Solving the Sequential Travel Forecasting Procedure with Feedback

David Boyce, Christopher R. O'Neill, and Wolfgang Scherr

Travel forecasters generally understand that an iterative solution of the sequential travel forecasting procedure is required to bring specific model inputs and outputs into consistent agreement. In particular, the congested interzonal travel time inputs to the trip distribution and mode choice steps should equal the user-equilibrium travel times obtained from the assignment step. The process of achieving consistency is called solving the sequential procedure with feedback. The Capital District Transportation Committee of Albany, New York, maintains a travel forecasting model with 1,000 traffic analysis zones. This model was used to evaluate feedback procedures for three applications drawn from its planning activities. Three alternative feedback solution procedures were applied to the model: (*a*) naïve or direct feedback (no averaging of trip matrices or link flows), (*b*) averaging of trip matrices with constant weights, and (*c*) the method of successive averages (MSA) applied to trip matrices. The convergence of the feedback procedures was measured by comparing the results as follows: total misplaced flow (trip matrices), relative gap (route-based user-equilibrium traffic assignments), and root squared error (travel cost matrices). The test results showed that (*a*) averaging of trip matrices by using constant weights converges to a single, stable solution with consistent travel costs; (*b*) a single pair of weights is most effective for all three applications; (*c*) neither naïve feedback nor MSA is as effective as use of constant weights; and (*d*) the relative gaps of the traffic assignment reach values of less than 10^{-7}. Tests with different models and software systems are needed to generalize the findings.

Travel forecasters have understood since the first conceptualization of the sequential travel forecasting procedure (also called the four-step procedure) during the 1950s that an iterative approach is required to bring the inputs and outputs of specific models into consistent agreement with each other (*1*). In particular, the congested interzonal travel costs (shortest route travel times, or skims) which are input to the trip distribution and mode choice steps should equal the user-equilibrium travel costs resulting from the solution of the assignment step. The iterative process of achieving consistency between these input and output travel costs is generally referred to as solving the sequential procedure with feedback.

Although the need to solve the four-step procedure iteratively was apparent, how to solve this problem has confounded numerous practitioners and researchers. At the 1993 Transportation Planning Applications Conference, the first author organized a session on this topic, which led to the computational experiments reported by Boyce et al. (*2*). Subsequently, Comsis undertook a study of how to incorporate feedback in travel forecasting (*3*); the results were inconclusive. Recently, however, Boyce and Xiong performed computational experiments that appeared to be more promising (*4*).

In all these investigations, the apparent difficulties stemmed from the question of just how to perform the feedback calculations in the sequential procedure. The experiments of Boyce et al. (*2*) and Comsis (*3*) demonstrated that it is ineffective to feed back the travel costs directly from the previous loop (naïve feedback, sometimes called direct feedback). Clearly, some sort of averaging between successive loops is required. But what should be averaged? Various investigators have averaged link flows, link costs, or even link speeds.

To gain a clearer view of the problem, the goal should be reconsidered. The goal is to find a multimodal trip matrix that depends in part on interzonal generalized modal travel costs (linearly weighted sums of travel times, operating costs, tolls or fares). For congested modes (typically, car), these travel costs should be either costs of the shortest routes or average costs of used routes, depending on whether a link-based or a route-based assignment algorithm is applied. To achieve this objective, a multimodal trip matrix, partly dependent on modal travel costs, is sought, which if assigned to the multimodal network yields those same costs. This problem statement suggests that one should focus on finding the multimodal trip matrix that satisfies this criterion rather than focus on link flows or costs. This line of thinking led to a feedback procedure that involves averaging of trip matrices. Persons with a mathematical background may recognize this as the statement of a fixed-point problem (*5, 6*).

This paper reports on tests of three ways of averaging successive trip matrices:

1. Averaging with constant weights (sometimes referred to as fixed weights);
2. Averaging with the method of successive averages (MSA), in which the weight on each new solution matrix decreases with each feedback loop; and
3. Naïve feedback (constant weight averaging with a weight of 1.0 on the new matrix).

