Forecasting Travel on Congested Urban Transportation Networks: Review and Prospects for Network Equilibrium Models[1]

David Boyce
Department of Civil and Environmental Engineering
Northwestern University
2149 Grey Avenue, Evanston, IL 60201 USA
dboyce@uic.edu

Revised June 21, 2004

1 Introduction

Fifty years ago, Beckmann, McGuire and Winsten (1956) completed their seminal formulation and analysis of an integrated model representing origin-destination, route and link flows on a congested road network, as a function of flow-dependent link costs. Their treatment of this fundamental problem, variously known as the traffic assignment problem with variable demand, the multi-commodity network equilibrium problem and the combined model of trip distribution and traffic assignment, sparked a vast literature of well over 1,000 references (Patriksson, 1994).

At the same time transportation engineers and planners in Detroit, and next in Chicago, were grappling with a way to solve computationally essentially the same problem with electromechanical accounting machines and first generation mainframe computers, in order to fashion future road and transit system plans. Unaware of the significance of the above formulation, they proposed a method, now known as the four-step or sequential procedure, which amounts to a simple heuristic for solving a version of that integrated model. By the time the significance of the model of Beckmann et al was understood, and a solution algorithm devised, the sequential procedure was so widely accepted that it is now effectively regarded as axiomatic.

The network equilibrium formulation in Part I of *Studies in the Economics of Transportation*

---

[1] This paper is part of a larger research and review activity of Huw C. W. L. Williams, Cardiff University, United Kingdom, and the author. A related paper was presented at the 43rd Congress of the European Regional Science Association, Jyväskylä, Finland, in August 2003 (Boyce and Williams, 2004). Sections of this paper initially drafted by Williams are denoted by *.

consists of a general origin-destination (OD) flow model (demand function) and a route choice model in which the flows between each origin and destination follow equal and minimal cost routes. These equilibrium origin-destination flows depend upon the travel costs of these used routes, which may include travel time and other appropriately weighted cost variables such as travel distance and tolls. The review presented in this paper focuses on this model formulation and its relation to the sequential procedure of travel forecasting practice. Extensive developments pertaining to the specification and estimation of demand functions, such as the multinomial and nested logit functions, as well as the roles of revealed and stated preference estimation procedures and urban activity analysis, are not treated in detail.

This paper examines the history of these developments in both research and practice, and seeks to understand, clarify and interpret why these events occurred as they did. Following a discussion of travel forecasting practice and software development, some thoughts about future prospects for the field are offered. The paper consists of five parts: Origins of Travel Forecasting Models; Origins and Development of the Sequential Procedure; Further Developments Related to Network Equilibrium Models; Travel Forecasting Practice and Software Development; and Future Prospects.

2 The Origins of Travel Forecasting Models

2.1 Context of Transportation Research in the Early 1950s

To understand and appreciate the extent of the accomplishments represented by Part I of Studies in the Economics of Transportation, one should attempt to imagine the state of knowledge in optimization, operations research and mathematical economics in 1950. Students of operations research know how efforts to manage the flow of war matériel during World War II led to early optimization procedures. These early advances stimulated broader interest in optimization methods in the U.S. Air Force, both at the Pentagon and at the recently founded Rand Corporation at Santa Monica, California. George Dantzig (2002) described his research at the Pentagon that led to the simplex method for solving linear programs, for example, and Saul Gass (2002) depicted the working environment and the role of computers in the Project on Scientific Computation of Optimal Programs. Among several persons, Dantzig contacted Tjalling Koopmans (1949) at the Cowles Commission for Research in Economics, then located at the University of Chicago. Koopmans had worked on a transportation model for the Allied Shipping Board during the war, research that was later recognized by his being awarded the Nobel Memorial Prize in Economic Science in 1975, together with L. V. Kantorovich.

In 1949, Koopmans and others organized the first conference on mathematical programming, and published selected papers in an edited volume entitled *Activity Analysis in Production and Allocation* (Koopmans, 1951). The title was carefully chosen to represent a field of inquiry broader than linear programming, as well as an emerging attitude towards the optimization of activities of private firms, as well as government. The contents of the volume, however, pertained almost exclusively to linear programming. The book was one of the very first in this new field, and by 1965, it had enjoyed six printings.

Another publication of lasting significance at that time was "Nonlinear Programming" by Harold Kuhn and Albert Tucker (1951), published in the Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. This seminal paper established the framework for the analysis of network equilibrium and network efficiency performed by Beckmann. A reference to the paper is given by Beckmann (1952).

These two works, then, formed the analytical basis for Part I of Studies. Moreover, they signify the intellectual fervor that must have pervaded the effort. The authors were clearly eager to contribute entirely new ideas from mathematical economics, optimization and probability theory to a field based on empirical relationships of traffic flow, speed and density. Although the authors were aware of traffic research, near the outset of their research in 1951-1952 they clearly could not have been aware of pages 344-348 of the paper, "Some Theoretical Aspects of Road Traffic Research," presented on January 24, 1952, and subsequently published in the *Proceedings of the Institution of Civil Engineers*. There John Wardrop (1952) stated his "two alternative criteria based on these journey times" for determining the distribution of flows on routes. In the four pages of his 38-page paper devoted to the subject of route choice, Wardrop explored a simple analysis of several alternative routes, stopping with the realization that "in the case of a network of roads the theoretical problem becomes very complicated." Wardrop was then a staff member of the Road Research Board of the UK Department of Scientific and Industrial Research.

2.2 Transportation Research at the Cowles Commission

Transportation research at the Cowles Commission originated with the appointment of Tjalling Koopmans as a staff member in 1944 and as director of research in 1948. As noted, Koopmans had earlier worked on transportation problems with fixed costs. With the support of the U.S. Air Force through the Rand Corporation, in 1951 research was initiated on the *Theory of Resources Allocation*, with applications to transportation, location and dispersal problems, among others. Rand's principal interest, it seems, was railway capacity analysis, motivated by a desire to estimate the capacity of the USSR railway system.

According to records on the Cowles Foundation website, http://cowles.econ.yale.edu/, Martin J. Beckmann joined the Commission in July 1951, C. B. McGuire in January 1952, and Christopher B. Winsten in October 1952. Beckmann received his Dr.rer.pol. in economics at the University of Freiburg, Germany, in 1950, after studying mathematics at the University of Göttingen, Germany, from 1945-1948. Prior to 1945, he served in the German army. In 1950 he came to the University of Chicago as a post-doctoral student. McGuire pursued undergraduate studies in economics and political science at the University of Minnesota, 1946-1949, and graduate studies in economic theory, statistics and mathematics at the University of Chicago until 1952. Following graduation from high school in 1943, he served in the U.S. Navy for three years as an electronic technician. In 1954, he had not yet begun his Ph.D. thesis. Background information on Winsten is not available, except for the fact that he came to Chicago from England.

A report for the period ending January 1952 stated Beckmann was constructing a model of "flows of transport to a given network of roads or railroads connecting a finite number of locations at which supply and demand is concentrated. … this introduces the new problem of congestion." Beckmann stated an interview in 1998, "According to Koopmans, McGuire's job was to convert our fuzzy ideas into a manuscript that could be read." McGuire later related, "Part of my job was to make the road stuff more practical." In a letter to Oskar Morgenstern at Princeton University dated April 1954, McGuire related: "Our original hope was that this work would give us some insight into the economics of city layout so that if a long-run policy of city dispersal were initiated, primarily for defense purposes, we could say something about where things should be dispersed to, and the costs or benefits thereof. While from this point of view I don't feel we have been very successful, I do think the work has led us to a better understanding of highway economics in general."

By late 1952 according to a report on the Cowles Foundation website, Beckmann produced two working papers, "Efficient Transportation in Networks," later revised into Discussion Papers 2049A and 2049B, and "Road Utilization Under Conditions of Individual Choice," the latter with McGuire. Copies of the first were sent to me by Beckmann about 1980, but the second has evidently not survived, or possibly was never issued. In July 1952, McGuire completed Discussion Paper 2048, "Highway Capacity and Traffic Congestion: A Preliminary Study."

In the following year, their work progressed to the point that an outline of a book was prepared by October 1953. Beckmann's proposal for Part I of the book largely conforms to the published version, but is somewhat more ambitious. In addition to Part II on Railway Transportation, he proposed Part III on Location, already one of his keen interests. A month later, he wrote a ten-page note describing his ideas for Part I, but not giving any specifics of the model. Evidently, the idea of publishing a book was problematic. On October 14, he wrote, "Even if the plan of a book is shelved, I should argue for an integrated, though perhaps not so broadly written, article or a series of articles, on highway transportation, following the outline" of Part I.

About the same time, Beckmann and McGuire completed a less technical discussion paper issued as Rand P-437, "The Determination of Traffic in a Road Network – an Economic Approach." The paper describes the problems of road transportation in relation to the present state of traffic engineering knowledge, and is remarkable for its sophistication and insights. Three levels of the problem of traffic forecasting are described: the structure of individual choices affecting traffic; repercussions of traffic conditions on these choices; and traffic equilibrium in a network. The basic principles of route choice and demand on a network are described verbally. The paper was submitted to *Traffic Quarterly* and rejected. No correspondence has survived except a memo by Koopmans puzzling over the outcome. At that time, *Traffic Quarterly* was arguably the leading journal in English serving multidisciplinary interests in road traffic.

By June 1954, the book manuscript had been hurriedly finished, as Rand had terminated funding for the project. Beckmann departed for Germany, his first trip home since arriving in Chicago in 1950. Correspondence between Beckmann and McGuire on technical and editorial points of the book continued throughout the summer and fall, augmented by many questions, comments and suggestions by Koopmans.

Of particular interest is a letter dated May 19, 1954 to William Vickrey offering $150 as "partial compensation" for a review of the manuscript. On July 15, Part I of the manuscript was mailed to Vickrey; Part II followed a week later. Detailed written comments from Vickrey have survived, indicating that he made extensive suggestions. Editing and revisions continued throughout the remainder of 1954, with McGuire coordinating the revisions and attending to details. The book was issued as a Rand Memorandum RM-1488-PR on May 12, 1955.

Submission of the book for publication by Yale University Press occurred in early 1955, as a result of Koopmans moving to Yale to become professor of economics. Yale University Press agreed to publish the book in May 1955, and McGuire spent much of the remainder of the year attending to myriad details. Beckmann was assigned responsibility for Part I. The book appeared early in 1956.

2.3 The Fundamental Contribution of Part I

Part I of *Studies in the Economics of Transportation* contains numerous innovations regarding how to conceptualize and formulate the problem of flows on a congested road network. As the authors were breaking entirely new ground, they clearly struggled to organize concepts in a way that provided for a tractable formulation. Among these were whether to treat delays at intersections of roads separately from delays on road segments, and how to separate directional flows from total flows on two-lane roads. With the advantage of accumulated research findings and hindsight, it seems clear the handling of some of these aspects could have been improved.

For the purpose of the present exposition and review, however, these details are best disregarded in order to focus on the fundamental contribution of the formulation of the two problems of network equilibrium with variable origin-destination flows, and network efficiency with variable origin-destination flows, the terms *equilibrium* and *efficiency* referring to Chapters 3 and 4. To be specific, these two formulations are now described using notation in current use. For historical reasons, I present the model using the authors' link-node notation, rather than the link-route convention generally used today. See Patriksson (1994, pp. 36-39), for details.

Consider a road network consisting of nodes $i \in \mathcal{N}$ and links $a = (i, j) \in \mathcal{A}$. Origin-destination flows between designated pairs of nodes, also known as commodities, are represented by $d_{pq}, (pq) \in \mathscr{C}$, the set of origin-destination pairs. A link flow from node *i* to node *j* is designated f_{ij}, and a link flow specific to a given origin-destination flow from *p* to *q* is designated f_{ij}^{pq}. Accordingly, $f_{ij} = \sum_{(p,q)\in\mathscr{C}} f_{ij}^{pq}$.

