Accuracy Evaluation Tests for Assignment Models of Large Traffic Networks

by **Matthew L. James**

This study was undertaken to determine for the large Greater Melbourne road network the appropriate traffic network assignment model from those available in either the United States Department of Transportation PLANPAC[1] or Urban Transportation Planning System[2] (UTPS) computer packages. The assignment types may vary by their dispersion of trips to alternative paths and by the loading of trips onto the selected paths. The traffic dispersion techniques used were the all-or-nothing or multipath types, while the loading algorithms were incremental capacity restraint or equilibrium loading.

The selection of the most appropriate assignment may be based on comparisons between observed count volumes and modeled prediction volumes, as well as measurement of the degree of convergence of the model. Volume comparisons may be based on algebraic formulations known as nonparametric and parametric tests and outlined in this article. Assignment convergence may be assessed using criteria tests that measure the ability of assignments to optimize distribution of traffic on a network. The following section outlines the tests used, commencing with nonparametric and followed by parametric and criteria tests.

Assignment Performance Tests

The nonparametric tests described here do not require conditions of the traffic volume distributions, but assume that

1. NON-PARAMETRIC TESTS

$$MR = \left[\sum_{i=1}^{N} (A_i/O_i)\right]/N. \qquad MAR = \left[\sum_{i=1}^{N} \left|1-A_i/O_i\right|\right]/N.$$

$$CC = \left[N\sum_{i=1}^{N} O_iA_i - \sum_{i=1}^{N} O_i \sum_{i=1}^{N} A_i\right] / \left\{\left[N\sum_{i=1}^{N} O_i^2 - \left(\sum_{i=1}^{N} O_i\right)^2\right]\left[N\sum_{i=1}^{N} A_i^2 - \left(\sum_{i=1}^{N} A_i\right)^2\right]\right\}^{\frac{1}{2}}.$$

$$MD = \left[\sum_{i=1}^{N} (O_i-A_i)\right]/N. \qquad SD = \left\{\left[\sum_{i=1}^{N} (O_i-A_i)^2 - \left[\sum_{i=1}^{N} (O_i-A_i)\right]^2/N\right]/(N-1)\right\}^{\frac{1}{2}}.$$

$$MAD = \left[\sum_{i=1}^{N} \left|O_i-A_i\right|\right]/N. \qquad MAE = \left[\sum_{i=1}^{N} \left|O_i-A_i\right|\right]/N^2.$$

$$RMSE = \left[\sum_{i=1}^{N} (O_i-A_i)^2/N\right]^{\frac{1}{2}}.$$

$$SADI = \left[\sum_{o=1}^{O}\sum_{d=1}^{D}\sum_{i=1}^{N} \left|O_{iod}-A_{od}\right|\right] / \left[\sum_{o=1}^{O}\sum_{d=1}^{D}\sum_{i=1}^{N} \left|O_{iod}+A_{iod}\right|\right].$$

$$SAD = \sum_{i=1}^{N} \left|O_i-A_i\right| / \sum_{i=1}^{N} \left|O_i+A_i\right|$$

$$TIC = \left[\sum_{i=1}^{N} (O_i-A_i)^2\right]/N \left\{\left[\left(\sum_{i=1}^{N} A_i^2\right)/N\right]^{\frac{1}{2}} + \left[\left(\sum_{i=1}^{N} O_i^2\right)/N\right]^{\frac{1}{2}}\right\}.$$

2. PARAMETRIC TESTS

$$LR = \prod_{i=1}^{N} (A_i/O_i)^{O_i} \text{ and } \exp(\ln LR) = \exp\left[\sum_{i=1}^{N} (\ln A_i - \ln O_i)O_i\right].$$

$$IG = \sum_{i=1}^{N} \left[O_i \ln(O_iT_A/A_iT_O)/T_O\right] = \sum_{i=1}^{N} O_i\left|\ln(O_i/T_O) - \ln(A_i/T_A)\right|/T_O].$$

Figure 1. Test formulae.

count and predicted volume observations are independent. Besides volumes, the tests could also be used to compare vehicle travel times or vehicle kilometers of travel.

