on two of these models -- The Highway Emulator (THE) and the Linear Programming (LP) model. These models use target/seed tables for guiding the development of output trip tables. In earlier research conducted by the Virginia Tech Center for Transportation Research for the Virginia Transportation Research Council, it was determined that the performance of these models could potentially be enhanced by using a superior target/seed table. The research in this report uses readily available socioeconomic data and link volume information to develop a methodology for obtaining an enhanced target/seed table through application of the trip generation and trip distribution steps of the four-step planning process. The enhanced table was then used as the target/seed to THE and LP models, and their performance evaluated. In addition to measuring the closeness of the output tables to surveyed tables and their capability to replicate observed volumes, their improvements over the case when a structural table is used as target was also studied. Tests showed that the use of the enhanced target/seed table significantly improved the performance of the LP model. However, mixed trends were obtained for THE model.

FINAL REPORT

A METHOD TO ENHANCE THE PERFORMANCE OF SYNTHETIC ORIGIN-DESTINATION (O-D) TRIP TABLE ESTIMATION MODELS

R. Sivanandan
Asst. Director of Research & Asst. Professor
Dhruv Nanda
Graduate Research Assistant

INTRODUCTION

An origin-destination (O-D) trip table is a two-dimensional matrix of elements whose cell values represent the number of trips made between various O-D zone pairs in a given region. The establishment of a current O-D trip table for an urban area through conventional surveys, such as home interviews, license plate surveys, and roadside surveys, is expensive, time consuming, and labor-intensive. In addition, most of these methods are conducted through sampling, and have associated sampling errors due to the reliance on only a small sample of trip makers. Even if all the trips on a particular day are recorded, the O-D table may not be stable over time due to daily variations (Willumsen, 1978). There are other inherent drawbacks associated with conventional techniques. One common problem is the changes in travel pattern due to changes in influencing factors. For instance, as the land use develops or changes rapidly, so will the trip table. Thus, the previously established trip table becomes outdated and obsolete. This will necessitate re-surveying, leading to further expenditures and efforts.

Recognizing the budgetary, time, and staff constraints faced by organizations needing O-D tables for planning and traffic operations purposes, researchers began exploring alternative techniques of establishing O-D tables, leading to the evolution of theoretical approaches in the early 1970s. Several approaches and models have been developed since then for establishing trip tables without the need for surveys by using link volumes.

In an earlier research effort sponsored by the Virginia Transportation Research Council (VTRC) and conducted by the Virginia Tech's Center for Transportation Research (Sivanandan, Narayanan & Peng, 1996), a comprehensive review of models that estimate trip tables from link volume information (hereinafter referred to as synthetic models) was performed. Two of these -- the Linear Programming (LP) model developed at Virginia Tech (Sivanandan, 1991; Sherali et al., 1994a,b; Narayanan, 1995) and The Highway Emulator (THE) model (Bromage, 1991) were selected for evaluation. See Appendix A for descriptions of these two models. Detailed and extensive tests were conducted to evaluate the validity of these models and to determine the sensitivity of the models to various percentages of link volumes available and target tables. The LP model was judged to be generally superior, both in terms of closeness of modeled trip tables to the "correct/surveyed" tables and in terms of replicating observed link volumes. One of the validation case studies was performed by comparing the models' output tables with the tables developed from O-D surveys conducted by the Virginia Department of Transportation (VDOT) for the town of Pulaski, Virginia. See Appendix B for a description of the development of the Pulaski trip tables.

Like most of the models in this family, the LP and THE models employ some form of old/prior trip table as a target/seed to guide the solution. However, such tables are not always available, leading to the questionable performance of some of these models. In fact, in the previously mentioned research effort, VDOT was primarily interested in the case where no prior trip table information was available. This has a very practical significance in that many of the urban areas for which a trip table is needed do not have a previously established table. Thus, the use of a structural target table, which is a table with 0 or 1 as cell value (0 signifying that the O-D interchange represented by the cell is not feasible, and 1 where it is feasible) is the only option. While THE model produced better results than the LP for this case, results from both the models were generally poor. However, it was determined in the study that the amount of information contained in the seed table played a key role in determining the quality of the output table.

It is possible to establish an O-D table based on easily accessible socioeconomic/census data that is a better representation of the travel patterns in the region than a structural table. This table can then be used as a target to selected O-D table estimation models.

PURPOSE AND SCOPE

The primary purpose of this research was to determine if the performance of THE and LP models could be improved through the use of a target/seed table developed from readily available and easily accessible socioeconomic data. A secondary purpose was to evaluate any improvement as well as the relative performance of the models by comparing their output with the results of VDOT's Pulaski survey (similar to the evaluation conducted in the aforementioned earlier study).

The Pulaski highway network and the O-D table developed by VDOT in a previous study were used for the research. Most of the required data were available from VDOT; however, some of the data on socioeconomic variables were obtained from officials in Pulaski.

RESEARCH DESIGN

The original intent in this research was to use the readily available and easily accessible socioeconomic data in the Census Transportation Planning Package (CTPP) to establish a trip table for use as the target in the synthetic models. However, since Pulaski is categorized under the statewide element of CTPP, the database contains only aggregated data, not the required disaggregated data at the zonal level. (Appendix C provides further information on the CTPP and why it was not usable in Pulaski.) Accordingly, the original approach was revised such that zone-specific socioeconomic data were used in a trip generation model, the conventional first step of the four-step planning process. The output of the trip generation model was then used in a trip distribution gravity model to derive the trip table to be input as the target/seed table for the synthetic models. This approach combined the conventional wisdom that socioeconomic characteristics (which are ignored by many synthetic models) generally influence trip-making behavior with the fact that observed traffic volumes on the network provide information on the actual trips being made during the period for which the trip table is being developed. Finally, the synthesized trip tables output by the models were evaluated to determine (1) if there were improvements using this approach and (2) how close they matched Pulaski's survey-derived tables. The overall approach is depicted in Figure 1.

METHODOLOGY

The basic steps in the methodology are shown in Figure 2 and are defined in the following five steps. A detailed description of how the performance of the synthetic models was evaluated follows the five steps.

1. The first step was to collect the data necessary to apply the trip generation and trip distribution models. The number of dwelling units and employment data by zone were obtained from VDOT and the town of Pulaski, respectively. Link volume data needed later to apply the synthetic models was available from the earlier study and were not collected as part of this effort.

2. This data was used to run the trip generation model in the transportation planning computer software package MINUTP. Output from the trip generation model was used as input to run MINUTP's gravity model that produced "enhanced" trip tables showing the distribution of the trips to the various zones for 24-hours and peak hour periods. A general discussion of how MINUTP processes trip generation and trip distribution is contained in Appendix D while the assumptions made and calculations used are shown in detail in Appendix E.

3. The link volumes and the enhanced target/seed trip tables from Step 2 were input to the LP and THE models to produce trip tables for Pulaski for the 24-hour and peak hour cases.

4. The trip tables from Step 3 were first evaluated by comparing their output with synthetically produced trip tables available from the earlier research study that were derived from running the same models using a structural target/seed table. This comparison was of particular interest since the use of a structural target table would be the choice in the absence of old or prior trip tables. It was hoped that by using the enhanced trip tables as input, the performance of the two synthetic models being used would be better than when the structural table was used. Also, the Step 3 trip tables were evaluated by comparing them with the 24-hour and peak hour trip tables derived from the VDOT surveys in Pulaski. Finally, the models (using