Conservation of flows of vehicles through nodes of the network is specified in three ways, depending upon whether the node is an origin *p*, a destination *q* or an intersection *i* of links:

$$\sum_{j\in\mathscr{A}_p} f_{pj}^{pq} - \sum_{h\in\mathscr{B}_p} f_{hp}^{pq} = +d_{pq}, \qquad \forall p \in \mathscr{P}$$

$$\sum_{j\in\mathscr{A}_q} f_{qj}^{pq} - \sum_{h\in\mathscr{B}_q} f_{hq}^{pq} = -d_{pq}, \qquad \forall q \in Q$$

$$\sum_{j\in\mathscr{A}_i} f_{ij}^{pq} - \sum_{h\in\mathscr{B}_i} f_{hi}^{pq} = 0, \qquad \forall i \in \mathscr{N}$$

where $\mathscr{A}_i$ is the set of nodes connected by links pointing out of *i* (after *i*), and $\mathscr{B}_j$ is the set of nodes connected by links pointing into node *i* (before *i*). The first condition states that the net outflow at node *p*, which corresponds to the total flow from *p* to *q*, must equal $+d_{pq}$. Likewise, the net inflow at node *q*, which corresponds to the total flow from *p* to *q*, must equal $-d_{pq}$. For all other nodes, total inflow equals total outflow of flow d_{pq}.

The origin-destination flow d_{pq} , or demand from *p* to *q*, is determined by a function $d_{pq} = g_{pq}(t_{pq})$, which is assumed to be decreasing and invertible, so that $t_{pq} = g_{pq}^{-1}(d_{pq})$, where t_{pq} is the minimal route travel time from node *p* to node *q*. The link travel time $t_{ij}(f_{ij})$ is a nondecreasing function of the link's own flow f_{ij}.

Given this description of origin-destination flows, route flows and link flows, Beckmann and his coauthors proposed to maximize the following function:

$$\sum_{(p,q)\in\mathscr{C}} \int_0^{d_{pq}} g_{pq}^{-1}(s)ds - \sum_{(i,j)\in\mathscr{A}} \int_0^{f_{ij}} t_{ij}(s)ds$$

By applying the theorem of Kuhn and Tucker (1951), they demonstrated that maximizing this function with respect to **d** and **f**, subject to the above node conservation of flow constraints and non-negativity constraints, results in the following properties:

1. All used routes from node *p* to node *q* have equal travel times, and no unused route has a lower travel time;

2. The origin-destination flow from node p to node q is determined by the function of this *minimum and equal*, or *equilibrium*, travel time.

Or, as the authors stated, “Demand refers to trips and capacity refers to flows on roads. The connecting link is found in the distribution of trips over the network according to the principle that traffic follows shortest routes in terms of average cost. The idea of equilibrium in a network can then be described as follows. The prevailing demand for transportation, that is, the existing pattern of originations and terminations, gives rise to traffic conditions that will maintain that same demand. Or, starting at the other end, the existing traffic conditions are such as to call forth the demand that will sustain the flows that create these conditions” (Beckmann et al, 1956, p. 59). On the following page, the authors state three principles corresponding to the two points stated above, and then proceed to define the above (artificial) objective function and prove the existence and uniqueness of their results.

Their accomplishment was not only novel; it was a remarkable example of how to use the new method of nonlinear programming to describe and investigate the behavioral properties of a highly complex, interactive phenomenon experienced daily by millions of travelers. The formulation seems to be one of the few cases of the representation of behavior by optimizing an artificial function, with the behavior corresponding to the optimality conditions. And, it provided the basis for forecasting the use of new roads and other facilities, if the resulting system of equations and inequalities could be solved for large networks.

Having solved the problem of equilibrium, the authors turned to the question of efficiency. By replacing the second term of the above function with the total time of travel on the network, $\sum_{(i,j)\in\mathcal{A}} t_{ij}\left(f_{ij}\right)\cdot f_{ij}$, they devised another maximization problem whose solution corresponds to the minimum total travel time with variable demand. In this case, travelers follow routes of equal marginal travel time; the authors interpreted the difference between these marginal times, and the average times normally experienced by the travelers, as *efficiency* tolls. As with their network equilibrium analysis, these *network efficiency* results were completely novel, although they had been understood in principle by Knight, as noted by the authors.

The exposition of these results required about 40 pages, plus introductory chapters on supply and demand, and a concluding chapter on unsolved problems. The presentation is neither straightforward nor simple to understand, even today when the results are clear, thanks to the efforts of several generations of researchers. I can readily attest to its difficulty, since as a graduate student in 1962 I was unable to penetrate the material. And, as Beckmann agreed in 1998, “The book was hard to read.”

Did the authors fully grasp the significance of their result? Clearly, they understood they had derived a new theory of traffic equilibrium and efficiency, which contained both behavioral and societal aspects. Did they understand it was potentially useful? Based on my interviews with Beckmann and McGuire, the answer is “frankly” no. They did not appreciate its practical significance for two reasons.

First, in 1954 as they completed their manuscript, there were no computers available to solve even the smallest example; therefore, they regarded the model to be theoretical. Beckmann expressed to me in this regard, "After all, there were no computers!" McGuire recounted that he explored solving the model computationally with developers of a computer at the University of Chicago, but this came to nothing. Moreover, there were no algorithms for solving nonlinear programs, although ironically the algorithm that was successfully applied in the early 1970s to solve the route choice problem was published in the same year as their book (Frank and Wolfe, 1956).

Second, despite their efforts to understand current research on traffic and to interact with traffic engineers, they were unaware that urban transportation studies were being undertaken to predict future road traffic and devise plans for its accommodation. Ironically, one of the most famous of these, the Chicago Area Transportation Study, was initiated in Chicago the next year after Beckmann and McGuire left that city.

Finally, what is the relation of Part I to route choice "criteria" of Wardrop (1952)? The Wardrop article is listed in the Bibliography of the monograph, but is not cited in Part I. Surviving correspondence to McGuire indicates that he was informed of the reference to Wardrop's paper in late 1952 . Beckmann agreed in 1998 that he was aware of Wardrop's criteria from McGuire. Beckmann's formulation of the network equilibrium model likely occurred concurrently and independently of Wardrop's short statement and discussion of his two criteria.

2.4 The Response to the Monograph

As noted, the Rand Corporation version of the book was released in May 1955, and the Yale University Press edition appeared in early 1956. By 1959, three printings of the book were issued in the US, and in the UK through the Oxford University Press. The price was $4 (US) and 32 shillings (UK). In 1959 a Spanish edition appeared. In 2003, the WorldCat List of Records showed 360 copies of the Yale edition held by libraries worldwide, 13 copies of the Rand version and 6 copies of the Spanish edition. The Rand version has been recognized as a *Rand Classic*, and the 359 page monograph may be downloaded from http://www.rand.org/ at no charge. Copies can also be purchased from University Microfilms, as well as from Internet book dealers, generally for prices in excess of $100.

Studies in the Economics of Transportation was the subject of nine book reviews in the leading economics and operations research journals of the day. All of the reviews praised the book. None, it seems, understood the significance of the formulations presented in Part I. Perhaps the reviewer best able to appreciate the work was William Prager, "an illustrious professor of applied mathematics at Brown University" (Newell, 2002). As described below, Prager had investigated one aspect of the network equilibrium problem. Even so, he doubted that the model would "yield valid numerical results." Perhaps the most noted reviewer in the transportation field was R. J. Smeed of the UK Road Research Laboratory, a colleague of John Wardrop. Untrained in advanced mathematics, he clearly did not grasp the essence of the equilibrium analysis, but stated "it is refreshing to read a book that attempts to tackle the subject of road transport in a comprehensive and fundamental way." The reviewer perhaps best able to understand the economic analysis was Edwin Mansfield. Although noting the use of the Kuhn-Tucker theorem to study equilibrium and efficiency, he did not seem to appreciate the significance of the results.

What is clear from today's perspective is that this was a difficult book. It was likely one of the first applications of the Kuhn-Tucker theorem to a real-world problem. The authors struggled conceptually with their model. While understanding the results, they did not do justice to it with their explanation. In a later paper written to "popularize" the ideas in the book, Beckmann (1967) did not connect the results with travel forecasting activities of urban transportation planning agencies, even though he was somewhat active in the professional meetings of these agencies through the Highway Research Board.

Nonlinear programming was a new field in 1950 when Kuhn and Tucker presented their paper at the Second Berkeley Symposium (Kuhn, 1976). It developed in the shadow of linear programming, which had a slight head-start and seemed more useful in applications, despite its limitations to linear problems. (See Dantzig (2002) for an interesting anecdote in this respect involving Hotelling and von Neumann.) The inability of the authors, or anyone else, to foresee that such problems might be solved computationally may have further limited interest in its serious study. Whatever discussions took place at the time were certainly of a theoretical nature.

When I obtained a copy of the book in 1962 and used it as the basis for a term paper in a graduate course on linear programming, neither I nor my professors saw a connection between the book and my keen interest in travel forecasting by the transportation studies of the day. Like others, I did not appreciate the prospects of using the model for forecasting. I return to this perspective below after reviewing early efforts to forecast travel.

2.5 Earlier and Concurrent Research on Network Equilibrium and Efficiency

The subject of equilibrium and efficiency in road networks was not entirely new when the authors took up these questions in Part I in 1952. In this section, I briefly review works that preceded the completion of Part I in 1954, and research undertaken subsequently without knowledge of its publication. The papers reviewed here pertain to partial formulations of the network equilibrium or network efficiency models of Chapters 3 and 4 of Part I, and share some of their attributes. The reviewed models consider the case of fixed origin-destination flows (demand), since no models that included variable flows were proposed during this period. In this respect the models of Part I are in a class by themselves.

The subject of route choice over a two-route road network was examined by A. C. Pigou in his seminal text, *The Economics of Welfare*, first published in 1918. Pigou made an error in an example (changed in subsequent editions, and now inaccessible), which attracted the attention of Frank Knight (1924). Knight succinctly described the equilibrium and efficiency conditions of route choice on a two-route road network, nearly 30 years before Wardrop (1952). Beckmann presumably found these descriptions by Knight in a reprinted version published in 1952, and used them to motivate his equilibrium analysis (p. 60) and efficiency analysis (pp. 86-87). Even earlier, Kohl (1841) evidently considered the same problem.

What was new in the treatment in Part I, of course, was consideration of route choice in a general road network. However, this idea was also not entirely novel. R. J. Duffin, later recognized for his contributions to geometric programming, wrote a series of papers on the behavior of electrical networks (in particular, Duffin, 1947). Although the context is unfamiliar and the notation difficult, it seems clear that Duffin did understand that the integral of a "conductivity function" defined on arcs connecting pairs of nodes of an electrical network was the key to proving that "a network of quasi-linear conductors has a stable state of currents, and this state is unique." Duffin used "quasi-linear" to contrast with the linear case, known as Ohm's law. He also described the functions as "nondecreasing." Duffin's paper was motivated by a desire to generalize the Kirchoff-Maxwell Laws.

Harold Kuhn (1976, 2002) related the following events, which are rather fascinating considering Duffin's insights. In 1948 during a chance meeting with George Dantzig, Albert Tucker noted that Dantzig's description of linear programming "sounded like Kirchoff's Laws." Later, while on leave at Stanford in 1949, Tucker returned to the question of the relation between linear programming and the Kirchoff-Maxwell treatment of electrical networks. At this point, "he recognized the parallel between Maxwell's potentials and Lagrange multipliers and identified the underlying optimization problem of minimizing heat loss." Tucker suggested to David Gale and Harold Kuhn that they write a sequel to a paper (later published as Gale et al, 1951) generalizing the duality of linear programs to quadratic programs. Kuhn actually wrote the paper with the emphasis "on the general nonlinear case and the properties of convexity that imply the necessary conditions for an optimum are also sufficient," given a constraint qualification added later. A revised version of this paper, "Nonlinear Programming," was presented at the Berkeley Symposium on Mathematical Statistics and Probability in 1950, the first presentation of the now classic Kuhn-Tucker theorem. Tucker (1957, pp. 254-256) later examined the relationship between quadratic programming and the classical Kirchoff-Maxwell Laws.