DESCRIPTION OF MODEL

The Capital District Transportation Committee (CDTC) is the metropolitan planning organization for the New York State counties of Albany, Rensselaer, Saratoga, and Schenectady, which have a total population of about 800,000, primarily in the central cities of Albany, Troy, Schenectady, and Saratoga Springs and the contiguous suburban development, as shown in Figure 1. CDTC has 25 members, including

D. Boyce, Department of Civil and Environmental Engineering, Northwestern University, Technological Institute, 2145 Sheridan Road, Room A236, Evanston, IL 60208. C. R. O'Neill, Capital District Transportation Committee, 1 Park Place, Albany, NY 12205. W. Scherr, PTV America, Inc., Suite 704, 1300 North Market Street, Wilmington, DE 19801. Corresponding author: D. Boyce, d-boyce@northwestern.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2077, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 129–135.
DOI: 10.3141/2077-17

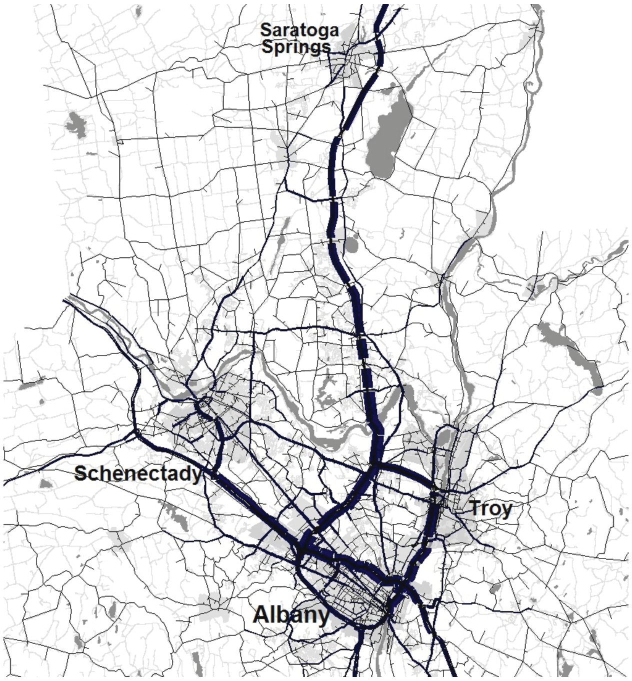

FIGURE 1 CDTC travel forecasting model network.

the New York State Department of Transportation (NYDOT), the New York State Thruway Authority, the Capital District Transportation Authority, the Capital District Regional Planning Commission, the Port of Albany, Albany International Airport, and elected officials from counties, cities, and towns. CDTC made extensive use of its regional travel forecasting model in preparing its New Visions Regional Transportation Plan, as well as for its Community and Transportation Planning Linkage Program, evaluation of Transportation Improvement Program project candidates, and development of performance measures. CDTC also provides traffic forecasts to NYDOT for project development. Forecasts assume steady progress in the integration of land use and transportation plans, urban reinvestment, demand management, and pedestrian, bicycle, and transit access.

The current travel model includes five trip purposes through trip generation and trip distribution, which are augmented by through traffic. The trip distribution models are doubly constrained gravity models with negative power functions. The only mode currently represented in the model is car travel; public transit is not yet included in the model. Vehicle assignments are computed for the evening peak hour utilizing capacities defined on Level of Service C volumes, not only on links but also on turns to emulate intersection delay. The network has 1,000 zones, 10,000 links, 4,000 nodes, and 21,000 turns with volume-delay functions. CDTC's modeling practice is based on VISUM; the tests reported here were computed with VISUM 10.0 (*7*). The computer used to solve the model was an off-the-shelf Windows PC with a 2.0-GHz processor and 2.0 GB of random-access memory, purchased in 2006.

Three applications of the model are considered:

1. Base—model calibration with the 2000 census;
2. Plan—2030 forecast for the planned network; and
3. Base_1.5—derived from the base by multiplying productions and attractions by a factor of 1.5 to obtain a more congested application for these tests.

FEEDBACK METHOD

The proposed feedback procedure is depicted in Figure 2 for a problem with a single congested mode (car) operating on a road network. The initial solution to the four-step procedure is represented by Box 2, consisting of trip distribution and assignment, given an assumed travel cost matrix and the zonal origin and destination totals by trip purpose in Box 1. Box 3 represents the general solution of the trip distribution model in loop k based on the travel cost matrix computed from the assignment of the averaged trip matrix in loop $k - 1$; in Loop 2, the trip distribution is solved by using the travel costs from the initial solution (Loop 1) in Box 2. In Box 4, this new trip matrix from Box 3 is averaged with the trip matrix from the previous loop $k - 1$, yielding the averaged trip matrix of loop k.

