Network Equilibration with Elastic Demands

NATHAN H. GARTNER

Elastic-demand equilibration (assignment) is an analytical model for travel forecasting in homogeneous and multimodal transportation networks in which the demand for travel between each origin-destination (O-D) pair is an elastic function of the service level offered by the network. The problem was formulated as a mathematical optimization program in 1956 and, since that time, a variety of iterative schemes have been proposed for its solution. In this paper, the mathematical-programming formulation of the network-assignment problem (NAP) with elastic demands is examined, an economic rationale for its optimization objective is derived, and an efficient method for its solution is presented. The method is based on modeling the NAP as an equivalent-assignment problem in an expanded network. The variable-demand NAP is thus transformed into a fixed-demand NAP that has a trip table that consists of the potential O-D travel demands and can therefore be solved by any fixed-demand assignment procedure available.

Conventional traffic assignment--the final phase in the travel-forecasting procedure--calculates loadings on a network of transportation facilities given the predicted interzonal travel demands. The result of the assignment is an estimate of user volumes and associated performance measures on each segment of the transportation network. The interzonal demands are usually assumed to be fixed and are estimated by earlier stages of the analysis. In the traditional urban transportation planning method, these stages consist of trip generation, trip distribution, and modal split. The user volumes may be determined by the number of vehicles, the number of persons, the number of transit riders, or any other measure that has an origin, destination, and some quantifiable trip-interchange characteristic (1).

A large variety of assignment techniques have been developed; those most frequently used are based on heuristic procedures, such as capacity restraint or probabilistic multipath assignment (2). During the last decade, a number of assignment methods have been introduced that are based on mathematical programming. In general, these methods model the assignment problem as a multicommodity convex cost-minimization problem in which each origin-destination (O-D) flow is considered to be a different commodity. Reviews and discussion of the methods may be found in papers by Gartner (3) and by Nguyen (4). The main advantage of these methods is that they provide access to efficient network-optimization techniques that are both mathematically rigorous and computationally predictable and therefore offer improved analysis capabilities.

A more-general class of problems in transportation-network analysis (one that has a sounder behavioral foundation) is to equilibrate (assign) traffic with elastic demands. The basic premise is that trips are undertaken by persons who (a) have a range of choices available to them and (b) are motivated by economic considerations in their decisions. Thus, the total amount of travel between any O-D pair and the mode chosen for the travel are considered to be a function of the perceived benefit (or disbenefit) to the potential travelers between this O-D pair. The problem was originally described in 1956 in a seminal study on the economics of transportation (5) in which it was also formulated as an equivalent mathematical optimization program. Over the years, this problem has attracted considerable attention, since it was recognized to have a wide range of applications in the analysis of transportation networks (6). A number of specialized techniques have been proposed for solution of the problem, all of which are based on various iterative schemes for equilibration of demand and supply in a network. I do not dwell in this paper on the various possible applications of the problem. Its main application recently has been in the development of multimodal equilibrium models in which the demand for each mode is an elastic function of the service levels offered by the mode (7-10). My purpose is to encourage use of the models and develop new applications through improved understanding of their formulation and the development of more-efficient computational techniques for their implementation.

In this paper, the formulation of the network-assignment problem (NAP) with elastic demands as a mathematical optimization program is reexamined, an economic foundation for its optimization objective is identified, and an efficient method for its solution is presented. The method is based on reformulating the problem as an equivalent-assignment problem in a modified network. The variable-demand NAP is thus converted into a fixed-demand NAP in which there is a trip table given by certain (fixed) potential demands. As a consequence, any technique available for fixed-demand network assignment becomes directly applicable to the more-general NAP with elastic demands.

