To the Graduate Council:

I am submitting herewith a dissertation written by Carlos Felipe Grangeiro Loureiro entitled "Modeling Investment Options for Multimodal Transportation Networks." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Arun Chatterjee, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor
and Dean of The Graduate School

MODELING INVESTMENT OPTIONS FOR MULTIMODAL TRANSPORTATION NETWORKS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Carlos Felipe Grangeiro Loureiro

August 1994

DEDICATION

This work is dedicated:

To my parents, Ronald and Carísia, in recognition of their love and devotion with which I have been touched and inspired throughout my life to keep fighting for my beliefs and dreams.

To my wife, Lídia, whose love, companionship, concerns, and all the sacrifices she has made toward the accomplishment of this old dream have showed me how stronger I am with her by my side.

ACKNOWLEDGMENTS

During the course of my graduate studies in the United States, a number of scholars have assisted me thoughtfully. Grateful thanks and appreciation are expressed to my dissertation committee chairman, Dr. Arun Chatterjee, for his guidance and encouragement throughout the course of this research. A very special appreciation is extended to Dr. Bruce Ralston for his support, enthusiasm, and valuable assistance to develop, improve, and complete this dissertation. Sincere thanks also go to the other members of my committee, Dr. Michael Bronzini and Dr. Frederick Wegmann, for their comments and suggestions in the preparation of this manuscript.

I would also like to express my immense gratitude to Dr. Frank Southworth, who introduced me to the topic of network design and devoted an enormous and generous amount of time to helping and guiding me during the development of this research effort. Special thanks are extended to Dr. Julian Ray, who allowed the use of his computer code for adaptation and expansion, and to Dr. Cheng Liu, who helped me during the stressing phase of debugging the computer program. Appreciation is also extended to the Department of Geography at The University of Tennessee for allowing the use of its GIS and computer facilities.

This research would not have achieved its complete objectives without the personal interest and valuable help of Mr. Milton Xavier and the staff of the Technological Research Institute (IPT-SP) in São Paulo, who provided the majority of the data used in the development of the Brazilian pilot application of the model presented herein. Special thanks also go to Dr. Theodor Crainic and Dr. José Eugênio Leal for contributing with the additional data used in this study.

The author gratefully acknowledges the financial support of the Brazilian Government through CAPES - Federal Agency for Post-graduate Education, and promises to devote all of his future efforts to return the investment made by the Brazilian people in his nine years of college and graduate education.

ABSTRACT

This research presents the development of a Multicommodity Multimodal Network Design (MCMND) model to be used as a planning tool when determining investment priorities for intercity freight networks. Such a model can be especially useful for the case of regions in developing countries that have experienced considerable growth and, consequently, have demanded high service levels in certain areas and corridors that did not have an adequate transportation infrastructure.

The MCMND model is designed to select the best set of investment options for a multimodal regional network, given a limited investment budget. The main component of the model comprises the solution of a nonlinear bilevel multimodal network design problem formulated to choose investments so as to minimize the transportation costs incurred by users (shippers) and the environmental impacts caused by the use of less efficient modes of transportation in moving freight. Investment options to be considered by the model may involve the addition of new physical links to the network, the improvement of existing links (i.e., increase in capacity), and the location of intermodal transfer facilities at specified nodes of the network. The representation of the multimodal network is performed at a level of detail appropriate for strategic planning for a large region. The demand for transportation services is fixed and exogenous to the model. Mode choice in

shipping freight is modeled in combination with traffic flow assignment, with the assumption that goods are shipped at minimum total generalized costs. Travel costs (shipping cost and delay) are considered to be dependent on link flows, according to an exponential delay function. A path-based stochastic user equilibrium assignment algorithm is used to distribute trips over the multimodal network according to a logit-type model.

The proposed solution algorithm consists of a heuristic procedure based on a column generating technique, which allows the solution of large scale problems in a reasonable time. The multimodal network design problem is called the master problem, with its subproblem being defined as a bi-criteria path finding problem. This solution procedure was tested for a small network and it was found to approximate the optimal solution to that test problem in a satisfactory way.

The Tietê-Paraná Valley, located in the Southeast region of Brazil, was used in the development of a pilot application of the MCMND model to assess its applicability and its efficiency when dealing with large networks. This pilot application also served to emphasize the importance of an interface between the MCMND model and a Geographic Information System (GIS) in solving real-life problems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

Background

It is widest accepted that transportation investment acts as a catalytic factor in the economic development process. However, determining the optimum amount of transport capacity and its rate of investment, for a given region at a specified stage of development, remains problematic. This issue is complex particularly when dealing with developing countries, the focus herein, where large expenditures in transportation should not be made to the detriment of other fundamental sectors. Under a scenario of severe budget constraints, the governments of these countries must weight carefully any capital investments in transportation infrastructure when doing their overall strategic planning.

In this context, transportation investment priorities must be set and frequently revised to reflect the impacts on transportation resulting from changes in the other sectors of the economy, as well as the impacts on the economy as a whole caused by physical and institutional changes to the transportation system. Aspects such as the influence of planned investments on the pattern of transportation activities, the change of transportation demand over time, and the irreversibility of certain transportation investment decisions due to the long-life and high-cost characteristics of transport projects should all be considered carefully when striving to provide an efficient transportation system.

One of the major tasks involved with the development of an efficient transportation system is the determination of an optimal and balanced modal mix, in which each mode performs the function for which it is best suited. A multimodal system is essential not only to guarantee the pluses of cooperation and competition among different modes, but also to make sure that lower generalized transportation costs will result from the use of more energy efficient and environmentally cleaner transportation options. The determination of an optimal modal mix is by no means an easy task, given the conflicting objectives and trade-offs involved with the use of a particular mode for a specific function. The consequent conflicts of having both public and private sectors in charge of providing, maintaining, and operating interconnecting parts of a multimodal transportation system is also a compounding factor in the case of developing countries, where governments still play a strong role in controlling economic factors.