The applied mathematician, William Prager (1954), described informally a formulation of the network equilibrium problem, which includes the integral of travel time as a linear function of opposing flows on a two-way road. Motivated by the criteria given by Wardrop (1952), Prager describes the equal journey times of routes in terms of the differences in potentials at nodes of the networks. Hence, he was also thinking along the same lines as Duffin and Tucker, as well as Beckmann, but without the advantage of the theorem of Kuhn and Tucker. Prager was aware of the earlier work of Koopmans (1949), and proposed a formulation for the efficient transportation of goods in which the total shipment cost is a quadratic function of the flows on the network, a special case of the problem considered by Beckmann in Chapter 4 (Prager, 1955). In both papers, Prager is unclear about the source of origin-destination flows. In his 1955 paper, it is unspecified, and in his 1954 paper, it appears to be specified in terms of net outflows and inflows of a single commodity, as in Koopmans's Transportation Problem of Linear Programming.

The next development was a model proposed by Abraham Charnes and William Cooper (1958). In this three-page paper they proposed a traffic network equilibrium formulation for the case of fixed origin-destination flows, drawing on ideas likely suggested to them personally by Duffin, who was their colleague at the Carnegie Institute of Technology (Cooper, 2002). They were also familiar with the formulation of Prager (1954) and with "the principles" of Wardrop (1952); they were evidently the first to recognize the relation of their model to the non-cooperative game theory of Nash (1951). The paper is written very much in the terminology of electrical networks with references to Kirchoff's Laws. As did Duffin, they proposed minimizing the sum of "integrated resistance functions" of total arc flow, which they refer to as "branch resistances." While they never specified an actual resistance function, it is clearly nondecreasing. They also proposed a piecewise linear formulation, which was a favorite device used repeatedly by the authors to convert nonlinear programs into linear programs, which they could then solve with known algorithms.

At the first Symposium on the Theory of Traffic Flow in 1959, Charnes and Cooper (1961) expanded on their 1958 paper by solving an example pertaining to a small Indiana town with 11 origin-destination pairs, 22 nodes and 27 links, using multicopy linear programming techniques. Interestingly, the example was provided by staff of the Chicago Area Transportation Study, indicating that Charnes, by then a professor at Northwestern University, was in contact with the first generation of travel forecasting professionals.

Of interest in the same volume is a paper by Wardrop (1961) concerning the theory of travel demand on a simple road system without congestion. Included is a reference to Beckmann et al concerning demand. In a short section near the end of his paper, Wardrop states: "The cost of traveling to a given destination is generally an increasing function of the flow of traffic on each part of the route, and this is a serious complication in the use of theoretical models. In principle it can be dealt with by starting with estimated degrees of congestion, calculating the resulting flow pattern, modifying the costs on the basis of the new flows, and repeating this process until it converges." Thus, Wardrop did not associate this problem with route choice, and his 1952 paper, and his thinking was not impacted by Beckmann et al.

The final chapter in this part of the story concerns the thesis research of a University of California, Berkeley graduate student from Denmark, Niels Jorgensen (1963). Jorgensen was a student in transportation engineering, and a research assistant of Robert Oliver, professor of industrial engineering and operations research. For this reason among others, Jorgensen studied queuing theory, network flow and traffic flow models with Oliver. Oliver suggested the problem of network equilibrium to Jorgensen as a M.S. thesis topic, evidently largely motivated by Wardrop's principles. Oliver also made Jorgensen aware of the papers of Charnes and Cooper, as well as the Kuhn-Tucker theorem, but not Beckmann et al (1956).

In his 39-page thesis, Jorgensen formulated the network equilibrium and network efficiency problems with fixed origin-destination flows. Using the Kuhn-Tucker theorem together with Wardrop's principle of user equilibrium, he derived the sum of the integrals objective function. Unlike all earlier papers on the subject, Jorgensen's treatment is a picture of clarity. He investigated several cases, and in particular explored a specific travel time-flow function based on the assumption that link speed is a linear, decreasing function of total link flow, or equivalently that travel time is a nonlinear increasing function of flow. Jorgensen did not attempt to devise a solution algorithm, but he did solve a few simple examples. In correspondence with me in 2003 in response to a question about the origins of his work, he stated, "I believe the integral formulation was my own. I recall that it gave me a kind of 'aha-feeling,' but in reality it was just a reinvention of the Beckmann formulation, which I should have known." Interestingly, he stated he was also working with an electrical analogy.

When Jorgensen returned to Denmark, he continued to work on the topic, and wrote a paper for the Third International Symposium on the Theory of Traffic Flow to be held in 1965 in New York, which was accepted. However, he could not raise the travel money, and eventually moved on to other research. Interest in his work declined because Danish practitioners did not like the model's property that route flows are not unique, a well-known property of network equilibrium models. Jorgensen shifted his interests to traffic safety research, completed a Ph.D. and recently retired as Professor following 22 years of service at the Technical University of Denmark. Jorgensen's thesis report is cited in early works on network equilibrium, and was one starting point for the research of Stella Dafermos (1968), discussed below.

The first academic research paper found to refer explicitly to Beckmann, McGuire and Winsten is a discussion paper on road pricing written in 1959 by the British economist, Alan Walters (1961). Walters drew extensively on the material in Chapter 4, as well as the earlier works of Pigou and Knight, to develop a two-road graphical analysis. Following the analysis of the effect of a bottleneck on one road, he presents an analytical analysis of a network of roads. In a footnote to this analysis, Walters states: "The Beckmann model distinguishes between the 'trip' and the 'route.' The 'trip' is from the origin to the destination, and this may be achieved by running over various roads, i.e. by various 'routes.' The demand function has the number of trips as the quantity variable, whereas the cost function relates to the quantity of traffic on particular roads. I think this is an excellent way of organizing the material. But I have not followed their lead in this paper, partly because I can add nothing to their treatment in this respect, and partly because it would introduce much complication into the notation."

From this note, we can see that Walters understood the essence of Beckmann's model, but was not inclined to explore further its mathematical properties. In his treatment, Walters implicitly appears to consider a model of several independent roads connecting two points. Unlike Beckmann et al, he considers cross-elasticities of road costs on demand, but does not consider that two routes might use the same road (link). Hence, this is not a network model in their sense.

Aside from a few papers in the economics literature (e.g. Johnson, 1964), there is little evidence of academic interest in Part I in the decade following its publication. Beginning in the late 1960s, however, further developments began to intensify, primarily in response to the travel forecasting requirements of urban transportation studies, as they were known. To provide background and insights, I review the beginnings of this field, and then consider Beckmann's influence on it.

3. Origins and Development of the Sequential Procedure

Urban transportation studies began in Detroit in 1953, then in Chicago in 1955 and later in Pittsburgh and the Tri-State (New York City) region, all under the leadership of Douglas Carroll, Jr. Similar studies were initiated during this period under the leadership of Glenn Brokke and others at the Bureau of Public Roads (BPR), US Department of Commerce. For the next decade, a friendly competition existed between the Carroll-led studies and the BPR with regard to development of methods and procedures for forecasting travel and traffic. Out of these early studies emerged a travel forecasting methodology known today variously as the sequential procedure or the four-step procedure in the United States (US), and the four-stage procedure in the United Kingdom (UK). The steps of the procedure came to be known as: trip generation (G); trip distribution (D); mode split (MS); and traffic assignment (A). The meaning of these terms will become clearer in the following historical account.

My objective in this section is to trace the development of these methods and their relationship, or lack of one, to the network equilibrium model of Beckmann et al. Initially, the account concerns developments in the US, but subsequently shifts to the UK, where Carroll also had some influence.

3.1 Traffic Assignment

I begin with an attempt to understand *traffic assignment*, the step most closely related to one of the key innovations of Beckmann's network equilibrium model. Although it is the fourth step of the sequential procedure, traffic assignment lies at the core of travel forecasting for two reasons. First, it is the mechanism for actually determining link flows on the network, which is the goal of the entire procedure. Although "desire lines" of origin-destination travel give some impression of future travel requirements, a forecast of future link flows is needed for design. Second, ultimately, congestion is represented in the traffic assignment model. If addressing congestion is the goal, then representing congestion and its effects are central to the task.

The origins of the term *traffic assignment* are obscure, but Earl Campbell (1952, p. iii), an early traffic engineer, described it as follows: "The estimated allocation of traffic to a proposed facility is commonly termed 'traffic assignment.' The estimated allocation may indicate annual average daily traffic volumes, periodic directional movements, and composition by types. Traffic assignment is fundamental to the justification of a proposed highway facility and to its structural and geometric design, to spotting points for access, and for planning of traffic regulation and control measures. As yet, traffic assignment is considered to be more of an art than a science." During the 1950s, traffic assignment evolved into two approaches: 1. *diversion curve assignment* of travel from the existing street network to a new highway, generally of a higher type with controlled access; 2. *all-or-nothing assignment* with or without *capacity restraints*. (The terms used reflect the usage at the time, and attempts are made to interpret their meanings in the following discussion.)

Diversion curve methods assumed the availability of existing zone-to-zone travel over the existing street network from an origin-destination survey or a forecast of the same travel pattern. The basic concept was to estimate the share of this travel that would divert to a proposed facility as a function of the proportion of the trip that could utilize the facility, measured in terms of distance. The relative speeds of travel on streets vs. the facility were incorporated into the calculation only by assumption. To be practical, the procedure was implemented on an electronic accounting machine or a first generation mainframe computer, such as the IBM 704 (Campbell, 1956). Underlying the diversion curve method was a simple *route choice* concept, although this term was not used in extensive descriptions found in Highway Research Board bulletins. Consideration of capacities of either the proposed or existing facilities and the corresponding effect on travel speeds were absent from the method.

The concept of all-or-nothing assignment seems to have evolved as an alternative to diversion curve methods, the advance being that all travel would be placed on a route, not just the travel diverted to the new facility. The core idea was that travel should be placed on the shortest route in terms of travel time, a simple, but implicit, behavioral assumption. The term, *all-or-nothing*, means *all* travel on the shortest route, and *nothing* on other routes. *Shortest route* assignment is a more apt term, and has occasionally been used.

Difficulties with early assignment methods occurred when some links of the network grossly exceeded their capacities, leading to the concept of "capacity restraints." Ad hoc methods to control for such excess flows involved the introduction of link travel time adjustments based on speed-flow relationships, and averaging of successive assignments. An innovative scheme is described by Carroll (1959), later called *tree-by-tree incremental assignment*. Carroll described traffic assignment as "the allocation of travelers moving between specified zones of an urban region onto particular travel routes. Having assigned every trip to a route, the level of service of the proposed network can be evaluated." Note the mechanical nature of the definition, and the lack of any mention of travel behavior. The method implemented, however, is based on shortest routes in terms of travel time, and on periodically updating link travel times as travelers are assigned from each origin to all destinations. Even though this procedure spread the route flows over more links of the network, the flow between a given origin-destination pair was restricted to only one route, in contrast to the concept of multi-route equilibrium described by Beckmann et al and Wardrop.

The shortest route calculations over the network were performed by the application of the shortest route algorithm of Moore (1957). Carroll acknowledges the contribution of the Armour Research Foundation of Chicago in proposing the method, and of Morton Schneider in devising a practical implementation for the computers of the day. Another type of *incremental assignment* was also utilized; in this case, the origin-destination trip table was divided into several subtables, and each was assigned to shortest routes after updating the link flows and link travel times. This procedure can allocate flows between individual OD pairs to more than one route. Most early assignments, and some still today, were made for a 24-hour period by transforming hourly capacities to 24-hour capacities based on the observed ratio of 24-hour flow to peak-hour flow. This shortcut results in all traffic being treated as if it occurs in the most congested period of the day, clearly an implausible notion. Yet, it is hardly ever questioned.