Figure 1 presents the formulae for all of the examined tests, which are mostly formulations of the average difference or ratio of assigned and observed volumes, as derived by various researchers. A common test is the mean ratio of assigned to observed volumes (MR), which tends to a value of unity when the volumes equalize each other.[3–5] If assigned volumes are both higher and lower than count volumes, a misleading perfect fit may be calculated for MR. Consequently, the mean absolute ratio (MAR) should be used, which indicates best agreement between volumes with a zero value. Similarly with a standard statistic, the mean difference (MD), a value of zero may be misleading, so the mean absolute difference (MAD) should be used, which also approaches zero for perfect agreement.

When comparing assignments according to road categories, it is necessary to account for the difference in average volumes by dividing by the average count volumes. The well-known standard deviation of differences (SD) follows the same trend. Traffic volumes agree more closely when the sample correlation coefficient (CC) approaches positive or negative unity.[6] This parameter tests the degree of linearity between two independent variables.

The mean absolute error (MAE) has been applied to trip distribution models, but may be used in assignment model evaluation, where a zero value represents best agreement.[7] It behaves like a related and popular test, the root mean square error (RMSE).[4,5,8,9] Note that when examining subgroups of selected road types, biases may occur by the squaring of roads with high volumes compared with those having low volumes. Division by the average observed volume for the road type would assist standardization of the RMSE. The standardized absolute difference link fit index (SADI) requires origin-destination tion paths to be enumerated.[10] Because the assignment versions used in this study do not store path information, a less complex expression—standardized absolute difference (SAD)—was developed and should be ideally of zero value. Finally, Theil's inequality coefficient (TIC), the ratio of the RMSE to the sum of root mean squares, also tends to zero for matching volumes.[11]

None of the preceding nonparametric tests evaluate the statistical significance of the differences between count and assigned volumes. Unfortunately, each link has only one sample observed and expected frequency pair, while many hundreds of links are to be compared for overall agreement. A number of statistical references and several statisticians were consulted to find appropriate tests of significance, but none were found.

However, it has been suggested that tests of significance such as the chi square and Kolomogorov-Smirnov one-sample tests be used.[9] These two tests compare the frequency distributions of assigned and count volumes.[12] But it is possible to have loaded networks with equivalent frequency distributions of volumes, when in fact individual link counts and assigned volumes differ greatly.[13] Therefore, these two tests would incorrectly suggest total agreement of volumes and thus should not be used.

The two parametric tests examined were the likelihood ratio (LR) and information gain (IG), both expressed in Figure 1. Parametric tests require condi-

3. CRITERIA TESTS

$$MD = \left(\sum_{i=1}^{N} C_i A_i\right) - \left(\sum_{o=1}^{O} \sum_{d=1}^{D} T_{Aod} C^*_{od}\right).$$

$$VD = \left[\left(\sum_{i=1}^{N} C_i A_i\right) - \left(\sum_{o=1}^{O} \sum_{d=1}^{D} T_{Aod} C^*_{od}\right)\right] / \left[\sum_{o=1}^{O} \sum_{d=1}^{D} T_{Aod}\right.$$

$$E2 = \left[\left(\sum_{i=1}^{N} C_i A_i\right) - \left(\sum_{o=1}^{O} \sum_{d=1}^{D} T_{Aod} C^*_{od}\right)\right] / \left[\sum_{i=1}^{N} C_i A_i\right].$$

$$S = \left\{\sum_{d}^{D} \sum_{o}^{O} \left[\left(\sum_{i=1}^{N} (A_{iod} I_i) / I_{odm} - A_{od}\right]\right\} / \left[\sum_{i=1}^{N} A_i\right].$$

LIST OF SYMBOLS

A_i	assigned volume on link i.
O_i	observed volume on link i.
T_A	total of assigned volumes.
T_O	total of observed (count) volumes.
T_{Aod}	total of assigned volumes from origin o to destination d.
N	number of links with both assigned and count volumes in the network.
C_i	cost of travel on link i.
C^*_{od}	minimum cost of travel along route from o to d.
O_{iod}	observed volume on route i from origin o to destination d.
A_{iod}	assigned volume on route i from origin o to destination d.
I_i	impedance for link i.
I_{odm}	impedance on the minimum origin destination path after all trips have been assigned.