The above discussion reinforces the importance of analyzing investments in transportation with a systematic approach, considering the conflicts and interactions between the various components of the transportation system, e.g., passenger vs. freight, urban vs. intercity movements, etc. The specific question of determining investment priorities for <u>freight intercity networks</u> seems to be a especially relevant part of such a systematic approach for transportation planning in developing countries. This relevance is justified by the recently rising interest of governments in these countries in improving their freight transportation system as a way of making their products more competitive in a global economy and, thus, tackling their trade balance problems - what, nowadays, appears to be viewed as the panacea for promoting economic development. The need for setting investment priorities and scheduling project implementation for the intercity freight system is also highlighted in the context of regions in developing countries which have experienced considerable growth and, consequently, have demanded high service levels in certain areas and corridors that did not have an adequate transportation infrastructure.

Purpose of This Study

The importance of planning infrastructure investments for intercity freight transportation networks was the main motivation for carrying out this research, which ultimately strived to develop a theoretical basis for determining network investment priorities in a multimodal environment. The scope of this research encompasses mainly the development of an optimization model for the selection of a set of alternative changes to a multimodal freight transportation network that will result in a desired optimal flow scenario. In contrast to models dealing with network analysis problems[†] in which a given network configuration is analyzed to

---

[†] For a discussion on models dealing with transportation network analysis and network synthesis refer to Chapter I of Elshaygi, 1983, pp. 1-2.

predict or dictate how a particular pattern of traffic demand will utilize the network (see, for instance, Ralston et al., 1994; and Florian and Crainic, 1989), the model developed in this study solves a typical network synthesis problem, searching for an optimal network configuration which satisfies a given set of requirements. Such an optimal configuration is achieved by improving the performance of the network, measured by a systemwide function, to its maximum (or minimum) amount.

The *Multicommodity Multimodal Network Design (MCMND)* model presented herein is designed to select the best set of investment options for a multimodal (highway, railway, and inland waterway) regional network, given a limited investment budget. The main component of the MCMND model involves the solution of a nonlinear bilevel multimodal network design problem formulated to choose investments so as to minimize the transportation costs incurred by users (shippers) and the environmental impacts caused by the use of less efficient modes of transportation in moving freight. It is assumed that the more efficient available modes are being underutilized because of capacity bottlenecks on the system or due to the lack of transshipment facilities necessary to promote intermodalism.

Investment options to be considered by the model involve the addition of new links to the network, the improvement of existing links (i.e., increase in capacity), and/or the location of intermodal transfer facilities at specified nodes of the network. The MCMND model thus establishes a link between two branches of the broad field of network optimization: discrete choice network design and facility location. Moreover, it does so in a multimodal, multicommodity environment, what appears to be one of its major contributions to the current state of knowledge in that field.

This model is intended to be used as a tool within a strategic planning framework. Although the decisions to be made at the strategic level often involve qualitative considerations and quantitative uncertainties not well suited for a deterministic optimization model, the methods presented in this study can provide useful insight to decision makers in developing countries willing to expand and improve their existing freight transportation system. If designed to represent reality in an appropriate way, a suitable model can serve as a mathematical tool to be applied in conjunction with project evaluation techniques, and the results of these two approaches then submitted to a broader qualitative set of objectives to achieve the searched for "optimal" system configuration.

As an essential component of any transportation modeling effort (Lewis et al., 1990), this research also encompasses the development of a graphic interface between the MCMND model and a Geographic Information System (GIS). This interface not only simplifies the data input-output process, but also allows users to visually evaluate the impacts of alternative design options.

In the next chapter, a literature review of the state of the art in the interrelated areas of Intercity Freight Network Models, Network Design, and Facility Location Problems is presented. This review is necessary for a better understanding of the needs for the model presented in this dissertation and the premises on which it is based on.

The subsequent chapters of this dissertation describe the MCMND model and its assumptions, emphasizing the formulation of the embedded multimodal network design problem (Chapter III). In Chapter IV, a heuristic algorithm based on column generation techniques is proposed, which allows the solution of large scale problems in a reasonable time. Chapter V presents a small test network, with 7 nodes and 18 arcs, which is used to assess the efficiency of the proposed solution algorithm.

Chapter VI includes the development of a pilot application of the MCMND model, schematically depicted in Figure 1, to a real world problem comprising the location of transshipment terminals along the Tietê-Paraná waterway, located in the Southeast region of Brazil. This region was selected as a case study due to the multimodal characteristics of its freight transportation system, its economic relevance within the continent of South America, and the recently expressed interest of both state and federal governments in Brazil to invest in infrastructure improvements to the multimodal network, fostering the usage of their inland waterway system. The application of the MCMND model to the Brazilian case also served to emphasize the importance of the interface between the model and a GIS when dealing with real-life problems.