What is missing from this development is any reference to a route choice or network equilibrium *model* of the type devised by Beckmann et al, Prager, or Charnes and Cooper up to that time. The solution procedure was strictly ad hoc, not even deserving of the term *heuristic*, since no objective function was available to assess the quality of the assignment. Also missing from all practitioner papers in the US was any reference to the principles of route choice as described by either Beckmann et al or Wardrop. Nevertheless, crude results were obtained for relatively detailed networks consisting of 600 zones and 5,000 links using mainframe computers with 32,000 words of memory and data storage on magnetic tape. Often the computers used were offsite, sometimes in a different city and available only on weekends. Schneider, for example, repeatedly traveled from Chicago to Cincinnati, a two-hour flight, to perform assignments over the weekend.

3.2 The Sequential Procedure

If traffic assignment was the culmination of the procedure, what preceded it? Clues to early concepts are found in an abstract, authored by Douglas Carroll and Roger Creighton, for a discussion held at the 1957 Highway Research Board Meeting. “A continuous, integrated analysis and planning process consists of three major parts: estimating traffic generation from land use; predicting future lines of travel desire; and predicting traffic flow on a transportation network.” At the time, such studies concerned only roads, so only auto travel was forecast. Expansion of the procedure to include travel by transit, and the “splitting” of those trips by mode, followed soon, but non-motorized travel was ignored, much as it generally is today.

Estimating traffic generation from land use became *trip generation*, essentially a linear statistical model. Predicting future lines of travel desire was performed by an origin-destination *trip distribution* procedure. Initially, growth factor methods were used, but later concepts of spatial interaction between points separated by travel distance, or even better travel time, were applied, which were analogous to the gravity law of spatial interaction, and became known by the term, gravity model. Early innovators in trip distribution modeling were Alan Voorhees (1955), who introduced the empirical gravity concept into urban travel forecasting, and Morton Schneider (1959), who derived and implemented a model based on intervening opportunities. Walter Isard (1960, Ch. 11) was also keenly interested in gravity models and provided a thorough synthesis of theory and empirical analysis prior to 1960. *Mode split* was initially a diversion curve concept, until replaced by mode choice functions of the logit or probit type.

Somewhat later, travel forecasters realized there is a conceptual flaw in this scheme if it is based on congested travel times. Origin-destination flows, desirably a function of travel time, are assigned to the road network, resulting in link flows, which determine link travel times as a function of flow. These link times in turn determine the shortest route times, which in turn determine origin-destination flows. In simple flowchart terms, a *feedback* from assignment to trip distribution is required. However, the question of whether such a feedback scheme, when iterated, converges to a stable equilibrium should be asked.

Brian Martin, Frederick Memmott and Alexander Bone (1961) depicted this procedure in their extensive review and synthesis of early practice in a large, complex diagram. Indeed, they show a feedback from assignment to interzonal transfers, their term for trip distribution; however, they do not discuss this feedback, except to note “these dashed lines indicate additional interactions which have not been included in past urban transportation studies.” However, later they added, “The process is repeated until equilibrium and consistency are reached between the input data and the final results. In the past this repetition has been confined to the assignment procedure, whereas in fact the demand as well as the supply of facilities for travel should have been considered.”

N. A. Irwin and H. G. Von Cube (1962) state a similar concept and concern in setting forth an ambitious view of the sequential procedure at the 1962 Highway Research Board Meeting. In an unusually clear diagram, they describe six blocks in a feedback circle. “In general, these new link travel times will be different from those on which the previous determination of routes and travel propensities (deterrence functions) were based. It is necessary, therefore, to turn again to block 1 (shortest routes), using this time the new link times. Weighted averages of the new route times via all available routes between each origin and destination are then used in block 2 to calculate new travel propensities, leading to a new calculation of zone interchange volumes in block 3 (trip distribution), and so on around the loop through blocks 4, 5 and 6. It is by this means that feedback occurs. The effects of road traffic congestion … are fed back within the model to effect the travel patterns as they do in real cities. This feedback procedure is repeated until equilibrium is reached; i.e. until link loads and travel times produced by one cycle are not appreciably different from those produced by a previous cycle.”

The description of this scheme in the balance of the paper is detailed and advanced, especially for its time. The authors were possibly the first to use the term *feedback* in this context. They also had a clear notion of an equilibrium solution in the presence of congestion. The authors described an application of their concepts to Toronto. In the sense of the network equilibrium model of Beckmann et al, however, their conceptual procedure was not a model, and like Beckmann, they proposed no convergent solution procedure.

3.3 Early British Experience with Travel Forecasting

The urban transportation studies that began in the US in the mid 1950s were initiated in many large metropolitan areas by 1960, and were further mandated by the 1962 Highway Act, which required the preparation of a transportation plan in order to receive federal government funds for road construction. Case studies of this period are found in Boyce et al (1970). Also during the 1960s, urban transportation studies were initiated in the UK, first in London and later in the Southeast Lancashire-Northeast Cheshire region (SELNEC) centered on Manchester. The latter sought to innovate with regard to some aspects of the travel forecasting procedure, but was also influenced by consultants experienced in the Carroll-led studies in the US. The principal departure and innovation of the SELNEC study concerned the trip distribution and mode split models innovated by Alan Wilson (1967, 1969) and Wilson et al (1969). In the manner of the intervening opportunities model of Schneider, Wilson sought to improve the theoretical foundation of the trip distribution model, in this case by adapting entropy methods from statistical mechanics.

In the UK context, mode choice models and transit assignment played a significantly larger role than in their US counterparts at this stage of development. With regard to the above points concerning road traffic assignment and feedback, however, there was little innovation. US concepts of all-or-nothing assignment with capacity restraint were applied. The inconsistency in the sequential model structure that requires an iterative solution, or feedback, is not mentioned in these papers. Curiously, neither is the paper of Wardrop (1952).

3.4 Relationships among the Steps of the Sequential Procedure*

Following the acceptance of the four-step paradigm, most researchers and professionals became engaged in the improvement of the models and methods described in the individual steps. Household-based category analysis replaced zone-based regression models (Wooton and Pick, 1967). Variously described utilities or generalized costs emerged from early studies of modal choice with models specified and estimated at the individual level. The incorporated generalized costs, specified as linear functions of objectively measured attributes with travel time suitably scaled to money units, served as an interface between policies, behavioral response and benefit evaluation. The numerical estimate of the *value of time* has proved to be one of the most important parameters in the whole of planning.

From an analytic viewpoint, the earliest distribution and modal split models, which involved apportioning trips between different locations and modes, adopted empirically derived functions, sometimes referred to as deterrence functions for spatial interaction, and diversion curves for modal shares; these were determined through *goodness-of-fit* criteria. By the late 1960s, these curves began to be replaced in academic discourse and some applications by analytic functions, and share models of the multinomial logit form became widely adopted (Wilson, 1970; Manheim, 1979). These were conceptually appealing, analytically tractable and consistent with a number of theoretical constructs that were starting to be used for interpreting dispersion associated with trip making (Erlander and Stewart, 1990). These analytic expressions more readily allowed for an appreciation that their *sensitivity* parameter(s) served a four-way role in determining the dispersion in travel patterns, the frequency distribution of trip lengths with respect to generalized cost, the response to transport policies, and economic benefit measures.

A problem that exercised the earliest modelers was the ordering of the steps and how they should be "linked" together. There were, from the start, informal behavioral assumptions underpinning the approach in terms of a sequence of decisions that mapped onto the individual models. Various alternative structures for the demand model were proposed reflecting the conditionality of a sequence of decisions, the most common being whether the distribution model preceded (G/D/MS/A), followed (G/MS/D/A) or was combined (G/D-MS/A) with the modal split model. The first two constructions involved the formation of "composite costs" that represented "average costs" at what were referred to as "later stages of the models." As late as the mid 1970s no detailed theoretical basis for the entire model existed; for a given ordering, which was suggested a priori from behavioral assumptions, the form of the composite cost was regarded as an extra degree of freedom for achieving improved "goodness-of-fit"; see Senior and Williams (1977) for a review of model structures applied in practice.

The derivation of the nested logit model within discrete choice theory provided one resolution to these ambiguities (Williams, 1977; Daly and Zachary, 1978; McFadden, 1978). This development endowed the whole model with a behavioral rationale in which the analytical structure of the demand function reflected underlying utility functions, imposing two important restrictions on the overall model. First, the composite costs that interfaced the different models needed to be formulated in a particular way; for logit-type models, these were in the form of a "log sum" function, a form that had already been implemented with microdata by Ben-Akiva (1974). Second, the parameters that determined the sensitivity of trip choices to changes in times or costs, had to decrease as one progressed from route choice, through mode to locational and frequency selection in the G/D/MS/A structure. Only then would it be ensured that the estimated direct- and cross-elasticity parameters had the appropriate sign, requiring the demand for an alternative to fall when its cost rose or the cost of a substitute fell.

The nested logit model thus provided a consistent way of combining the various constituent choices with differential cross substitution between alternatives, and made the ordering of associated logit share functions subject to empirical test. It is important to note that the specification of the demand model with empirically derived functions for locational and/or modal shares is not immune from this strict requirement for appropriate response properties derived from the calibrated model. Williams and Senior (1977) reconfigured the four-stage procedure as a nested logit structure, experimented with different orderings of the distribution and modal split models, and showed that many models in UK practice did not satisfy the necessary parameter inequalities implied by the chosen structure. The specific implication of this finding was that such models could have produced counter-intuitive results. The more general implication was that calibration and the traditional notion of validation based on goodness-of-fit and held-back samples was not a sufficient test for the validity of such cross-sectional models for policy appraisal.

4 Further Developments Related to Network Equilibrium Models

Having introduced and reviewed concepts, innovations and implementations of travel forecasting methods in the US and UK, I next resume tracing the evolution and impact of the network equilibrium model devised by Beckmann et al and the independent contributions of Prager, Charnes and Cooper, and Jorgensen up through 1963. The first group of works reviewed below identified the relationship between Beckmann's formulation of route choice on a congested road network and the road traffic assignment problem confronting early practitioners. The second group concerns efforts to find a unified formulation of the sequential procedure, which led back to Beckmann's variable demand formulation.

4.1 Network Equilibrium with Fixed OD Flows

As described above, early network equilibrium research concurrent with Beckmann et al considered only fixed OD flows. Research on this problem continued during the 1960s and 1970s, but at some point the connection with Beckmann's formulation was recognized. One of the first papers to relate the traffic assignment problem to Beckmann et al and Wardrop was Almond (1967) presented at the Third International Symposium on the Theory of Traffic Flow in 1965. In considering the problem of traffic assignment on congested roads, she recognized that a solution method involved finding shortest routes, and used graphical methods to explore cases with two and three links. Although aware of Beckmann et al, she did not use their diagrams found on pp. 83-85. Instead, she appears to have been influenced by the graphical approach of Wardrop.

At the same symposium, Knud Overgaard (1967) presented a test of a traffic assignment algorithm, which is effectively a linearization method with averaging of route and link flows by the method of successive averages. As implemented, the algorithm required that route flows be stored and averaged, indicating his relatively primitive understanding of the problem. Jorgensen's formulation of the network equilibrium problem is stated, but not explored further. In particular, the objective function is not used to monitor the convergence. As with Jorgensen (1963), the relation of the research to Beckmann et al was not realized.

Beckmann (1967) also presented a paper at this symposium, the first in this series of symposia initiated in 1959 that he attended. His paper is closely related to the Efficiency chapter of Part I, but does not present the network formulation found there. Instead, he was content to discuss problems of two competing roads with variable demand, after developing the general principles governing the analysis. Hence, over ten years after the completion of Part I, we can see that he continued to view this as a theoretical model from which policy insights might be gained.

Two papers directly related to Beckmann's formulation and the traffic assignment problem of practice appeared in the late 1960s: Brynooghe, Gibert and Sakarovitch (1969) and Dafermos and

Sparrow (1969), the latter based on Dafermos (1968). Brynooghe et al's paper, presented in 1968 and published in French in 1969, synthesized Beckmann's formulation for the fixed OD flow case, Wardrop's principles and the practitioners' traffic assignment problem. The formulation is the link-node type proposed by Beckmann. Following an optimality analysis, the authors proposed two solutions algorithms, one related to the general method of Frank and Wolfe (1956), and presented convergence proofs. They also briefly discuss the variable OD flow problem, and refer to a note by Murchland (1967), with whom Gibert (1968) was collaborating. Their paper also cites the thesis of Jorgensen and related work of Jewell (1967). Perhaps because of its somewhat obscure place of publication, and the language of the publication, the paper had relatively little impact, at least outside of French-language universities and research centers.

