
GENERAL PLANNING CONSULTANT

TECHNICAL MEMORANDUM 88.3.13

HNET VALIDATION

Prepared for:

SOUTHERN CALIFORNIA RAPID TRANSIT DISTRICT

Prepared by:

SCHIMPELER-CORRADINO ASSOCIATES

June 1988

1. INTRODUCTION

Task 3.1 involved preparing a regionwide highway network for use by the Southern California Rapid Transit District (SCRTD), and performing a highway assignment to validate the network. The main purpose of this task was to bring the highway assignment capability in-house to RTD. Until now, the highway assignment phase of the UTPS process was conducted by SCAG and Caltrans for their respective purposes; there was little reason for RTD, as the transit agency, also to deal with the highway side. However, the decision to implement INET at RTD gave rise to the need to know how to perform highway assignments.

INET, which stands for integrated network, is to be implemented in the 1988-89 fiscal year. Current practice for regional transportation modeling at all key agencies in Southern California is to maintain separate networks for highway and transit assignment purposes. That is, bus and rail lines are coded with a separate set of nodes and links than the highway network, even though buses use the same streets that cars do. These two different networks are built through UTPS programs HNET and UNET, respectively. The practical implication of having separate networks is that conditions on the highway network are not reflected in the transit network when transit assignment is done (the converse is also true, but that is a much smaller effect). In particular, transit assignment is performed under the assumption of free-flow speeds, which is unrealistic since buses are traveling on the congested street system.

As the name suggests, INET involves building a single, integrated, highway and transit network, which explicitly permits information such as congested speeds to be shared between highway and transit assignments. An understanding of, and the ability to use, the highway assignment procedure (UROAD), then, is a prerequisite to the implementation of INET.

This memorandum documents the process used to perform and validate a regional highway assignment, and serves as a guideline to RTD and GPC staff in preparing future highway assignment runs. Section 2 briefly describes the highway assignment procedure to be used. Section 3 discusses the preparation of the input files required by UROAD: the network files (supply side of the simulation), and the trip tables (demand side of the simulation). Section 4 presents the validation checks used to determine that the model output is consistent with SCAG's baseline assignment and reasonably reflects reality.

2. HIGHWAY ASSIGNMENT PROCEDURE

To understand the preparation of the input files discussed in the next section, it is first necessary to understand the specific highway assignment procedure being used. The UROAD assignment program allows a number of options, and each agency adopts slightly different assignment methodologies and parameters. The procedure used for this report is intended to replicate the highway assignment process used by SCAG as closely as possible.

UROAD requires as input a matrix of vehicle trips (not person trips) in origin-destination (not production-attraction) form. It is standard practice to split total daily demand by time of day (typically "A.M. peak," "P.M. peak," and "Off-peak" periods), and perform assignments separately by time period. In addition, carpools are allowed to use certain facilities that single-occupant vehicles are not: specifically, high occupancy vehicle (HOV) lanes and freeway ramp meter bypass lanes. Thus, in preparing the input network, we must be able to distinguish differences in the network depending on time of day and mode of travel. This is explained further below. In practical terms, however, the consequence is that six separate assignment runs are performed: Non-carpool trips for each of three time periods, and carpool trips for each time period.

In general, UROAD calculates impedances, builds minimum-impedance paths between origin and destination, and assigns trips to the minimum paths. The assignment is typically performed in several iterations, with the desired outcome being an equilibrium condition wherein travelers cannot reduce their travel time by altering their paths. In other words, at equilibrium, all paths between any O-D pairs are loaded such that the travel time on each of the paths is the same. The assignment is capacity-restrained; that is, at the end of each iteration, travel times on links are adjusted to account for congestion, and trips in the next iteration are loaded based on these adjusted travel times. The final output assignment is the weighted average of assignments from previous iterations. In this analysis, "impedance" is equivalent to travel time and thus minimum paths are simply the traditional minimum time paths. UROAD allows a flexible definition of impedance, however, as a user-specified combination of time, cost, and distance.