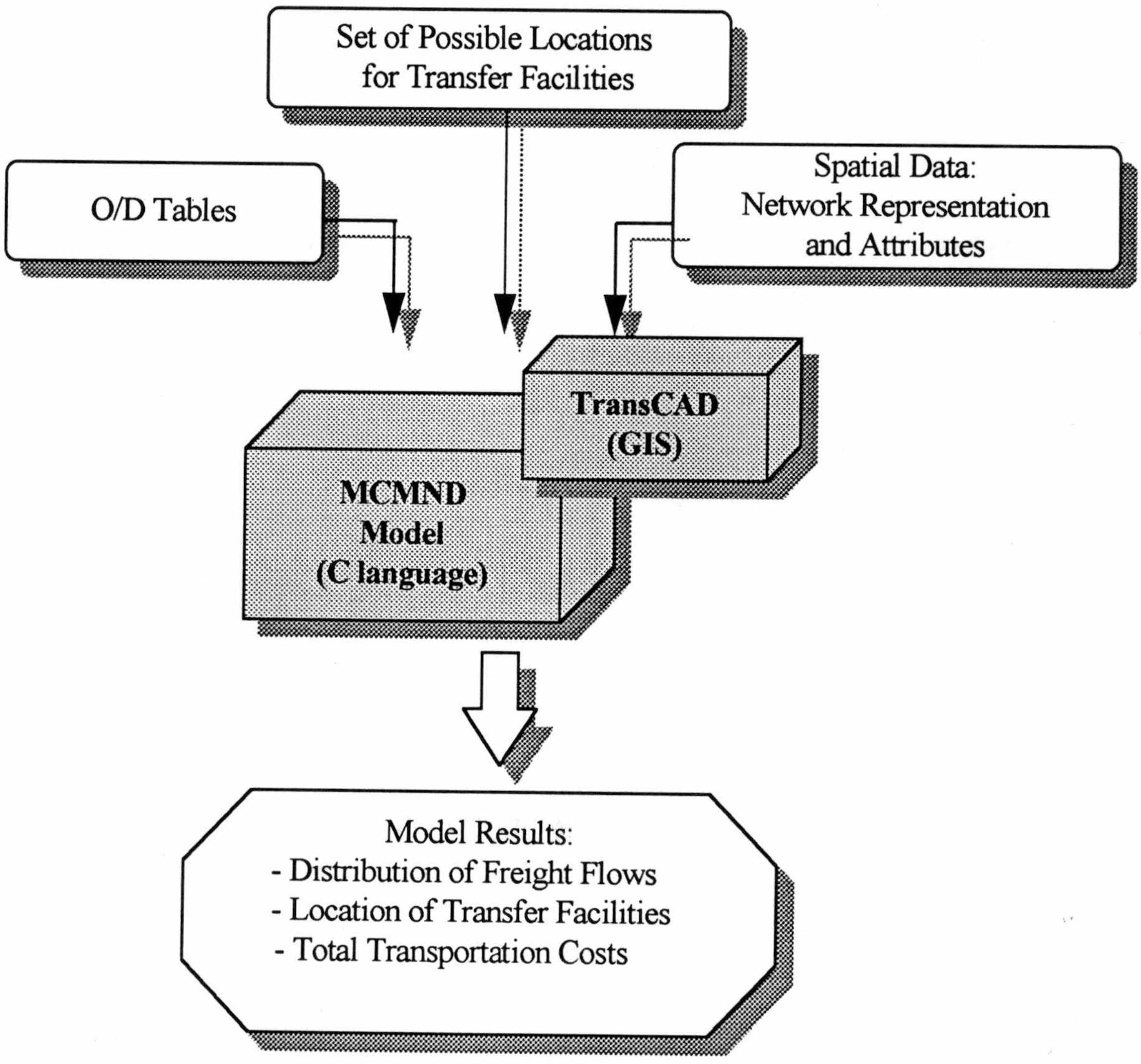

Figure 1

Schematic Representation of the MCMND Model

The concluding Chapter VII reviews the MCMND model, summarizing its contributions and the range of applications it may be suitable for, and discussing future extensions to this research.

CHAPTER II

REVIEW OF RELATED LITERATURE

Intercity Freight Network Models

Models to simulate the flow of freight over multimodal networks have attracted the attention of researchers in the past two decades. These optimization models can be generally classified as transportation network analysis models (Elshaygi, 1983), bearing the particular characteristic that, given a specified network configuration, they try to predict, and, in some cases, to dictate how a particular set of traffic demands will utilize the network. Crainic (1987) provided an extensive review of the evolution of this class of network optimization models. As Crainic stated, "... these models take a global, aggregated, often schematic, view of the transportation system and are centered on the estimation of the commodity and, eventually, service flow distributions over the transportation network, given the specification of the actors, of their behavioral rules and of their relationships".

The first published development in this class of models, also called predictive freight network models, was the Harvard Model (Kresge and Roberts, 1971). Their work, developed for a study of freight flows in Colombia, was based on fairly simple and restrictive assumptions, such as "direct link" representation of the physical network and no consideration of congestion effects. Later, the Multi-State Transportation Corridor Model (Jones and Sharp, 1977; and Sharp, 1979) took a step further in representing explicitly the multimodal network. However, this model also did not include any consideration of congestion in the network.

A similar type of model was also developed by Bronzini (1980). The so called Transportation Network Model was defined as a fixed demand, multimodal freight network model that worked based on the assumptions that, first, freight routes were decided by shippers so as to minimize transportation and inventory costs and, secondly, the cost of a path was a linear combination of transportation cost, time, and energy. In the original version, commodities were sequentially assigned to a capacitated network according to an iterative all-or-nothing assignment. A later review of the model introduced an equilibrium-seeking traffic assignment routine with a single-commodity formulation considering aggregate instead of commodity-specific cost functions. This was probably the first attempt to consider the congestion phenomena in freight flow models, by using flow dependent link cost functions.

A different modeling approach was adopted by Friesz, Gottfried, and Morlok (1986) in the development of the Freight Network Equilibrium Model. They produced a sequential model which used two network representations: an aggregate network that was perceived by the users, which served to determine the carriers chosen by the shippers, and then more detailed separate networks for each carrier, where commodities were transported at least total cost. The model assumed that shippers acted according to Wardrop's first principle of user optimization, while carriers behaved according to Wardrop's second principle of system optimization (Wardrop, 1952). Their major contribution was the combination of a detailed description of the behavior of shippers and carriers, and the network itself, with the variable demand modeling approach of spatial equilibrium models (Harker and Friesz, 1986).