tions of the data. Likelihood methods of evaluation are used to assess the accuracy of gravity model calibration, but may be adapted to assignment model assessment.[14] They assume that flows between any origin and destination zone pair are multinomially distributed. Because drivers may choose from a variety of routes, this condition is satisfied. As the likelihood ratio approaches unity, the assignment is more effective. The information gain test is used to assess alternative trip distribution models, but can also be used for assignment models, noting that the expression approaches zero for best fit.[7] Figure 1 shows that this test may be converted to a form incorporating absolute differences, so similarity to results for the comparable nonparametric tests can be expected. The information gain concept is derived from Bayes's theorem, relating prior and posterior probabilities to a monotonic likelihood function.

The assignment criteria tests are concerned with estimating model algorithm convergence, based on the impedance to all trips. Because these tests do not compare count and assigned volumes, they avoid errors inherent in measuring count volumes. The tests involve the assignment impedance function, and so any performance comparison is only useful when the impedance is constant for all assignments.

Murchland's delta (MD), Van Vliet's delta (VD), and the error term (E2) all have a similar formulation to estimate the difference between the total network travel cost and the cost if all trips use the link paths of minimum impedance.[1,15,16] They should all be ideally as small as possible. Another proposed criteria test is the sigma technique(s), which is concerned with minimizing travel times over origin/destination paths.[17] However, because the models used in this study do not enumerate paths, it could not be applied.

Study Results

The 19 assignments run in this study were applied to the road network incorporating the 1978 Melbourne statistical division covering an area of 6,110 square kilometers. The total 1978 population was 2.67 million, including 1.18 million employed, for which an appropriate trip table has been prepared. The network consists of 736 zones, 4453 nodes, and 5056 two-way links. Each link has been categorized as either a minor road (14%), undivided main road (50%), divided main road (33%), or freeway (3%). Some 1506 links have traffic counts taken from the 1978 24-hour counts.[18]

The assignments vary by the number of assignment iterations, the traffic dispersion type, delay and toll factors for freeway exits and entries, and impedance function weightings. Generally, PLANPAC assignments involved all-or-nothing path dispersion and four successive capacity-restrained loadings of one-quarter of the trip table. UTPS assignments enable an initial multipath route selection for the first iteration and equilibrium trip loading of the full trip table, with capacity restraint.

The top section of Table 1 indicates the nature of each assignment, various

Table 1. Test Values and Rankings of Best Assignment Performance for the Whole Network

Assignment Type				*PLANPAC*			
Assignment No.	1	2	3	4	5	6	7
Dispersion*	A	A	A	A	A	A	A
Iterations**	4I	4I	4I	4I	4I	4I	5I
Impedance***	0:1	1:1	1:2	1:3	1:4	1:3	0:1
Toll							
Delay****						D = 1.0	
Test Type							
Rejected							
MR	1.032	1.022	1.028	1.028	1.028	1.048	1.034
	5	1	3	2	4	8	6
MD	1326	1463	1411	1414	1403	1252	1327
	3	17	12	13	10	1	4
Non-Parametric							
MAR	0.439	0.428	0.432	0.433	0.433	0.444	0.434
	16	11	12	13	14	18	15
CC	0.844	0.842	0.849	0.843	0.849	0.835	0.846
	12	16	9	14	8	17	11
MAD	5391	5433	5328	5338	5311	5409	5321
	16	18	13	14	11	17	12
SD	7093	7106	6934	6961	6943	7230	7038
	15	16	9	12	11	17	14
MAE	3.58	3.61	3.54	3.54	3.53	3.59	3.53
	16	18	14	13	11	17	12
RMSE	40.36	40.58	39.57	39.73	39.61	41.15	40.06
	15	16	8	11	10	17	14
TIC	0.170	0.171	0.167	0.167	0.167	0.173	0.168
	15	16	9	11	8	17	13
SAD	0.1564	0.1578	0.1547	0.1550	0.1541	0.1566	0.1542
	15	17	13	14	11	16	12
Parametric							
LR	46781	47156	46323	46641	46815	45601	45499
	15	17	13	14	16	11	10
IG	2.243	2.216	2.180	2.190	2.192	2.239	2.192
	18	16	11	12	14	17	15
Assignment Criteria							
MD		22976	72085	115381	165513	552048	31906
	—	6	14	16	17	18	8
VD		0.024	0.049	0.057	0.065	0.023	0.060
	—	9	15	16	18	8	17
E2		0.023	0.046	0.054	0.061	0.189	0.057
	—	8	14	15	17	18	16
System Impedance							
C1 × 10000	42.3	98.3	155.5	212.6	269.8	290.9	56.4
C2 × 10000	—	96.0	148.2	201.1	253.2	235.7	53.2