SCAG uses a stochastic assignment (that is, trips are assigned to various competing paths with probabilities that are inversely related to the path impedance) in the first iteration, in the expectation that a stochastic assignment will provide an initial solution relatively close to equilibrium. Subsequent iterations, however, are the faster all-or-nothing assignments (that is, all trips between an origin and a destination are assigned to the single shortest path between those two points). SCAG has found that seven iterations (one stochastic, followed by six all-or-nothing) are sufficient to produce an equilibrium assignment, and that is the procedure that was followed for this report. Caltrans, on the other hand, uses 10 all-or-nothing iterations without an initial stochastic assignment.

The assignment is performed in the following manner:

1) Assign the **Drive alone (DA) and Shared ride** (2-person carpool) trips by time of day to the highway network, neglecting the carpool links (i.e., HOV links, which are reserved for carpools of three or more persons).

2) At the end of the combined Drive alone and Shared ride carpool assignment for each time period, UROAD outputs a congested travel time value and volumes for each link. The carpool links of the network contain missing values for the congested times, since they were not used in the assignment.

3) To assign the **carpool** (3 or more person) trips by time of day, the initial travel times now used to compute impedances on standard network paths are taken as the congested travel time values obtained from the previous step. The "missing values" for the carpool links are changed to free-flow times. A single all-or-nothing assignment is used. The travel times on all these paths represent equilibrium conditions, and therefore an all-or-nothing assignment is acceptable.

3. INPUT PREPARATION

3.1 HIGHWAY NETWORK

The base highway network used in this task was the 1980 network provided to RTD by Caltrans. A 1984 network was also available, but it was coded in much less detail than the 1980 network. There are few differences in the highway system between 1980 and 1984, so use of the earlier network is preferable because of the greater level of detail.

Two steps were needed to prepare this original network for RTD use:

1) The first step was to translate the node coordinates from the "State plane" scale (miles) used by SCAG and Caltrans to the metric scale used by RTD. This was done by means of conversion equations previously developed by RTD staff.

2) The Caltrans network was developed for the 1555-analysis zone system used by SCAG and Caltrans; specifically, centroid connectors linked zone centroids in the 1555-zone system to the actual street network. Accordingly, the major step in preparing the network for use at RTD was to replace those links with centroid connectors generated for the 1628-zone system which is the RTD standard.

A FORTRAN program was written to automate the generation of centroid connectors. Two input parameters are required by the program: MXLINK is the maximum number of connectors to be generated per centroid, and DSTMAX is a cutoff distance. For each zone centroid, the program searches all valid (i.e. existing) non-freeway nodes (i.e. nodes greater than or equal to 4000) and finds all nodes (up to MXLINK) of distance DSTMAX away from the centroid or less. If all valid nodes are more than DSTMAX away, the closest one is chosen to form the centroid connector for that zone. It was originally planned that up to four connectors would be generated for each zone (i.e. MXLINK = 4). However, the underlying highway network was coded in such detail that the maximum number of links permitted by UROAD (32,766) would have been exceeded had more than one connector per zone been created. Thus, each zone was connected only to the single nearest existing non-freeway node in the highway network. The new network containing these centroid connectors was plotted, and the connector links were manually checked and approved before proceeding.

As mentioned in Section 2, the highway network characteristics differ by time of day and mode of travel. However, it is not necessary to build six separate networks; the differentiation is achieved by changing certain variables in the assignment set-up. In terms of **mode of travel**, the only difference for our purposes* is that only three or more person carpools are allowed to use HOV lanes**. The HOV links in the network file have a distinct "use code" which enables UROAD to ignore them when doing a non-carpool assignment and include them when doing a carpool assignment.

As for **time of day**, the only change necessary is to account for the length of each of the time periods. The link capacities provided in the look-up tables are hourly capacities, and therefore the peak period volumes are converted into hourly volumes for calculating the V/C ratios and congested travel times. This is accomplished with the CONFAC parameter in UROAD. For the 2-hour A.M. peak period, CONFAC = 0.5, i.e., a one-hour link volume is half the total volume. For the three-hour P.M. peak, CONFAC = 0.33. For the off-peak period, consistent with Caltrans practice, CONFAC = 0.09, or 1/11. The assumption is that travel is negligible outside the 16 hours covered by both peaks and an 11-hour off-peak period.

