U.S. Bureau of the Census, May 1979.
9. J. R. Baker. GeoModel: Integrated Data Structures for Representing Geographic Entities. In Proc., Fourth International Symposium on Computer-Assisted Cartography (Nov. 4-8, 1979), American Congress on Surveying and Mapping and American Society of Photogrammetry, Falls Church, VA, 1980, pp. 275-282.

Introduction to Aggregate Data Analysis by Using UTPS: UMATRIX

ROBERT B. DIAL AND LAWRENCE F. QUILLIAN

The effectiveness of any transportation planning operation depends largely on its ability to acquire, manipulate, and present data. The Urban Transportation Planning System program UMATRIX, a powerful data manager, provides the transportation planner with capability in summarization, modification, and display of data. As numerous examples ranging from simple data preparation to application of complex travel-demand-forecasting models demonstrate, the UMATRIX user can access a variety of data forms and can accomplish analytical chores that have been difficult and time consuming. UMATRIX's ease of operation provides an effective and flexible tool for the transportation planner.

The effectiveness of any transportation planning operation depends largely on its data-management ability. In the evaluation of long-range strategies, the analysis of short-range tactics, or merely the observation of the status quo, a planner collects, refines, forecasts, summarizes, and reports data. The Urban Transportation Planning System (UTPS) consists of computerized and manual planning techniques developed, maintained, and distributed by the Urban Mass Transportation Administration (UMTA) and the Federal Highway Administration. New UTPS software that readily creates and manipulates particular types of data considerably reduces the time and cost of transportation planning. This paper describes one of these new programs--UMATRIX.

DATA TYPES

The planner must understand and forecast both the transportation and the demographic characteristics of the region. The data consist initially of network data (which describe the highway and transit infrastructure), survey data (household interviews, on-board surveys, cordon counts, etc.), and data from secondary data sources (Census Bureau information, marketing service summaries, etc). As the planner processes this large data base, he or she creates additional data for purposes of forecasting, plan evaluation, and decision making. Although they are of enormous variety in substance, a large portion of the planner's data falls into two categories: zonal data and interzonal data.

Zonal Data

Zonal data are collected and aggregated into some predetermined geographic "areal unit" configuration. Zonal data (whether forecast or observed) provide such information as population, average income per household, dwelling units, trip attractions or productions stratified by trip purpose, transit route miles, vehicle miles traveled, lane miles, carbon monoxide emissions, energy consumption, and total congestion time for each unit of local geography. Zonal data are read and written by numerous UTPS programs. Among these are UCEN70, MBUILD, AGM, UMODEL, and UMATRIX.

Figure 1 shows a hypothetical five-zone region; population figures are printed along with the zone number. Each zone's specific zonal population value is shown in Figure 2 as an indexed list for the zonal attribute "population". This can be imagined as a single array that has two components: an implicit row index and an explicit cell value. Stored in the array is a list of attribute values (LAV) in which each cell value contains the data for the zone whose number corresponds to the cell's row index (e.g., in cell 5 resides the value of population for zone 5--5900 persons).

Interzonal Data

While zonal data attributes are related to individual entities, interzonal or matrix data are related to a pair of entities. These matrix data relate various characteristics to a given set of origins and destinations. Often these values are summed from the individual link data that constitute the network, and each cell value represents a value for time, trips, fares, modal shares, or distances from one zone to another. Interzonal data constitute the input and output of numerous UTPS programs, notably UROAD, UPATH, AGM, and UMATRIX.

Interzonal data provide a matrix that gives a value for a trip from origin zone I to destination zone J. Figure 3 shows peak-hour home-based-work (HBW) person trip values from zone 5 of a hypothetical region to all other zones. Trips from zone 5 and all other regional trip interchanges are shown in Figure 4. In this case, this matrix of interchange values can be shown as a table of cells that has three components: a row index (I), a column index (J), and a cell value. Thus, for example, for a five-zone region, the matrix contains 25 cell values of trips between each origin and destination zone pair. Row 5 and column 3 refer to the number of peak-hour HBW person trips (230) from origin zone 5 to destination zone 3.