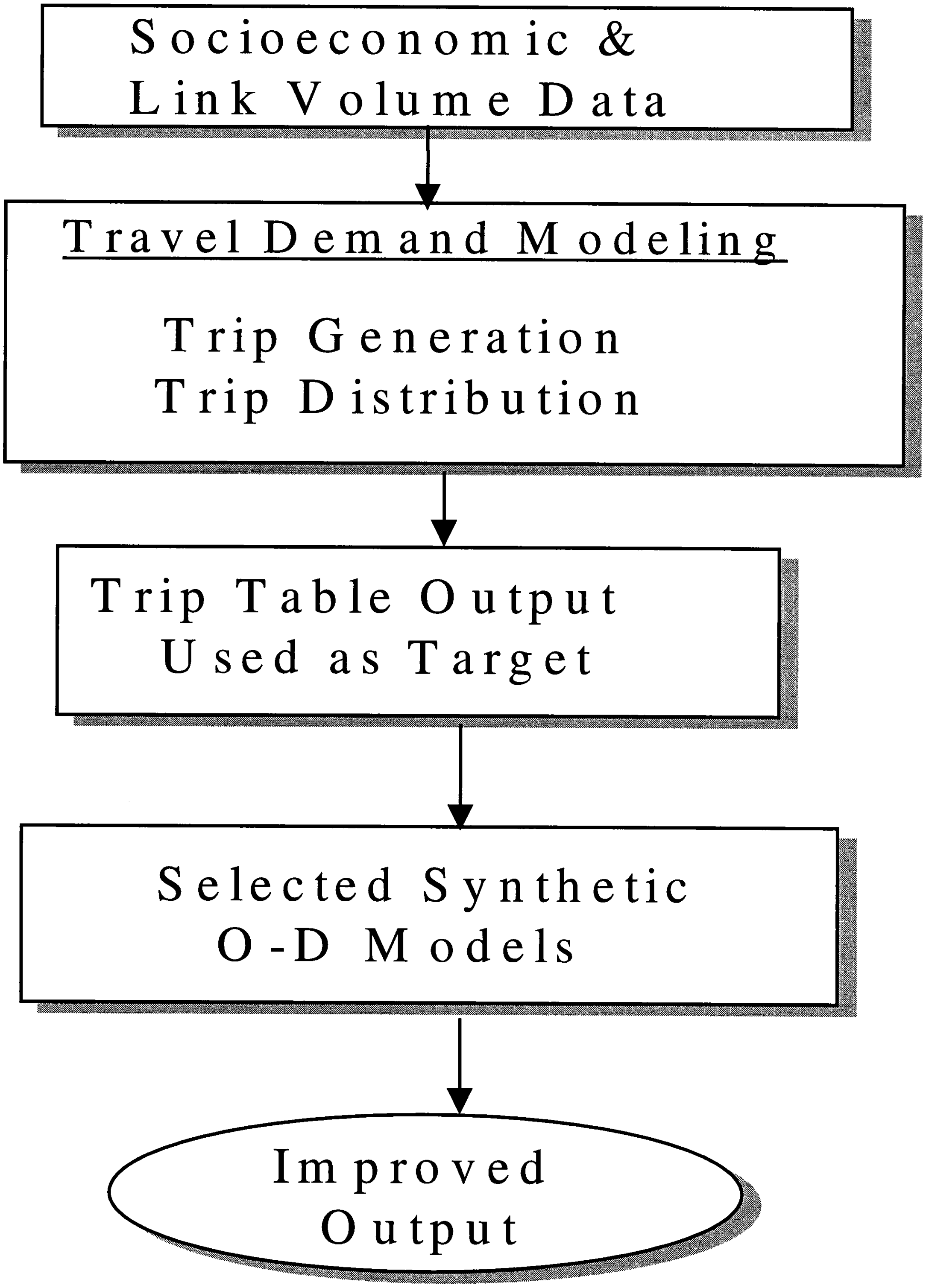


Figure 1 The Research Approach

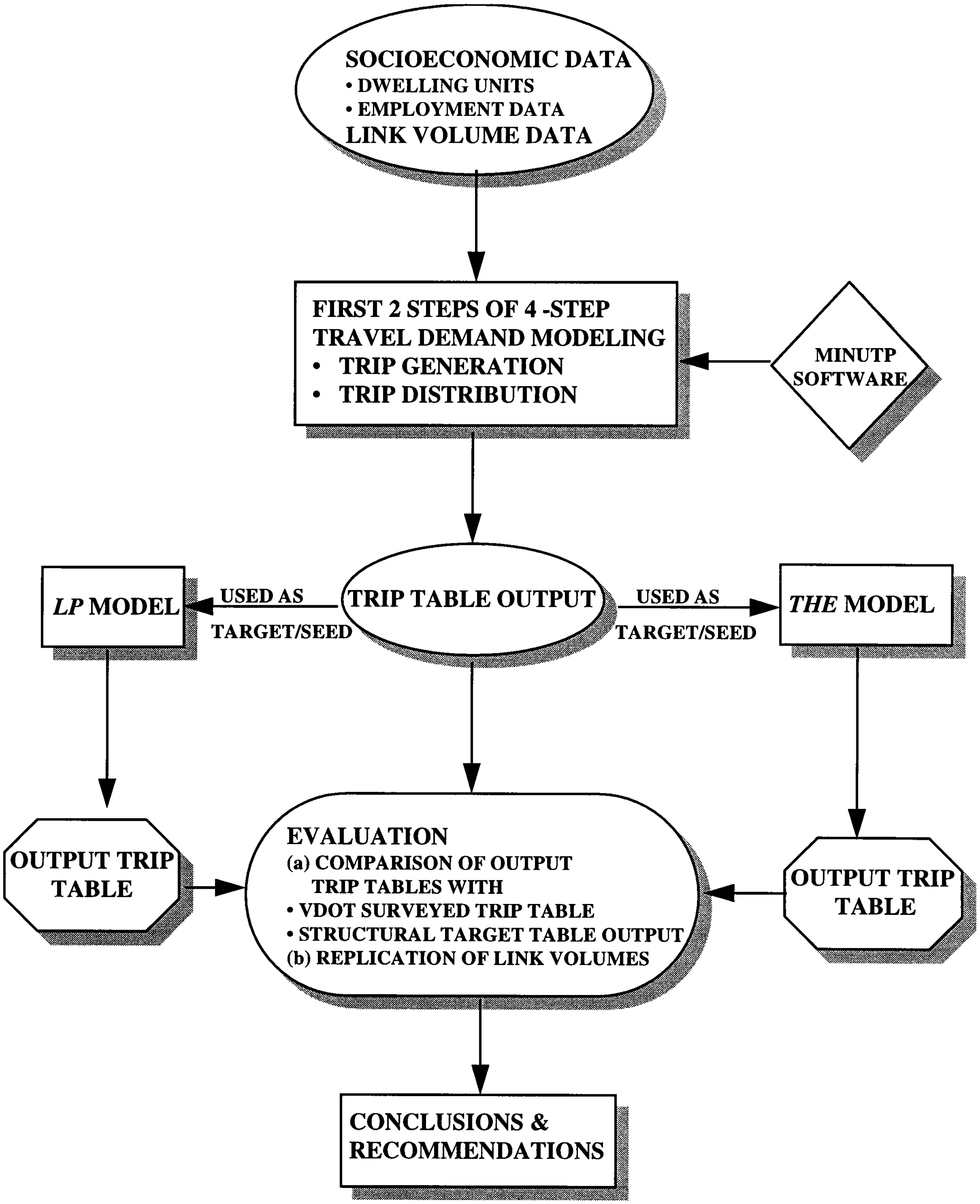


Figure 2 Methodology

the Step 3 trip tables) were tested for their sensitivity to varying percentages of available link volumes. (The evaluation is described in detail in a later section.)

5. Finally, conclusions and recommendations were derived from these analyses for both models. Recommendations for future research and potential areas for enhancing this research were also included.

Evaluation of the Synthetic Models with Improved Target/Seed Tables

In order to evaluate the synthetically derived trip tables generated by the THE and LP models, two measures of closeness were used -- the closeness of the generated table to VDOT's survey-derived table (assumed to be the "true" table) and the replication of the observed link volumes. These were obvious choices since the objective of trip table estimation is to develop a table that is as close to the true table as possible and replicates observed link volumes.

Closeness of Generated Trip Tables to VDOT's Survey-Derived Trip Tables

The validity of a trip table generated from a synthetic model is best evaluated by comparing how close it is to the true or correct trip table; however, obtaining such a table may be impossible. In practice, an O-D table is derived from a survey that is based on a sample of interviews taken on a given day. The sample's characteristics are then expanded based on the population of the study area with the assumption that the sampled travel characteristics are representative of the area's entire population. Further, it is assumed that travel during the time of the survey is representative of the area's daily travel, and that this travel is stable over time. It can be argued that an O-D table so derived cannot be the true or correct table due to these assumptions. As a compromise, however, a table derived in such a way is used in transportation planning and is considered reasonably good. Thus, for purposes of this research, VDOT's survey-derived trip tables were considered as true of correct tables for comparison purposes.

There are various statistical measures of closeness for comparing trip matrices. In addition to the percentage root mean square error (% RMSE) and percentage mean absolute error (% MAE), Smith and Hutchinson (1981) evaluated different closeness of fit statistics for trip distribution models and concluded that the PHI statistic (ϕ) was one of the most appropriate measures to use. Consequently, the %RMSE, %MAE, and ϕ were the statistical measures selected the comparison of trip table matrices. These measures of closeness are defined below:

$$\% RMSE = \sqrt{\frac{\sum (t_{ij} - t_{ij}^*)^2}{n_{OD}}} * \frac{100 * n_{OD}}{\sum t_{ij}^*}$$

$$\% MAE = \frac{\sum | t_{ij} - t_{ij}^* |}{\sum t_{ij}^*} * 100$$

$$\phi = \sum \max(1, t_{ij}^*) \left| \ln \frac{\max(1, t_{ij}^*)}{\max(1, t_{ij})} \right|$$

(Note: The above definition of ϕ has been slightly modified from Smith and Hutchinson [1981].)

where

t_{ij}^* = true/correct/reasonably good/surveyed number of trips for O-D interchange (i, j)
t_{ij} = estimated or modeled number of trips for O-D interchange (i, j)
n_{OD} = number of feasible O-D interchanges

Since the above statistics are measures of error in estimation, the smaller the values of these measures, the closer the generated tables are to the true or correct tables. Ideally, values of zero for each of these statistics would mean that the generated tables are the same as the survey tables.

Replication of Observed Link Volumes

In general, one of the most important measures of the quality of a trip table is its ability to replicate observed volumes on the network links once it is assigned to the network. The link volumes that were compared to the observed were obtained as byproducts of running THE and the LP models for Pulaski. This comparison was applied only to links for which observed volumes were provided as input.

The % RMSE and % MAE were selected as the statistical measures to compare the closeness of modeled link volumes to the observed link volumes. These measures are defined as follows:

$$\%\,RMSE = \sqrt{\frac{\sum_{a \in A_v} (V_{assign}^{a} - V_{obs}^{a})^2}{n}} * \frac{100 * n}{\sum_{a \in A_v} V_{obs}^{a}}$$

$$\%\,MAE = \frac{\sum_{a \in A_v} | V_{assign}^{a} - V_{obs}^{a} |}{\sum_{a \in A_v} V_{obs}^{a}} * 100$$

where

V_{assign}^{a} = assigned volume on link a
V_{obs}^{a} = observed volume on link a
n = number of links with available volumes
A_v = set of links with available volumes

The smaller the values of these measures, the better the replication of observed link volumes. Ideally, values of zero for each of these measures mean perfect replication.

Test Cases

In many practical cases, a number of the link volumes on a network are unknown. Thus, the LP and THE models were tested separately for cases in which 50%, 60%, and 75% of the network link volumes were available. These links were selected randomly and allowed for the study of sensitivity of the models to volume information. Additionally, 24-hour and peak hour trip tables were available. Accordingly, in this research, the statistical measures described above were calculated for both the 24-hour and peak hour time periods at each of the three percentages of link volume availability for the case when the MINUTP-based target/seed table was used. The same statistics for the same combination of time period and link volume availability were obtained from the earlier study for the case when a structural target/seed table was used. These values were also reported and used in the evaluation.

FINDINGS

24-Hour Trip Tables

Closeness of Generated Trip Tables to VDOT's Survey-Derived Trip Tables

The performance of the LP and THE models in replicating the surveyed trip tables are depicted in detail for the %MAE, %RMSE, and PHI statistics in Figures 3 through 5, respectively. The following observations can be made:

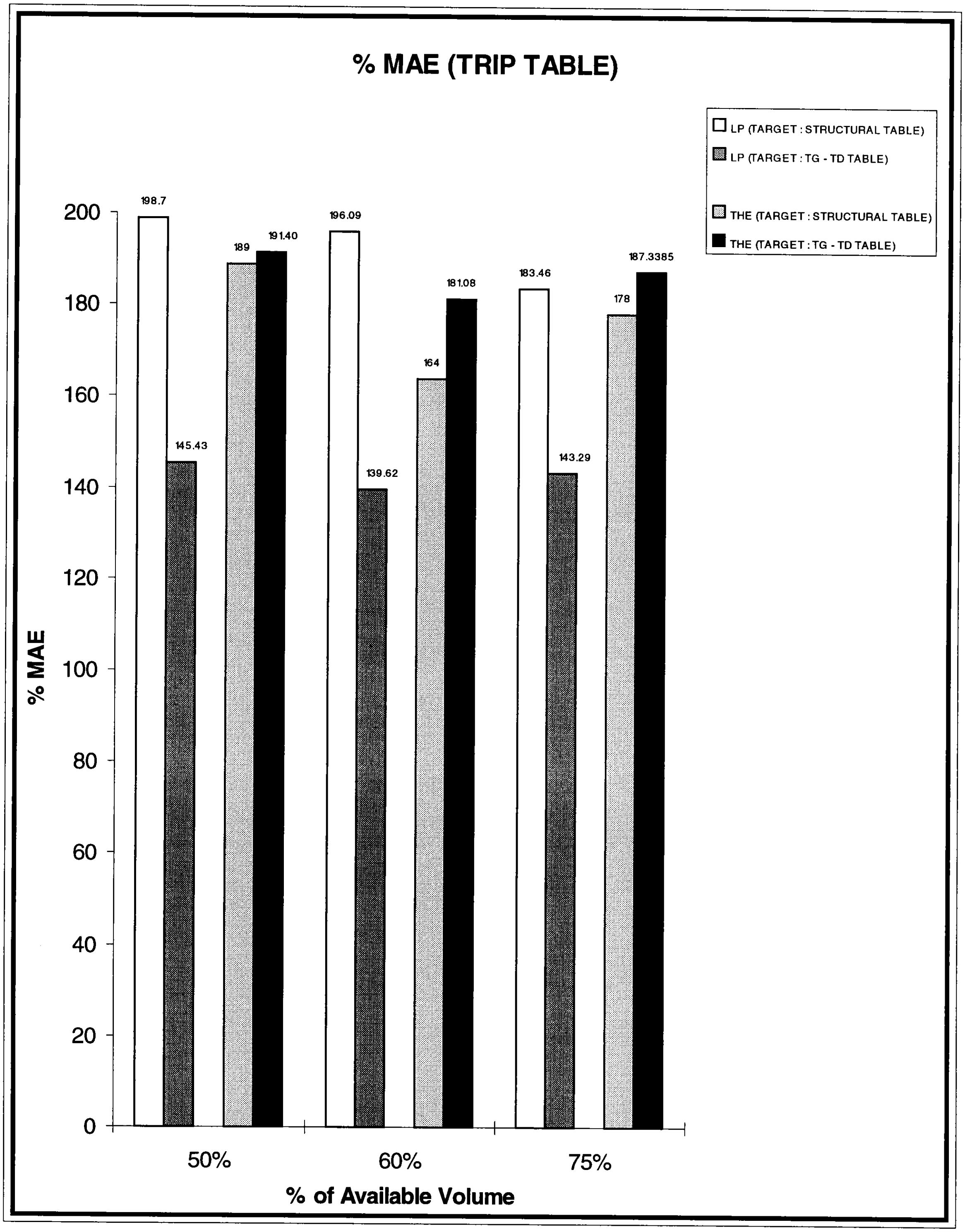


Figure 3 Trip Table Comparisons (Modeled vs Surveyed)
24-Hour Case (% MAE)