---

*SCAG distinguishes between carpool (2-person and 3+ -person) and drive alone network characteristics in another way: ramp metering. Carpools are allowed to bypass ramp meters while single-occupant autos must wait in line. SCAG models this situation by providing much slower look-up speeds for ramp meter links when assigning drive alone trips than for carpool trips (8 mph versus 30 mph, respectively). Caltrans, however, does not believe the assignment model is sufficiently sensitive to the small time savings for carpool vehicles to justify the effort it takes to identify ramp meter links in the network. Accordingly, the base network RTD received from Caltrans did not contain a distinct code for ramp meter links, and consequently it was not possible to incorporate ramp meter bypass into the carpool assignments.

**In 1980, the year for which the network was developed, the only HOV facility in the region was the El Monte Busway, which was restricted to buses and carpools of 3 or more persons. HOV facilities on SR 91 and SR 55 lanes (opened in 1985) permitted two-person carpools, but these did not exist at the time the network was developed and therefore are not identified in the network as HOV links. Future-year highway assignments, however, may need to be done in three stages for each time period instead of two: drive alone trips first, then 2 person carpool trips (opening those HOV links which 2 person carpools are allowed to use), then 3 or more carpool trips (opening those links which only 3 or more person carpools are allowed to use).

3.2 TRIP TABLES

The steps in preparing trip tables for input to UROAD were as follows:

The auto trip tables produced by a previous mode choice run on 1985 data were obtained. These trips are production-attraction (P-A) vehicle trips and are segregated by trip purpose. Home to work (H-W) trips are split among three auto sub-modes: drive alone , two-person carpool, and three-or-more person carpool. Other trip purposes do not distinguish auto sub-modes. Thus, only H-W carpool trips can be assigned separately.

The P-A auto trip tables, segregated by trip purpose and (for H-W trips) by sub-mode, were, in a single step, converted to O-D trip tables segregated by time of day. Results from the 1976 Origin-Destination Survey were used which show the percentage of daily trips (by trip purpose) occurring in each time period. The percentages used for this report are shown in Table 3.1. With the exception of the percentages for carpool trips, these are the factors used by SCAG (Travel Forecast Atlas, 1985). (Carpool factors were not available from SCAG, so the Caltrans numbers were used for those trips.) Caltrans uses identical factors for H-W trips and similar factors for O-W and N-W trips (LARTS Travel Forecast Summary, 1987).

Using these percentages, Production-Attraction trips (by trip purpose) were converted to Origin-Destination trips by time of day. This is done using the UTPS program UMATRIX. For example, if matrix X contains Home-to-Work trips, and it is known that 30% of Home-to-Work trips occur in AM from P to A, while 0.5% Home-to-Work trips occur in AM from A to P, then the Origin Destination trips would be

$$0.3(X) + 0.005(X^T) \quad \text{(Where } X^T \text{ is the transpose of X)}$$

Thus, internal O-D trips by time of day were obtained. These internal trips (non-carpool) were then combined with external trips. The external trip table available was for the 1325-zone system, and had to be expanded to the 1628-zone system. This was done using the USQUEX program along with zone-district fraction cards.

Once the 1628-zone external trip table was developed, it was split among the three time periods based on percentages obtained from SCAG. These percentages are shown in Table 3.2. The external trips by time period are combined with the internal trips by time period to yield total trips for each time period.