1. Input data:
   (O_i) and (D_j) by trip purpose for zones i, j
   Road network

2. Compute the initial solution (Loop 1):
   Initialize travel times/costs $\Rightarrow c_{ij}(1)$
   Solve trip distribution $\Rightarrow e_{ij}(1) = d_{ij}(1)$
   Assign $d_{ij}(1)$ to road network $\Rightarrow f_a(1)$

3. Compute the solution for Loop k:
   Compute costs of used routes $\Rightarrow c_{ij}(k)$
   Solve trip distribution $\Rightarrow e_{ij}(k)$

4. Average trip matrices $d_{ij}(k-1)$ and $e_{ij}(k)$:
   CW: $d_{ij}(k) = (1-w) \cdot d_{ij}(k-1) + w \cdot e_{ij}(k)$, or
   MSA: $d_{ij}(k) = \left(\frac{k-1}{k}\right) \cdot d_{ij}(k-1) + \left(\frac{1}{k}\right) \cdot e_{ij}(k)$

5. Assign $d_{ij}(k)$ to the road network to the desired level of convergence $\Rightarrow f_a(k)$

6. Check convergence of $e_{ij}(k)$ to $d_{ij}(k-1)$:
   $\text{TMF} = \sum_{ij} \left| d_{ij}(k-1) - e_{ij}(k) \right| \le \text{E}$, or
   $\text{RSE} = \left(\sum_{ij} \left(d_{ij}(k-1) - e_{ij}(k) \right)^2 \right)^{1/2} \le \text{E}$
   If converged, STOP; if not, set $k = k + 1$.

k: Loop index w: Constant weight
E: Feedback convergence target
$d_{ij}(k)$: Averaged matrix for assignment
$e_{ij}(k)$: New matrix from distribution
TMF: Total misplaced OD flow
RSE: Root squared OD error
MSA: Method of successive averages
CW: Constant weights

FIGURE 2 Procedure for averaging matrices: (O = origin, D = destination).

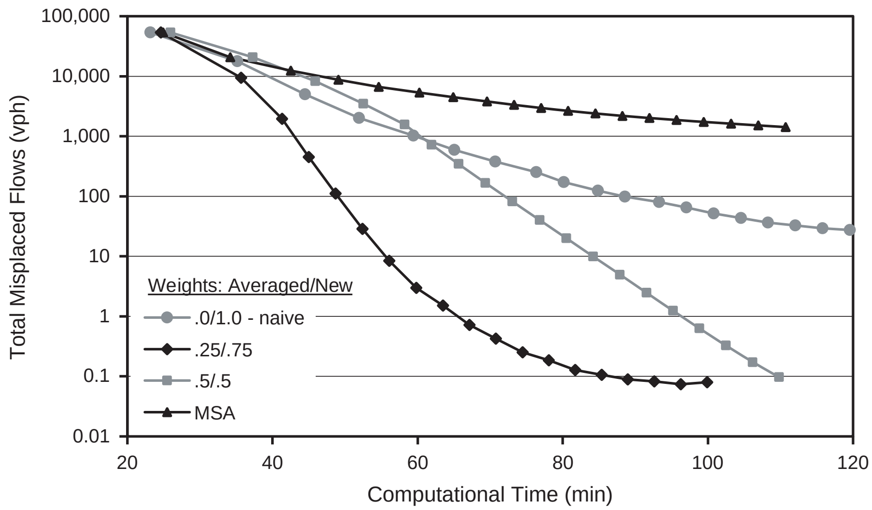

FIGURE 3 Convergence of trip matrices for base.

(Technical terms for the averaged and new matrices are main problem solution and subproblem solution, respectively.)

Averaging may be performed with constant weights with a weight of $(1 - w)$ on the averaged matrix and a weight of w on the new matrix. Alternatively, MSA can be applied; the weight on the averaged matrix is then $(k - 1)/k$, and the weight on the new matrix is $1/k$, which is called the step size. Naïve feedback is a special case of constant weights with $(1 - w) = 0$ and $w = 1$. The averaged matrix is assigned to the road network (Box 5), and a convergence test is performed in Box 6 to test whether the two matrices averaged in Box 4 are effectively equal. MSA can be mathematically proved to converge to the desired consistent solution, but convergence may be very slow. Averaging with constant weights has not been proved to converge.