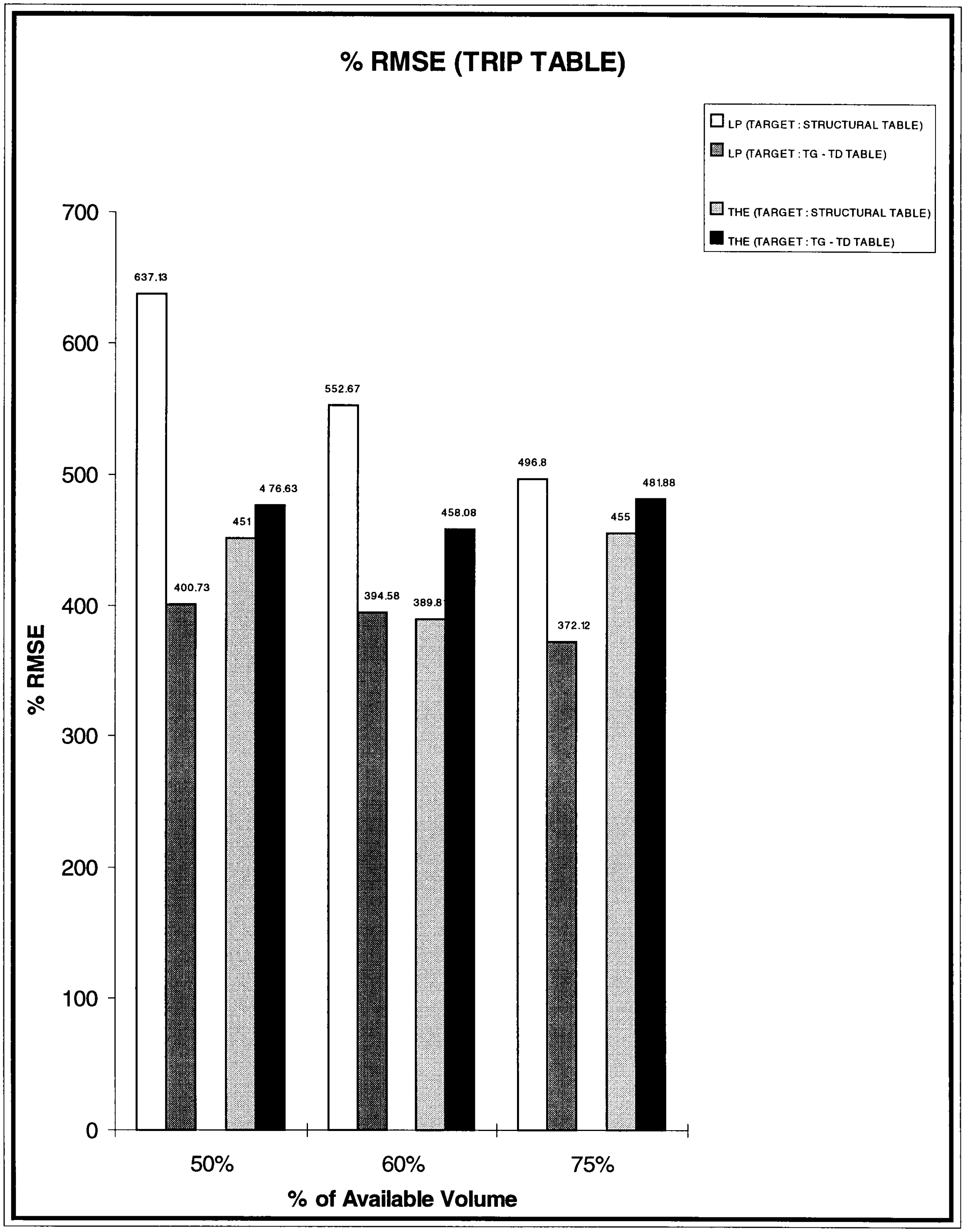


Figure 4 Trip Table Comparisons (Modeled vs Surveyed)
24-Hour Case (% RMSE)

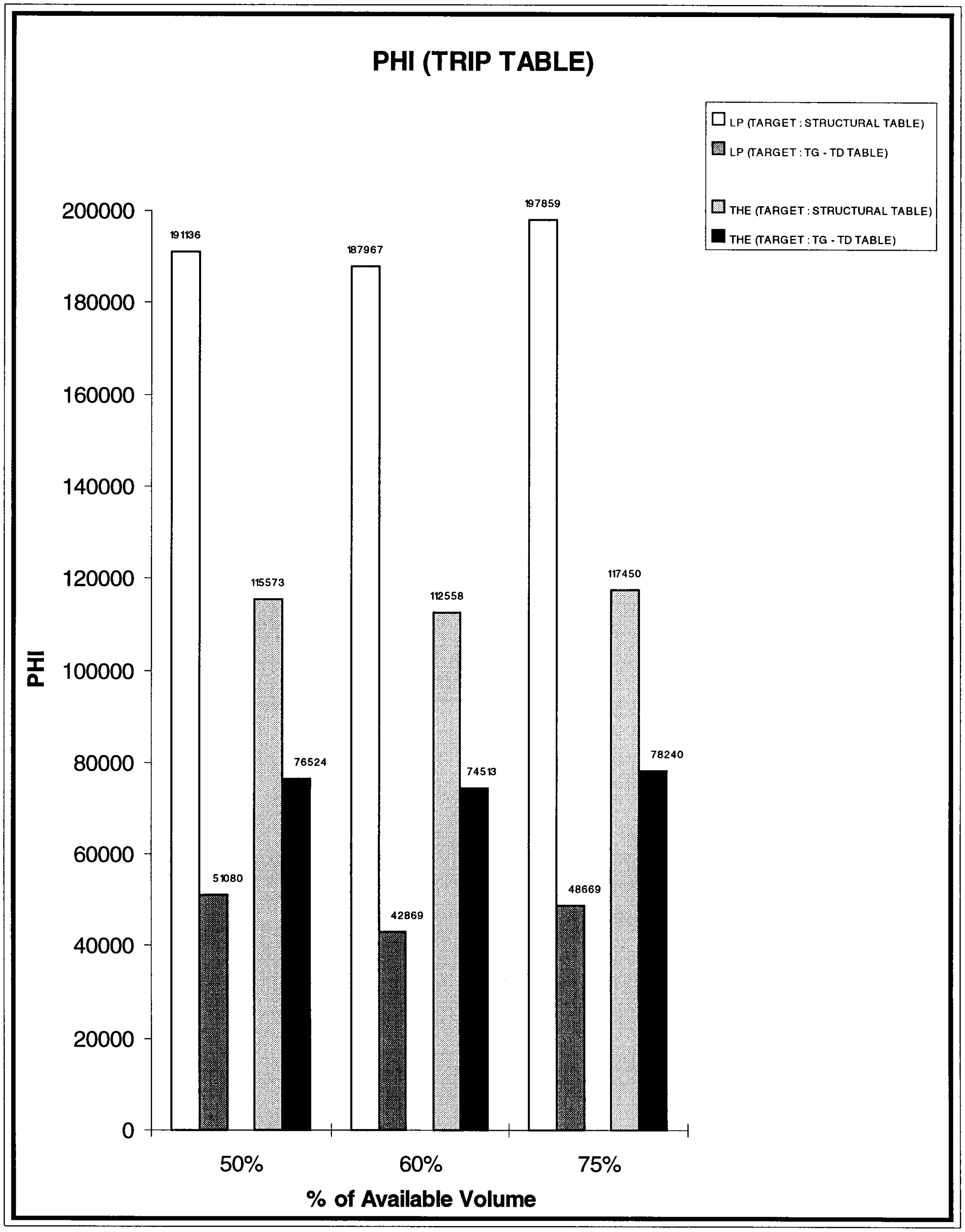


Figure 5 Trip Table Comparisons (Modeled vs Surveyed)
24-Hour Case (PHI)

- Applying the LP model with the socioeconomic based MINUTP-derived (referred to hereafter and in the figures as the TG-TD table) target table resulted in consistently improved performance in terms of the closeness of the output tables to the VDOT survey table when compared to the case using the structural target table. This was true for all three cases of available link volumes and for all the three measures of closeness. The decrease in %MAE was between 40 and 56, the decrease in %RMSE was between 125 and 236, and % decrease in PHI was between 73% and 77% for the various percentages of available link volumes input to the model. This represented a significant reduction in error rates.

- The THE model, on the other hand, showed improved performance only with respect to the PHI statistic, which decreased by 33% to 34% for the three percentages of volume availability. The increase in %MAE was between 2 and 17 and the decrease in %RMSE was between 26 and 68 for the various percentages of available link volumes.

- While the values of the statistics decreased significantly with the use of the TG-TD table for the LP model, the absolute values were still relatively high. For instance, the %MAE ranged from 140% to 145% and the %RMSE from 372% to 401% for the various percentages of volume availability. The absolute values of the statistics for THE model were likewise still relatively high. The %MAE ranged from 181% to 191% and the %RMSE from 458% to 482% for the various percentages of volume input.

- The LP modeled tables were significantly better than the THE modeled tables in terms of their closeness to the VDOT survey table when the TG-TD table was used as the target. This was true for all three cases of volume availability.

- The values of the closeness statistics showed a mixed trend with respect to variation in available link volumes for both the LP and THE models; however, the variations were small.

Replication of Observed Link Volumes

The performance of both the LP and THE models in replicating link volumes are depicted in detail for the %MAE and %RMSE in Figures 6 and 7, respectively. The following observations can be made:

- Applying the LP model with the TG-TD target table resulted in consistently poorer performance in terms of replicating measured link volumes when compared to using the structural target table. This was true for all three cases of available link volumes and for both measures of closeness. The increase in %MAE was between 0.07 and 3.02 and the increase in %RMSE was between 0.83 and 9.45 for the various percentages of available link volumes. (It should be noted that the high percentage increase was for the 60% volume availability for both statistics. The increase was much smaller for the 50% and 75% volume availability.)