One motivation for the research of Dafermos and Sparrow (1969) was also the practitioners' traffic assignment problem, as they received financial support for their research from the US Bureau of Public Roads. Despite this practical objective, Stella Dafermos attacked the problem in a highly mathematical manner, building on earlier results by Jorgensen, Wardrop, Beckmann, and Charnes and Cooper. She proposed the terms *user-optimized* and *system-optimizing* to describe "the two criteria proposed by Wardrop," explored the properties of the two problem formulations extensively, and proposed solution algorithms with proofs of convergence. She used the link-route form of the conservation of flow conditions, evidently for the first time in the network equilibrium literature. Small computational examples were also included.

Although the algorithms used the concept of shortest routes, it also required knowing all routes between each OD pair, or the shortest and longest used route in the solution. Both sets might be computationally difficult to identify, or impractical to store. Therefore, the algorithms proposed were not practical for the problems of several thousand links being considered at that time. Subsequently, Leventhal, Nemhauser and Trotter (1973) enhanced the algorithm proposed by Dafermos by showing how shortest routes can be generated as needed. Dafermos (1971, 1972, 1973, 1980, 1982, 1983) and Dafermos and Sparrow (1971) published many additional findings; her research became widely recognized as providing a foundation for more advanced models.

4.2 Combined Models of Trip Distribution and Assignment

At the same time as Alan Wilson was proposing a method for deriving the gravity model, and applying it in the SELNEC study, a British applied mathematician, John Murchland, showed that the doubly-constrained gravity model with the negative exponential deterrence function is equivalent to a certain nonlinear programming problem (Murchland, 1966). In his unpublished note, he offered a conjecture that the trip distribution and traffic assignment problems might be "combined" into a more general problem. With this conjecture, Murchland connected the sequential procedure with the network equilibrium model of Beckmann et al, although he may not have realized it at the time. Murchland was an expert on shortest route algorithms. Like others of that period, he probably did not yet understand that the traffic assignment problem could be formulated as a nonlinear programming problem. The connection between these two problems, once identified, is direct and straightforward. In a paper presented in Germany in 1969, Murchland (1970) proposed such a combined model, perhaps having absorbed the basics of the traffic assignment problem from Gibert (1968). When I met Murchland in 1970, he had a working computer code. However, the formulation was somewhat awkward, and certainly was not transparent.

Another formulation of a combined model of trip distribution and assignment was described by Tomlin (1967, 1971). In this case, the assignment part was formulated as the minimization of total travel costs; however, link costs and therefore OD costs were fixed. A solution algorithm was also proposed. The findings of Murchland, Tomlin and others set the stage for two major contributions that effectively solved the problem of formulating the origin-destination, mode choice and road traffic assignment problems as a single integrated model, rather than a sequential procedure. By proposing algorithms these authors also sought to solve their models efficiently, thereby also answering the question of how to iterate the sequential procedure to equilibrium. During the same period of the early 1970s, new and more practical results on equilibrium assignment of fixed OD flows were offered. I first review the fixed flow models, and then return to combined models.

Two Ph.D. students in operations research, Larry LeBlanc (1973, 1975) and Sang Nguyen (1973, 1974, 1976) proposed and implemented algorithms for the traffic assignment problem with fixed OD flows. LeBlanc examined the formulation of Beckmann et al, and noticed that the quadratic programming algorithm of Frank and Wolfe (1956) was potentially an efficient approach because the linearized subproblem could be solved as a shortest route problem. He demonstrated the potential of the problem on a small test network he devised for the city of Sioux Falls, South Dakota. Nguyen proposed an adaptation of the convex-simplex method, taking advantage of the node-link network structure. Later he compared his algorithm with LeBlanc's but found that neither was dominating for his the test networks and the computers of the day.

Nguyen and his advisor, Michael Florian, formulated a combined model of trip distribution and assignment, and proposed an algorithm based on generalized Benders decomposition (Florian et al, 1975). Later they expanded the formulation to include mode choice, and proposed a full linearization algorithm based on Frank and Wolfe (Florian and Nguyen, 1978). In both papers they strongly related the sequential procedure to the formulation of Beckmann et al, and used the term "combined model" to describe their result.

During the same period, Suzanne Evans (1973, 1976) independently assembled the formulations proposed by Beckmann et al, Gibert, Murchland, Tomlin, Wilson and others, and analyzed the properties of the model using the convex analysis tools of Rockafellar (1967). She systematically worked through all known formulations and analyzed their properties. Then she proposed a partial linearization algorithm, which may be seen to be a generalization of the linearization algorithm of Frank and Wolfe (1956); see Patriksson (1994) for mathematical details. Unlike LeBlanc and Nguyen, her thesis and subsequent papers were theoretical, without computational implementations or tests. That said, her 311-page thesis contains an extensive synthesis of formulations and their mathematical properties for the period 1950-1970; since it was written concurrently with the research of LeBlanc, and of Florian and Nguyen, she does not consider their findings.

The findings of Evans, Florian and Nguyen, together with earlier insights by Murchland and Tomlin, complete the circle from Beckmann et al to the sequential procedure of Carroll and other early practitioners, and back to Beckmann. Given the interest of practitioners in better ways to solve the traffic assignment problem, and in solving the sequential procedure with feedback, these combined models could have provided the breakthrough needed to improve the validity and precision of applications in practice. In fact, they were largely ignored. The story of what did happen to implement and test these algorithms is considered in the next sections of this saga.

4.3 Implementation and Validation of Network Equilibrium Models

With the reformulation of the original model of Beckmann et al in the context of urban travel forecasting procedures of the 1970s, and the prospects for convergent algorithms for solving these "combined models," experimental implementation and validation soon followed. This section briefly reviews these developments.

Carolyn Frank (1978) implemented and compared the algorithms proposed by Evans (1976) and Florian and Nguyen (1978) for the doubly-constrained, negative exponential gravity model with flow-dependent travel times. She showed that the partial linearization algorithm of Evans was quite superior to the full linearization algorithm advocated by Florian and Nguyen, even for a small, but realistic network for Hull, Ontario. Moreover, Larry LeBlanc built on his thesis research to consider combined models of mode and route choice (Abdulaal and LeBlanc, 1979; LeBlanc and Farhangian, 1981). In this context, he also performed tests of full vs. partial linearization algorithms, coming to similar conclusions as Frank.

The studies by Frank and LeBlanc on the Evans algorithm became the principal basis for research with my students during the 1980s, often in collaboration with LeBlanc. This research was initiated in 1979 using 1975 estimates for the Chicago region prepared by the Chicago Area Transportation Study (CATS), and later the 1990 home interview survey of CATS in conjunction with 1990 census data. Combined models incorporating the auto and transit modes were implemented and solved for a single-class, singly-constrained model for a 400 zone system, as presented in Boyce et al (1983) and Boyce (1984, 1990), and for a two-class, doubly- constrained model estimated and validated on a 1790 zone system, as reported by Boyce and Bar-Gera (2003). Reviews related to this work are found in Boyce et al (1988) and Boyce and Daskin (1997).

Lars Lundqvist and I implemented a single class, two-mode model of Stockholm in 1983 solved by the Evans algorithm; preliminary findings are found in Boyce and Lundqvist (1987). Optimization methods for solving and calibrating this model were evaluated by Torgil Abrahamsson (1996); see Abrahamsson and Lundqvist (2000). Alan Horowitz (2000) also implemented and devised codes for the Evans algorithm, incorporating its features into his *Quick Response System II* (*QRS II*).

Florian and his colleagues at the University of Montreal engaged in an implementation of a two-mode equilibrium model with fixed origin-destination trip tables (Florian et al, 1979), following a validation effort for the traffic assignment problem for Winnipeg (Florian and Nguyen, 1976) and a formulation of a two-mode model (Florian, 1977). These related works became the basis for the creation of *The EMME/2 Transportation Planning Software*, a system of tools for solving the sequential procedure and combined models programmed by the user. A combined model that included trip generation was proposed by Nabil Safwat (Safwat and Magnanti, 1988), and was implemented with data for Texas cities and Riyadh, Saudi Arabia by Mohamad Hasan (Safwat and Hasan, 1989; Hasan and AlGadhi, 1998; Hasan and Safwat, 2000).

Most of the combined models described above were for a single class of travelers, so that travel for all purposes was aggregated into a single class. William Lam and Hai-Jun Huang (1992a, 1992b, 1994) appear to be the first to formulate and solve a combined model representing more than one class, although various types of multiclass models have been proposed, one as early as 1954 by Prager; see Patriksson (1994, pp. 51-54). Lam and Huang found a novel way to formulate their model as an optimization problem; however, in their model, classes correspond to modes, rather than trip purposes or socio-economic groups, as is more common. Jonquín de Cea and Enrique Fernández (2001, 2003) extended these models to include route choice on congested transit networks, in contrast to the fixed transit travel times assumed in earlier formulations. Their models were implemented in several Chilean cities, including Santiago where they have been extensively applied in redesigning the bus and rapid transit network. The model has also been offered as a commercial software system called *ESTRAUS*. Additional details about multiclass combined models are reviewed by Boyce and Bar-Gera (2004).

Finally, Hillel Bar-Gera (2002) introduced his origin-based traffic assignment algorithm, the first algorithm able to solve the traffic assignment problem for large-scale networks to fine convergence with reasonable computing effort. Bar-Gera and Boyce (2003) extended the origin-based algorithm to solve a single-class, multimodal combined model, formulated as a fixed point problem. The algorithm rapidly solves this model for the Chicago regional road network. Although the algorithm has not yet been applied to a multiclass model, it has the potential of solving a detailed travel forecasting model to the fine convergence necessary for thorough comparisons of alternative land use and transportation scenarios.

4.4 Generalization of the Optimization Formulation

Coincidently with model implementations, calibration and algorithm testing that began in the late 1970s, several new and more general formulations were identified. In the following brief review, I only mention the principal papers. Reviews of the mathematics are available in Michael Patriksson (1994) and Michael Florian and Donald Hearn (1995). As with the identification of solution algorithms in the early 1970s, several of these works appeared in a short period.

Dickson (1977) was perhaps the first to state a variational inequality formulation corresponding to

Wardrop's criteria, although his model has separable link travel costs; see Patriksson (1994, p. 84). Michael Smith (1979) and Stella Dafermos (1980) independently formulated the network equilibrium problem as a variational inequality. While Smith focused on the existence, uniqueness and stability properties of the solution, Dafermos presented a computational algorithm based on her knowledge of algorithms for variational inequalities in economics and applied physics. Aashtiani and Magnanti (1981) generalized the formulation of the mode and route choice equilibrium model using a nonlinear complementarity formulation. Florian and Spiess (1983) contributed more specific models on mode choice to this area. Terry Friesz (1985) and his colleagues also made extensive theoretical investigations. Anna Nagurney (1993) synthesized and extended these results and others in her text on variational inequalities with applications to transportation networks and related problems.

Finally, a few research monographs and texts should be listed as more general references, in addition to Patriksson (1994), which serves as a general guide to the field. Yossi Sheffi (1985) synthesized earlier works on network equilibrium and his own findings with regard to stochastic route choice. Sven Erlander and Neil Stewart (1990) provided a mathematical synthesis of the mathematics of the gravity model and trip distribution. Norbert Oppenheim (1995) synthesized random utility models, not otherwise considered here, with network equilibrium approaches. Michael Bell and Yasunori Iida (1997) offered a synthesis of their work to date on network models including reliability aspects. Finally, Juan de Dios Ortúzar and Luis Willumsen (2001) provided an up-to-date text on network equilibrium, random utility and practitioner travel forecasting models.