Two useful convergence measures, total misplaced flow (TMF) and root squared error (RSE), are defined in the flowchart. (As both measures yield essentially the same pattern, mainly TMF is shown in the results.) If the convergence measure selected is not sufficiently small, the procedure returns to Box 3 and begins again. Note that only trip matrices are averaged; link flows, link costs, and zone-to-zone costs are computed from the assignment in Box 5 and used directly in solving the trip distribution model again.

COMPUTATIONAL RESULTS

Convergence of Solution Procedure

For each of the three applications, solutions with various weights were computed: constant weights with values of w ranging from 1.0 down to 0.3 plus MSA. TMF in vehicles per hour (vph) is plotted in log scale versus computational time in minutes for the base application in Figure 3; equivalent results were obtained for the two other applications and therefore are not shown. Various symbols depict results for each loop. Figure 4 shows a comparison of the base application with the plan and base_1.5 applications for the most effective constant weights ($w = 0.75$) as well as MSA and naïve feedback. For more values of w, see www.trb-appcon.org/2007conf/program.html#s2.

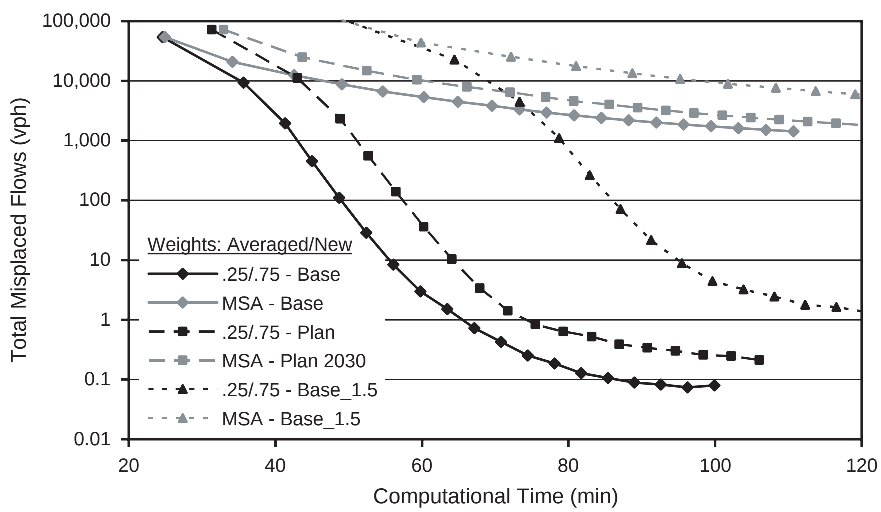

FIGURE 4 Convergence for three applications.

The patterns in the reduction of TMF are similar for the three applications for values of *w* of 0.5 to 0.8. A value of 1.0 is much less effective, especially in the most congested application, Base_1.5. MSA is ineffective in all three applications. Note that a *w* value of 0.5 is quite effective after 20 loops but somewhat ineffective earlier. In addition to the stability of the overall patterns, it is important that a *w* value of 0.75 is effective for all three applications, implying that the same constant weights can be applied to all future plan scenarios, adjustment of the constant weights for each scenario being unnecessary.

MSA was adopted by many practitioners, evidently because of its mathematically proved convergence properties. However, few tests of convergence appear to have been conducted with models used in practice. Moreover, MSA has often been applied to average successive link flows, which is essentially a form of smoothing of the inputs to the trip distribution model, and may not yield a consistent trip matrix. Why MSA is ineffective for these applications is unknown and requires further study.