This latter type of model, which has been extensively used for analyzing interregional commodity flows, was stated initially by Samuelson (1952) and lately extended by Friesz, Tobin, and Harker (1983). Spatial equilibrium models are usually applied for determining simultaneously the flows between production and consumption centers as well as the selling and buying prices of commodities. The transportation network is usually modeled in a very simplistic way, as a bipartite network, with these models relying to a large extent on the supply and demand functions of producers and consumers respectively. The main advantage of using spatial equilibrium models is the explicit consideration of the impact caused by changes in the transportation system to the economy as a whole, translated into the resulting variations in market prices and trade patterns, and the consequent changes in travel demand over time. However, due to the need of modeling all the micro-circumstances governing the shipment of freight and the interactions among the planning steps of travel demand generation, distribution, and traffic assignment, these models tend to be extremely difficult to formulate and solve when representing the physical network in a way appropriate for strategic transportation planning. Although Harker and Friesz (1986) presented a mathematical formulation for a combined spatial equilibrium network flow model, their theoretical approach is yet to be tested in a real-life application.

The latest efforts in developing network analysis models for evaluating the impact of policy and investment alternatives for transport development can be credited to Florian and Crainic (1989), and Ralston and Liu (1992). Florian and Crainic carried out a thorough joint study with the Pontificia Universidade Católica of Rio de Janeiro (PUC-RJ) and the Brazilian Federal Transportation Planning Agency (GEIPOT) aiming at the development of a strategic planning methodology that could be used at the level of the entire country and would permit the evaluation of lower energy consuming transportation options.

The project involved the development of the STAN - Strategic Transportation Analysis interactive-graphic system, which has embedded a multimode multiproduct network assignment model (Guelat et al., 1990) that has been used in Brazil to simulate flows, service levels, costs, and fuel consumption on the national multimodal freight network for different planning alternatives. The STAN model assumes that goods are shipped at the least total generalized cost, without considering shippers and carriers as distinct actors in the decisions made in shipping freight. Passenger movements are not considered in the model's formulation. It takes as input the origin to destination demand, the allocation of products to particular modes, the transportation infrastructure with link capacities, transshipment terminals with their delay and cost characteristics, and predicts the way in which the demands give rise to link flows for each mode and to transshipment flows through terminals. The multimode multiproduct model was formulated as an equilibrium assignment problem and solved using a decomposition-by-product algorithm based on the linear approximation method proposed by Frank and Wolfe (1956).

Ralston and Liu (1992) developed a transportation modeling system for analyzing transportation policies and investments in Bangladesh. Similar to the modeling approach used for the STAN system, the Bangladesh Transportation Modeling System (BTMS) is also an interactive-graphic, micro-computer based, software package designed to simulate the movements of freight and passengers over the Bangladesh multimodal transportation network, and estimate changes in travel costs and times relative to different investment scenarios. However, in addition to the modeling capabilities of the STAN system, the BTMS incorporates some new features such as routines to calculate internal rates of return on investments and a calibration package for estimating model parameters based on observed flows. It merges the technologies of Geographical Information Systems (GIS) with the analytical power of mathematical models to yield a user-friendly, interactive, Spatial Decision Support System (SDSS).

Embedded in the BTMS, there are two different routines provided to perform a multimode multiproduct flow assignment. The first, referred to as a logit-based assignment model, relies on linear cost and delay functions (fixed costs and times for a given arc) to assign flows to the least cost paths. The second, assumes that link cost and time are flow dependent, using a stochastic equilibrium assignment algorithm to load flows to the network. As for the case of the STAN system, the network is characterized as a set of nodes, physical and logical links, with the latter being used to represent pickups, deliveries, and intermodal transfers.

At this point, it is important to emphasize a common characteristic intrinsic to network analysis models. All the models reviewed so far take as input a given network and, then, simulate the flow of commodities over this network. The comparison of alternative investment options is performed by defining future network configuration scenarios, simulating the traffic flow on each one of these different scenarios, and then comparing their respective network performance measures and/or economic indicators. Thus, flows, link costs and delays, intermodal shipments, and other performance indicators are compared to those of a base case in order to carry out cost/benefit analyses and other economic evaluations.

It should be noted that all of these models rely on the experience of the transportation analyst who defines the alternative investment strategies, and then reaches his/her decisions (or provides the basis to the decision maker) based on the comparative results of the model. Under no circumstances, do these models have the capability of suggesting whether any of the analyzed scenarios represents an "optimal" configuration for the network. The model can only indicate the relative advantages of one network over the others. If the analyst could not identify all the "potentially good" investment options while developing alternative networks, the optimal configuration may not be known and it will be left out of the analysis.

Another major disadvantage of network analysis models is the practical difficulty of comparing a large number of alternative scenarios due to the time required to setup each scenario, analyze it, and compare its results to all of the others. Although in most of the real-world applications the set of infrastructure investment options is restricted and well-defined, for some cases it may happen that the analyst may be interested in getting a better feeling of which scenarios should be analyzed in further details. Thus, he/she may want to run a relatively large series of different scenarios before setting up the ones which would be considered the best candidate options. A good example of this situation is provided by the analysis, in Chapter VI, of the Brazilian case in which we search for the best places to locate transfer/loading terminals along the Tietê-Paraná waterway. As if the set of candidate locations were not large enough, we still have to decide on the optimal number of facilities to be located. This task, if we had to use one of the models reviewed so far, would involve the analysis of an extremely large number of possible configurations, since we would have to look at all possible combinations of locating one, two, three, and so on, facilities.