5. Travel Forecasting Practice and Software Development

5.1 States of Practice in the US

Fragmented development of travel forecasting computer codes in the US led in the early 1970s to a more coordinated effort within the US Department of Transportation. The Urban Transportation Planning System (*UTPS*), whose development was led by Robert Dial, introduced into more widespread practice several model advances, as follows: the multinomial logit function for forecasting mode choice; a user-equilibrium algorithm for assigning auto trips to congested road networks; improvements in the coding of transit networks and the assignment of transit trips; and recognition of the need to achieve consistency among the four stages, with some concern about solving the procedure with feedback. The creation and distribution of *UTPS* also led to the preparation of manuals and training courses (US DOT, 1977).

Changes in US policy in 1981 led to a decision to terminate the development of *UTPS* and to encourage its dissemination for the IBM Personal Computer (PC) by Comsis, a consultant involved in prior code development, resulting in *MinUTP*. A somewhat similar product called *TranPlan* had been under development for some years, supported initially by the Control Data Corporation, an earlier competitor to IBM. These two software systems, either directly or indirectly encompassing the model and code development efforts of US DOT, were the initial versions of PC-based travel forecasting models and software; recently, these systems were merged into *CUBE* (Citilabs, 2004). Other software systems from that period have not survived.

In parallel with these developments, as noted above, Florian and colleagues at the University of Montreal, implemented an equilibrium-based multi-mode travel forecasting model, which became the basis for *EMME/2*, a commercial software system first released in the late 1980s for linking and solving the models of the four-stage procedure. In addition to a rigorous implementation of a user-equilibrium road assignment algorithm and a probabilistic transit route choice algorithm, *EMME/2* includes tools for solving the doubly-constrained trip distribution model, stochastic mode choice models, network coding and related utilities; see INRO (2004).

Building on the capabilities of the emerging field of Geographical Information Systems, another group of travel modelers developed *TransCAD* based on PC technology (Caliper Corporation, 2004). This system sought to embrace and incorporate various research advances throughout the US and beyond, including an early version of the combined model described above. Another US-developed, research-based software system, which has been found a market in smaller regions, is *QRS II* developed by AJH Associates (2004). More recently, the German software developer PTV (2004) has introduced its *ptv vision* software system into US practice. *EMME/2* and *TransCAD* enjoyed considerable success in the US during the 1990s, and *EMME/2* developed an international following in the UK, Sweden, Canada, Australia and New Zealand, and South Africa. From its strong base in Germany, PTV has expanded into other European countries.

The late 1980s were a period of change in the US, as environmental-based interest groups successfully sought to challenge the status quo of transportation planning in the courts (Garrett and Wachs, 1996). The Bay Area lawsuit led to the Clean Air Act Amendments of 1990, and the requirement of conformity analysis for metropolitan planning organizations (MPO). To qualify a transportation plan for US funding, each MPO has to demonstrate that building the proposed system would not result in the deterioration in air quality. One implication of this requirement is the need to determine road link flows and speeds by time periods during the typical weekday, in order to forecast atmospheric emissions. Since available travel forecasting methods were based on a 24-hour period, an ad hoc factoring method, based only on surveys, was devised to allocate flows by time periods of the day (Deakin, Harvey, Skabardonis 1993).

Attempts to reform transportation planning practice also led to new provisions in the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). Many new requirements were placed on planners and MPOs by this legislation, but the one most pertinent to this discussion is the requirement to solve the four-stage procedure with “feedback.” During the early 1990s many MPOs updated their travel forecasting models for the first time in many years, which often resulted in adoption of a different software system.

In response to these legislative mandates, model updates and related technical requirements, the US Department of Transportation and the US Environmental Protection Agency created the Travel Model Improvement Program http://tmip.tamu.edu/. The agencies had begun to devise a series of short and longer run model improvement tasks, when they were directed by the US Congress to fund a large-scale systems simulation project at the Los Alamos National Laboratory (2004) dubbed TRANSIMS. This ambitious effort set out to apply microsimulation techniques to represent interrelated travel and location-based decisions of each simulated inhabitant of a metropolitan area throughout the 24-hour weekday. The result is not a new, integrated model so much as an ambitious system of computer programs related to choice of when, where, and how to travel, and the associated computation of atmospheric emissions. An attempt to implement the system was initially made for the Portland, Oregon, metropolitan area, but was not completed by 2001, as intended. At that time, a transportation modeling consultant was retained to make a final attempt to implement and validate the system. This effort remains a work in progress.

5.2 States of Practice in the UK*

By the middle of the 1970s, the era of the large urban transport studies in the UK was ending and few cities had the resources or inclination to maintain large transport models. With few exceptions, most notably in London, model systems constructed a decade or more earlier, and the databases that supported them, were allowed to atrophy. Much local expertise dispersed, and the under-resourced and lonely task of local authority modelers fell on fewer and fewer shoulders. Where necessary, international consultants, some with the founders of the discipline at their helm, were called upon to address strategic and tactical issues, often in the form of urban development projects and traffic studies. Throughout the 1990s, there was also an increasing interest in demand restraint; several metropolitan areas conducted modeling exercises involving public transport, traffic restraint and, in some cases, limited highway investment.

For such purposes, several software packages are available in the private sector. Indeed, a key feature of the 1980s and early 1990s was one of fragmentation of travel forecasting software, and its organization within management information systems as a collection of models that were refined and integrated on a "pick-and-mix" basis (see Williams, 2004, for an overview). In turn, large consultancies found it in their interest to join forces with smaller specialist companies, particularly in the context of implementation of land use models, microeconometric studies of discrete choice (typically multimodal studies), and stated preference exercises.

A prominent example, and one of the most widely used and innovative in the UK, was Martin, Voorhees and Associates's software system *TRIPS*, now a component of *CUBE* (Citilabs, 2004). The various models of the four-stage approach were enhanced and offered in both synthetic and incremental (pivot point) forms (Bates et al, 1987). In the former, travel behavior is modeled at the cross section and elasticity parameters estimated prior to forecasting, while in the latter, changes from a given state (e.g. the base state) are estimated utilizing given elasticity parameters. Both are available for application at the micro (individual data) or aggregate (grouped data) level.

Since the mid-1980s, *SATURN* (Van Vliet, 1982) was extensively applied in and beyond the UK. Matrix updating was a standard part of the package and was widely applied to breathe new life into dated trip matrices. Initially promoted as a "modern" assignment program, it was the first to incorporate a rigorous approach to equilibrium assignment in UK travel forecasting. *SATURN* had the capability of working at different levels of network resolutions, requiring different and compatible specifications, and in the 1990s became a framework within which research was conducted on consistently integrating demand and assignment models (Atkins, 2004). For example, Williams, van Vliet and Kim (2001) applied the *SATURN* software system with variable demand to investigate the contribution of suppressed and induced traffic to emissions and user benefits derived from new roads and capacity changes under different pricing regimes.

Current research initiatives on specification and equilibration are being undertaken in conjunction with *SATURN,* whose latest version includes a wide variety of demand functions, including nested logit forms. More broadly, a software system *DIADEM* (Dynamic Integrated Assignment and Demand Modelling) has been constructed by lead consultants Mott MacDonald (2003), which will permit an interface between a wide range of demand models and assignment procedures. It is intended that great flexibility will be permitted in the specification of demand functions, with respect to the market segmentation, their functional form, and the parameter values included in hierarchical models.

In the late 1980s and early 1990s there was considerable interest in both the US and UK in the effect of congestion on travel, and of additional traffic (and vehicle miles) that might be induced by capacity expansion and new roads. In the US the effect of induced traffic on atmospheric emissions and energy consumption was a matter of particular concern (TRB, 1995) while in the UK it was the effect on congestion, traffic growth and economic benefits. The report of the Standing Advisory Committee for Trunk Road Assessment (SACTRA, 1994) had a substantial impact on the methods by which travel forecasting and the appraisal of schemes were conducted.

In guidance to be issued in 2004 by the UK Department for Transport in relation to infrastructure investment, the advice on variable demand analysis is likely to have significant implications for both strategic and tactical model development. Following earlier advice (Highways Agency, 1997), the emphasis will be on model design and will be much enhanced in terms of the discussion of: model complexity, proposed structures, model parameters, the design and convergence of equilibrium-seeking algorithms, model validation and sensitivity analysis. The implications of suppressed and induced traffic on scheme appraisal will likely form a particularly important aspect of this advice, given their effect on user benefits over the lifetime of a scheme.

As part of the research underpinning the ongoing scrutiny of impact models and appraisal methods by the UK Department for Transport, Bly et al (2001) reviewed the structure, parameters and validation of 24 of those models that have been implemented in the last ten years or so, drawing largely but not exclusively from UK applications. These models exhibit wide variation in their explicit representation of responses and range from five-level nested structures in the form G/D/MS/T/A (with T representing the time-of-day period choice) to 2-stage models.

Many models are in the form of nested (logit) functions, some in incremental form (Bates et al, 1987). It is sobering to note the finding by Bly et al that several of the models with nested structures are endowed with estimated parameter values that are inconsistent with the structures selected, suggesting that model validation is deficient and some of the predictions may have been unreliable. Where incremental nested logit forms have been adopted, parameter values have been inserted that are consistent with the selected specifications. The above UK experience suggests that current US practice, in which distribution models often include empirically derived deterrence functions, ought to be scrutinized to confirm that the demand functions are endowed with acceptable elasticity properties.

6. Future Prospects

Combined or integrated models of origin-destination, mode, route and time period choice can be implemented and solved with practitioner software systems such as *CUBE*, *EMME/2*, *ptv vision*, *QRS II*, *SATURN* or *TransCAD*, if the travel forecaster has a detailed understanding of the models and the solution algorithm. My ongoing discussions with practitioners reveal that few practitioners have this requisite expertise. The observed inability of many practitioners to solve the sequential procedure with feedback in a convergent manner is one indication of this dilemma. An alternative approach is to design a software system which directly solves such an integrated model, including providing a number of options to the user. This approach was taken by the Chilean software vendor, MCT, in the development of its product *ESTRAUS* (MCT, 2004).

The distinction between these two software development philosophies may be more subtle than is generally appreciated. The former approach offers the forecaster a "tool kit" with which to "build" a model. Then the practitioner must acquire the expertise to use it, an uncertain and arduous process with many pitfalls. The latter provides a "canned" model and an algorithm for solving it. When presented with a description of ESTRAUS, a typical user's comment is, "This is interesting, but I don't know how we would use it to solve *our* model."

This comment may offer a clue to the dilemma of vastly improving or revolutionizing practice, which presumably is one of the aims of academics as well as software vendors. Would a standard set of models with numerous user options provide a more effective pathway to improving travel forecasting practice than attempting to upgrade practice model by model? Certainly, this option is worth considering, so long as several of the vendors can effectively participate in the software market. That is, I am not suggesting a "one size fits all" approach that seems to characterize the ill-advised *TRANSIMS* effort. Rather, I think, what this line of thinking suggests is *standards* for travel forecasting software, as well as for travel forecasting practice. Reaching agreement on such standards for software, as well as its use, may well be equally arduous, but it does have the merit that the participants (vendors and practitioners) must confront and examine the quality of the methods in use, and how they apply them.

In the meantime, academics have other responsibilities to this field, in addition to getting on with their research. One, which has been badly neglected, is the preparation of textbooks and related teaching materials. Considering the size of this field, it is a scandal that there is only one current textbook on travel forecasting accessible to practitioners and students in professional degree programs (Ortúzar and Willumsen, 2001). The Internet provides an inexpensive and innovative way to publish teaching materials. Few instructors post their class notes or book manuscripts on their websites, however; two examples of such postings are Nagurney (2004) and Train (2003).

An attitude that travel forecasting ought to be simple, whereas in fact it is actually quite complex, together with missed opportunities, have conspired to yield a state of practice of travel forecasting that is substantially inferior to the state of the art resulting from 50 years of solid research. Since excellent practice inspires and financially supports excellent research, our field may be far behind where it might have been, given its impressive beginning in a small study by mathematical economists and operations researchers. Even so, intellectual curiosity about urban travel, together with genuine societal problems related to energy use, atmospheric pollution, and safety suggest that our field will continue to attract some of the best scientific talent. Let us hope that their efforts will be as fruitful as those of some of the leading innovators examined in this review.