UMATRIX

The UTPS program UMATRIX is a powerful zonal and interzonal data-management system. It provides the planner with complete data-analysis capabilities. It accommodates the large range of data forms required for transportation planning, including matrix data, household surveys, zonal-based demographics, and network characteristics.

The UMATRIX user can access various input data

types and modify, update, display, and create new data required for demand estimation, traffic assignment, and system evaluation. Its simple, algebraic command language provides the user with arithmetic (+, -, *, /) and logical (if-then-else) operations to transform and create aggregate and disaggregate data. These transformations are specified by using algebraic expressions in which the left-hand side of the expression names new data created by the expression on the right-hand side. Special functions (log, exp, transposition, etc.) are also available. In addition, "look-up tables" provide an efficient way to transform, edit, and screen survey data.

The best way to present UMATRIX's capability is by means of examples. From basic to complex, they are

1. Arithmetic operations,
2. Logical operations,
3. Look-up tables,
4. Special LAV functions, and
5. Special matrix functions.

Figure 1. Zonal data: UTOWN zone population.

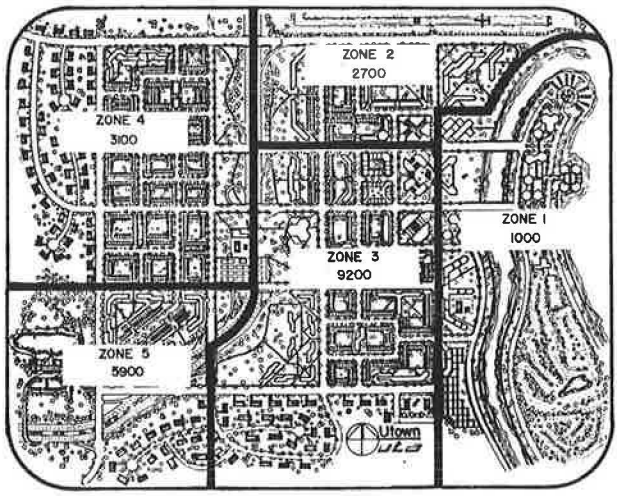


Figure 2. Zonal data for the attribute "population".

$$\text{POPULATION} \begin{cases} 1 & 1000 \\ 2 & 2700 \\ 3 & 9200 \\ 4 & 3100 \\ 5 & 5900 \end{cases}$$

Figure 3. Interchange data: peak-hour HBW trips from zone 5.

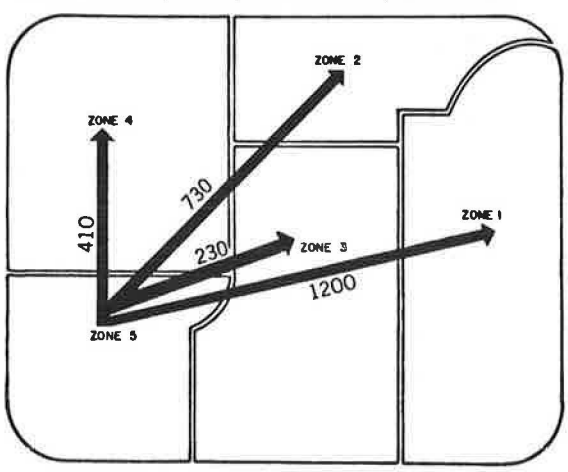


Figure 4. Interchange matrix for regional peak-hour HBW trips.

ORIGIN \ DESTINATION	1	2	3	4	5
1	200	440	270	330	120
2	1700	180	450	310	60
3	1150	1000	220	150	170
4	720	890	510	310	280
5	1200	730	230	410	250

Arithmetic Operations

Arithmetic operations allow the planner to specify simple matrix or zonal data transformations. For example, creation of a new zonal data item, named POPDIFF [the difference between year-2000 forecast population (POP2000) and base-year population (POPBASE)], is created by means of the following expression (for simplicity the mnemonic naming convention used in the examples has been slightly altered from that actually used in UMATRIX):