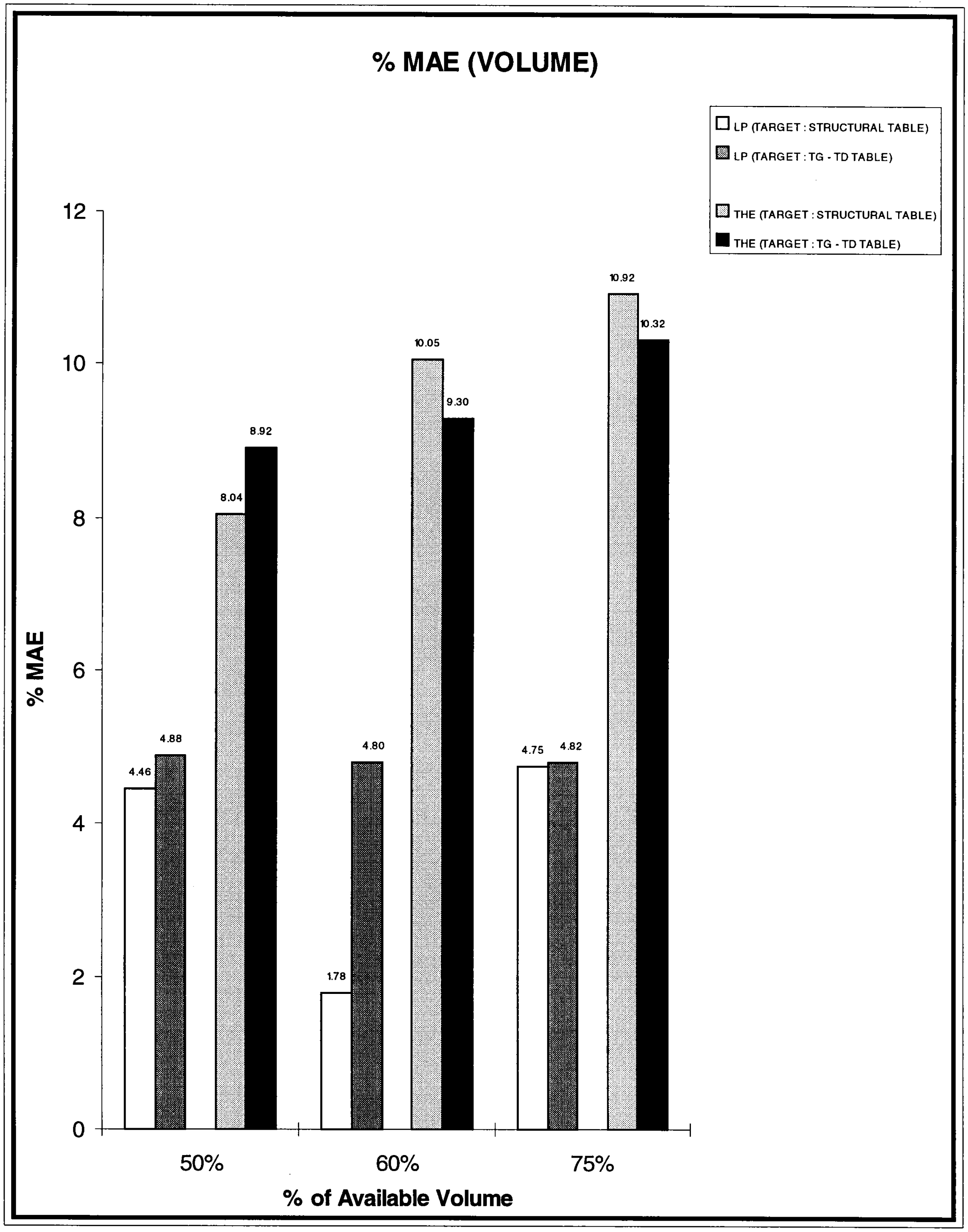


Figure 6 Volume Comparisons (Modeled vs Observed)
24-Hour Case (% MAE)

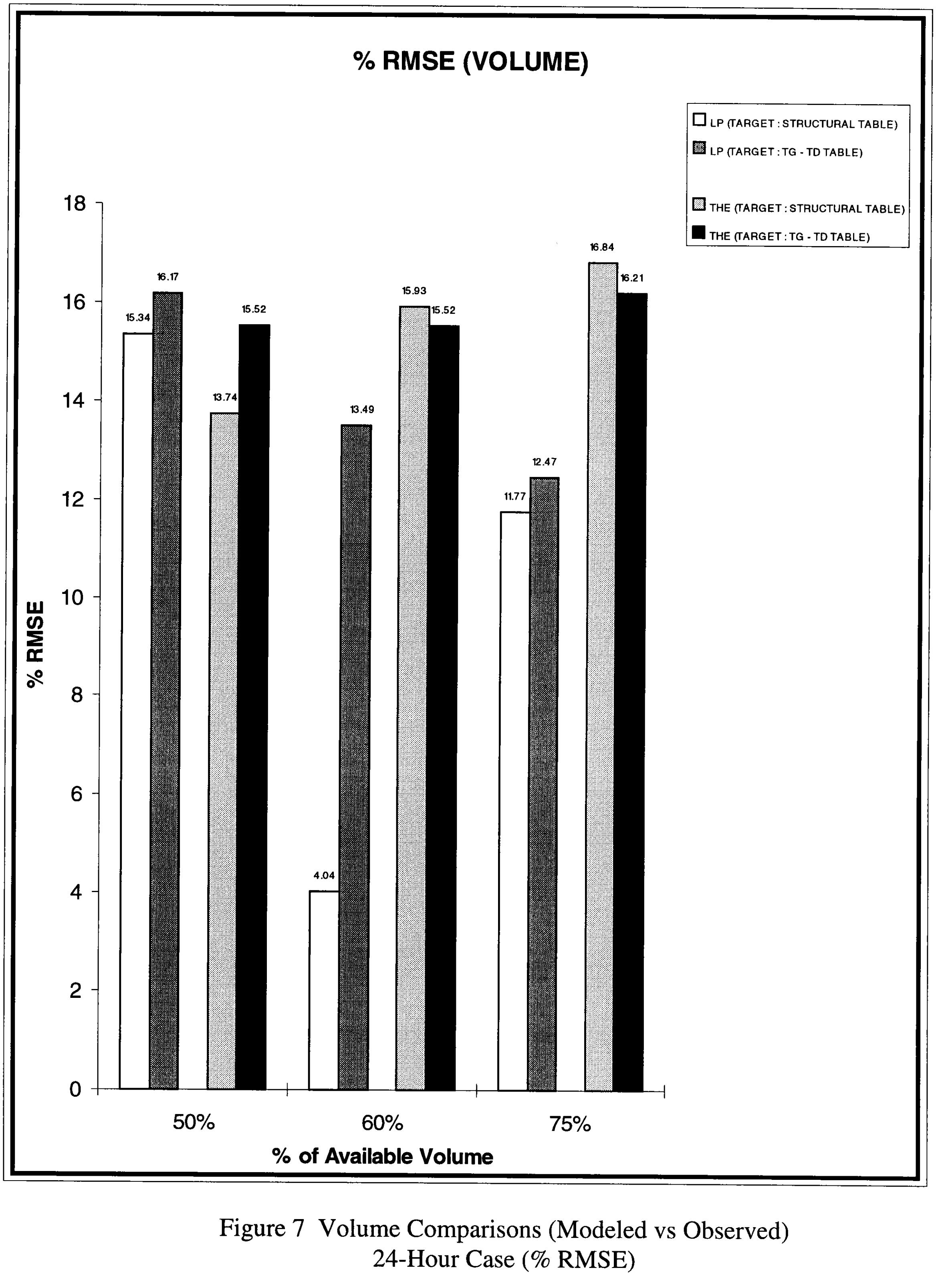


Figure 7 Volume Comparisons (Modeled vs Observed)
24-Hour Case (% RMSE)

- THE model, on the other hand, showed improved performance with 60% and 75% volume availability but poorer performance for the 50% volume availability. The decrease in %MAE was 0.60 and 0.75 and the decrease in %RMSE was 0.41 and 0.63 for the two higher volumes; however, the increase was 0.88 (%MAE) and 1.78 (%RMSE) when 50% of available volumes were input.

- Even though the values of the statistics increased for the LP model, their absolute values were still relatively low. The %MAE was around 5% while the %RMSE ranged from 12% to 16% for the various percentages of volume availability. The absolute values of the statistics were also relatively low for THE model. The %MAE ranged from 9% to 10% while the %RMSE was around 16% for the various percentages of volume availability.

- Except for the %RMSE at the 50% volume availability, the LP model's ability to replicate volumes was better than that of THE model.

Peak-Hour Trip Tables

The performance of the LP and THE models in replicating the surveyed trip tables are depicted in Figures 8 through 10 and in replicating link volumes in Figures 11 and 12. Many of the trends in results were, in general, similar to those described in the 24-hour trip table case.

Use of the TG-TD target table with the LP model resulted in improved performance with regard to closeness of the output table to the VDOT survey table as compared to use of the structural target table, though to a lesser degree than that for the 24-hour table. Also, as compared to the THE model, the error rates for the LP model were lower for all the cases of available link volumes. Thus, the LP model's output was again better than that of the THE model in terms of closeness to the VDOT survey table. In terms of replicating the measured link volumes, the THE model results were better than those of the LP model for two of the three cases of volume availability. For both models, the error rates were relatively small.

CONCLUSIONS

With regard to the primary purpose of the research of enhancing the target/seed table used as input to the LP and THE models in the case when a prior trip table is not available (rather than by using a structural table with cell values of "0" or "1"), the following specific conclusions were reached from the Pulaski case study. The conclusions are generally similar for both the 24-hour trip tables and the peak hour trip tables.

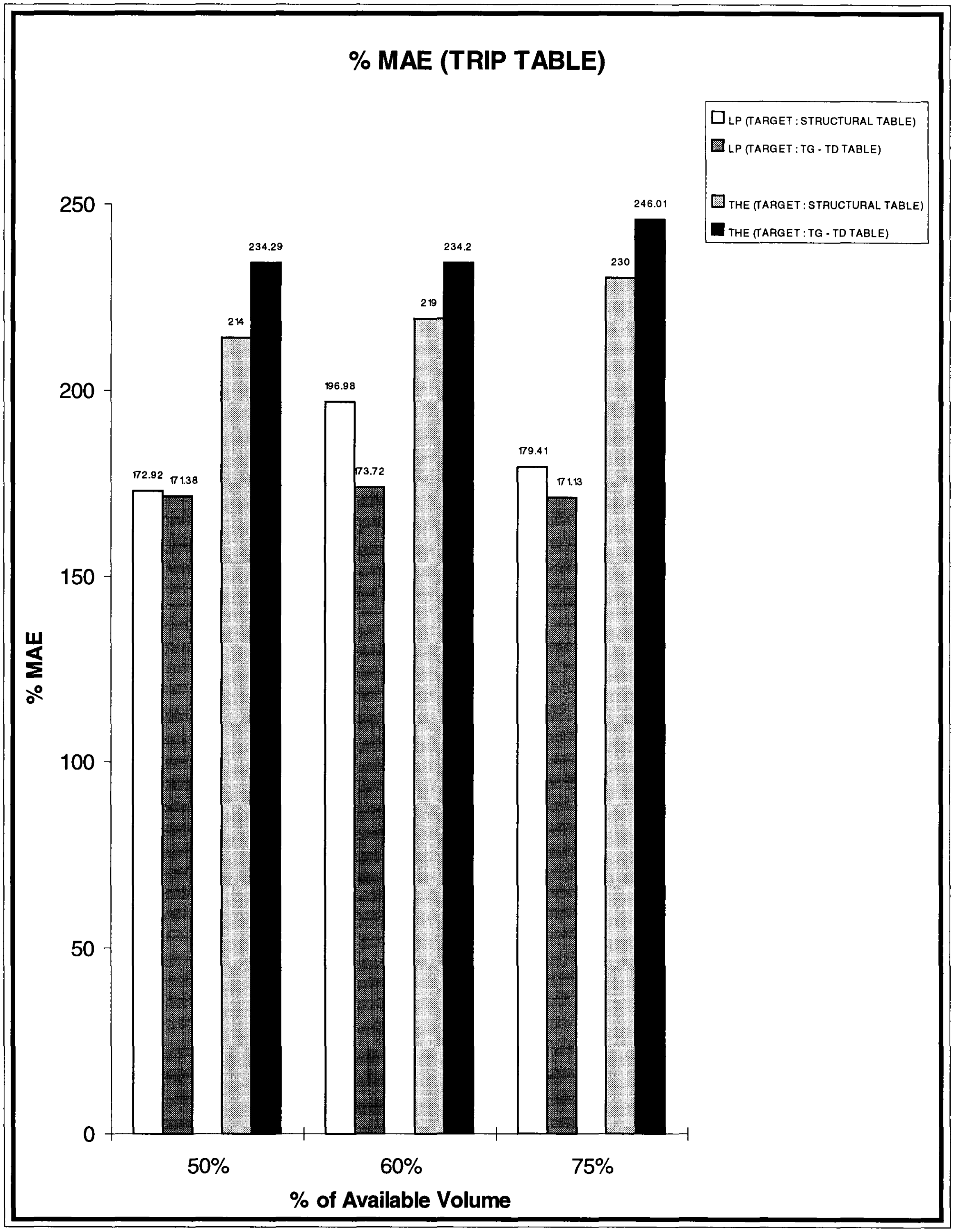


Figure 8 Trip Table Comparisons (Modeled vs Surveyed)
Peak-Hour Case (% MAE)