MATHEMATICAL FORMULATION

In this section the NAP with elastic demands is formulated as a mathematical optimization problem. A transportation network is considered that consists of N nodes and L links. Some of the nodes represent centroids, i.e., origins and destinations of traffic. Between each O-D pair (i,k) there exist, in general, P_{ik} distinct possible paths. M denotes the set of all O-D trip interchanges (i,k) in the network. The following variables are used:

f_j = flow on link j;
$c_j(f_j)$ = average cost of travel (or, in general, the level of service) on link j;
$m_j(f_j)$ = marginal cost of travel on link j;
g_{ik} = trip rate from i to k;
h_p = flow on path p;
C_{ik} = average path travel cost from i to k;
$G_{ik}(C_{ik})$ = demand function for travel from i to k, i.e., trip rate as function of interchange travel cost; and
$W_{ik}(g_{ik})$ = $G_{ik}^{-1}(C_{ik})$ = inverse of demand function, i.e., interchange travel cost as a function of trip rate.

The following integral functions are defined:

$$Z_j = \int_0^{f_j} c_j(z)dz$$

$$Y_{ik} = \int_0^{g_{ik}} W_{ik}(y)dy$$

If, for convenience, a link-path formulation is used, the elastic-demand NAP consists of the following equivalent mathematical optimization program.

Determine link flows f_j and O-D trip rates g_{ik} so that

$$\max\left(\sum_m Y_{ik} - \sum_L Z_j\right) \tag{1}$$

is subject to

$$\sum_{P_{ik}} h_p = g_{ik} \qquad h_p, g_{ik} \geq 0 \tag{2}$$

The link flows are related to the path flows by means of

$$f_j = \sum_M \sum_{P_{ik}} a_{jp} h_p \tag{3}$$

where a_{jp} = 1 if link j is on path p, or 0 otherwise.

According to the theory of mathematical programming, an optimal solution to the NAP (indicated by an asterisk) is characterized by the Kuhn-Tucker necessary conditions:

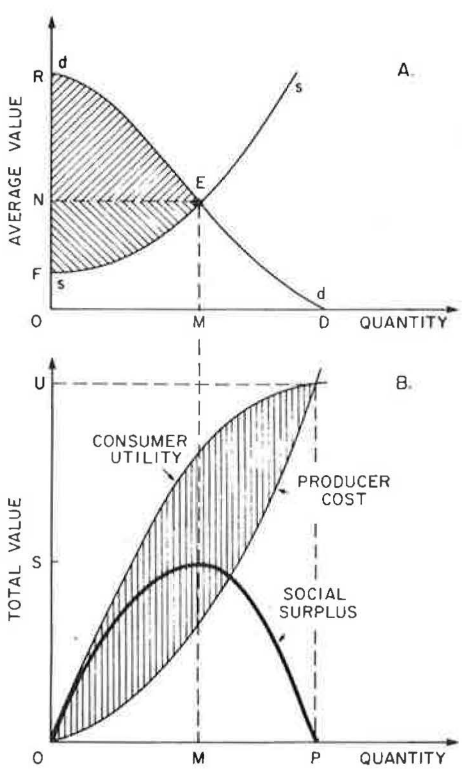
$$\sum_j a_{jp} c_j^*(f_j^*) \left\{\begin{matrix} = \\ \geq \end{matrix}\right\} C_{ik}^* \qquad \text{if } h_p^* \left\{\begin{matrix} \geq \\ = \end{matrix}\right\} 0 \tag{4}$$

$$W_{ik}^* = C_{ik}^* \tag{5}$$

Equation 4 represents the network-equilibrium condition that corresponds to Wardrop's first principle; i.e., travel costs on all routes used between any O-D pair are equal to or less than those on unused routes. W_{ik}^* is the cost (level of service) that generates the demand g_{ik}^* that, at optimality, has to be equal to the average path costs. When the objective function is convex, the necessary conditions are also sufficient. Commonly used link performance functions (such as the Bureau of Public Roads volume-delay function) and O-D demand functions (such as those of the simple gravity type) are, in general, convex with respect to cost.

If demand is inelastic (i.e., if it is given by a fixed value rather than by a function), the first term in expression 1 is a fixed quantity and can be eliminated from the optimization objective. The fixed-demand NAP objective is then simply $\min \Sigma_j Z_j$. The term "user optimization" has been coined for this problem (11).

Figure 1. Market equilibrium paradigm.