To consider how many feedback loops are required for practical applications of the method, see Figure 5. For any method, the minimum number of feedback loops is three: the initial solution, a second loop to determine the level of disparity from the initial solution, and a third loop to check whether averaging of the first two solutions is adequate. The real question, then, is how many more loops are needed. For a *w* value of 0.75, at Loop 4 TMF is reduced to 4,500 vph, or 1.4% of the total origin–destination (O-D) flow of 313,310 vph. Loop 5 reduces TMF to 1,100 vph or 0.4%, which appears sufficient for comparisons of alternative plans. In contrast, MSA results in a TMF of more than 17,600 vph or 6% of total flow after five loops with a similar solution time for a *w* value of 0.75. Naïve feedback ($w = 1.0$) has a TMF of 27,200 vph, or 9% of total flow with a higher solution time. Naïve feedback is much slower because more assignment iterations are required to reach the specified level of assignment convergence. TMF should be less than 1% of total O-D flow.

For all three applications, five feedback loops with a *w* value of 0.75 are sufficient to achieve a reduction in TMF to less than 1% of total flow. At this point in the solution process, however, the rate of improvement in TMF from additional feedback loops is very high, so that one or two more loops may well be worthwhile for this best value of *w*. In contrast, the rate of improvement in TMF is slow for MSA and naïve feedback.

Unique Trip Matrix Solution

An important question raised about these results concerns whether the proposed averaging procedure converges to the same trip matrix for the various weights applied. To investigate this matter, the feedback procedure was solved with a *w* value of 0.5 for 100 loops, yielding a precisely converged solution with TMF = 0.003. Then, the procedure was applied with values of *w* ranging from 1.0 down to 0.3, as well as MSA for 20 loops. The sum of the absolute differences in cell values between each trip matrix and the precisely converged matrix was computed at the conclusion of each loop.

Figure 6 shows the results of this computational experiment. For all of the constant weights, except *w* values of 1.0 and 0.3, the trip matrix converges to a very close approximation of the precisely converged matrix. For naïve feedback and MSA, convergence is not achieved; however, it might be achieved after a very large number of feedback loops. In Figure 6, the *x*-axis shows the number of feedback loops, rather than computational time, because the computational times are not comparable to Figures 3 through 5. Moreover, plotting the result in this manner facilitates the comparison of the total differences in flow at each feedback loop. It is concluded from this experiment that all constant weight methods between values of *w* of 0.4 and 0.9 enable the sequential procedure to converge quickly toward the same unique trip matrix.

Convergence of Travel Cost Matrices

In addition to examining the convergence of trip matrices, which partly depends on the zone-to-zone travel costs, whether the successive travel cost matrices are converging to a stable value also should be determined. To examine whether this criterion is satisfied, the RSE of pairs of successive travel cost matrices were computed. The results for the plan are shown in Figure 7. Similar plots were obtained for the other two applications. Note that the cost matrices computed from the route-based assignment are the flow-weighted average costs

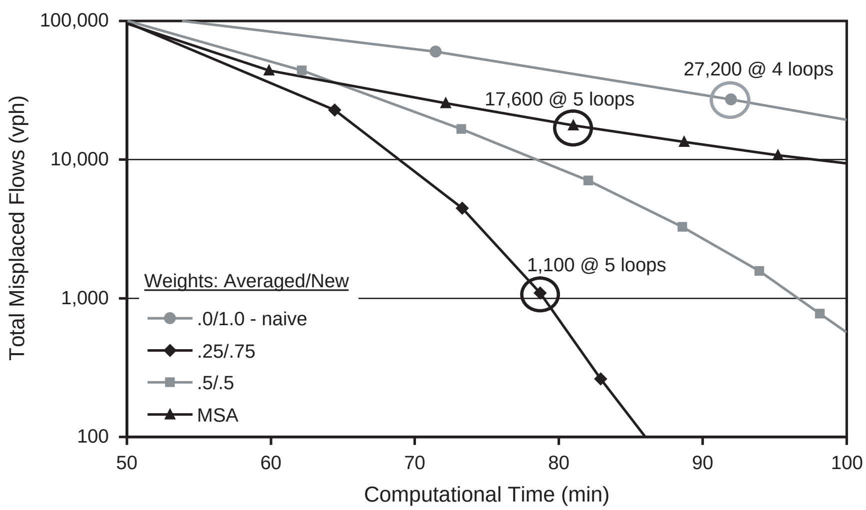

FIGURE 5 Proposed stopping point for Base_1.5.