To overcome the problems mentioned above, inherently present in models of the type discussed so far, a different class of models, named network synthesis models, have been researched and developed in the last few decades. As stated in the introductory chapter, these models seek an optimal network configuration which satisfies a given set of requirements. Such requirements may be met, for instance, by increasing the capacity of some links in the network, by adding new links to it, or by locating service facilities in the existing network. Among the models that belong to this category, two are of especial interest to us: network design and facility location. As pointed out before, they are directly related to the multicommodity multimodal freight network flow problem which is the focus of this dissertation.

Network Design Models

As Janson, Buckels, and Peterson (1991) so well synthesized, "the basic network design problem is to determine a set of link additions or improvements that are feasible within the resource constraints specified, so that the performance of the network is improved by the maximum amount". By performance of the network, they meant a measure that could be approximated by some systemwide objective function, such as the total travel cost of all travelers or commodity shipments. Link additions were described as new roads or transit routes, whereas link improvements represented upgrades to existing facilities or services.

In other words, network design models are used to select the arcs, and their respective capacities, to add to a transportation network accounting for the effects that the design decision will have on the operating characteristics of the transportation system. Therefore, this type of model can consider trade-offs among various design alternatives, striking a balance between increased design expenditures and resulting improvements in the system's operation (e.g., decreases in operating costs).

In a review of several types of network design problems for transportation planning, Magnanti and Wong (1984) formulated a generic discrete choice network design model as the following linear mixed integer program:

Minimize :

$$\sum_{k \in K} \sum_{(i,j) \in A} c_{ij}^{k} \cdot f_{ij}^{k} + \sum_{(i,j) \in A} F_{ij} \cdot y_{ij} \tag{1.1}$$

Subject to :

$$\sum_{j \in N} f_{ij}^{k} - \sum_{l \in N} f_{li}^{k} = \begin{cases} R_k & \text{if} \quad i = O(k) \\ -R_k & \text{if} \quad i = D(k) \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

$$\sum_{k \in K} f_{ij}^{k} \leq K_{ij} y_{ij} \qquad \forall (i,j) \in A \tag{1.3}$$

$$(f, y) \in S \tag{1.4}$$

$$f_{ij}^{k} \geq 0, \quad y_{ij} = 0 \text{ or } 1 \qquad \forall (i,j) \in A, \quad k \in K \tag{1.5}$$

In the formulation above, **N** is the set of nodes and **A** is the set of directed arcs that are available for designing a network. **K** denotes the set of commodities and, for each $\mathbf{k} \in \mathbf{K}$, $\mathbf{R_k}$ denotes the required amount of commodity **k** to be shipped from its point of origin **O(k)**, to its point of destination **D(k)**. $\mathbf{y_{ij}}$ is a binary variable that indicates whether ($\mathbf{y_{ij}} = 1$) or not ($\mathbf{y_{ij}} = 0$) arc (i,j) is chosen as part of the network design and $\mathbf{F_{ij}}$ defines the fixed improvement (or design) cost for this same arc. $\mathbf{f^k_{ij}}$ is a continuous variable denoting the flow of commodity **k** on arc (i,j), while $\mathbf{c^k_{ij}}$ defines the cost of moving one unit of commodity **k** on this arc.

The objective function (1.1) minimizes the total cost of moving the commodities over the network plus the investment cost involved in building the selected arcs. The first constraint (1.2) is the usual flow conservation constraint. The inequality (1.3) states that the total flow $\mathbf{f^k_{ij}}$ of all commodities on arc (i,j) cannot exceed the capacity $\mathbf{K_{ij}}$ of the arc if it is chosen as part of the network design ($\mathbf{y_{ij}} = 1$), and must be zero if arc (i,j) is not selected ($\mathbf{y_{ij}} = 0$). The set **S**, in equation (1.4), embraces any side constraints imposed either singly or jointly on the flow and design variables, including limitations imposed upon resources shared by several arcs - in other words, any kind of budget constraint.

In their review, Magnanti and Wong (1984) also pointed out the importance of the network design problem, not only for its practical significance as a design tool, but also because of the range of problems that it models. In the paper, they listed many of the most noted problems that arise in network optimization - such as the minimal spanning tree and the shortest path problems, vehicle routing and the traveling salesman problem, facility location models, and traffic equilibrium - as either special versions or close relatives of the network design problem. In addition, they summarized reported experiences with various optimization and heuristic solution techniques to solve those problems.

Among the available optimization-based techniques, the authors quoted Benders decomposition and branch-and-bound as the two most effective procedures. Both methods develop and exploit lower bounds on the objective function of the problem. These two methods have been used extensively to solve small to medium-sized network problems (Hoang, 1973; Boyce et al., 1973; LeBlanc, 1975; Boyce and Soberanes, 1979; Dionne and Florian, 1979; Rothengatter, 1979; Los and Lardinois, 1980; Minoux, 1989). Lagrangian relaxation and linear programming methods were also mentioned as successful techniques for deriving lower bounds for network design problems.

Since the general network design problem is NP-hard (Johnson et al., 1978) and past empirical experiences have suggested that large scale (50-100 nodes) problems are extremely difficult to solve exactly (Minoux, 1989; Magnanti and Wong, 1984; Dionne and Florian, 1979; among others), heuristic algorithms have often been explored as alternative solution methods (Poorzahedy and Turnquist, 1982; Janson and Husaini, 1987). The three most commonly used heuristics are the add, delete, and interchange procedures. All three methods start with some initial feasible design and, respectively, add, delete, or interchange candidate arcs at each step until no further improvement in cost is possible.

Magnanti and Wong (1984) reported that these types of heuristics are fairly effective for solving small to medium-size problems. However, for large scale problems they could still require excessive amounts of computer time.