INTERPRETATION OF OPTIMIZATION OBJECTIVE

Several economists have studied the effects of transportation costs on equilibrium prices in spatially separated markets in the early 1950s, notably Nobel laureates Koopmans and Samuelson. However, Beekmann, McGuire, and Winsten (BMW) (5) adapted their results to the travel market by considering trip making itself as the commodity that is traded. They discuss the computational aspects of their mathematical formulation but neglect to furnish an economic justification for the optimization objective. This has led some analysts to argue that there is no such justification and that the formulation is an artificial construct (12), whereas other analysts (10,13,14) believe that the equilibrium-NAP objective implies the maximization of consumer surplus. BMW specifically warn against the adoption of this simple interpretation, which is valid only in capacity-free networks, i.e., when link costs are independent of volumes, a rather restrictive assumption that is of little practical value. In this section, the BMW formulation is reexamined and it is shown that its optimization objective can be rationalized on the basis of accepted economic criteria.

Market Equilibrium

The equilibrium market price is where the demand (d-d) and supply (s-s) curves intersect (point E, Figure 1a). At this point, consumers buy and producers supply quantity OM at price ON. If we assume that money provides a firm measuring rod of utility, the areas in Figure 1 represent the following values:

OMEN = total revenue paid by consumers to producers,
OMER = total use to consumers,
OMEF = total cost to producers,
NER = OMER - OMEN = consumer surplus,
NEF = OMEN - OMEF = producer surplus = economic rent, and
OMER - OMEF = NER + NEF = social surplus.

It is easy to verify that, at equilibrium,

Social surplus = consumer surplus + producer surplus = consumer utility - producer cost,

which is maximized with respect to the rate of consumption (see Figure 1b).

An analogy is now drawn in a transportation system with due consideration for the inherent differences between the consumer-product market and the distribution of trips among given facilities of a transportation system. A major difference is that traffic routing is a short-term problem that has an objective of optimal use of facilities that already exist and not a long-term one that has an objective of optimal investment. (Therefore, the notion of performance rather than supply should be used.) Travel costs are presumed to include only those short-term costs that users perceive in deciding whether or not to transport, when and how to do so, which mode and route to use, and so forth. Those costs paid by users but considered by them only on some longer-term basis are not included. The period considered is also a short one, e.g., a typical daily peak period. Thus, the operators of the system do not expect to recover investments by affecting routing, and fixed costs can be disregarded in the analysis.

Transport Network Equilibrium

A simple transportation system is considered below that consists of one link (j) and a single O-D pair (i,k); it is related to the paradigm described above. Complex networks can be similarly analyzed when the summations over links and O-D pairs are restored. As stated in the section on mathematical formulation, Y_{ik} represents the total utility to travelers between i and k measured by the maximal cost they are willing to expend for making the trip.

Figure 2. Demand-performance equilibration in a transportation system.

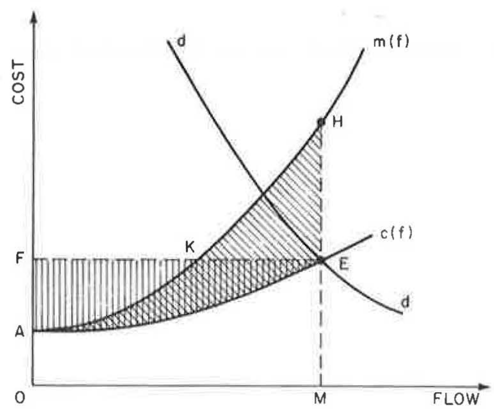
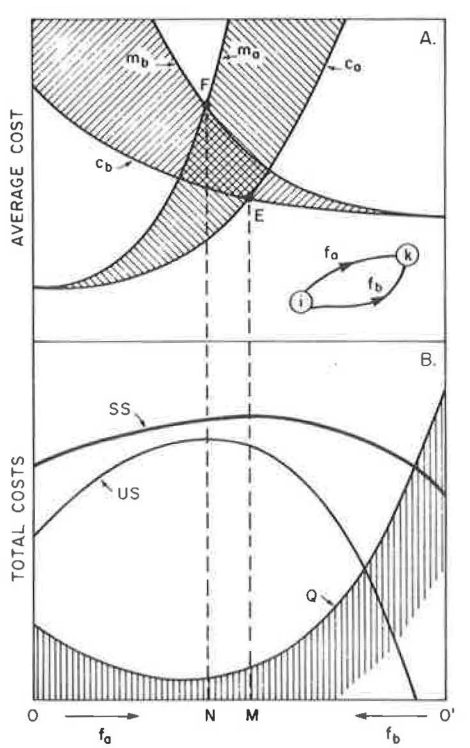