TABLE 3.1

PEAK/OFF PEAK FACTORS

PERIOD	DIRECTION	H-W	O-W	N-W	CARPOOL
AM PEAK	P-A	28.98	11.81	5.51	34.37
(6:30-8:30)	A-P	0.53	-	0.51	0.63
PM PEAK	P-A	3.17	30.32	12.61	3.85
(3:00-6:00)	A-P	29.8	-	11.45	36.15
OFF-PEAK	P-A	21.12	57.87	46.99	14.07
	A-P	16.40	-	22.93	10.93

TABLE 3.2

Time-of-Day Conversion Factors

for External Trips (%)

Period	Direction	External Trip %
AM Peak	P-A	2.17
	A-P	8.65
PM Peak	P-A	31.11
	A-P	15.91
Off-Peak	P-A	66.72
	A-P	75.44

The outcome of these steps is the creation of six O-D trip tables for input to the six assignment runs:

a) Drive alone and 2 person AM trips.

b) Three or more person carpool AM trips

c) Drive alone and 2 person PM trips

d) Three or more person carpool PM trips

e) Drive alone and 2 person off-peak trips

f) Three or more person carpool off-peak trips

4. VALIDATION

4.1 TOTAL O-D TRIP CHECK

This highway assignment was performed using the 1985 trip table. Since this task endeavored to replicate SCAG's procedures as closely as possible, key intermediate and final results were compared with the corresponding SCAG results to ensure consistency.

The initial trip table used by GPC contained about 41 million (1985) person trips, while the SCAG trip table had about 40 million (1984) trips. These trips are production-attraction trips, and had to be converted to origin-destination trips by time of day. For this purpose, factors representing the percentage of each trip in each period were used. These were obtained from SCAG.

On conversion of P-A person trips to O-D vehicle trips, it was found that the total O-D trips produced by GPC were lower than the SCAG numbers. This was not logical since the GPC trip table initially contained a greater number of person trips. On closer examination, it was found that SCAG had used a factor of 1.44 for auto occupancy in non-work trips, while the GPC model used 1.55 as the auto occupancy.

To be consistent with SCAG, the GPC non work trips were adjusted to a auto occupancy of 1.44. The total number of O-D trips calculated by SCAG was 28,410,000 while the adjusted O-D trips used by the GPC model was about 29,176,000. The GPC O-D trips are about 2.7% higher than the SCAG O-D trips, which is attributed to the growth in travel between 1984 and 1985.

4.2 ASSIGNMENT CHECK

To test the accuracy of the assignment results, travel time and distance values for the three periods (AM, PM and off-peak) were checked between some well-known zones such as the El Monte Bus Station to Downtown, Universal City to Downtown, etc. These were found to be consistent with the expected values.

4.3 COMPARISON OF TOTAL SCAG AND GPC TOTAL VOLUMES BY SCREENLINES

At the end of traffic assignment, UROAD produces a link by link assigned volume report, and also aggregates these volumes by screenline. This section examines screenline totals, while Section 4.4 analyzes link-level volumes for screenline links.

A major step in the validation was to compare the screenline volumes obtained with the GPC model against those produced by the SCAG model. In both cases, volumes were added across all six assignments to obtain total daily predicted vehicle trips by screenline. Before this comparison was made, however, it was necessary to account for any differences between the two models arising from differences in initial internal trips.

Differences between SCAG and GPC screenline totals were traced to two main causes. The first is the 2.7% growth in travel between 1984 (SCAG) and 1985 (GPC) as discussed in Section 4.1. The second cause of the variation is more subtle, and lies in the difference in zone structures underlying the analysis. UROAD only assigns interzonal trips to the network; intrazonal trips are not loaded. Because the 1628-zone system had more and smaller number of zones than the 1325-zone system used by SCAG, a smaller proportion of our trips are intrazonal than for SCAG. Accordingly, a higher number of our interzonal trips are assigned to the network than for SCAG. Specifically, SCAG's baseline data contained 81.8% interzonal trips, while GPC 's data contained 83.0% interzonal trips.

When the GPC model results are adjusted to account for the two differences, they are relatively close to the SCAG model results, as shown in Table 4.1. The total across all screenlines is within 1% of the SCAG total and within 6% of the ground count total. More importantly, the directions of bias are nearly identical between SCAG and GPC. That is (except for screenline 3), when the SCAG model overestimates the screenline total, the GPC model does the same, and vice versa. This is additional confirmation that the SCAG procedure is being closely replicated.