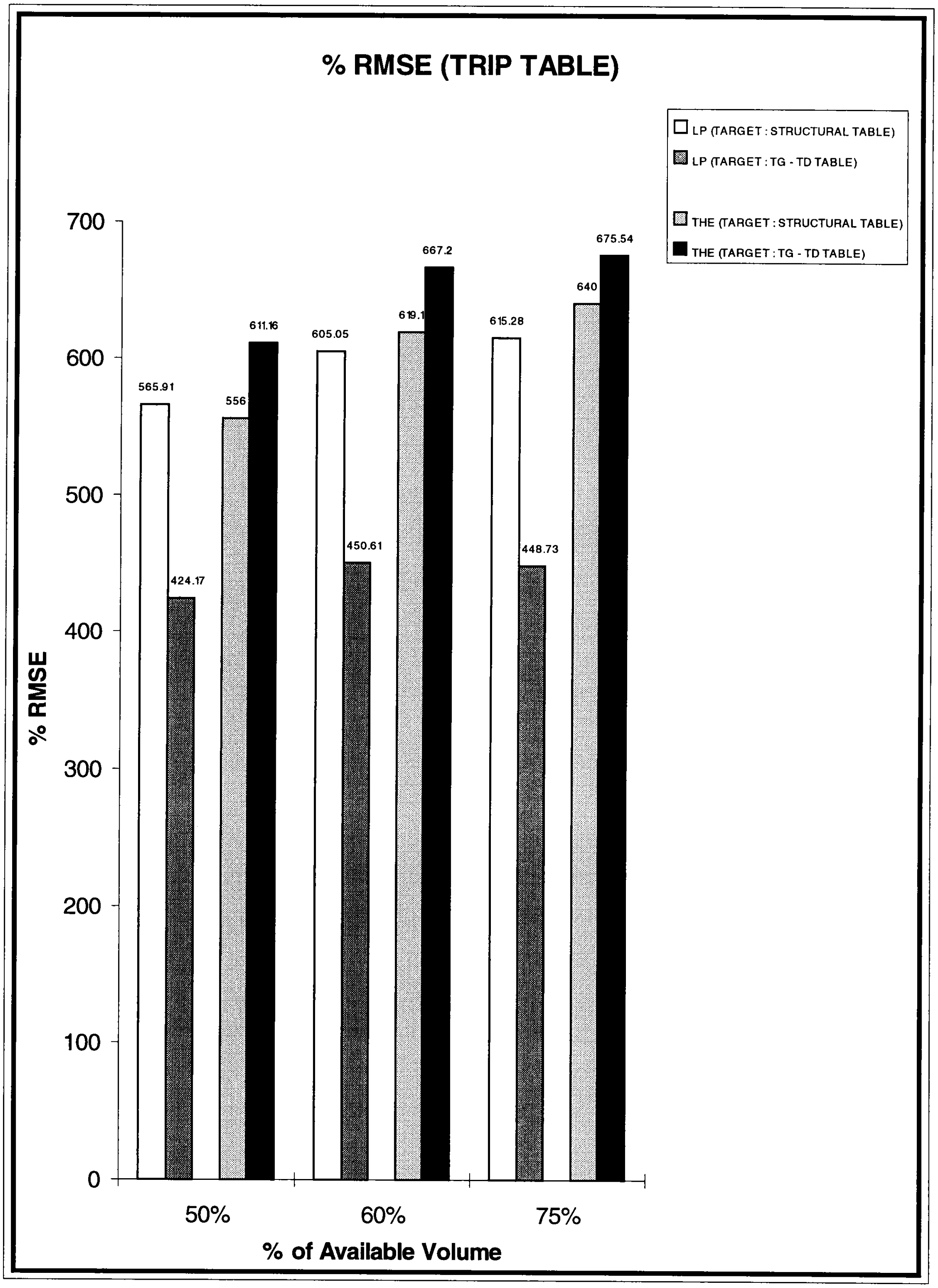


Figure 9 Trip Table Comparisons (Modeled vs Surveyed)
Peak-Hour Case (% RMSE)

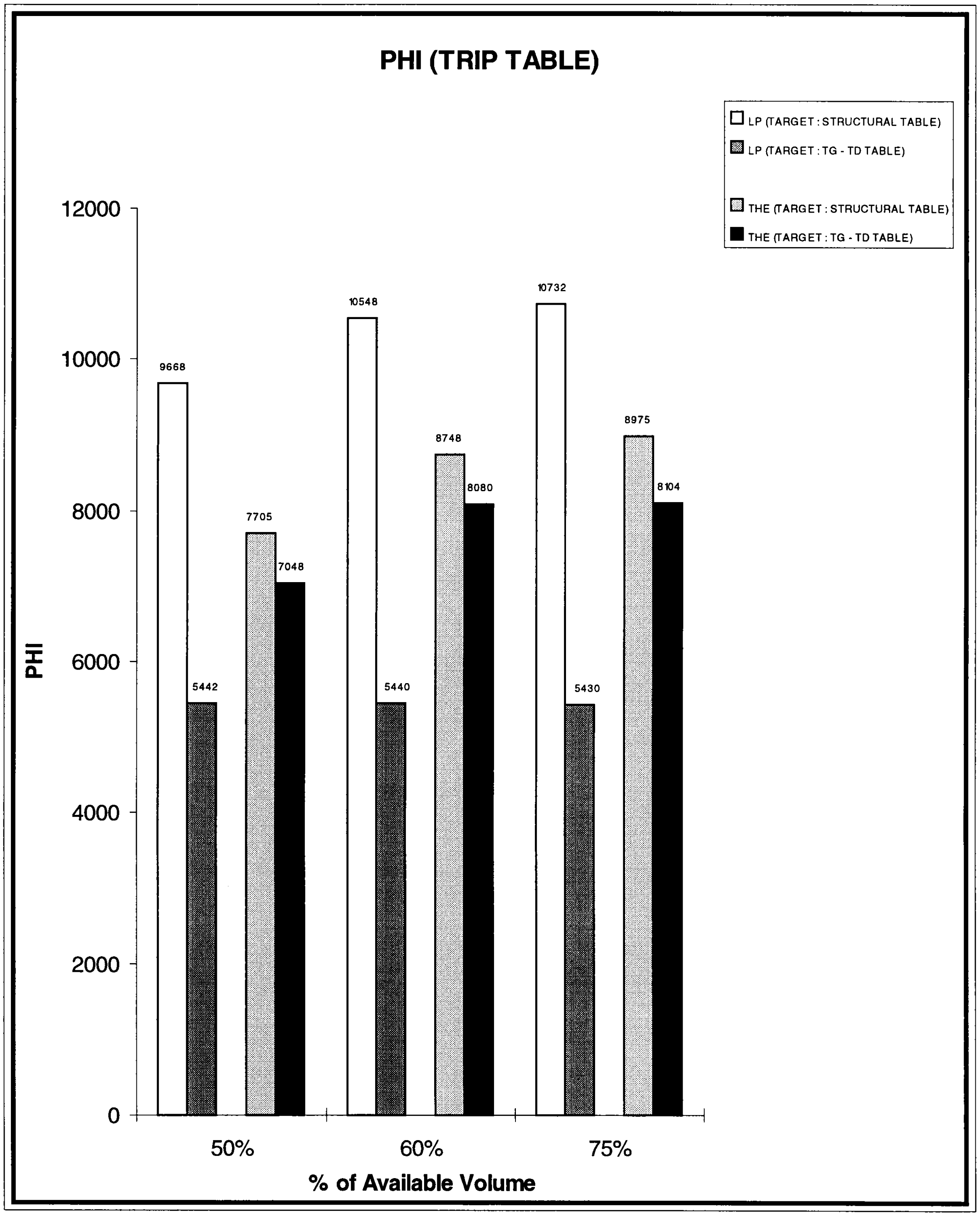


Figure 10 Trip Table Comparisons (Modeled vs Surveyed)
Peak-Hour Case (PHI)

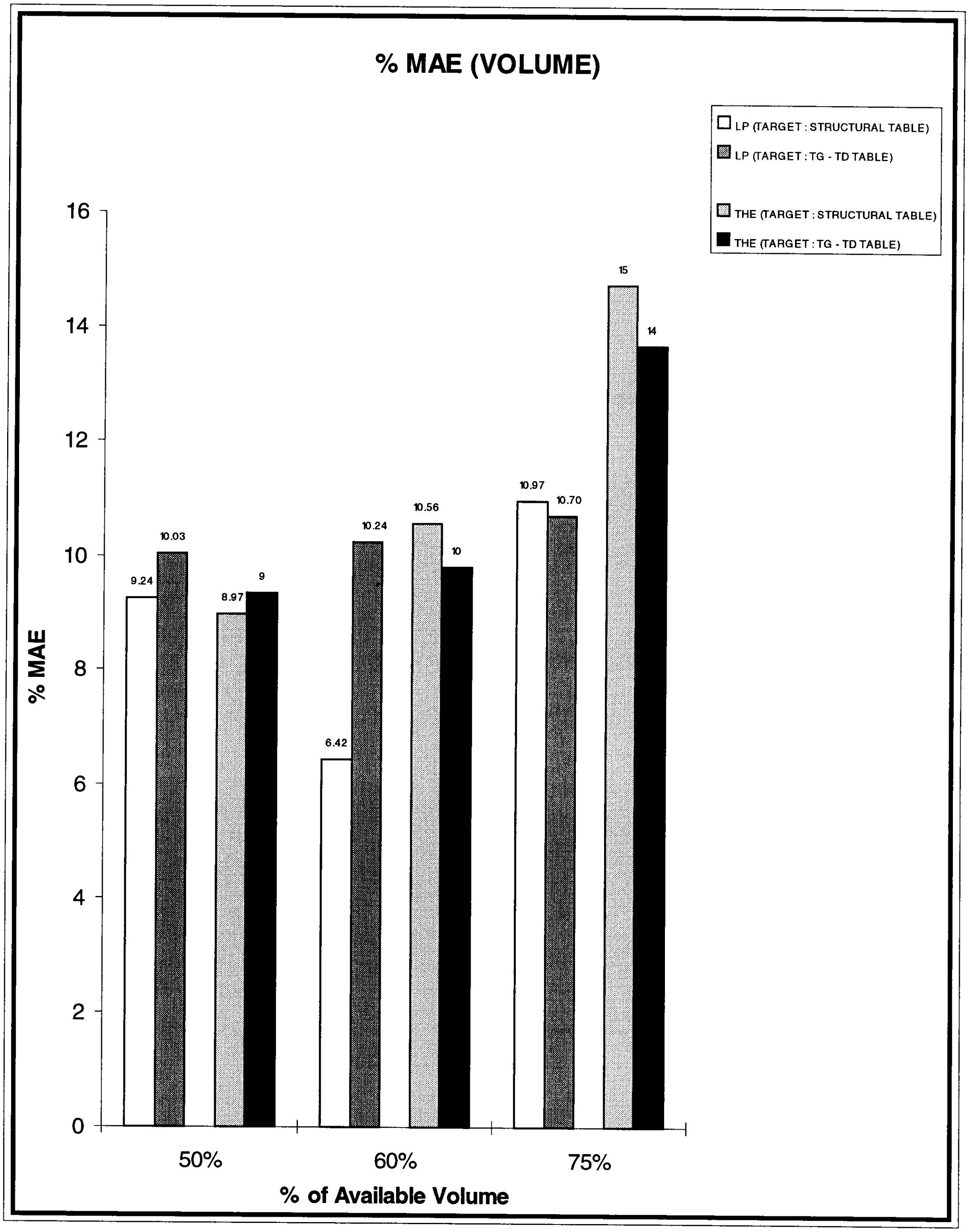


Figure 11 Volume Comparisons (Modeled vs Observed)
Peak-Hour Case (% MAE)