Up to this point, we have been referring to a formulation of the network design problem which assumes that the unit shipment cost of each link is fixed, regardless of flow volume. In spite of this formulation being usually adequate for regional network improvement analyses, which is the case herein, there might be the need for considering congestion effects in the network. In fact, when the consideration of congestion phenomena becomes essential for an appropriate representation of reality, volume/delay functions are usually calibrated as nonlinear functions, or linear approximations of nonlinear functions, to represent the operation of facilities which are subjected to capacity constraints.

Different formulations and solution algorithms to network design problems in which travel costs are dependent on link flows have been developed by several researchers (LeBlanc, 1975; Dantzig et al., 1979; Boyce and Janson, 1980; Poorzahedy and Turnquist, 1982; and LeBlanc and Boyce, 1986). These models are especially useful when modeling flows on railroad lines and yards. They can also be applied to the case of waterways, since, for instance, congestion may be present due to a limited number of barges on some service lines and/or due to a limited handling capacity on ports and lock facilities.

A slightly different formulation of the network design problem, closer to the problem discussed on this research, was proposed by Current, ReVelle, and

Cohon (1986). In their paper, the authors introduced the hierarchical network design problem (HNDP), which sought to identify the least cost, two-level hierarchical network. Such a network would have to include a primary path from a predetermined starting node to a predetermined terminus node. In addition, each node not on the primary path would have to be connected to some node on the path by means of a secondary path.

Two years later, Current (1988) proposed some modifications to the HNDP formulation to include the fixed charges for transshipment facilities in the hierarchical network. The new formulation (FC-HNDP) designed a hierarchical transportation network and simultaneously located transshipment facilities at the intersections of the primary and secondary paths. The total network cost to be minimized consisted of the primary arc costs, the secondary arc costs, and the transshipment facility costs. Exact and heuristic solution procedures were presented and compared for sample problems. Lately, Pirkul, Current, and Nagajaran (1991) developed a more efficient heuristic solution procedure based upon Lagrangian relaxation of a reformulation of the problem. They also presented a branch-and-bound algorithm, which proved to be very efficient due to the tight bounds generated by the heuristic, to identify the optimal solution for the 5% of the test problems for which the Lagrangian-based heuristic failed to do so.

Recently, applications of artificial intelligence based techniques, such as neural networks and genetic algorithms, have been applied to find optimal solutions to the discrete transportation network design problem (see Xiong and

Schneider, 1992). These procedures were found to search for and to identify system optimal designs extremely fast. Cumulative genetic algorithms were also proved to handle well the processing of multiple objective functions simultaneously.

Regarding the recent uses of network design models to address real-world problems, it is worth mentioning, first, the work of Janson, Buckels, and Peterson (1991), who developed a design model for the selection and programming of alternative route improvements to the US highway network of interstates, four-lane urban highways, and principal rural arterials. In their model, route investment strategies were specified as mutually exclusive sets of link improvements considered along a multi-decade planning period. Improvement strategies were candidates for inclusion in 10-year increments of a 30- to 50-year period according to the reduction caused by each strategy to the total interregional commodity shipment cost for the entire planning period. Thus, the strategy programming criterion was the ratio of the objective function reduction per unit of construction cost for a given route improvement.

The authors presented two different formulations for fixed versus variable interregional shipment distributions. The first case was formulated assuming fixed O/D demand matrices, while the second incorporated a typical trip distribution model to redistribute shipments in each decade on the basis of path cost changes resulting from network improvements. In solving the US application for the two distinct formulations, they applied a rank-add-and-swap (RASH) heuristic solution procedure to a network comprising 5,620 nodes and 7,775 bi-directional links. A total of 536 improvement strategies were considered for routes connecting adjacent regions covering the continental US.

Although their model fit the purpose of the US highway improvement program, some of its characteristics are too restrictive for the type of analysis aimed by this research. First, the model is intrinsically unimodal, assuming fixed unit shipment costs and, thus, not considering congestion on the highway network. Despite this approach being appropriate for modeling highway link flows, it cannot be applied realistically to the case of railways, inland waterways, and transfer terminals. The main drawback of that model, however, relates to the RASH solution algorithm. At each iteration, the algorithm ranks and selects the improvement strategies based on their B/C ratios which are calculated with improvements being added individually to the solution. Therefore, the benefits of each improvement project are considered to be independent of all other projects, disregarding the interrelated impacts of different projects to the flow pattern over the network.

Another interesting application of a network model for formulating investment priorities for road improvements is the work by Kim, Park, and Rho (1985). They built their model to be tested on South Korea's multiregional-multicommodity flow data, thus, addressing the specific case of developing countries. Although similar in its scope and its objectives, their model differs substantially from the model developed for the US (Janson, Buckels, and

Peterson, 1991) in terms of its formulation and solution approach. The model's formulation considered explicitly the interaction between transportation and the other economic sectors by finding the amounts of commodities to be produced in each zone, and also exported from each export zone, such that total transportation costs, export costs and production costs were minimized, subject to export constraints and the conservation of flow. The estimated travel demand was assigned to the network according to a user-equilibrium criterion, thus considering road congestion by the use of BPR-type link cost functions. Investment priorities were determined implicitly through the ranking of travel disutility values for all links in the network. In other words, the link that showed the highest travel disutility was supposed to receive the highest priority for the improvements. In this sense, their model differed from the typical network design formulation since it was neither formulated in terms of discrete links to be considered for addition to an existing network, nor in terms of continuous increases in link capacity. Rather, it defined investment priorities by identifying the magnitude of capacity bottlenecks in the existing network, based on aggregate estimates of interregional shipment patterns.

From the models reviewed up to this point, one can notice that the major emphasis in the past has been on the development of network design applications for unimodal networks only, more specifically, highway networks. This was probably the result of the past trend of looking at the transportation system as a set of independent parts, corresponding to different modes of transportation. Only in the last few years, interest has been rising for intermodal issues, with a better perception that the efficiency in the transportation system can be more easily reached through an optimal and balanced modal mix, in which each mode performs the function for which it is best suited.