POPDIFF = POP2000 - POPBASE

The following is an example of LAV subtraction for a three-zone study area:

$$\text{If} \quad \text{POP2000} = \begin{bmatrix} 12\,500 \\ 17\,500 \\ 21\,000 \end{bmatrix} \text{ and } \text{POPBASE} = \begin{bmatrix} 11\,000 \\ 15\,000 \\ 22\,000 \end{bmatrix} \text{ then}$$

$$\text{POPDIFF} = \text{POP2000} - \text{POPBASE} = \begin{bmatrix} 1500 \\ 2500 \\ -1000 \end{bmatrix}$$

Subtraction, like all LAV operations, is performed on an element-by-element basis. Thus, the first element of POPDIFF is 12 500 - 11 000, or 15 000; the second element is 17 500 - 15 000, or 2500. Whether a 3-zone system or a 1300-zone system is being analyzed, the same expression is used.

For matrix data, arithmetic operations are defined in the same manner. For example, the addition of two existing trip matrices, HBW and home-based other (HBO), to create a new matrix named total trips (TT) is illustrated in the following example of matrix addition:

$$\text{If} \quad \text{HBW} = \begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix} \text{ and } \text{HBO} = \begin{bmatrix} 5 & 15 & 30 \\ 10 & 20 & 40 \\ 30 & 60 & 120 \end{bmatrix} \text{ then}$$

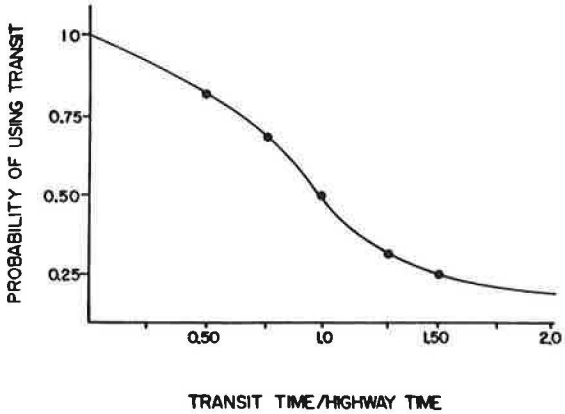
$$\text{TT} = \begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix} + \begin{bmatrix} 5 & 15 & 30 \\ 10 & 20 & 40 \\ 30 & 60 & 120 \end{bmatrix} = \begin{bmatrix} 15 & 35 & 60 \\ 50 & 70 & 100 \\ 100 & 140 & 210 \end{bmatrix}$$

Just as for zonal data, matrix operations are performed on a cell-by-cell basis, so that the row index (I) references the current row and the column index (J) references the current column. One statement performs all operations on the two input matrices regardless of whether they are 3 x 3 or 2500 x 2500.

Logical Operations

Comparisons of zonal and matrix data are accomplished through the use of logical operations.

Figure 5. Mode split to transit.



The "IF a THEN b ELSE c" statement can discriminate values or cells (zones or interchanges) that meet or fail certain criteria. Through the use of relational operations, such as less than (<) or less than or equal to (≤), the planner can identify where certain cell values occur. For instance, if the planner wishes to identify zones in which total employment exceeds 1000 workers, the following expression could be used:

EMPGT1K = IF EMP > 1000 THEN 1 ELSE 0

Thus,

$$\text{If EMP} = \begin{bmatrix} 1500 \\ 500 \\ 1750 \\ 990 \\ 1010 \end{bmatrix} \text{ then EMPGT1K} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Zones 1, 3, and 5 meet the criterion, and zones 2 and 4 fail.