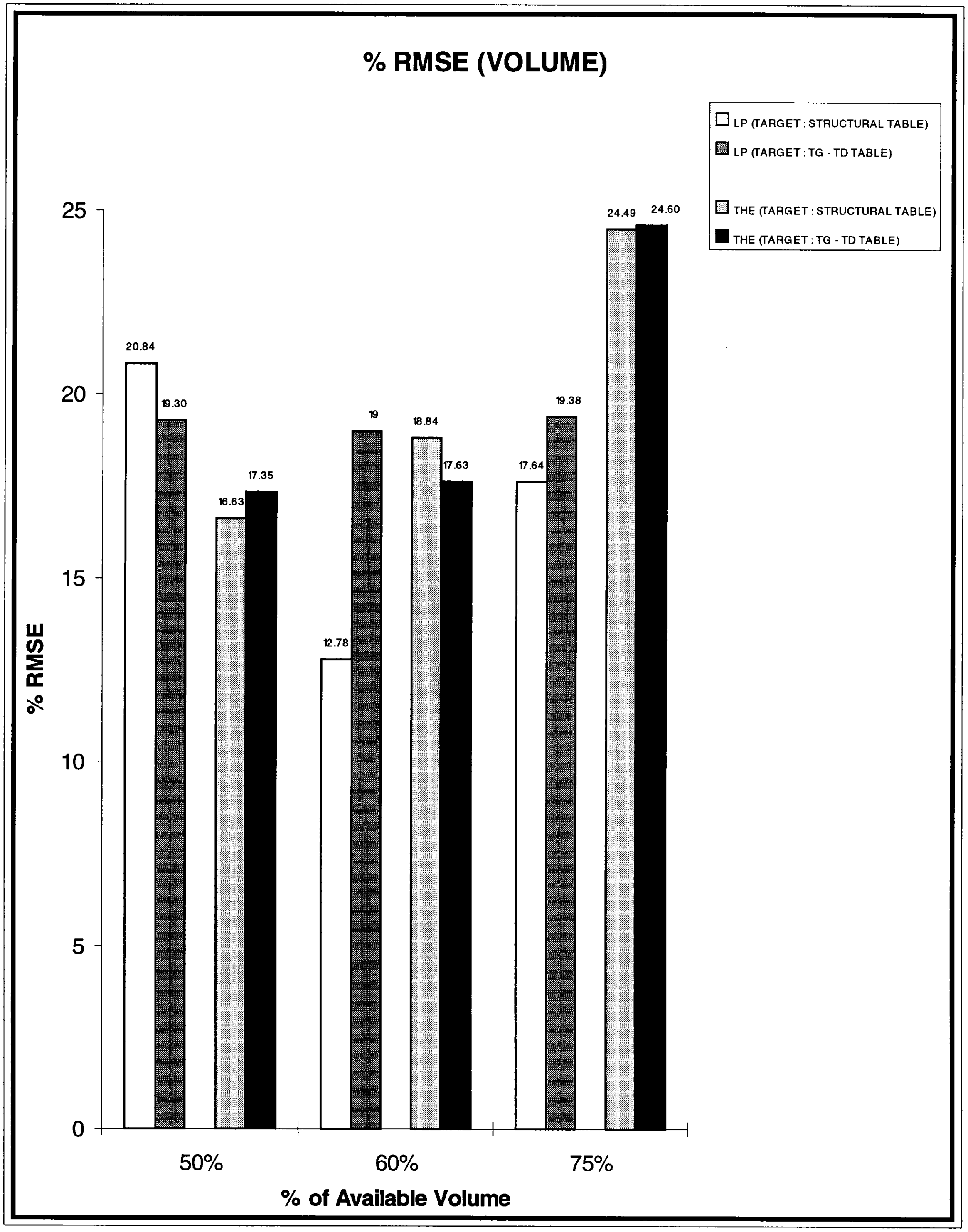


Figure 12 Volume Comparisons (Modeled vs Observed)
Peak Hour Case (% RMSE)

1. The use of an improved target/seed table developed by inputting local socioeconomic data to a traditional planning model (MINUTP in this study) has resulted in the following findings with regard to the quality of the output tables generated by the THE and LP models.

   - The performance of the LP model clearly improved. That is, the LP model generated a trip table that was closer to the survey table than was the table generated by the use of a structural table as the target/seed. While the values of the measures of closeness statistics decreased significantly, the absolute values were still relatively high.

   - The performance of the THE model worsened for two of the three evaluative statistics. That is, for these two statistics the THE model generated a trip table that was not as close to the survey table as was the table generated by the use of a structural table as the target/seed. The absolute values of the statistics were again still relatively high.

2. The use of an improved target/seed table developed by inputting local socioeconomic data to a traditional planning model (MINUTP in this study) has resulted in the following findings with regard to the replication of link volumes by the THE and LP models.

   - The performance of the LP model worsened marginally. That is, the trip table generated by the LP model did not replicate the observed volumes as well as the table generated by the use of a structural table as the target/seed. While the values of the statistics increased, the absolute values were relatively low.

   - The performance of the THE model improved marginally for two of the three percentages of available input volume. That is, the trip table generated by the THE model replicated observed link volumes better than the table generated by the use of a structural table as the target/seed. Again, the absolute values of the statistics were still relatively low.

With regard to the secondary purpose of the research of evaluating the absolute and relative performances of the THE and LP models by comparing their output with the trip tables derived from VDOT's Pulaski survey, the following specific conclusions were reached from the Pulaski case study. The conclusions are generally similar for both the 24-hour trip tables and the peak hour trip tables.

1. When using an enhanced target/seed table, the LP model produced better results than the THE model when comparing the closeness of the generated trip table to the survey trip table. However, the absolute values of the comparative statistics were relatively high for both the models. It should be noted that these findings are based on only the Pulaski case study and the assumption that the data used in the validation and comparison processes are in fact "correct". This suggests that the use of the LP and THE models to generate a trip table using the proposed methodology and link volumes for cases in which there is no prior trip table for use needs further validation.

2. When using an enhanced target/seed table, the LP model generally performed better than THE model in replicating input link volumes. Further, the absolute values of the comparative statistics (%MAE and %RMSE) were relatively low. However, the caveats and suggestion for further validation mentioned should be noted.

3. There was no clear evidence to indicate that either the LP or THE model performed better in terms of their ability to match VDOT survey tables as the percentage of input network link volumes increases.

RECOMMENDATIONS FOR FURTHER RESEARCH

Further tests and validation of the models and ways to establish even more superior target/seed tables can be potential areas of further research. The results were encouraging for the Pulaski case study. Further tests on more real networks will help confirm the findings presented in this report. Other ways of establishing target tables can also be tested in the context of the methodology presented in this research. For a credible validation, the model results must be compared with tables that are known to be correct or reasonably good. More time and resources are worth investing in continuing this research due to the potential benefits, in terms of money, time, and manpower that this approach can offer.

ACKNOWLEDGMENTS

This research project was made possible through the sponsorship and funding provided by the Virginia Transportation Research Council (VTRC). Support of the project and provision of data by Virginia Department of Transportation (VDOT) is also much appreciated. In particular, the authors would like to thank Mr. Nelson Newton and Mr. Herman Hollins of VDOT's Transportation Planning Division. Mr. John Hawley of the Town of Pulaski deserves a special mention here for assisting the project team in data collection. Finally, the authors would like to express sincere gratitude to Mr. Eugene Arnold of VTRC for serving as the technical monitor for this project and for his support throughout the project.

REFERENCES

Bromage, E. J. (1991). *The highway emulator model - release 4.0.* Central Transportation Planning Staff. Boston, Massachusetts.

Census Summary Tape Files (STF) 3A and 1B. (1990).

Center for Survey Research. (1994). *Town of Pulaski origin-destination mail survey.* Blacksburg: Virginia Tech.

COMSIS Corporation. (1991). *Technical reference manual for MINUTP.*

Institute of Transportation Engineers. (1991). *Trip generation.* 5th Edition and Update. Washington, DC: Author.

Narayanan, A. (1995) A *linear programming approach for synthesizing origin-destination (o-d) trip tables based on a partial set of link traffic volumes.* Unpublished Masters Thesis, Virginia Polytechnic Institute & State University, Blacksburg.

Sherali, H. D., Sivanandan, R., & Hobeika, A. G. (1994a). A linear programming approach for synthesizing origin destination (o-d) trip tables from link traffic volumes. *Transportation Research-B, 28B*, pp. 213-233.

Sherali, H. D., Sivanandan, R., Hobeika, A. G., & Narayanan, A. (1994b). Estimating missing link volumes in a traffic network - a linear programming approach. Presentation at the 1994 TRB Annual Meeting, Washington, DC.

Sivanandan, R. (1991). *A linear programming approach for synthesizing origin - destination (o-d) trip tables from link traffic volumes.* Unpublished Doctoral Dissertation, Virginia Polytechnic Institute & State University, Blacksburg.

Sivanandan, R., Narayanan, A. & Lei, Peng. (1996). *Review and evaluation of models that produce trip tables from ground counts* (VTRC Report No. 96-IR2). Charlottesville: Virginia Transportation Research Council.

Smith, D. P. & Hutchinson, B. G. (1981). Goodness of fit statistics for trip distribution models. *Transportation Research-A, 15A*, pp. 295-303.

Sosslau, A. (1984). *Model-related uses of census data for transportation planning* (Report No. TRR 981). Washington, DC: Transportation Research Board.

Transportation Research Board. (1978). *Quick-response urban travel estimation techniques and transferable parameters: User's guide.* (NCHRP Report 187). Washington, DC: Author.

U.S. Department of Transportation, Federal Highway Administration. (1995). *CTPP handbook (an instructional guide to the 1990 census transportation planning package and software.* Washington, DC: Author.

Van Zuylen, H. J., & Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic-counts. *Transportation Research-B, 14B*, pp. 291-293.

Willumsen, L. G. (1978). *Estimation of an o-d matrix from traffic counts - A Review* (Working Paper 99). England: University of Leeds, Institute of Transportation Studies.

APPENDIX A

DESCRIPTION OF THE HIGHWAY EMULATOR (THE) AND THE LINEAR PROGRAMMING (LP) MODELS

THE HIGHWAY EMULATOR (THE) MODEL

THE model (Bromage, 1991) is a microcomputer highway traffic simulation model for modeling of individual communities, corridors, and sections of counties, as well as analysis of small sections of major cities. Two distinct modeling approaches are incorporated in THE model. The first approach utilizes the traditional four-step urban transportation planning methodology. The second approach utilizes the maximum entropy algorithm for estimating trip tables from link traffic volumes. It extracts a most likely trip table that will produce observed traffic counts. The trip table estimation program is based on the maximum entropy formulation and algorithm detailed by Van Zuylen and Willumsen (1980). A maximum entropy algorithm is one that attempts to define a trip table with the maximum degree of disorder or random exchange between zones.

Trip Table Estimation Procedure

The program first assigns a seed trip table to the specified network. While the assignment is in progress, the program creates a link use probability file. These probabilities are used to make adjustments to the trip exchange matrix in order to duplicate the observed volumes. The next step is the actual application of the trip table estimator algorithm. The algorithm is iterative and may take up to one hour for each iteration, for large networks. The output of the trip table estimation program is a new trip table which closely duplicates the observed traffic counts for the given probabilities. The above steps are repeated to satisfy the number of calibration iterations. For microcomputer applications, with respect to handling large networks, the trip table estimation program can handle a maximum of 300 traffic zones and up to 500 links (Bromage, 1991). Through a personal communication with Mr. Bromage, it is understood that this program can handle greater number of links.