The only work dealing with multimodal networks that could be identified by this literature review was the one carried out by Rothengatter (1986) on scheduling investments for interstate road and rail networks. Due to the close relationship of Rottengatter's work to the model presented herein, a more detailed review of the former will be presented in the following chapter when discussing the MCMND problem formulation.

Facility Location Models

Since modeling investment options for multimodal networks intrinsically involves the problem of locating intermodal transshipment facilities, a review on the topic of facility location models is appropriate.

Locating facilities on a network is a problem that has received much attention in the past three decades. The literature on network location problems started growing with the appearance of Hakimi's paper (Hakimi, 1964) on the "absolute center and median" location problems. Tansel, Francis, and Lowe (1983) provided an extensive bibliography on the subject.

In general terms, the network location problem can be defined as one of locating service facilities on a network so as to serve demands, generated at a finite number of nodes, in an optimal way. Thus, service facilities can be seen as destination points for demands arising at nodes in the network. According to the optimality criterion, network location models are usually subdivided into two classes: medians or centers. Median models minimize an average measure of spatial separation (e.g., travel time, cost, or distance) between vertices and the closest facility on the network, whereas center models minimize the maximum measure of separation from any vertex to its closest facility. The solution of median models was highly simplified due to Hakimi's (1964) finding that at least one set of optimal location points consists exclusively of a subset of the nodes in the network. For the case of center models, however, such a rule is not always applicable. Center models are then subdivided into vertex restricted (also called covering problems) and unrestricted center problems, according to whether or not the requirement for locating facilities only at nodes in the network is in place. Since this research addresses the issue of locating transshipment facilities at the intersection of arcs representing different modes, this literature review will concentrate on the solution of median and vertex restricted center problems.

Hakimi's result for median problems was further extended and generalized by Goldman (1969), Hakimi and Maheshwari (1972), and Wendell and Hurter (1973) to concave cost functions and to various distribution problems arising in deterministic transportation networks - deterministic in the sense that demands on the nodes are fixed and the travel impedance values (distances, times, or costs) on the links are deterministic and have fixed values. Later on, this deterministic assumption was relaxed and it was shown that node locations were still optimal for a class of probabilistic networks (Mirchandani and Odoni, 1979).

Among the many approaches for solving vertex-restricted facility location problems, two of the most common are node substitution and mathematical programming. The most used heuristic node substitution technique was proposed by Teitz and Bart (1968). Their algorithm starts by selecting a set of nodes as its initial solution. It then proceeds by examining the nodes that are not in the current solution set and checking if the objective function can be improved by replacing nodes already in the solution set. This process continues until no further improvement in the objective function occurs. The location problem can also be formulated and solved as either a mixed or a pure integer program (ReVelle and Swain, 1970). Branch-and-bound and relaxation techniques are usually applied in these cases.

In addition to the single objective function formulation, network location problems have also been analyzed in a multiple criteria decision environment (Current, ReVelle, and Cohon, 1985; Current and Min, 1986). Multiple objective functions have the advantage of allowing a more realistic modeling of real-world decisions in which decision makers have to compromise on conflicting objectives. The recent progress in computer technology and accessibility has even made possible the development of interactive decision support systems where human decision makers interact with embedded mathematical models, inputting their preferences and aspiration levels (Densham and Rushton, 1988). A recent application of these systems can be found in the work of Ogryczak, Studzinski, and Zorychta (1992). In their paper, the authors present an implementation of an aspiration-based interactive procedure to solve multiobjective transshipment problems with facility location, formulated as a mixed integer programming problem. Another development in this area was carried out by Hu (1992) for the selection of waste disposal sites.

The problem of locating intermodal transfer facilities differs somewhat from the conventional facility location models described above, in the sense that transfer facilities are neither origins nor destinations for commodities moving over the network. Rather, these facilities have to be located at intermediate vertices along the paths between O/D pairs. Moreover, there must be a transshipment node at the intersection of arcs representing different modes whenever those arcs belong to the same feasible path.

A similar type of problem, named the Flow-Resource Facility Location Problem (FRFLP), was formulated and solved by Ray (1990). He defined flow-resource facilities as those used in systems where the volume of commodities or services that can be moved directly between any two vertices decreases with increased distance. A trade-off must be made, therefore, between the distance moved and the amount of goods that can be shipped. A good example of this type of problem is the operation of a fleet of aircraft freight carriers which must compromise on the balance between cargo and fuel volumes. The aircraft can reduce the length of a non-stop flight by diverting to an enroute airbase (flow-resource facility) to refuel, thus decreasing the maximum weight of fuel required for the longest arc on the route.

The solution to the FRFLP is a distribution of flow-resource facilities and a set of paths that maximize the amount of commodities or services moved in the network. As in the case of intermodal transfer facilities, for a particular route to be feasible, flow-resource facilities are required at the intermediate vertices. However, as opposed to the problem of locating intermodal transfer terminals, the FRFLP is inherently unimodal and, therefore, the selection of intermediate nodes to locate facilities does not depend on the characteristics of the incident arcs. In his research, Ray (1990) formulated the FRFLP as a mixed integer program and proposed two specialized algorithms for solving that problem, both based on a column generation technique (Cullen et al., 1981) which allows the FRFLP to be decomposed into a facility location problem and a routing problem submodel.

Summary

Except for the work developed by Rothengatter, which is reviewed in the following chapter, this literature review could not identify any published research that dealt exclusively with the development of optimization models for either designing multimodal networks or locating intermodal transfer facilities. The closest connection to this topic was found in the work of Current (1988), which presented the problem of designing a hierarchical transportation network with transshipment facilities. However, Current considered only a unique predetermined pair of starting/terminus nodes for defining the primary network. Moreover, Current's formulation deals solely with unimodal networks, where transshipment facilities are used to transfer commodities from a secondary path to the primary path in the network.