Figure 3. Surplus maximization in a two-route network.



The user surplus (which replaces consumer surplus) aggregates the excess of this utility over the actual costs incurred in making the trips. The notion of social surplus is also replaced by system surplus, while the optimized quantity that corresponds to producer surplus will be given a new interpretation. In analogy to market equilibrium, the NAP objective corresponds to the maximization (with respect to flow) of

$$\text{SS (system surplus)} = \text{US (user surplus)} + Q \tag{6}$$

where Q is given by

$$Q = fc(f) - \int_0^f c(z)dz \tag{7}$$

The marginal travel cost is defined as follows:

$$m(f) = (d/df)\,[fc(f)] = c(f) + fc'(f) \tag{8}$$

i.e., it is equal to the average (private) cost plus the increment in cost to all other users imposed by an additional user, which is termed the marginal social cost (MSC). When Equations 7 and 8 are combined, the following results:

$$Q = \int_0^f m(z)dz - \int_0^f c(z)dz = \int_0^f zc'(z)dz \tag{9}$$

Economists believe that economic efficiency is achieved when every user pays the full social cost of his or her travel. Therefore, the cost increment fc'(f) should be charged as a toll by the operators of the transportation system. This argument is critical to this analysis; however, I shall not elaborate on it here, since it has been discussed extensively in the literature (5,15). By using the terms of economists, Equation 9 aggregates the difference between the social costs and the private costs when flows are considered incrementally, i.e., the summation of the MSC. If no tolls are charged, the value of Q represents an undercharge to the users or, equivalently, a lost revenue for the operators. The assigned flow pattern maximizes this quantity together with US, as indicated by Equation 6. Economists also suggest another meaning for Q: Since the existence of congestion creates an obligation to pay, the failure to price the social costs of congestion amounts to an outright subsidy to motorists (16, p. 49). This reinforces the notion of user optimization for describing equilibrium flows in a transportation network.

The concepts discussed here are illustrated in Figure 2 for the single link. The Q-value is represented by area AEF, which (according to Equation 9) is equal to area AEH, the congestion undercharge. Since area AEK is common to both quantities, the two triangular areas AFK and HEK are equal.

Example

Consider a system of two parallel links a and b that have flows f_a and f_b and connect one O-D pair (illustrated in Figure 3). Total demand is represented by the baseline OO' (assumed to be of variable length). At user equilibrium the flow distribution is determined by the intersection of the two average link-cost functions at point E. Average travel cost on each link is then ME (Figure 3a) and the system surplus is maximal at M (Figure 3b). US is calculated as the difference between the total utility (a fixed quantity) and the travel cost and is not maximized in this pattern. Its maximum occurs at N, the nonequilibrium situation in which marginal costs are equalized (17).