Data Requirements

THE model requires the following data as input for running the trip table estimation program:

- Node numbers at each end
- Link length in miles
- Free flow speed on the link
- Link delays (if any)
- Whether link is one-way or two-way

- Any additional impedance (such as a zone connector terminal time)
- Hourly capacity for each direction
- Link volumes (preferably for all links, but partial counts also accepted)
- Prior (target) trip table, if desired to guide the solution (a structural table could also be used)

The coding of the network is based on traditional method, and is also outlined in NCHRP Report # 187 (1978). It is also consistent with FHWA's UTPS coding scheme. It must be noted that it cannot accept a node with more than four connections. In such cases, dummy links are needed. However, turn links can be coded. In terms of user parameters to be supplied, THE requires the specification of number of iterations for maximum entropy process, assignment, and calibration. However, default values for these can also be used.

THE LINEAR PROGRAMMING (LP) MODEL

This new approach to estimate trip tables from ground counts has been developed at Virginia Tech (Sherali et al., 1994a,b). The model employs the non-proportional assignment assumption, and finds a user equilibrium solution that reproduces the observed link flows whenever such a solution exists. The model recognizes that due to incomplete information, although the individual user is driven by the choice of a least impedance path, the actual flow may not exactly conform to a user equilibrium solution. Moreover, due to inherent inconsistencies in the link traffic data, there might not exist a trip table that can exactly duplicate the link flows. Accordingly, these features are accommodated into the model through suitable artificial variables and objective penalties. However, if there exists a user equilibrium solution that reproduces the link flows, the model, with suitable penalty parameters, will determine such a solution along with the corresponding O-D trip table. Additionally, due to the potentially large number of alternative paths to be considered between the different O-D pairs, an efficient column generation technique that utilizes shortest-path sub-problems in order to determine an optimal solution to the linear programming model has been developed. The model is also designed to handle the situation in which a prior target trip table is specified, and it is required to find a solution that, in addition to the foregoing considerations, has a tendency toward reproducing this table as closely as possible.

This model has been programmed in FORTRAN for computer runs. One of the weaknesses of the original linear programming model was that it required the specification of volume data on all the links of the network. This data, however, is hard to obtain for real networks. Realizing the constraint, the model was enhanced to accommodate missing volume data, and to estimate O-D tables even when only a partial set of link traffic counts is available. This enhancement was also programmed for computer applications (Sherali et.al., 1994; Narayanan, 1995).

The notation "LP" is used in this report to refer to the linear programming model, in general, and not to the specific formulation named "LP" by Sivanandan (1991) and Sherali et al. (1994a). Several versions and different formulations of the model evolved over the course of the development and enhancement of the approach. Thus, more than a single version/formulation has been used in this research.

Data Requirements

The linear programming model (LP) requires the following data input for running the trip table estimation program:

- Zones, nodes, and link numbers (traditional network coding method)
- Hourly capacities, free-flow speeds, and lengths of links (alternatively, if current travel times on links are known, they will suffice)
- Whether link is one-way or two-way
- Link volumes (preferably for all links, but partial counts also accepted)
- Prior (target) trip table data (optional) (a structural target could also be used)
- Any link delays and additional impedance, if appropriate, may be included in the current link travel times.

For this model, the user also needs to input a value for a parameter to reflect the relative degree of importance in minimizing the trip table deviations (modeled vs. targeted) versus the link flow deviations (modeled vs. observed). These parameter values for the tests were chosen based on judgment.

COMPUTER REQUIREMENTS

THE runs satisfactorily on an IBM-compatible PC. Due to the higher computational demands of the LP model, more powerful machines are beneficial. Accordingly, the SUN/SPARC server 1000 and 20 machines were used for the LP runs for the Pulaski network.

APPENDIX B

DESCRIPTION OF THE PULASKI O-D SURVEY

PULASKI NETWORK

Located in the central area of Pulaski County, Southwestern Virginia, the town of Pulaski had a population of around 10,000 in 1990. The network, as defined by VDOT, consists of 21 internal zones and 11 external stations. These internal zones have been divided according to the density of the population and the activity centers in and around the area, in the original network provided by VDOT. This network was reduced by the Center for Transportation Research by eliminating redundancies and other information that was not necessary for test purposes. The test network as used in this study had 32 zones, 57 intersection nodes, and 230 links (Figure B-1). Data on network characteristics, such as link lengths, free flow speeds, and capacities were also provided by VDOT. In order to validate the O-D models with real data, VDOT, with the help of Virginia Tech's Center for Survey Research (CSR) conducted an O-D survey, and established a trip table, for an earlier study conducted by the Virginia Tech Center for Transportation Research. The details of this survey are also included in this chapter.

NETWORK VOLUME DATA COLLECTION

In order to facilitate volume input for the O-D models, VDOT surveyed 24-hour traffic volumes on 175 out of the 230 links of the Pulaski network (defined for this study) using counters, in 1994. These volumes were collected for 15-minute intervals, which also helped in determining peak-hour volumes. Since VDOT was interested in estimating both the daily (24-hour) and the peak-hour trip tables, the corresponding volume data were used in the models. There were some variations/inconsistencies in volume measurements for some of the stations, hence some cleaning-up of the data was undertaken before inputting the link volumes. The period between 9:00 AM on June 14, 1994 and 9:00 AM on June 15, 1994 was chosen as the 24-hour data for consideration. Since complete data for the above period was not available for some stations, data collected during a second measurement were used for the missing time periods. Since not all the links of the network had the peak flow during the same hour, the peak hour was chosen as 3:30 - 4:30 PM, based on the occurrence of peak flow on the majority of the links.

It must be noted that volume data was not available for 55 links. These links were mostly centroid connectors. Since these connectors are abstractions of several minor streets on the field, single volume measurements cannot be performed for these. Hence, in several modeling applications, centroid connector volumes are generally unknown. In addition, there were some links where construction activities hampered data collection. Thus, observed volumes were available for only 75% (approximately) of the links. This may be considered reasonable and realistic, since similar situations could be expected for such real-life applications. This also

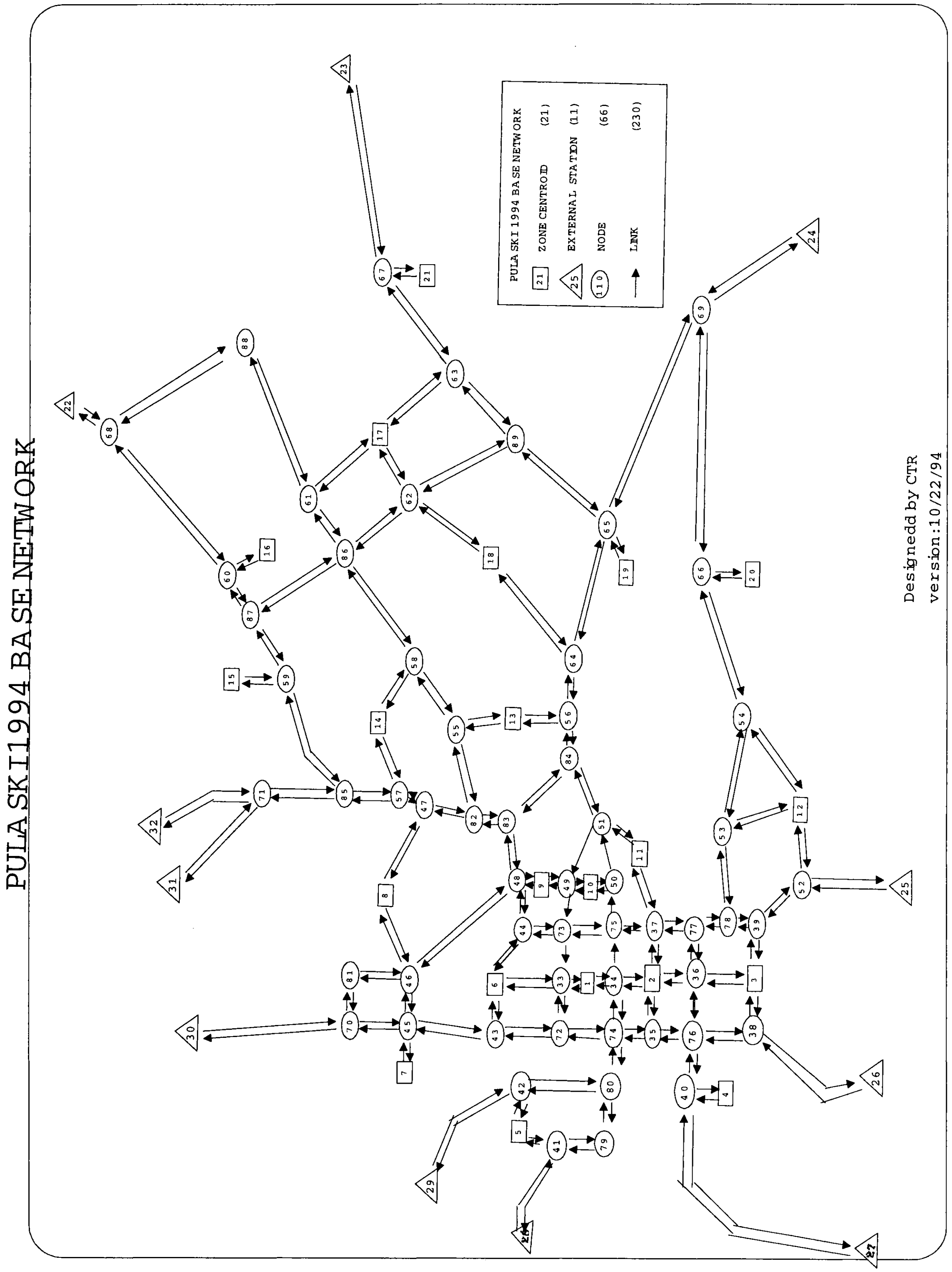


Figure B-1 Pulaski Base Network (Source: VDOT) created an opportunity to evaluate the performance of the models in the case of 25% missing volumes. For the present study, these link volumes have been used directly for the model runs.

APPENDIX C

DESCRIPTION OF THE CENSUS TRANSPORTATION PLANNING PACKAGE (CTPP)

THE USE OF CENSUS DATA

There are several elements of data contained in the census that have relevance to urban transportation planning. Following are some of the uses of census data in the application of urban transportation planning models (Sosslau, 1984; ITE 1991):

- Current socioeconomic data, such as population, dwelling units, and income, can be used as input to determine current trip generation with existing models.
- Census data can serve as a benchmark for checking updated long-and short-range land use and socioeconomic data.
- Journey-to-Work (JTW) census questions can yield responses that can be used as a secondary source for checking the validity of trip length frequency distributions, trip ends, and work trip tables.
- The data can be used for the calibration and development of urban transportation planning models.

A good source of census data for transportation planning applications is the 1990 Census Transportation Planning Package (CTPP) (FHWA, 1995). The CTPP is a collection of summary tables that contain information about population and household characteristics, worker characteristics, and characteristics of the Journey-to-Work (JTW). These tables have been designed specifically for transportation planning analyses.

The CTPP is organized into the following two elements: Statewide and Urban. The Statewide Element of the CTPP consists of data summaries for all places of 2,500 or more population, the balance of the county, the county as a whole, and entirety of the state. The Urban Element of CTPP consists of data summaries for urbanized areas with a population of 50,000 or more. This data is available at the level of Traffic Analysis Zones (TAZ) or the Census tract. Data is also included for census-defined urbanized areas, the MPO defined study area, and the Metropolitan Statistical Area in which the MPO is located.