NOTE: Values are rounded.
* Traffic dispersion is either entirely all-or-nothing (A), or multipath (M) in the first iteration.
** The number of iterations is given with either the number of capacity restraint increments (I) or equilibrium iterations (E).
*** The ratio of distance weight to time weight is given. Impedance = distance × distance-weight + time × time-weight.
****The toll value adjusts link times for queuing at toll facilities. The delay is imposed on drivers using freeway ramps.

input parameters, the test values, and rankings. The various nonparametric statistical measures performed as expected. For instance, the mean ratio is seen to give misleading results, favoring incremental assignments, when compared with the mean absolute ratio, which prefers equilibrium assignments. Similarly, the biased mean difference favors incremental loading, whereas the mean absolute difference prefers equilibrium loading. The correlation coefficient, standard deviation of differences, mean absolute error, root mean square error, inequality coefficient, and the standardized absolute difference are all consistent, favoring equilibrium assignments. The test statistics vary by around 9% between best and worse values, so all of the assignments are reasonably consisent. However, the best MAD of 5069 suggests a fairly large difference between observed and assigned volumes. An average CC^2 value of 0.72 indicates that a 28% variation exists.

The parametric information gain statistic also prefers equilibrium loading, as does the likelihood ratio. However, this ratio does have slightly different results, because the ratio values were very large and varied little, making ranking awkward.

Table 1 also presents the assignment

UTPS												Percentage Variation Between Assignments	
1 A 5E 20:1	2 M 4E 1:3	3 M 8E 1:3	4 M 5E 1:3 TOLL	5 A 5E 1:3 TOLL	6 A 6E 1:3 TOLL D=1.5	7 A 6E 1:3 TOLL D=3.0	8 A 6E 1:3 TOLL D=1.5	9 M 8E 1:2	10 A 4E 1:2	11 M 8E 1:4	12 A 4E 1:4	1st to 2d	1st to Last
1.047	1.049	1.048	1.051	1.052		1.060	1.058	1.053	1.052	1.054	1.058		
7	10	9	11	12	—	18	17	14	13	15	16	—	—
1422	1343	1446	1411	1367		1384	1411	1593	1384	1451	1319		
14	5	15	8	6	—	9	11	18	7	16	2	—	—
0.426	0.418	0.425	0.419	0.420		0.423	0.426	0.443	0.421	0.425	0.423		
9	1	7	2	3	—	6	10	17	4	8	5	0.2	6.2
0.850	0.853	0.851	0.854	0.853		0.843	0.843	0.829	0.851	0.847	0.850		
7	3	4	1	2	—	13	15	18	5	10	6	0.1	2.9
5180	5087	5143	5069	5084		5139	5169	5382	5097	5125	5094		
10	3	8	1	2	—	7	9	15	5	6	4	0.3	7.2
6840	6756	6787	6709	6735		6936	6962	7261	6781	6873	6795		
7	3	5	1	2	—	10	13	18	4	8	6	0.4	8.2
3.44	3.38	3.42	3.37	3.38		3.41	3.43	3.58	3.39	3.41	3.38		
10	3	8	1	2	—	6	9	15	5	7	4	0.3	7.1
39.37	38.82	39.11	38.60	38.73	42.63	39.86	40.04	41.89	38.99	39.59	39.00		
7	3	6	1	2	19	12	13	18	4	9	5	0.3	10.4
0.166	0.163	0.165	0.163	0.163		0.168	0.169	0.177	0.164	0.167	0.164		
7	3	6	1	2	—	12	14	18	5	10	4	0.3	8.6
0.1520	0.1478	0.1499	0.1475	0.1477		0.1482	0.1510	0.1585	0.1476	0.1481	0.1493	0.1	7.5
10	4	8	1	3	—	6	9	18	2	5	7		
44793	44951	42898	43292	43168		44399	44924	50680	43458	42340	45654		
7	9	2	4	3	—	6	8	18	5	1	12	1.3	19.7
2.072	2.027	2.051	2.016	2.024		2.067	2.073	2.191	2.037	2.060	2.038		
9	3	6	1	2	—	8	10	13	4	7	5	0.3	11.0
10256	14984	9872	20994	24090	76489	67848	57815	21246	33745	37190	66702	—	—
2	3	1	4	7	15	13	11	5	9	10	12		
0.012	0.022	0.007	0.012	0.013	0.039	0.039	0.027	0.015	0.025	0.015	0.030	—	—
3	7	1	2	4	13	14	11	5	10	6	12		
0.012	0.021	0.007	0.012	0.013	0.037	0.038	0.027	0.015	0.024	0.015	0.029	—	—
2	7	1	3	4	12	13	10	5	9	6	11		
893.0	212.8	211.1	212.2	212.1	216.7	217.7	215.6	156.7	158.6	271.4	273.9	—	—
892.0	211.3	210.1	210.1	209.7	209.1	210.9	209.8	154.5	155.2	267.7	257.3	—	—