Logical operators are used on matrices for criteria-based system evaluations. For example, assume that TTT = transit travel times, HTT = highway travel times, TRIPS = peak-hour person trip table, and TDTRIPS = transit-deficient person trips. If we want a trip table that contains only those trips for which the transit travel time is more than twice the highway travel time, we form the statement TDTRIPS = IF TTT >2 * HTT THEN TRIPS ELSE 0. Thus,

$$\text{If TTT} = \begin{bmatrix} 27 & 54 & 52 & 96 \\ 60 & 16 & 32 & 30 \\ 30 & 26 & 30 & 32 \\ 65 & 45 & 27 & 15 \end{bmatrix} \text{ and HTT} = \begin{bmatrix} 30 & 28 & 25 & 47 \\ 25 & 15 & 17 & 19 \\ 19 & 14 & 12 & 15 \\ 30 & 16 & 12 & 12 \end{bmatrix}$$

$$\text{and TRIPS} = \begin{bmatrix} 1000 & 1500 & 1750 & 1900 \\ 1200 & 210 & 410 & 500 \\ 1350 & 650 & 720 & 550 \\ 1200 & 230 & 410 & 190 \end{bmatrix} \text{ then}$$

$$\text{TDTRIPS} = \begin{bmatrix} 0 & 0 & 1750 & 1900 \\ 1200 & 0 & 0 & 0 \\ 0 & 0 & 720 & 550 \\ 1200 & 230 & 410 & 0 \end{bmatrix}$$

Look-Up Tables

The UMATRIX look-up tables enable the user to either create new data or alter old. Look-up tables allow the conversion of existing data values to new values, which facilitates transforming, updating, and editing. Look-up tables are lists of arguments and their corresponding converted values. In addition, the planner specifies a missing or "no-hit" conversion value for each table. This value is used every time the referenced value does not match any identified argument.

The special function LOOKUP references these tables and the data to be converted through two components--the look-up table number and an argument. Thus, for example, LOOKUP(1,POP) uses look-up table 1 to convert values drawn from zonal data in the POP category.

In the special use of look-up tables to update zonal data values, the look-up table often requires as an argument a special parameter I, the row indicator, for the cells to be updated, as well as their corresponding new values. In this case the no-hit value is usually set to zero. Therefore, if the planner wishes to update zones 3 and 4 of EMP from the example above, let LOOKUP 1 be as shown below:

Argument	Value
3	2000
4	2200
No-hit	0

The expression EMPUPD = IF LOOKUP (1,I) NOT = 0 THEN LOOKUP (1,I) ELSE EMP creates EMPUPD (which is a slight modification of EMP) by substituting the values in look-up table 1 for the zones that appear as explicit arguments in the table.

In this case, when I = 1, the look-up operation results in a no-hit, since 1 is missing from the look-up table. The value of 0 results and the value for EMPUPD is derived from EMP. The same process occurs for zone 2 (I = 2). However, when I = 3 and I = 4, a hit occurs, and the values for EMPUPD come from the look-up table (i.e., 2000 and 2200, respectively). When I = 5, another no-hit occurs, and the value for EMPUPD comes from EMP. Thus EMPUPD, which represents updated employment values, is shown below:

$$\text{If EMP} = \begin{bmatrix} 1500 \\ 500 \\ 1750 \\ 990 \\ 1010 \end{bmatrix} \text{ then EMPUPD = IF LOOKUP (1,I) NOT = 0}$$

$$\text{THEN LOOKUP (1,I) ELSE EMP creates EMPUPD} = \begin{bmatrix} 1500 \\ 500 \\ 2000 \\ 2200 \\ 1010 \end{bmatrix}$$

Look-up-table data depend on application and consist of four basic optional types: (a) positional: one conversion value exists for every explicit value of the numeric argument from 1 to the maximum argument value; (b) keymatch: conversion values exist only for explicit numeric argument values; (c) range: one specific conversion value exists for a range of argument values; and (d) linear interpolation: if no direct match is found with the argument, the two successive values that bracket it are used to interpolate for the conversion value.