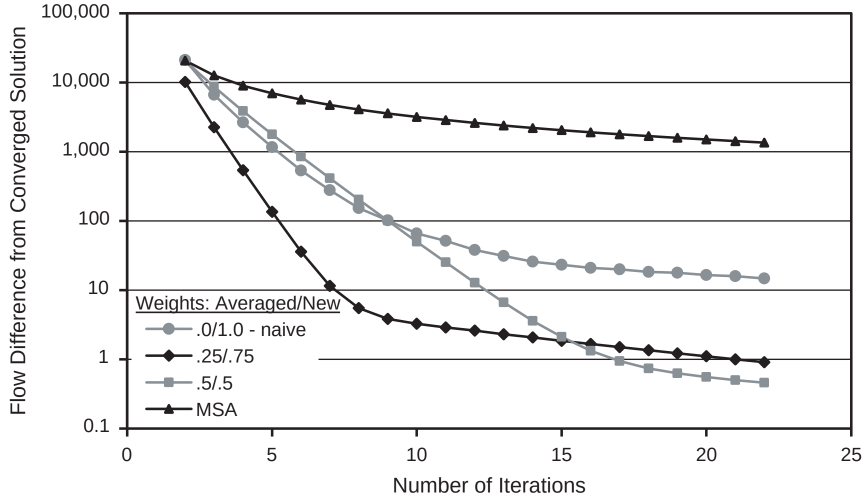

FIGURE 6 Convergence to the same matrix for plan.

over all used routes, not the costs of the current shortest routes as found in link-based assignments.

In the initial solution, free-flow travel times are used to compute the first trip distribution. The second trip matrix is then based on the average travel costs from a multipath assignment of the initial trip matrix. The RSE for these two travel cost matrices is 17,300 min, as shown by the first cluster of points in the upper left of Figure 7. For the best w value of 0.75, RSE decreases sharply to a value less than 1. For MSA and naïve feedback, RSE decreases more slowly, again suggesting these methods do not converge as quickly as the best constant weights.

Convergence of Traffic Assignments

Each loop of the iterative sequential procedure requires the solution of an assignment problem, which is an iterative procedure in itself. In these tests the assignment is solved with a route-based user-equilibrium method described by Bothner and Lutter (*8*) and Schittenhelm (*9*), rather than the link-based methods generally found in other travel forecasting software systems.

The route-based algorithm, in contrast to link-based methods, stores all routes that belong to the current solution. The knowledge of routes, and of the relationships between routes and links, is used to shift O-D flows among alternative routes until the route costs are equal and minimal for each O-D pair; this shifting of route flows is called balancing. In its outer iteration the assignment algorithm performs (*a*) a shortest route search for all O-D pairs; (*b*) several inner iterations of balancing over all O-D pairs, including simultaneously updating the link and turn travel times; and finally (*c*) a convergence test defined on the maximum relative gap (RG). The user of the route-based assignment controls the maximum number of outer loops, the maximum number of balancing iterations per main loop, and the level of RG as a stopping criterion. The settings for these tests are unlimited number of outer loops; maximal number of balancing iterations less than or equal to 3; and RG $< 10^{-5}$ (for the initial solution, RG $< 10^{-1}$).

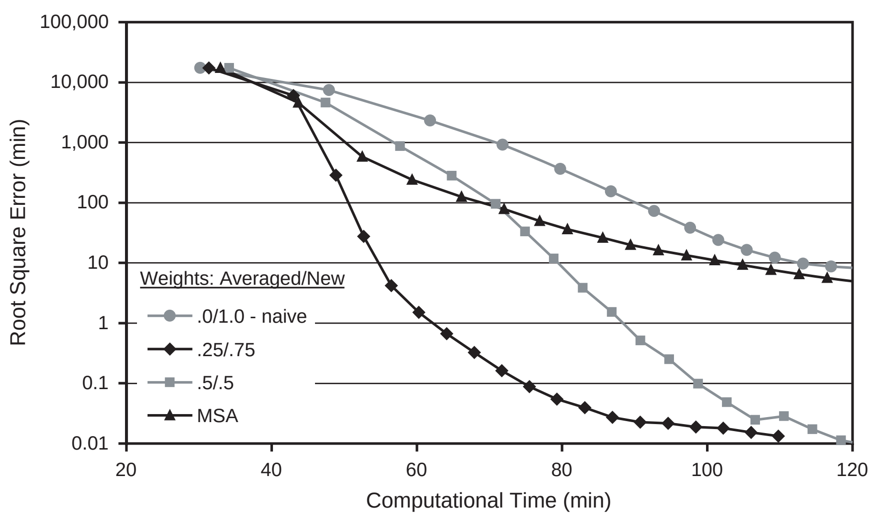

FIGURE 7 Convergence of travel costs for plan.