Acknowledgements

My recent analyses of the impact of *Studies in the Economics of Transportation* were initiated while an Erskine Fellow at the University of Canterbury, Christchurch, New Zealand, April-June 2003, and continued while a Visiting Professor at the University of Pennsylvania during the Fall Term 2003. I am grateful to both institutions for their support. Initial findings were reported at the 50th North American Meetings of the Regional Science Association International, Philadelphia, PA, November 22, 2003 (see Boyce, Mahmassani and Nagurney, 2004). I am please to acknowledge comments on earlier drafts of the paper by Anna Nagurney, Michael Patriksson and Huw Williams.

References

H. Z. Aashtiani and T. L. Magnanti (1981) “Equilibria on a congested transportation network,” *SIAM Journal of Algebraic and Discrete Methods* 2, 213-226.

T. Abrahamsson (1996) Network Equilibrium Approaches to Urban Transportation Markets, Ph.D. thesis, Regional Planning, Royal Institute of Technology, Stockholm, Sweden.

T. Abrahamsson and L. Lundqvist (1999) “Formulation and estimation of combined network equilibrium models with applications to Stockholm,” *Transportation Science* 33, 80-100.

M. Abdulaal and L. J. LeBlanc (1979) “Methods for combining modal split and equilibrium assignment models,” *Transportation Science* 13, 292-314.

AJH Associates (2004) *Quick Response System II*, http://my.execpc.com/~ajh/

J. Almond (1967) "Traffic assignment with flow-dependent journey times," *Vehicular Traffic Science, Proceedings of the Third International Symposium on the Theory of Traffic Flow*, L. C. Edie, R. Herman and R. Rothery (eds), American Elsevier, New York, pp. 222-234.

Atkins (2004) *SATURN*, http://www.its.leeds.ac.uk/software/saturn/

H. Bar-Gera (2002) "Origin-based algorithms for the traffic assignment problem," *Transportation Science* 36, 398-417.

H. Bar-Gera and D. Boyce (2003) "Origin-based algorithms for combined travel forecasting models," *Transportation Research* 37B, 405-422.

J. J. Bates, D. J. Ashley and G. Hyman (1987) "The nested incremental logit model: Theory and application to modal choice," *Proceedings*, 15th PTRC Summer Annual Meeting, Seminar C, Bath, UK.

M. Beckmann (1952) "Efficient transportation in networks," Cowles Commission Discussion Paper: Economics No. 2049, with the assistance of C. B. McGuire, Chicago, August 5.

M. J. Beckmann (1967) "On optimal tolls for highways, tunnels and bridges," *Vehicular Traffic Science*, *Proceedings of the Third International Symposium on the Theory of Traffic Flow*, L. C. Edie, R. Herman and R. Rothery (eds), American Elsevier, New York, pp. 331-341.

M. J. Beckmann (1967) "On the theory of traffic flow in networks," *Traffic Quarterly* 21, 109-116.

M. Beckmann, C. B. McGuire and C. B. Winsten (1956) *Studies in the Economics of Transportation*, with an Introduction by T. C. Koopmans, Yale University Press, New Haven, CT; published as Rand-RM-1488-PR, Rand Corporation, Santa Monica, CA, 1955.

M. Bell and Y. Iida (1997) *Transportation Network Analysis*, Wiley, Chichester.

M. E. Ben-Akiva (1974) "Structure of passenger travel demand models," *Transportation Research Record* 526, 26-42.

P. Bly, P. Emmerson, T. Van Vuren, A. Ash and N. Paulley (2001) User-friendly Multistage Modelling Advice, Phase 2: Modelling Parameters, Calibration and Validation, Project Report for the Integrated Transport Economic Appraisal Division, UK Department for Transport, Local Government and the Regions, London.

D. E. Boyce (1984) "Network models in transportation/land use planning," *Transportation Planning Models*, M. Florian (ed) North-Holland, Amsterdam, pp. 475-498.

D. E. Boyce (1990) "Network equilibrium models of urban location and travel choices: A new research agenda," *New Frontiers in Regional Science*, M. Chatterji and R. E. Kuenne (eds) Macmillan, New York, pp. 238-256.

D. Boyce and H. Bar-Gera (2003) "Validation of urban travel forecasting models combining origin-destination, mode and route choices," *Journal of Regional Science* 43, 517-540.

D. Boyce and H. Bar-Gera (2004) "Multiclass combined models for urban travel forecasting," *Network and Spatial Economics* 4, 115-124.

D. E. Boyce, K. S. Chon, Y. J. Lee, K. T. Lin and L. J. LeBlanc (1983) "Implementation and computational issues for combined models of location, destination, mode and route choice," *Environment and Planning A* 15, 1219-1230.

D. E. Boyce and M. S. Daskin (1997) "Urban transportation," Chapter 7, *Design and Operation of Civil and Environmental Engineering Systems,* C. ReVelle and A. McGarity (eds.) Wiley, New York, pp. 277-341.

D. E. Boyce, N. D. Day and C. McDonald (1970) *Metropolitan Plan Making*, Regional Science Research

Institute, Philadelphia.

D. E. Boyce, L. J. LeBlanc and K. S. Chon (1988) "Network equilibrium models of urban location and travel choices: A retrospective survey," *Journal of Regional Science* 28, 159-183.

D. E. Boyce and L. Lundqvist (1987) "Network equilibrium models of urban location and travel choices: alternative formulations for the Stockholm region," *Papers*, *Regional Science Association* 61, 91-104.

D. E. Boyce, H. S. Mahmassani and A. Nagurney (2004) "A retrospective on Beckmann, McGuire and Winsten's *Studies in the Economics of Transportation*," paper based on a panel at the 50th North American Meetings of the Regional Science Association International, Philadelphia, November 2003.

D. E. Boyce and H. C. W. L. Williams (2004) "Urban travel forecasting in the USA and UK," *Methods and Models in Transport and Telecommunications*: *Cross-Atlantic Perspectives*, A. Reggiani and L. Schintler (eds), forthcoming.

M. Bruynooghe, A. Gilbert and M. Sakarovitch (1969) "Une methode d'affectation du trafic," *Proceedings of the 4th International Symposium on the Theory of Road Traffic*, W. Leutzbach and P. Baron (eds), Bundesminister fur Verkehr, Abt. Strassenbau, Bonn, Germany, pp. 198-204.

Caliper Corporation (2004) *TransCAD*, http://www.caliper.com/

E. W. Campbell (1956) "A mechanical method for assigning traffic to expressways," *Bulletin* 130, Highway Research Board, Washington, DC, pp. 27-46.

M. E. Campbell (1952) "Foreword," *Traffic Assignment*, *Bulletin* 61, Highway Research Board, Washington, DC, pp. iii-iv.

J. D. Carroll, Jr. (1959) "A method of traffic assignment to an urban network," *Bulletin* 224, Highway Research Board, Washington, DC, pp. 64-71.

A. Charnes and W. W. Cooper (1958) "Extremal principles for simulating traffic flow in a network," *Proceedings of the National Academy of Sciences* 44, 201-204.

A. Charnes and W. W. Cooper (1961) "Multicopy traffic networks," *Theory of Traffic Flow*, R. Herman (ed), Elsevier, Amsterdam, pp. 85-96.

Citilabs (2004) *CUBE*, http://www.citilabs.com/

W. W. Cooper (2002) "Abraham Charnes and W. W. Cooper (et al.): A brief history of a long collaboration in developing industrial uses of linear programming," *Operations Research* 50, 35-41.

S. C. Dafermos (1968) Traffic Assignment and Resource Allocation in Transportation Networks, Ph.D. thesis, Operations Research, Johns Hopkins University, Baltimore.

S. C. Dafermos (1971) "An extended traffic assignment model with applications to two-way traffic," *Transportation Science* 5, 366-389.

S. C. Dafermos (1972) "The traffic assignment problem for multiclass-user transportation networks," *Transportation Science* 6, 73-87.

S. C. Dafermos (1973) "Toll patterns for multi-class user transportation networks," *Transportation Science* 7, 211-223.

S. Dafermos (1980) "Traffic equilibrium and variational inequalities," *Transportation Science* 14, 42-54.

S. Dafermos (1982) "The general multimodal network equilibrium problem with elastic demand," *Networks* 12, 57-72.

S. Dafermos (1983) "An iterative scheme for variational inequalities," *Mathematical Programming* 26, 40-47.

S. C. Dafermos and F. T. Sparrow (1969) "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards* 73B, 91-118.

S. C. Dafermos and F. T. Sparrow (1971) "Optimal resource allocation and toll patterns in user-optimized transportation networks," *Journal of Transport Economic and Policy* 5, 184-200.

A. J. Daly and S. Zachary (1978) "Improved multiple choice models," *Determinants of Travel Choice*, D. A. Hensher and M. Q. Dalvi (eds) Saxon House, Westmead, UK.

G. B. Dantzig (2002) "Linear programming," *Operations Research* 50, 42-47.

Deakin, Harvey, Skabardonis, Inc. (1993) Manual of Regional Transportation Modeling Practice for Air Quality, National Association of Regional Councils, Washington, DC.

J. de Cea and J.E. Fernández (2001) "ESTRAUS: A simultaneous equilibrium model to analyze and evaluate multimodal urban transportation systems with multiple user classes," *Proceedings of the Ninth World Conference on Transport Research*, Seoul, Korea.

J. de Cea, J. E. Fernández, V. DeKock, A. Soto and T. L. Friesz (2003) "ESTRAUS: A computer package for solving supply-demand equilibrium problems on multimodal urban transportation networks with multiple user classes," presented at the Annual Meeting of the Transportation Research Board, Washington, DC.

T. J. Dickson (1977) "Traffic assignment," unpublished note, Traffic Studies Group, University College London, London.

R. J. Duffin (1947) "Nonlinear networks. IIa" *Bulletin of the American Mathematical Society* 53, 963-971.

S. Erlander and N. F. Stewart (1990) *The Gravity Model in Transportation Analysis*, VSP, Utrecht, The Netherlands.

S. P. Evans (1973) Some Applications of Optimisation Theory in Transport Planning, Ph.D. thesis, Civil Engineering, University College London, London.

S. P. Evans (1976) "Derivation and analysis of some models for combining trip distribution and assignment," *Transportation Research* 10, 37-57.

M. Florian (ed) (1976) *Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems* 18, Springer-Verlag, Berlin.

M. Florian (1977) "A traffic equilibrium model of travel by car and public transit modes," *Transportation Science* 11, 169-179.

M. Florian and H. Spiess(1983) "On binary mode choice/assignment models," *Transportation Science* 17, 32-47.

M. Florian, R. Chapleau, S. Nguyen, C. Achim, L. James-Lefebvre, S. Galarneau, J. Lefebvre and C. Fisk (1979) "Validation and application of an equilibrium based two-mode urban transportation planning method (EMME)," *Transportation Research Record* 728, 14-23.

M. Florian and D. Hearn (1995) "Network equilibrium models and algorithms," *Network Routing, Handbooks in Operations Research and Management Science* 8, M. O. Ball, T. L. Magnanti, C. L. Monma and G. L. Nemhauser (eds), Elsevier Science, Amsterdam, pp. 485-550.

M. Florian and S. Nguyen (1976) "An application and validation of equilibrium trip assignment methods," *Transportation Science* 10, 374-389.

M. Florian and S. Nguyen (1978) "A combined trip distribution modal split and assignment model," *Transportation Research* 12, 241-246.

M. Florian, S. Nguyen and J. Ferland (1975) "On the combined distribution-assignment of traffic," *Transportation Science* 9, 43-53.

M. Frank and P. Wolfe (1956) "An algorithm for quadratic programming," *Naval Research Logistics Quarterly* 3, 95-110.

C. Frank (1978) A Study of Alternative Approaches to Combined Trip Distribution-Assignment Modeling, Ph.D. thesis, Regional Science, University of Pennsylvania, Philadelphia.

T. L. Friesz (1985) "Transportation network equilibrium, design and aggregation: Key developments and research opportunities," *Transportation Research* 19A, 413-427.