Often it is necessary to input a curve, or a series of curves, to aid in the application of modal split or automobile-pollution emission models. Look-up tables can be used for these calculations. For example, let mode split to transit for HBW trips be a function of the ratio of transit time to highway time (graphically shown in Figure 5). The graph is entered in UMATRIX as a linear interpolation look-up table in the following manner:

Argument	Value
0.25	0.85
0.50	0.80
0.75	0.70
1.0	0.50
1.25	0.30
1.50	0.20
1.75	0.15
2.0	0.12

If the input trip table is named HBW and the travel time ratio is TTR, then the mode-split application is expressed as follows (where the matrix TTRIPS contains total transit trips).

$$\text{If TTR} = \begin{bmatrix} 0.75 & 0.25 & 1.25 \\ 0.50 & 0.50 & 1.30 \\ 1.25 & 1.40 & 1.30 \end{bmatrix} \text{ and HBW} = \begin{bmatrix} 1000 & 200 & 400 \\ 980 & 1000 & 900 \\ 560 & 500 & 800 \end{bmatrix}$$

then TTRIPS = LOOKUP (1,TTR) * HBW creates

$$\text{TTRIPS} = \begin{bmatrix} 700 & 170 & 120 \\ 784 & 800 & 252 \\ 168 & 120 & 224 \end{bmatrix}$$

Special LAV Functions

For purposes of summarization, the planner may wish to aggregate all data values or combine certain values according to zonal-district equivalencies. The functions SUM, COMPRESS, and EXPAND allow the planner to aggregate, replicate, or disaggregate zonal data values.

1. SUM: The SUM function simply aggregates all values within a data item and then replicates that value in the resultant:

$$\text{If POP} = \begin{bmatrix} 500 \\ 990 \\ 1750 \end{bmatrix} \text{ then TOTALPOP = SUM(POP) creates}$$

$$\text{TOTALPOP} = \begin{bmatrix} 3240 \\ 3240 \\ 3240 \end{bmatrix}$$

A common use of SUM is to create percentages, as shown below:

$$\text{If HBWPROD} = \begin{bmatrix} 450 \\ 250 \\ 400 \end{bmatrix} \text{ then HBWTOTAL = SUM(HBWPROD) cre-}$$

$$\text{ates HBWTOTAL} = \begin{bmatrix} 1100 \\ 1100 \\ 1100 \end{bmatrix}$$

For a zonal data item of total trips, data containing percentages of trip productions or attractions are created in one step:

$$\text{If HBWPROD} = \begin{bmatrix} 450 \\ 250 \\ 400 \\ 250 \end{bmatrix} \text{ and HBWATTR} = \begin{bmatrix} 250 \\ 425 \\ 425 \end{bmatrix} \text{ then PERPROD =}$$

HBWPROD ÷ SUM(HBWPROD) and PERATTR = HBWATTR ÷ SUM(HBWATTR)

$$\text{creates PERPROD} = \begin{bmatrix} 450 \\ 250 \\ 400 \end{bmatrix} \div \begin{bmatrix} 1100 \\ 1100 \\ 1100 \end{bmatrix} = \begin{bmatrix} 0.409 \\ 0.227 \\ 0.364 \end{bmatrix} \text{ and}$$

$$\text{PERATTR} = \begin{bmatrix} 250 \\ 425 \\ 425 \end{bmatrix} \div \begin{bmatrix} 1100 \\ 1100 \\ 1100 \end{bmatrix} = \begin{bmatrix} 0.227 \\ 0.386 \\ 0.386 \end{bmatrix}$$

2. COMPRESS: The function COMPRESS has two components: the LAV to be compressed and a "mapping" LAV item. In this process, a LAV of one size usually has its elements combined in such a manner as to result in a LAV that has a smaller number of elements. The mapping LAV has the same number of elements as the LAV to be compressed. Each of its cells holds the new cell index (I) of the recipient LAV, into which values from the compressed LAV are to be entered or aggregated. COMPRESS is very useful for summarizing zonal data to a more aggregate (district) geography. The example below makes this clear:

COMP = COMPRESS(POP, EQUIV) will compress POP, creating COMP by using the mapping in EQUIV thus:

$$\text{If POP} = \begin{bmatrix} 10\,000 \\ 20\,000 \\ 30\,000 \\ 25\,000 \\ 17\,000 \\ 18\,000 \end{bmatrix} \text{ and EQUIV} = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 1 \\ 3 \\ 2 \end{bmatrix} \text{ then COMP =}$$

$$\begin{bmatrix} 20\,000 + 25\,000 \\ 30\,000 + 18\,000 \\ 10\,000 + 17\,000 \end{bmatrix} = \begin{bmatrix} 45\,000 \\ 48\,000 \\ 27\,000 \end{bmatrix}$$

The first element of POP (10 000) has a mapping index of 3 from EQUIV; thus, it is entered in the third position of COMP. The second element of POP (20 000) is entered in the first position of COMP. The fourth element of POP (25 000) is also entered in the first position of COMP. However, the value 20 000 is already there; thus, the two values are added together. Any of COMP's elements not represented in EQUIV are set to zero (i.e., a missing zone). The size of COMP corresponds to the highest zone number that appears in EQUIV.