4.4 STATISTICAL TESTS ON THE SCREENLINE LINK VOLUMES

Section 4.3 focused on the comparison between SCAG's model output and GPC's model output. It is equally (if not more) important to compare the GPC model output to observed volumes, i.e., ground counts, to see how well the assignment reflects reality. This section reports the results of three statistical tests performed to compare link level model volumes with link level ground counts for 244 screenline links. The ground counts were obtained from Caltrans, since SCAG did not yet have link level counts available for the 1980 network used in this analysis.

4.4.1 Sources of Differences between Model Volumes and Ground Counts

In comparing GPC model volumes to SCAG model volumes in Section 4.3, it made sense to adjust for all known systematic differences between the two sets of numbers (i.e., growth and zone structure). Comparing GPC model volumes to ground counts, however, particularly at the link level, introduces numerous other differences which are not as readily accounted for. Rather than make a superficial adjustment for some of these sources of difference and not deal

TABLE 4.1

COMPARISON OF SCREENLINE VOLUMES

SCREENLINE NUMBER	SCAG GROUND COUNTS (SG)	SCAG MODEL (SM)	GPC MODEL (GM)	SG vs SM (% DIFF)	SG vs GM (% DIFF)	SM vs GM (% DIFF)
1	1214784	1324021	1298362	9%	7%	-2%
2	1767744	1939090	2011306	10%	14%	4%
3	1079206	1013289	1083734	-6%	0%	7%
4	1248636	1256943	1396589	1%	12%	11%
5	1049024	1176870	1278221	12%	22%	9%
6	573170	564813	469958	-1%	-18%	-17%
7	423155	349604	319219	-17%	-25%	-9%
8	839801	975076	902765	16%	7%	-7%
9	224081	216188	175027	-4%	-22%	-19%
10	243796	280813	245280	15%	1%	-13%
11	139394	160660	142550	15%	2%	-11%
TOTAL	8802791	9257367	9323011	5%	6%	1%

with others because their effects are unknown, we simply list below some causes of differences between model volumes and ground counts and, when known, the direction and approximate magnitudes of bias. For the tests reported in succeeding subsections, no adjustments were made to the model volumes or to the ground counts.

1. First and most obvious, the model is only an approximation to reality, and represents that reality to varying degrees under varying conditions. The magnitude and direction of any differences due to model validity are **unknown**.

2. Even if the model were an extremely faithful representation of reality, it must be remembered that model volumes and ground counts are each just single realizations of underlying random processes. Model-produced link volumes will vary somewhat from one iteration to the next. Link ground counts are typically factored up (approximately) to daily levels from observations over a few hours, and even daily traffic will fluctuate widely by day of week and time of year. So there is not a single "correct" link volume, either for the model or for observed reality. The direction and magnitude of these differences are **unpredictable** as well.

3. The model assigns a 1985 trip table, whereas the ground counts were taken in 1984. Based on comparing our model input to SCAG's 1984 model input, we might expect 1985 model volumes to be 2-3% **higher** than 1984 ground counts.

4. The model output does not include truck traffic, whereas the ground counts do. According to Caltrans, trucks are estimated to account for 3-4% of the volumes on surface streets, and 7-8% of freeway volumes, so this difference suggests that the model output will be **lower** than ground counts by 5-7%.

5. The model does not assign intrazonal trips, whereas most of those trips <u>are</u> actually using the street network. Many of them are crossing screenlines (and therefore are being included in the ground counts). Intrazonal trips were 17% of the total for GPC assignment. This difference will tend to **lower** the model output by an amount which is probably less than 17%.

6. Finally, the ground count data are incomplete. Not every network link that was identified as a screenline link had an available ground count. Further, many streets crossed by screenlines were too small even to be coded in the network. The demand represented by the cumulative traffic on these streets will be assigned by the model (to nearby streets that <u>are</u> in the network), but will not appear in link or total ground counts. This will tend to make the model volumes **higher** than the ground counts by an unknown amount.

4.4.2 Chi-Squared Test

The first, and least informative comparison test performed was a chi-squared significance test between the model-produced link volumes and the observed ground counts, to determine whether the two sets of data were statistically different.