EFFICIENT TECHNIQUE FOR SOLUTION

The first computational attempt at predicting flows in a network by means of elastic demands was made by BMW (5). They proposed a heuristic procedure conceived to emulate user behavior: Given existing (nonoptimal) traffic conditions, a fraction of the users (who have or can obtain adequate knowledge of these conditions) will divert during the upcoming period to a route that is optimal at the present transportation cost and will set their demand for transportation at levels that correspond to the present average trip costs. The responsive fraction of road users in each period is regarded as an independent random sample drawn from the total population of users; its size is assumed to decrease as time proceeds. Martin and Manheim (18) developed an iterative assignment procedure based on a different heuristic. Assuming an unloaded transportation network at the outset, they incrementally assign fractions of the potential O-D demands onto current shortest routes until equilibrium is approached. This, too, is believed to emulate user choices as they gradually load up the network. The procedure was later incorporated into the DODOTRANS analysis package (19). Bruynooghe, Gilbert, and Sakarovitch (20) use a technique in which shortest and longest routes between each O-D pair need to be calculated. Flows and demands are iteratively adjusted until they converge. Wigan (21) uses a simple iterative procedure in which the variable-demand functions are simply looped with a fixed-demand traffic-assignment algorithm (20). Wilkie and Stefanek (22) present a constrained-gradient algorithm and a modified Newton-Raphson procedure for the same problem. Although these algorithms can (potentially) provide rigorous solutions, they fail to exploit the specialized structure of the transportation network problem and are computationally unwieldy. Florian and Nguyen (13) developed an iterative scheme based on interlacing the variable-demand function with a fixed-demand traffic-assignment algorithm via generalized Benders decomposition. Dantzig, Maier, and Lansdowne (23) also proposed use of fixed-demand assignment by introducing an additional slack variable for each commodity. A more-detailed review of these algorithms may be found elsewhere (16,24).

The technique for solution described in this section is based on representing the O-D variable-demand function by an auxiliary link that augments the network model of the physical transportation system. This artificial link is termed a demand link (as opposed to the ordinary supply links). The resulting formulation, called the excess-demand formulation, is discussed below.

Consider expression 1, the objective function of the elastic-demand NAP. The first term in this expression is given by the integral of the inverse-demand function. Referring to Figure 4, it may be seen that this integral may be decomposed as follows:

$$\int_0^{g_{ik}} W_{ik}(y)dy = \int_0^{G_{ik}^m} W_{ik}(y)dy - \int_{g_{ik}}^{G_{ik}^m} W_{ik}(y)dy \quad (10)$$

where G_{ik}^m is a fixed upper bound. The first term on the right-hand side of Equation 10 is a constant (say, J_{ik}) and is unaffected by the optimization procedure. The maximizing objective of expression 1 may therefore be replaced by a minimizing objective:

$$\min \sum_{i,k} \int_{g_{ik}}^{G_{ik}^m} W_{ik}(y)dy \quad (11)$$

Defining the excess-demand $e_{ik} = G_{ik}^m - g_{ik}$, the following is obtained for expression 11:

$$\min \sum_{i,k} \int_0^{e_{ik}} W_{ik}(z)dz \quad (12)$$

The new function $[W_{ik}(e_{ik})]$ is obtained from $W_{ik}(g_{ik})$ by flipping the inverse-demand function about a vertical axis that passes through $g_{ik} = G_{ik}^m$. It may easily be seen that this function is similar in shape to the average link-travel-cost functions (Figure 3a) and the elastic-demand NAP can now be restated as follows:

$$\min\left(\sum_j Z_j + \sum_{i,k} X_{ik}\right) \quad (13)$$

subject to

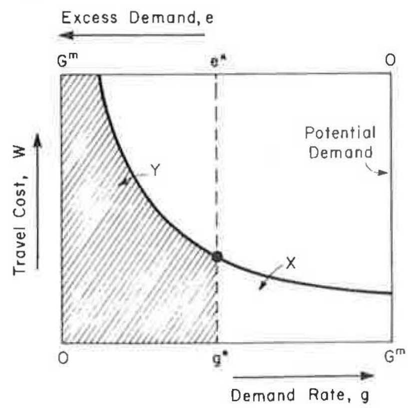
$$\sum_{p_{ik}} h_p + e_{ik} = G_{ik}^m \qquad h_p, e_{ik} \geq 0 \quad (14)$$

where

$$X_{ik} = J_{ik} - Y_{ik} = \int_0^{e_{ik}} W_{ik}(z)dz \quad (15)$$

The elastic-demand NAP now becomes a fixed-demand NAP on a network that is modified by forward-demand links that connect each O-D pair (i,k) and carry the excess-demand e_{ik}. The cost associated with the link is $W_{ik}(e_{ik})$. The resulting configuration is illustrated in Figure 5. The fixed O-D demands are G_{ik}^m, which are termed the potential demands. Thus, after the modified network has been created, there need not be a distinction between demand links and ordinary links and any fixed-demand network-assignment algorithm can be used to solve this problem. It is important to choose G_{ik}^m large enough to prevent binding the solution too low and so that there will always be (at optimality) a positive excess demand.