3. EXPAND: The EXPAND function is used to map elements of data into larger-sized data through element replication. It has two components: the LAV to be expanded and a mapping LAV that is different in size. In this case, the mapping LAV and the resultant will have the same size, i.e., the same number of cells. It is useful for transforming zonal data in order to reference a more disaggregate geography, e.g., subzone.

$$\text{If POP} = \begin{bmatrix} 45\,000 \\ 48\,000 \\ 27\,000 \end{bmatrix} \text{ and MAP} = \begin{bmatrix} 3 \\ 1 \\ 3 \\ 2 \\ 1 \\ 2 \end{bmatrix} \text{ then EX = EXPAND}$$

$$\text{(POP, MAP) creates EX} = \begin{bmatrix} 27\,000 \\ 45\,000 \\ 27\,000 \\ 48\,000 \\ 45\,000 \\ 48\,000 \end{bmatrix}$$

In this case, the mapping LAV (MAP) describes the position of the value from POP that is entered in the appropriate position of the resultant. Thus, the first element of EX contains the third value from POP (27 000), the second element contains the first value from POP (45 000), etc. Note that, although they use the same mapping LAV, COMPRESS and EXPAND are not mutual inverses.

Special Matrix Functions

It is often necessary to transpose a matrix or to summarize matrix elements across rows and columns. To facilitate these operations, two special functions exist in UMATRIX--TR and ROWSUM. The transposition function, TR, transposes a matrix, as shown:

$$\text{If MATRIX} = \begin{bmatrix} 10 & 50 & 70 \\ 20 & 90 & 30 \\ 35 & 95 & 15 \end{bmatrix} \text{ then OUTPUT = TR(MATRIX)}$$

$$\text{creates OUTPUT} = \begin{bmatrix} 10 & 20 & 35 \\ 50 & 90 & 95 \\ 70 & 30 & 15 \end{bmatrix}$$

That is, rows become columns and vice versa. A typical application of transposition occurs when a trip table of productions (P) and attractions (A) must be split to an origin-destination (OD) trip table:

PA = P and A work-trip table; OD = O and D work-trip table; and splitting factors = 70 percent of work trips are P to A and 30 percent of work trips are A to P.

$$\text{If PA} = \begin{bmatrix} 100 & 400 & 250 \\ 700 & 550 & 320 \\ 1000 & 300 & 700 \end{bmatrix} \text{ then OD = 0.70 * PA + TR(PA)}$$

$$\text{OD} = 0.70 * \begin{bmatrix} 100 & 400 & 250 \\ 700 & 550 & 320 \\ 1000 & 300 & 700 \end{bmatrix} + 0.30 * \begin{bmatrix} 100 & 700 & 1000 \\ 400 & 550 & 300 \\ 250 & 320 & 700 \end{bmatrix}$$

$$\text{OD} = \begin{bmatrix} 70 & 280 & 175 \\ 490 & 385 & 224 \\ 700 & 210 & 490 \end{bmatrix} + \begin{bmatrix} 30 & 210 & 300 \\ 120 & 165 & 90 \\ 75 & 96 & 210 \end{bmatrix}$$

$$\text{OD} = \begin{bmatrix} 100 & 490 & 475 \\ 610 & 550 & 314 \\ 775 & 306 & 700 \end{bmatrix}$$

The ROWSUM operator aggregates all matrix elements across a row and then replicates that single sum across the row of the output table:

$$\text{If MATRIX} = \begin{bmatrix} 100 & 150 & 200 \\ 50 & 75 & 125 \\ 100 & 100 & 100 \end{bmatrix} \text{ then OUTPUT = ROWSUM}$$

$$\text{(MATRIX) creates OUTPUT} = \begin{bmatrix} 450 & 450 & 450 \\ 250 & 250 & 250 \\ 300 & 300 & 300 \end{bmatrix}$$

In order to obtain column sums, the transposition function must be used. In this case, transposition creates an output matrix in which the rows are the input columns and the columns are the rows of the input matrix.