D. Gale, H. W. Kuhn and A. W. Tucker (1951) "Linear programming and the theory of games," *Activity Analysis of Production and Allocation*, T. C. Koopmans (ed), Wiley, New York, pp. 317-329.

M. Garrett and M. Wachs (1996) *Transportation Planning on Trial*, Sage, Thousand Oaks, CA.

S. I. Gass (2002) "The first linear-programming shoppe," *Operations Research* 50, 61-68.

A. Gibert (1968) "A method for the traffic assignment problem when demand is elastic," LBS-TNT-85, Transport Network Theory Unit, London Business School, London.

M. K. Hasan and S. A. AlGadhi (1998) "Application of simultaneous and sequential transportation network equilibrium models to Riyadh, Saudi Arabia," *Transportation Research Record* 1645, 127-132.

M. K. Hasan and K. N. Safwat (2000) "Comparison of two transportation network equilibrium modeling approaches," *ASCE Journal of Transportation Engineering* 126, 35-40.

Highways Agency (1997) *Design Manual for Roads and Bridges, Volume 12, Traffic Appraisal of Road Schemes. Section 2, Part 2: Induced Traffic Appraisal*, The Stationery Office, London.

A. J. Horowitz (2000) Quick Response System II, Version 6.0, travel forecasting software distributed by McTrans, University of Florida, and PC-Trans, University of Kansas.

INRO Consultants, Inc. (2004) *EMME/2*, http://www.inro.ca/

N. A. Irwin and H. G. Von Cube (1962) "Capacity restraint in multi-travel mode assignment programs," *Bulletin* 347, Highway Research Board, Washington, DC, pp. 258-287.

W. Isard (1960) *Methods of Regional Analysis*: *an Introduction to Regional Science*, Technology Press of MIT and Wiley, New York.

W. S. Jewell (1967) "Models for traffic assignment," *Transportation Research* 1, 31-46.

M. B. Johnson (1964) "On the economics of road congestion," *Econometrica* 32, 137-150.

N. O. Jorgensen (1963) Some Aspects of the Urban Traffic Assignment Problem, ITTE Graduate Report #1963:9 (M.S. thesis, Civil Engineering), University of California, Berkeley; available from the Institute of Transportation Studies Library.

F. H. Knight (1924) "Some fallacies in the interpretation of social cost," *Quarterly Journal of Economics*, 38, 582-606.

J. E. Kohl (1841) "Der verkehr und die ansiedelungen der menschen in ihrer abhangigkeit von der gestaltung der erdorberflache," Dresden, Leipzig, Germany.

T. C. Koopmans (1949) "Optimum utilization of the transportation system," Econometrica 17 (supplement), 136-146.

T. C. Koopmans (ed) (1951) *Activity Analysis of Production and Allocation*, Wiley, New York.

H. W. Kuhn (1976) "Nonlinear programming: A historical view," *Nonlinear Programming*, R. Cottle and C. E. Lemke (eds), *SIAM-AMS Proceedings* 9, 1-26.

H. W. Kuhn (2002) "Being in the right place at the right time," *Operations Research* 50, 132-134.

H. W. Kuhn and A. W. Tucker (1951) "Nonlinear programming," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (ed), University of California Press, Berkeley, 481-492 (presented summer of 1950).

W. H. K. Lam and H.-J. Huang (1992a) "A combined trip distribution and assignment model for multiple user classes," *Transportation Research* 26B, 275-287.

W. H. K. Lam and H.-J. Huang (1992b) "Calibration of the combined trip distribution and assignment model for multiple user classes," *Transportation Research* 26B, 289-305.

W. H. K. Lam and H.-J. Huang (1994) "comparison of results of two models of transportation demand in Hong Kong: CDAM and a version of MicroTRIPS," *Journal of Advanced Transportation* 28, 107-126.

L. J. LeBlanc (1973) Mathematical Programming Algorithms for Large Scale Network Equilibrium and Network Design Problems, Ph.D. thesis, Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

L. J. LeBlanc and K. Farhangian (1981) "Efficient algorithms for solving elastic demand traffic assignment problems and mode-split assignment problems," *Transportation Science* 15, 306-317.

L. J. LeBlanc, E. K. Morlok and W. P. Pierskalla (1975) "An efficient approach to solving the road network equilibrium traffic assignment problem," *Transportation Research* 9, 309-318.

T. Leventhal, G. L. Nemhauser and L. Trotter, Jr. (1973) "A column generation algorithm for optimal traffic assignment," *Transportation Science* 7, 168-176.

Los Alamos National Laboratory (2004) *TRANSIMS*, http://www-transims.tsasa.lanl.gov/.

B. V. Martin, F. W. Memmott, 3rd, and A. J. Bone (1962) Principles and Techniques of Predicting Future Demand for Urban Area Transportation, Research Report No. 38, M.I.T., Cambridge, MA.

M. L. Manheim (1979) *Fundamentals of Transportation Systems Analysis,* MIT Press, Cambridge, MA.

D. McFadden (1978) "Modeling the choice of residential location," A. Karlqvist, L. Lundqvist, F. Snickars, J. Weibull (eds) *Spatial Interaction Theory and Planning Models*, North- Holland, Amsterdam, 75-96.

MCT, Ltda. (2004) *ESTRAUS*, http://www.MCTsoft.com/

E. F. Moore (1957) "The shortest path through a maze," International Symposium on the Theory of Switching, *Proceedings, Part II*, Harvard University, Cambridge, MA.

Mott MacDonald (2003) DIADEM Overview of the Algorithms, Software and Convergence Requirements, presented at the SATURN User Group Meeting, Epsom, UK.

J. D. Murchland (1966) "Some remarks on the gravity model of traffic distribution, and an equivalent maximization formulation," Report LSE-TNT-38, Transport Network Theory Unit, London Business School, London.

J. D. Murchland (1967) "Two remarks on congested assignment," Report LBS-TNT-53, Transport Network Theory Unit, London Business School, London.

J. D. Murchland (1970) "Road network traffic distribution in equilibrium," Mathematical Models in the Social Sciences 8, Anton Hain Verlag, Meisenheim am Glan, Germany, pp. 145-183 (in German translation).

A. Nagurney (1993) *Network Economic*, 2nd edition in 1999, Kluwer, Boston, MA.

A. Nagurney (2004) Network Economics and Sustainable Transportation Courses, http://supernet.som.umass.edu/austria_lectures/fulmain.html

J. F. Nash (1951) "Noncooperative games," *Annals of Mathematics* 54, 286-298.

G. F. Newell (2002) "Memoirs on Highway Traffic Flow Theory in the 1950s," *Operations Research* 50, 173-178 (published posthumously).

S. Nguyen (1974) "An algorithm for the traffic assignment problem," *Transportation Science* 8, 203-216.

S. Nguyen (1976) "A unified approach to equilibrium methods for traffic assignment," *Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems* 18, M. Florian (ed), Springer-Verlag, Berlin, pp. 148-182.

K. R. Overgaard (1967) "Testing a traffic assignment algorithm," *Vehicular Traffic Science*, *Proceedings of* *the Third International Symposium on the Theory of Traffic Flow*, L. C. Edie, R. Herman and R. Rothery (eds), American Elsevier, New York, pp. 215-221.

N. Oppenheim (1995) *Urban Travel Demand Modeling*, Wiley, New York.

J. D. Ortúzar and L. G. Willumsen (2001) *Modelling Transport*, 3rd Ed., Wiley, New York.

M. Patriksson (1994) *The Traffic Assignment Problem - Models and Methods*, VSP, Utrecht.

A. C. Pigou (1918) *The Economics of Welfare*, Macmillan, New York.

W. Prager (1954) "Problems of traffic and transportation," *Proceedings of the Symposium on Operations Research in Business and Industry*, Midwest Research Institute, Kansas City, MO, pp. 105-113.

W. Prager (1955) "On the role of congestion in transportation problems," *Zeitschrift für Angewandte Mathematik und Mechanik* 35, 264-268.

PTV AG (2004) *ptv vision* http://www.ptv.de/

R. T. Rockafellar (1967) "Convex programming and systems of elementary monotonic relations," *Journal of Mathematical Analysis and Applications* 19, 543-564.

K. N. A. Safwat and M. K. Hasan (1989) "Computational experience with a simultaneous transportation equilibrium model under varying parameters," *Transportation Research Record* 1251, 17-23.

K. N. A. Safwat and T. L. Magnanti (1988) "A combined trip generation, trip distribution, modal split, and trip assignment model," *Transportation Science* 22, 14-30.

M. Schneider (1959) "Gravity models and trip distribution theory," *Papers*, *Regional Science Association* 5, 51-56.

M. L. Senior and H. C. W. L. Williams (1977) "Model-based transport policy assessment; Part I: The use of alternative forecasting models," *Traffic Engineering and Control* 18, 402-406.

Y. Sheffi (1985) *Urban Transportation Networks*, Prentice-Hall, Englewood Cliffs, NJ.

M. J. Smith (1979) "Existence, uniqueness and stability of traffic equilibria," *Transportation Research* 13B, 295-304.

Standing Advisory Committee for Trunk Road Assessment (1994) *Trunk Roads and the Generation of Traffic*, The Stationery Office, London.

J. A. Tomlin (1967) Mathematical Programming Models for Traffic Network Problems, Doctoral dissertation, Mathematics, University of Adelaide, Adelaide, Australia.

J. A. Tomlin (1971) "A mathematical programming model for the combined distribution-assignment of traffic," *Transportation Science* 5, 122-140.

Transportation Research Board (1995) *Expanding Metropolitan Highways: Implications for Air Quality and Energy Use*, *Special Report* 245, National Research Council, Washington, DC.

K. E. Train (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK, http://emlab.berkeley.edu/users/train/.

A. W. Tucker (1957) "Linear and nonlinear programming," *Operations Research* 5, 244-257.

US DOT (1977) *An Introduction to Travel Demand Forecasting: A Self-instructional Text*, Federal Highway Administration and Urban Mass Transportation Administration, Washington, DC.

D. Van Vliet (1982) "SATURN: A modern assignment model," *Traffic Engineering and Control* 23, 575-581.

A. M. Voorhees (1955) "A general theory of traffic movement," *1955 Proceedings*, Institute of Traffic Engineers, New Haven CT, pp. 46-56.

A. A. Walters (1961) "The theory and measurement of private and social cost of highway congestion," *Econometrica* 29, 676-699; reprinted in *Transport*, D. Munby (ed), Penguin Books, Harmondsworth, England, 1968.

J. G. Wardrop (1952) "Some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers, Part II* 1, 325-378.

H. C. W. L. Williams (1977) "On the formation of travel demand models and economic evaluation measures of user benefit," *Environment and Planning A* 9, 285-344.

H. C. W. L. Williams (2004) "Themes in the development and application of transport planning models," *Urban and Regional Transportation Modeling*, D.-H. Lee (ed.) Edward Elgar, Northhampton, MA.

H. C. W. L. Williams and M. L. Senior (1977) "Model-based transport policy assessment; Part II: Removing fundamental inconsistencies from the models," *Traffic Engineering and Control* 18, 464-469.

H. C. W. L. Williams, D. van Vliet and K. S. Kim (2001) "The contribution of suppressed and induced traffic in highway appraisal, part 1: Reference states, and part 2: Policy tests," *Environment and Planning A* 33, 1057-1082 and 1243-1264.

A. G. Wilson (1967) "A statistical theory of spatial distribution models," *Transportation Research* 1, 253-269.

A. G. Wilson (1969) "The use of entropy maximising models in the theory of trip distribution, mode split and route split," *Journal of Transport Economics and Policy* 3, 108-126.

A. G. Wilson (1970) *Entropy in Urban and Regional Modelling*, Pion, London.

A. G. Wilson, A. F. Hawkins, G. J. Hill and D. J. Wagon (1969) "Calibrating and testing the SELNEC transport model," *Regional Studies* 3, 337-350.

H. J. Wooton and G. W. Pick (1967) "A model for trips generated by households," *Journal of Transport Economics and Policy* 1, 137-153.