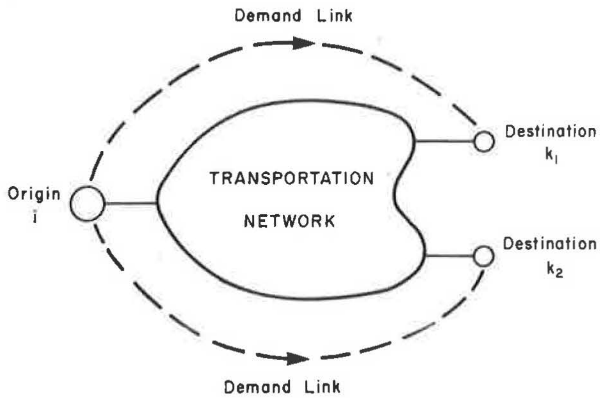
Figure 4. Travel cost versus demand representation.



CONCLUSION

This paper derives an economic rationale for the NAP with elastic demands and presents an efficient

Figure 5. Equivalent network for excess-demand formulation.



method for its solution. The optimization objective of the NAP implies the maximization of user surplus + Q, in which Q represents an undercharge to the users due to the social costs of congestion. The method of solution is based on modeling the problem as an equivalent network in which the elastic-demand functions are represented by appropriate demand links. This transforms the variable-demand NAP into an equivalent fixed-demand NAP that has the (fixed) O-D trip table given by the potential O-D demands.

The equivalent network model has the obvious advantages of convenient representation and efficiency in data handling, which thereby renders unnecessary the specialized iterative schemes inherent in all other methods of solution. The model is amenable to solution by efficient fixed-demand network-assignment algorithms without modification to those algorithms. Most important, in terms of computation, the model requires no additional nodes in the expanded network. Since network-assignment algorithms, which are based on the calculation of shortest-path trees, are more sensitive to the number of nodes in the network than to the number of links (25), this model requires only a moderately larger computational effort than that for a fixed-demand assignment on the same physical network. This effort is estimated to be only 25-75 percent larger than a comparable fixed-demand assignment. The most important conclusion, however, is that there are no inherent computational differences between fixed-demand and elastic-demand network-assignment problems, and the same algorithms can be used in both cases.

As noted above, the method described in this paper can be extended to consider more-general demand (cost) functions and is also applicable to other transportation analysis problems that involve choice situations that can be modeled as an equivalent-assignment (path-choice) problem in an expanded network. Such problems include, for example, the combined distribution-assignment problem (which involves origin or destination choice) and assignment in multimodal transportation networks (which may also include simultaneous modal choice).

ACKNOWLEDGMENT

Preparation of this paper was supported by an Intergovernmental Personnel Act Research Fellowship from the Office of Research, Federal Highway Administration (FHWA), Washington, D.C. The contents of this paper reflect my views and not necessarily those of FHWA.