Operations that combine both matrix and zonal data allow the user to perform a multitude of planning analyses, such as trip generation and mode-split model application, determination of regional accessibility, and criteria-based system evaluation.

Sometimes the planner needs matrix output that uses zonal data input for manipulative purposes. In this case, since matrix calculations are performed on a cell-by-cell basis, a matrix can be created from a zonal data item by means of column replication. For example, the expression MATRIX = POP will create a matrix in which there are as many rows as there are elements in POP and every column is the LAV POP, thus:

$$\text{If POP} = \begin{bmatrix} 1500 \\ 2000 \\ 750 \end{bmatrix} \text{ then MATRIX = POP creates MATRIX =}$$

$$\begin{bmatrix} 1500 & 1500 & 1500 \\ 2000 & 2000 & 2000 \\ 750 & 750 & 750 \end{bmatrix}$$

Similarly, the transposition of a LAV when equated to a matrix name yields a matrix whose rows are all duplicates of the LAV. Replication of columns or rows when matrices are being output is quite useful in the performance of accessibility analyses. For example, the planner may wish to estimate the percentage of regional blue-collar employment opportunities that are within 30 min (on the existing transit system) of each production zone. To accomplish this, a percentage blue-collar employment LAV (BCEMPPCT) must be created from BCEMP (zonal blue-collar employment data). This is done by means of BCEMPPCT = BCEMP/SUM(BCEMP). Once BCEMPPCT is created, a transit travel time matrix (TTR) is used to determine, for each production zone, whether the time exceeds 30 min. If it does, then none of the blue-collar jobs are accessible; if it does not, then the percentage employment figure must be used:

$$\text{If TTR} = \begin{bmatrix} 10 & 20 & 25 \\ 15 & 8 & 35 \\ 27 & 38 & 9 \end{bmatrix} \text{ and BCEMPPCT} = \begin{bmatrix} 0.10 \\ 0.25 \\ 0.30 \end{bmatrix} \text{ then}$$

$$\text{TR(BCEMPPCT)} = \begin{bmatrix} 0.10 & 0.25 & 0.30 \\ 0.10 & 0.25 & 0.30 \\ 0.10 & 0.25 & 0.30 \end{bmatrix}$$

ACCESSIBILITY = IF TTR > 30 THEN 0 ELSE TR(BCEMPPCT)

$$\text{creates ACCESSIBILITY} = \begin{bmatrix} 0.10 & 0.25 & 0.30 \\ 0.10 & 0.25 & 0 \\ 0.10 & 0 & 0.30 \end{bmatrix}$$

The transposition function is used because, as the destinations change for a given origin, the employment percentages must change accordingly. The statement ACCESS = ROWSUM(ACCESSIBILITY) creates zonal accessibility-to-employment percentages:

$$\text{ACCESS} = \begin{bmatrix} 0.65 \\ 0.35 \\ 0.40 \end{bmatrix}$$

Thus, zone 1 has accessibility to 65 percent of employment opportunities, zone 2 has only 35 percent, and zone 3 has 40 percent.

Gravity Model

Certain demand models can be readily implemented within UMATRIX, e.g., the gravity model. Let PROD be zonal productions, and ATTR be zonal attractions; TIMES has travel times and KFACT has K-factors. The traditional gravity model is effected by means of a single UMATRIX statement:

```
DISTTRIPS = PROD * TR(ATTR) * LOOKUP(1,TIMES) *
   K(FACT)/ROWSUM[LOOKUP(1,TIMES)] * TR(ATTR) *
   KFACT.
```

Look-up table 1 has friction factors for corresponding travel times. Note that the transposition of the ATTR LAV is used, since the attractions change as destinations change.