The test statistic was calculated as

$$T = \text{SUM} \frac{(Y_p - Y_o)^2}{Y_o} \quad \text{where } Y_p = \text{predicted value}, \; Y_o = \text{observed value}$$

$$= 9{,}850{,}000 \, .$$

This statistic belongs to a chi-squared distribution. For n>30, the $T_{critical}$ for the 0.05 probability level is approximated as (Ref: CRC, Standard Mathematical Tables) :

$$T_{critical} = \frac{1}{2} \left(x_{0.05} + (2n-1)^{1/2} \right)^2 ,$$

where $x_{0.05}$ is the critical value for the normal distribution. In our case,

$$T_{critical} = \frac{1}{2} \left(1.96 + (487)^{1/2} \right)^2 = 288.6,$$

which is far less than the calculated T value.

This indicates that the two sets of data are significantly different at the 0.05 level. However, this is not surprising since the chi-squared test is notoriously easy to fail. That is, it is very difficult for two sets of data to be found statistically indistinguishable according to the chi-squared test.

4.4.3 Root Mean Squared Error

The second test was an analysis of root mean squared error (RMSE) by volume class. The RMSE is simply the square root of the mean squared difference between the observed and the predicted values. That is,

$$\text{RMSE} = \left(\text{SUM} \, (Y_p - Y_o)^2 / N \right)^{1/2} \quad .$$

Table 4.2 displays the RMSEs by volume class in absolute terms and as a percentage of the class mean. The RMSE ranges from 28% of the mean for the highest volume class (freeway class) to 91% of the mean for the lowest volume class. Overall, the model performs more accurately for higher volume links than for the lower volume links, which is intuitively expected.

TABLE 4.2

ROOT MEAN SQUARE ERROR BY VOLUME CLASS

VOLUME CLASS	ROOT MEAN SQUARE	ROOT MEAN SQUARE/CLASS MEAN
0-10,000	6293	0.91
10,000-20,000	8407	0.51
20,000-40,000	12040	0.47
40,000-100,000	29520	0.39
100,000 +	52280	0.28

4.4.4 Regression Analysis

The third and final test was to regress the model output link volumes against ground count link volumes. If the two sets of numbers were equal, the slope of the regression line would be one and the intercept zero. It is possible to test whether the estimated values of these coefficients differ significantly from their hypothesized values.

Table 4.3 presents various statistics obtained from the regression analysis. The correlation between the dependent and the independent variables is 0.96, and the regression R^2 is 0.92, both quite high. The slope is close to 1, as expected, but it is significantly different from 1. The intercept is not significantly different from 0 at the 95% confidence level ($T_{critical}$=1.96), but it is different from 0 at the 90% level ($T_{critical}$=1.65).

A second regression was performed for which the intercept was constrained to be 0. R^2 for that regression was 0.915, barely lower than the unconstrained R^2. The slope coefficient was 1.17, still statistically different from 1 (t-value=8.73).

Taken together, these results indicate a strong relationship between model volumes and ground counts which is close, if not statistically identical, to the theoretical relationship. Overall, the model overestimates the ground counts, but for reasons discussed in subsection 4.4.1, this does not necessarily mean that the model is "wrong" and the ground counts are "right".

TABLE 4.3

REGRESSION OF MODEL LINK VOLUMES VS LINK GROUND COUNTS

Number of observations = 244

Correlation coefficient (model volumes, ground counts) = 0.96

R^2 of regression = 0.916

	estimated coefficient	null hypothesis - value	t-value for H_o
Slope	1.19	1	8.15
Intercept	-2428.05	0	1.72

REFERENCES

Caltrans (October 30, 1987). **Los Angeles Regional Transportation Study Travel Forecast Summary: 1984 Base Model.** Transportation Analysis and LARTS Branch, Caltrans District 07, Los Angeles.

CRC, **Standard Mathematical Tables** (1973), CRC Press, Cleveland.

SCAG (August, 1985). **Travel Forecast Atlas.** Southern California Association of Governments, Los Angeles.

SCAG (July, 1987). **HOV Modeling.** Southern California Association of Governments, Los Angeles.