criteria test results, which favor equilibrium loading. However, note that the assignment with a distance to time weighting of 20 to 1 is ranked second behind that with a weighting of 1 to 3. This result arises because the iterative loading of the equilibrium algorithm alters travel time only, using a capacity restraint technique. Thus, these tests should only be used when analyzing assignments using similar impedance values.

The total system impedance (C1) and minimum system impedance (C2) are presented in Table 1. These values increase with larger impedance function weighting. Because Van Vliet's delta and the error term statistics use C1 and C2 as their denominators, respectively, it is not surprising that the assignment algorithm with the greater impedance will result in lower statistic value rankings.

The results indicate that a combination of the tests is the most appropriate means of assessing assignment performance. Murchland's delta is a suitable test to initially assess assignment convergence, assuming similar impedance weightings. Link count and assigned volumes should then be compared using information gain, mean absolute difference, mean absolute ratio, root mean square error, mean absolute error, correlation coefficient, or the inequality coefficient according to the ease of computation. However, information gain is the most rigorous test to use.

The study has not determined any test of error significance. Although the statistics indicate preferred assignment types, they still do not suggest the precision of the best assignment. Rather, the acceptability of an assignment results remains a subjective judgment. Possibly, a 10% error is acceptable, considering that the final volumes are an outcome of many different calculations and data inputs. However, some modelers may be content with a 30% error, based on their own experiences. In this study, the CC parameter suggests a 28% error for the Melbourne assignments. Consequently, efforts were made to improve assignment accuracy by checking link coding and revising trip tables. The study has identified appropriate tests for all future assignment accuracy checks.

In this study, the preferred assignments have equilibrium loading with a one to three impedance function. Although multipath traffic dispersion is preferred, the dispersion type has a lesser effect on assignment performance. Following the study, Melbourne assignments were changed to the equilibrium type.

An analysis by each link road category was also undertaken to permit comparison of links with similar traffic volumes. The normalizing procedure for MAD and RMSE described earlier were verified. Roads with low volumes were best modeled by equilibrium assignment, though not as well as main roads. All road types were found to prefer equilibrium assignments, except for freeways, which had been coded to attract vehicles using the incremental algorithm.

Summary

Alternative traffic assignment performance tests have been presented and applied on UTPS and PLANPAC assignments of a large urban traffic network model. A range of these tests should be used to analyze assignment accuracy. The relevance of various assignment techniques shall continue to be investigated in future work. Equilibrium, multipath assignment was found to be preferred for the congested Melbourne network, and has now been adopted.