Trip-Generation Model

UMATRIX first calculates trips from each zone by using a simplistic cross-classification trip-generation model (stratified by two income levels) for HBW and then adjusts trip productions to sum to the attraction total. Zonal production estimates are based on trips per dwelling unit for each income type. The rates for income level 1 are 2.12 trips/dwelling unit; for income level 2, they are 2.54 trips/dwelling unit.

Attraction estimates are developed by using trip-attraction rates and measures of zonal activity, primarily employment. The daily work-trip attraction formula is

Person trip attractions = 1.70 * (total zonal employment).

Let DU1 and DU2 represent zonal dwelling units stratified by income, EMP represent zonal employment, and P1, P2, and ATTR represent data for productions and attractions. The generation equations can be expressed by

P1 = 2.12 * DU1.

P2 = 2.54 * DU2.

ATTR = 1.70 * EMP.

These yield three new data items that represent unbalanced trip-production levels for each zone for each income category, as well as a trip-attraction total for each zone. In this case, after applying the rates, the productions need to be summed to create regional production totals from which a normalization factor can be determined. Then, this factor is multiplied by the initial production estimates, stratified by income, to create balanced (normalized) trip values. The attraction sum-balancing procedure is shown below:

```
      TP = SUM (P1 + P2)
      TA = SUM (ATTR)
  NORMAL = TA/TP
   PROD1 = NORMAL * P1
   PROD2 = NORMAL * P2
```

Impedance Calculations

For demand modeling purposes, this example can be used to create a series of HBW highway and transit impedances. The components of in-vehicle time, out-of-vehicle time, and cost are combined according to the following formula:

Impedance = in-vehicle time + (2.5 x out-of-vehicle time) + [(3600 x cost)/average zonal income],

where highway cost = ($0.15/mile x highway distance) + (0.5 x parking cost) and transit cost = out-of-pocket fare.

Assume that matrix TIV contains transit in-vehicle times, TOV contains out-of-vehicle times, and FARES contains transit fares. Matrix HIV contains highway in-vehicle times, HOV contains out-of-vehicle times, and HDIST contains distance. AVEINC has average zonal income, and PARKC has zonal parking costs in cents. Since each interchange will have a unique set of impedances, matrix output is created by using mixed zonal and matrix operations. The impedance operations are

TIMPS = TIV + (2.5 * TOV) + [(3600 * FARES)/AVEINC].

HIMPS = HIV + (2.5 * HOV) = (3600 * {(0.15 * HDIST) + [1/2 * TR(PARKC)]}/AVEINC).

Transit impedances are stored on TIMPS; the highway impedances are stored on HIMPS. Note that the transposition of PARKC appears because the parking cost is a destination charge.

Application of a Logit Mode-Split Model

To use the transit and highway impedances created above to forecast HBW mode split by means of a logit model, the following UMATRIX statement is all that is needed:

PT = 1/{1 + EX [-0.1 * (HIMPS - TIMPS)]}

where PT = probability of using transit.

SUMMARY AND CONCLUSIONS

This paper has introduced the reader to the UTPS program UMATRIX, a powerful data-base manager. UMATRIX's straightforward algebra-like command language provides the planner with an efficient and flexible planning tool. The UMATRIX user can access a large variety of data forms and can accomplish analytical chores that, prior to now, were difficult and time consuming. By allowing a large range of complex transportation planning tasks to be performed with less effort, UMATRIX gives the planner more time to devote to the important functions of system evaluation, impact assessment, and planning recommendations.

ACKNOWLEDGMENT

Many individuals contributed significantly to the development of UMATRIX and to the preparation of this paper, specifically Ernest Luther (Planning Research Corporation), Stan Yutkins (Kentron Hawaii Limited), and Fred Dickson and Rich Regan (National Institute for Community Development). Also, Yehuda Gur merits acknowledgment for the idea of zonal files. This paper and the software discussed are the products of research and development that UMTA funded but they do not necessarily reflect UMTA policy.