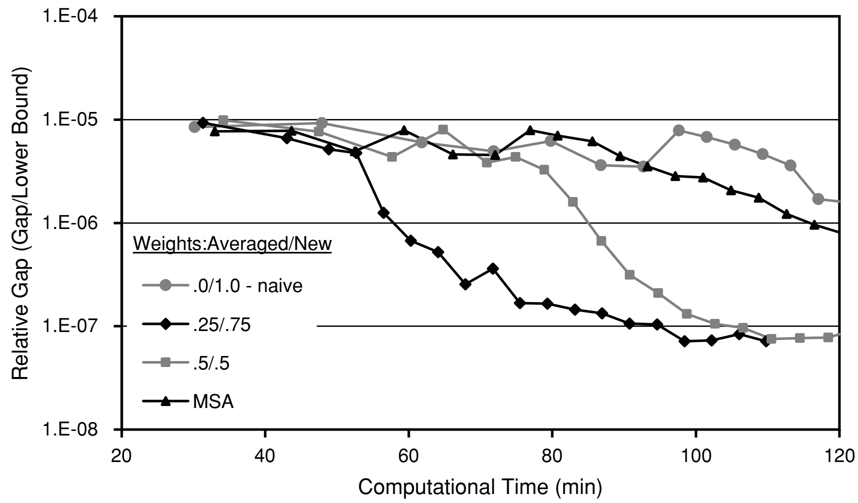

FIGURE 8 Convergence of relative gaps for plan.

The convergence of the assignment steps is shown in Figure 8. The level of assignment convergence (measured by the RG) achieved is far better than what most practitioners are able to achieve with link-based methods [typically on the order of 10^{-3}, as described by Boyce et al. (*10*)]. For the most efficient feedback methods (CW with *w* values between 0.5 and 0.9), the RG decreases to 10^{-7} after 15 to 20 feedback loops in all three applications. After six feedback loops, a very stable assignment with an RG less than 10^{-6} is reached, which means that relatively little noise would be found if two plan alternatives were compared.

In the later feedback loops the RG decreases well beyond the stopping criterion of 10^{-5}. During each loop, the route flow from the previous assignment is adjusted with the updated averaged trip matrix (called a reload). After each reload the algorithm performs a minimum of two route searches and a several balancing steps, which is why the assignments continue to improve below the stopping criterion.

Effect of Initial Travel Cost Assumption

In all of the reported results, free-flow travel costs were used to find the initial trip matrix. Suppose a travel cost matrix from a related problem was used instead. Would the convergence of the feedback procedure be improved? To explore this question, the feedback procedure was initiated with a previous solution to a similar problem. For the best constant weights, relatively little improvement was observed with this approach.

Effects of Level of Congestion

The applications considered in this paper were performed on a zone system and road network for a medium-size metropolitan region (800,000 population) with moderate levels of congestion during the peak hour in the base year. Practitioners from larger metropolitan areas have questioned whether the findings are applicable to more congested conditions. Although unable to answer this question, the congestion levels for the three applications are documented here for comparison with future findings. Table 1 shows some key indicators of the size of the problem, the number of links and turns with volumes (*v*) in excess of 1.25, and the space–mean–speed (vehicle miles traveled–vehicle hours traveled). The capacities (*c*) in the volume-delay functions in this model are set for level-of-service C. Therefore, a volume greater than or equal to 1.25 times the capacity is equivalent to level-of-service E or F.

In the CDTC model, travel times are not only based on link delays, but also on intersection delays, where intersection delays are modeled by turn-based volume-delay functions. Table 1 shows that turn delays contribute more to overall network delay than link delays. Experience with other link- and turn-based models has shown that during the equilibrium assignment and feedback iterations, the volume-to-capacity ratios are higher on the turns than on the links.