REFERENCES

1. FHWA Computer Programs for Urban Transportation Planning. Federal Highway Administration, U.S. Department of Transportation, July 1974.
2. Comsis Corporation. Traffic Assignment. Federal Highway Administration, U.S. Department of Transportation, Aug. 1973.
3. N.H. Gartner. Analysis and Control of Transportation Networks by Frank-Wolfe Decomposition. Seventh International Symposium on Transportation and Traffic Theory, Institute of Systems Science Research, Kyoto, Japan, 1977.
4. S. Nguyen. A Unified Approach to Equilibrium Methods for Traffic Assignment. In Traffic Equilibrium Methods (M. Florian, ed., Lecture Notes in Economics and Mathematical Systems, Vol. 118), Springer Verlag, New York, 1976, pp. 382-295.
5. M.J. Beckmann, C.B. McGuire, and C.B. Winsten. Studies in the Economics of Transportation. Yale Univ. Press, New Haven, CT, 1956.
6. M.J. Beckmann. On the Theory of Traffic Flow in Networks. Traffic Quarterly, Vol. 21, Jan. 1967, pp. 109-117.
7. M. Florian. A Traffic Equilibrium Model of Travel by Car and Public Transit Modes. Transportation Science, Vol. 11, 1977, pp. 166-179.
8. M. Florian, R. Chapleau, S. Nguyen, C. Achim, L. James-Lefebvre, S. Galarneau, J. Lefebvre, and C. Fisk. Validation and Application of an Equilibrium-Based Two-Mode Urban Transportation Planning Method (EMME). TRB, Transportation Research Record 728, 1979, pp. 14-23.
9. I. Hasan and A. Talvitie. An Equilibrium Mode-Split Model of Work Trips Along a Transportation Corridor. Proc., 3rd World Conference on Transportation Research, Martinus Nijhoff, the Hague, Netherlands, 1977, pp. 129-136.
10. D. McFadden and others. Demand Model Estimation and Validation. Univ. of California, Berkeley, Rept. UCB-ITS-SR-77-9, June 1977.
11. S.C. Dafermos and F.T. Sparrow. The Traffic Assignment Problem for a General Network. Journal of Research of the National Bureau of Standards, Vol. 73B, 1969, pp. 91-118.
12. G.F. Newell. Traffic Flow on Transportation Networks. MIT Press, Cambridge, MA, 1980.
13. M. Florian and S. Nguyen. A Method for Computing Network Equilibrium with Elastic Demands. Transportation Science, Vol. 6, 1974, pp. 321-332.
14. L.J. LeBlanc. The Use of Large-Scale Mathematical Programming Models in Transportation Systems. Transportation Research, Vol. 10, 1976, pp. 419-421.
15. J.R. Meyer and M.R. Straszheim, eds. Techniques of Transport Planning, Volume 1: Pricing and Project Evaluation. Brookings Institution, Washington, DC, 1971.
16. Traffic Congestion as a Factor in Road-User Taxation. HRB, Highway Research Record 47, 1964.
17. N.H. Gartner. Optimal Traffic Assignment with Elastic Demands: A Review. Transportation Science, Vol. 14, 1980, pp. 174-208.
18. B.V. Martin and M.L. Manheim. A Research Program for Comparison of Traffic Assignment Techniques. HRB, Highway Research Record 88, 1965, pp. 69-84.
19. M.L. Manheim and E.R. Ruiter. DODOTRANS I: A Decision-Oriented Computer Language for Analysis of Multimode Transportation Systems. HRB, Highway Research Record 314, 1970, pp. 135-163.

20. M. Bruynooghe, A. Gilbert, and M. Sakarovitch. Une Methode d'Affectation du Traffic. In Beitrage zur Theorie des Verkehrsflusses (W. Leutzbach and P. Baron, eds.), Strassenbau and Strassenverkehrstechnik Heft 86, Bonn, 1969, pp. 198-204.
21. M.R. Wigan. Benefit Assessment for Network Traffic Models and Application to Road Pricing. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 417, 1971.
22. D.F. Wilkie and R.G. Stefanek. Precise Determination of Equilibrium in Travel-Forecasting Problems Using Numerical Optimization Techniques. HRB, Highway Research Record 369, 1971, pp. 239-252.
23. G.B. Dantzig, S.F. Maier, and Z.F. Lansdowne. The Application of Decomposition to Transportation Network Analysis. Control Analysis Corporation, Palo Alto, CA, Rept. DOT-TSC-OST-76-26, Oct. 1976.
24. E.R. Ruiter. The Prediction of Network Equilibrium--The State of the Art. Proc., International Conference on Transportation Research (Bruges, Belgium, June 1973), Transportation Research Forum, pp. 717-726.
25. S.E. Dreyfus. An Appraisal of Some Shortest-Path Algorithms. Operations Research, Vol. 17, 1969, pp. 395-412.