The CTPP can be used in the Urban Transportation Planning Process (UTPP) by utilizing census data to estimate new or calibrate pre-existing trip generation models based on socioeconomic data and travel characteristics stratified by TAZ. Once available at the TAZ level, trip generation and trip distribution models can use the data to calculate trip tables.

EXAMINATION OF CTPP FOR USE IN THE PULASKI CASE STUDY

The use of CTPP data for the town of Pulaski was examined for this research. Since it is a small town, the CTPP data summaries fall under the Statewide Element. The available data is shown aggregated for the whole town, and is not amenable to the research approach initially intended for this study. In other words, the socioeconomic data required at the TAZ level is not available for Pulaski in the CTPP's database. Hence, even though the CTPP was the preferred choice for accomplishing the objective of establishing a target/seed table in this research approach, it could not be used for the Pulaski case study. Hence, investigation of other potential data sources was undertaken.

APPENDIX D

USE OF MINUTP SOFTWARE FOR TRIP GENERATION AND TRIP DISTRIBUTION APPLICATIONS

The four-step Urban Transportation Planning Process (UTPP) can be conveniently carried out on a computer. The interest in this research was only on the first two steps of the planning process, namely, trip generation and trip distribution. MINUTP is a library of programs that provides the capability to perform the usual functions of traditional transportation planning. Inter-program chaining is provided by a batch file (COMSIS, 1991).

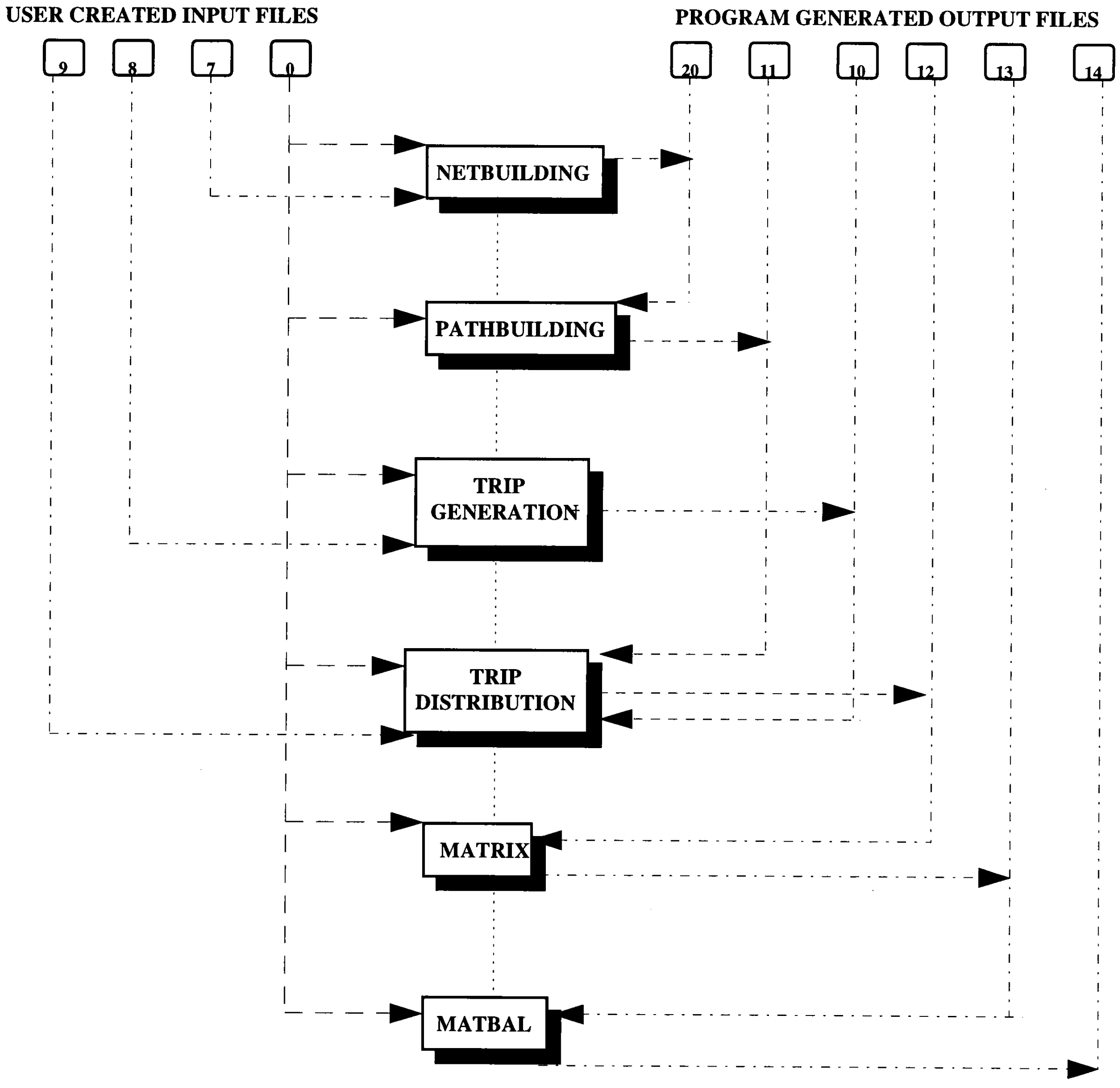
The normal application of MINUTP requires the link data, zonal data, and friction factor data records. These data can be stored in accessible files for MINUTP runs. These data for Pulaski case study were provided by VDOT. A job control stream is created and submitted to the computer for batch execution; it begins with building the network and ends by predicting the link volumes.

Figure D-1 provides an overview of MINUTP program elements and their flow as used in this research application. The input and the output files for each of the individual program elements have also been shown. The chart shows the flow of data and the execution of program elements in MINUTP. Only two steps of the four-step planning process were required to develop the target/seed table. A brief description of the various MINUTP elements adopted in this research is presented below.

The Network Building (NETBLD) module performed in MINUTP is the first step in running MINUTP. NETBLD reads link data records describing the street sections of the transport network and forms an internal network. The Pulaski study area has been divided into 32 zones, of which 21 are internal and 11 are external stations. There are 230 directional links in the study network. The necessary data for NETBLD module was provided by VDOT.

The Pathbuilding (PTHBLD) module determines the minimum paths between all selected zone pairs. This was the second module that was run after NETBLD. The values of impedance, distance, and time can be stored in three different matrices. Basically, the PTHBLD module makes use of the network data (from the NETBLD module) and applies the traditional Moore's algorithm for determining the minimum paths. The PTHBLD module finally gives the minimum time or impedance matrix (from each origin to all the destination zones). The impedance matrix file generated by this module is used as input for the trip distribution step.

The Trip Generation (TRPGEN) module estimates the total number of trip ends for different zones in the region. For the case study, trip productions and attractions were input directly into MINUTP. Three purposes, namely, Home Based Work (HBW), Home Based Other (HBO), and Non Home Based (NHB) were considered. The percentage split of trips for these purposes were specified based on judgment and through consultation with VDOT. Attractions were factored to match productions for two purposes-HBW and HBO. For the third purpose, NHB, productions were factored to match attractions. The final results are trip productions and



0 - SYSTEM CONTROL
7 - LINK DATA FILE
8 - ZONAL DATA FILE
9 - FRICTION FACTOR FILE
10 - PRODUCTION/ATTRACTION TRIP ENDS FILE
11 - IMPEDANCE MATRIX FILE
12 - PRODUCTION/ATTRACTION TRIP MATRICES
13 - PRODUCTION/ATTRACTION TRIP MATRICES
14 - ORIGIN-DESTINATION TRIP MATRICES

Figure D-1 Program Elements of MINUTP Used in the Research
(Source: Based on Normal System Flow Diagram (COMSIS 1991))

attractions for each zone in the study area, stratified by the three purposes. The TRPGEN module creates the output file containing production/attraction trip ends for each zone. This file was saved in unit 10 for further use as input for the trip distribution (TRPDST) module. The distribution of trips was carried out using the gravity model.

The Trip Distribution (TRPDST) module of MINUTP reads the zonal productions and attractions trip ends file, and distributes trips to all the zones using the standard gravity model. The TRPDST module also requires the impedance matrix (from PTHBLD) stored in unit 11 and the friction factors stored in unit 09 for carrying out the distribution. The output from the TRPDST module are the production/attraction trip matrices (stored in unit 12) for each of the purposes.

The MATRIX module combines the production/attraction trip ends for the three purposes (output from TRPDST) into one production/attraction trip matrix. The output from the MATRIX module gets saved in unit 13. This output is used as input to MATBAL, to obtain an origin-destination trip matrix. This module converts the production/attraction trip table to an origin-destination trip table, based on the balancing factors for rows and columns specified by the user. The output from MATBAL is the O-D trip table of interest. This table is then used as the seed/target for the synthetic O-D models.

APPENDIX E

APPLICATION OF MINUTP FOR PULASKI CASE STUDY

As noted previously, socioeconomic variables at the TAZ level are needed primarily for use in trip generation analysis. Based on VDOT's recommendation, MINUTP software was selected as the tool for the trip generation and trip distribution applications. The 1991 Version of MINUTP (Comsis, 1991), already available at the Center for Transportation Research, was used. (See Appendix D for a general discussion on MINUTP.)

For trip generation analysis, the Institute of Transportation Engineers (ITE) trip generation equations/rates (ITE, 1991) were used for determining the vehicle trip ends (both productions and attractions). These trip generation equations/rates require information about the values of the independent variables and, in this case, their values at the TAZ level. After much consideration, and based on the availability of data, it was decided to use dwelling units as the independent variable for trip production estimates. The dwelling unit data for each TAZ was supplied by VDOT. Since this data represented the total number of dwelling units, it was necessary to account for unoccupied units. This was accomplished through use of information on unoccupied dwelling units in the Census Summary Tape Files 3A and 1B (Census Summary Tape Files (STF), 1990). ITE land use code 210 was used for generating production trip ends. All the dwelling units were assumed as single family, detached housing since the data on dwelling units by various types at the TAZ level was not readily available. Since the majority of the dwelling units were in fact single family, detached housing, the assumption was deemed reasonable. ITE guidelines on using the equations versus the rates for estimating vehicle trip end productions were followed. Production estimates were carried out for both 24-hour and peak hour cases.

Obtaining data for estimating trip attractions was a more challenging task. Several agencies and sources were contacted to obtain data suitable for this application; that is, the identification of individual trip attraction activities, their locations with respect to the TAZs, and the values of the independent variables needed for application of the ITE trip generation equations/rates. This process, though laborious and time-consuming, was made possible by the cooperation and assistance of officials in the Town of Pulaski and the Pulaski Chamber of Commerce. In some cases, the needed data was directly solicited from the owners of the attraction activities. Appropriate ITE trip generation equations/rates were then chosen for estimating attractions at the TAZ level for both the 24-hour and peak hour cases.

While the above procedure was adopted for internal zones, external stations were treated differently. The traffic count data provided by VDOT as part of the network-wide link volume information was directly used as either productions or attractions.

The output of trip generation analysis was then used as input to MINUTP's use of the gravity model for the trip distribution step. This was a fairly straight forward application, except for the treatment of external stations. Since the required data to analyze the external stations as external zones was unavailable, they were treated as internal zones. Based on VDOT's recommendation, the friction/travel time factors for Lynchburg, Virginia, were used. The socioeconomic adjustment factors (K-factors) were assumed to be 1.0.

It should be noted here that the basic philosophy in establishing the target/seed trip tables is only to guide the synthetic O-D models' solutions. The link volume information is the primary data that these models use to develop current O-D tables. Thus, the motivation in establishing the target/seed trip tables was to start with a table that is superior to the structural table, and which will represent realistic travel patterns to a reasonable extent. Hence, some simplifications, assumptions, and approximations in the trip generation and trip distribution steps were considered acceptable.