Acknowledgments

This article is presented with the permission of Mr. I. Stoney, chairman and managing director of the Road Construction Authority, Victoria, Australia. The views in the article are those of the author and do not necessarily represent those of the Authority.

References

1. Urban Mass Transportation Administration and the Federal Highway Administration. "Urban Transportation Planning System." Washington, DC: U.S. Department of Transportation, 1979.
2. Federal Highway Administration. "Urban Transportation Planning Program Documentation." Washington, DC: United States Department of Transportation, 1974.
3. Boyce, D. E.; Janson, B. N.; and Eash, R. W. "The Effect on Equilibrium Trip Assignment of Different Link Congestion Functions." *Transportation Research,* Vol. 15A, 1981, pp. 223-232.
4. Edwards, J. L., and Robinson, F. O. "Multipath Assignment Calibration for the Twin Cities." *Proceedings of the American Society of Civil Engineers,* Vol. 103, No. TE6, 1977, pp. 791-806.
5. Smith, R. L., and Brennan, T. S. "Traffic Assignment Techniques for Smaller Cities." *Proceedings of the American Society of Civil Engineers,* Vol. 106, No. TE 1, 1980, pp. 85-98.
6. Nie, N. D. *SPSS: Statistical Package for the Social Sciences.* 2d ed. New York, NY: McGraw-Hill Book Company, Inc., 1975.
7. Smith, D. P., and Hutchinson, B. G. "Goodness of Fit Statistics for Trip Distribution Models." *Transportation Research,* Vol. 15A, 1981, pp. 295-303.
8. Black, J. A., and Salter, R. J. "A Statistical Evaluation of the Accuracy of a Family of Gravity Models." In *Proceedings.* Institution of Engineers, Australia, Part 2, 59, March 1975, pp. 1-20.
9. Oxlad, L. M. "A Study of Traffic Assignment Models." M. E. thesis, University of N.S.W., 1976.
10. Judge, E. J. "Tests of Assignment Accuracy: An interurban Case Study." *Transportation* (3), 1974, pp. 25-44.
11. Theil, H. "Economic Forecasts and Policy." North Holland Publishing Company, 1965.
12. Siegel, S. *Non Parametric Statistics.* New York, NY: McGraw-Hill Book Company, Inc., 1956.
13. Apelbaum, J. P., and James, M. L. "Testing and Comparison of Alternative Traffic Assignment Techniques for the Greater Melbourne Road Network." In *Proceedings.* Eighth Australian Transport Research Forum, Canberra, May 1983, Vol. 2, pp. 171-205.
14. Wilson, S. R. "Statistical Notes on the Evaluation of Calibrated Gravity Models." *Transportation Research,* Vol. 10, 1976, pp. 343-345.
15. Murchland, J. D. "Road Network Traffic Distribution in Equilibrium." In *Proceedings: Mathematical Models in the Economic Sciences.* Mathematisches Forshungs—Institut, Oberwolfach, 1969.
16. Van Vliet, D. "The Choice of Assignment Techniques for Large Networks, Traffic Equilibrium Methods." *Proceedings of the International Symposium.* Universite de Montreal, 1974, pp. 396-412.
17. Lucas, C., and Davidson, K. B. "A Method for Measuring the Effectiveness of a Given Assignment Method and an Application to Three Different Methods. *Australian Road Research* Vol. 5, No. 5, June 1974, pp. 92-99.
18. Country Roads Board, Victoria. "Traffic Volumes—Estimated 1978 Annual Average Daily Traffic." Advance Planning Division, 1978.

Matthew L. James is an engineer involved in road planning at the Victorian State Road Construction Authority. He has a B. E. Hons. in civil engineering from Monash University and an M.E. in transportation engineering from the University of Florida. James is an associate member of the Institute and is currently editor of ITE's Australian Section newsletter. He is also concerned with aerospace development, serving on the National Committee on Space Engineering of the Institution of Engineers, Australia, among other organizations.