CONCLUSIONS AND RECOMMENDATIONS

From the tests conducted for three applications, the following conclusions are drawn:

1. Averaging the trip matrix by using constant weight values of *w* in the range of 0.5 to 0.8 yields stable and highly converged solutions to the problem of solving the sequential procedure with feedback. This result agrees with the findings of Bar-Gera and Boyce (*6*) for a research model with well-defined convergence properties.
2. The same *w* values of 0.75 were highly satisfactory for three applications with quite different congestion levels. This finding is significant because it suggests weights that are good for one application can be transferred to another application without extensive testing.

TABLE 1 Characteristics of Networks and Solutions

Indicator Variable	Base	Plan	Base_1.5
Total O-D flow (vph)	207,540	233,100	311,310
No. of links with capacities	9,760	9,790	9,760
No. of links with v/c >1.0	280	420	830
No. of links with v/c >1.25	80	140	340
No. of turns with capacities	21,300	21,390	21,300
No. of turns with v/c >1.0	250	350	660
No. of turns with v/c >1.25	100	300	150
Space–mean–speed (mph)	34.1	31.9	27.4

3. Performing feedback without averaging (naïve feedback) is relatively ineffective and should not be used. The method of successive averages is much less effective than use of constant weights in these tests. MSA should be used only if constant weights have been shown to be ineffective, as could happen with another model.

4. For the tests conducted, performing five feedback loops was effective in reaching convergence, as measured by TMF. Additional loops improved the convergence to some extent; divergence of the solution was not observed.

5. Tests are required for each practitioner's model to determine the effective number of feedback loops. As a general guideline, the ratio of TMF to the total flow among all origins and destinations should be less than 1%.

6. Procedures using link-based algorithms for vehicle assignment may need more feedback loops because the assignment often is unable to achieve precise convergence levels.

The experience accumulated to date with VISUM pertains to three applications solved with the CDTC model. Additional tests with more complex models and other software systems are needed to generalize these findings further. Practitioners are urged to perform their own tests and report them such that findings across models, networks, and software systems can be compared.

ACKNOWLEDGMENTS

The authors thank Hillel Bar-Gera for advice on the design of the proposed feedback method and for comments on the findings. Comments of several practitioners and three reviewers are also acknowledged.

REFERENCES

1. Carroll, J. D., Jr., and H. W. Bevis. Predicting Local Travel in Urban Regions. *Papers and Proceedings, The Regional Science Association,* Vol. 3, 1957, pp. 183–197.
2. Boyce, D. E., Y.-F. Zhang, and M. R. Lupa. Introducing "Feedback" into Four-Step Travel Forecasting Procedure Versus Equilibrium Solution of Combined Model. In *Transportation Research Record 1443,* TRB, National Research Council, Washington, D.C., 1994, pp. 65–74.
3. Comsis Corporation. *Incorporating Feedback in Travel Forecasting: Methods, Pitfalls and Common Concerns.* DOT-T-96-14. FHWA, U.S. Department of Transportation, 1996.
4. Boyce, D., and C. Xiong. Forecasting Travel for Very Large Cities: Challenges and Opportunities for China. *Transportmetrica*, Vol. 3, 2007, pp. 1–19.
5. Bar-Gera, H., and D. Boyce. Origin-Based Algorithms for Combined Travel Forecasting Models. *Transportation Research,* Vol. 37B, 2003, pp. 405–422.
6. Bar-Gera, H., and D. Boyce. Solving a Nonconvex Combined Travel Forecasting Model by the Method of Successive Averages with Constant Step Sizes. *Transportation Research,* Vol. 40B, 2006, pp. 351–367.
7. *VISUM 10.0 Manual.* PTV AG, Karlsruhe, Germany, 2007.
8. Bothner, P., and W. Lutter. *Ein direktes Verfahren zur Verkehrsumlegung nach dem ersten Prinzip von Wardrop.* University of Bremen, Germany, 1982.
9. Schittenhelm, H. On the Integration of an Effective Assignment Algorithm with Path and Path-Flow Management in a Combined Trip Distribution and Traffic Assignment Algorithm. Presented at 18th Petroleum Technology Research Centre Summer Annual Meeting, Sussex, United Kingdom, 1990.
10. Boyce, D., B. Ralevic-Dekic, and H. Bar-Gera. Convergence of Traffic Assignments: How Much Is Enough? *ASCE Journal of Transportation Engineering*, Vol. 130, 2004, pp. 49–55.

The Transportation Planning Applications Committee sponsored publication of this paper.