Another work worth mentioning due to the similarity to our problem is the above described FRFLP. The solution procedure adopted by Ray (1990) suggests formulating and solving the facility location problem as a network design problem, in which the inclusion of new nodes containing flow-resource facilities to the network is modeled by a node design variable incurring a fixed construction cost.

Given the intrinsic characteristics of the intermodal facility location problem and its distinctions from the other location problems already pointed out, we decided for an approach to our MCMND model somewhat parallel to the one discussed to solve the FRFLP. Such an approach encompasses, first, the representation of intermodal facilities as logical arcs in the network and, then, the solution of a discrete network design problem to identify, among other investment options, which of those logical arcs should be included in the design, given a limited budget. This is the main concept behind the model described in Chapter III.

CHAPTER III

THE MULTICOMMODITY MULTIMODAL NETWORK DESIGN MODEL

Introduction

This chapter presents the development of a *Multicommodity Multimodal Network Design* (MCMND) model that embeds a discrete choice network design algorithm capable of accounting for the intrinsic characteristics of multimodal transportation networks. The MCMND model was developed to serve as an analytical tool, to be used within a strategic planning framework, for the selection of infrastructure investment strategies for intercity freight multimodal transportation networks. The set of investment options to be considered would involve the addition of new links to the network, the improvement of existing links, and the location of intermodal transsshipment terminals. These investment projects are intended to achieve the goal of providing an efficient freight transportation system by promoting intermodalism and the usage of more energy efficient and environmentally cleaner modes of transportation.

The following sections describe the assumptions behind the MCMND model and present the corresponding mathematical formulation of the combined network design, intermodal transfer facility location problem.

Model Description and Assumptions

In conceptualizing the MCMND model, the first and basic assumption was that such a model would be used as an analytical tool in the process of planning transportation infrastructure investments at a strategic level in developing countries. As emphasized by Florian and Crainic (1989), the strategic level of planning implies a medium to long term time horizon and a rather aggregate level of detail for the representation of the transport services provided. The contemplated alternatives usually reflect anticipated (or predicted) changes in the demand patterns for the commodities considered and major changes to the transportation infrastructure. The MCMND model was conceived to be used within this strategic framework, dealing specifically with major infrastructure projects to be implemented in the intercity freight multimodal transportation network.

The focus on applying the model to developing countries resulted from the assumption that these countries would need major infrastructure investments to their not well developed transportation systems. In addition, it was also considered the fact that the public sector in these countries still play the major role in supplying the transportation infrastructure, controlling most of the decisions involved in this process.

As is the case of the Southeast region of Brazil (object of analysis in Chapter VI), which has experienced a considerable population and economic growth in the last decades, and where the growth is expected to continue, large expenditures will be required to improve the efficiency and accessibility of the region's multimodal network, especially considering the investments being currently planned for its inland waterway system. A large set of alternative investment options is now available, with a good example being the decisions on where to locate new ports and transfer terminals along the rivers. Therefore, a model like the one proposed herein could bring useful insight to transport analysts and decision makers in charge of defining the infrastructure changes to be carried out in the network.

It is important to point out, however, that the range of applications for the MCMND model is not restricted only to the case of developing countries. The model can be applied to any situation in which a planning agency is in charge of strategically defining investment priorities and has to face a large number of alternative options before deciding on the ones to be analyzed at a greater level of detail. Thus, the key aspects of application scenarios are twofold: first, a strategic level of analysis and, secondly, the selection of investment options out of a very large set of possible alternatives.

Network Characterization

The multimodal network, comprising the three main modes of freight transportation (highways, railways, and inland waterways), is characterized inside the MCMND model as a set of nodes, physical arcs, and logical arcs. The level of detail for network characterization should be appropriate for strategic planning.

Physical arcs represent the guideways used by each mode, such as roads, rail lines, and inland waterway routes. These arcs connect nodes representing cities, rail yards/stations, river ports, other types of transfer facilities, and, sometimes, just the intersection between two different road segments.

Logical arcs, on the other hand, result from the explosion (or expansion) of zone centroids as well as transfer nodes located at the intersection of arcs representing different modes. They are incorporated into the analytical network to allow the modeling of pickups, deliveries, and, most importantly, intermodal transfers. The inclusion of logical arcs is especially useful in defining investments in transfer facilities at specific nodes, since they allow the representation of all possible combinations of transfer movements occurring at each node. A more detailed explanation about the process of creating logical links is carried out in the section to follow.

Both physical and logical arcs are defined in the model as triplets (i,j,m), where i is the from-node, j is the to-node, and m is the mode allowed on the arc. Thus, the model considers parallel links between nodes, one for each available mode. In this representation, the mode is more than just an attribute of the link; rather, it is an integral part of the network. This concept of network characterization has been successfully applied, with small variations, in some recent studies (see Ralston et al, 1994; and Florian and Crainic, 1989).

An important concept behind the MCMND model is the definition of a *mode* as a means of transportation that has its own characteristics, such as technology and capacity, and its unique cost functions. Therefore, a broad class of modes such as rail can be sub-divided, for instance, into several different modes representing different gauges or technologies. The decisions on defining the set of modes should be made by the analyst according to the scope of the analysis and the availability of data. For the case of logical links representing transfer movements, a special mode named "intermodal transfer mode" is defined.

Following the same rationale, the model also defines a *commodity* as a group of products, presenting similar characteristics in terms of storage and shipment. Among the shipping characteristics to be considered when grouping different products, the most important are the types of vehicles used in transporting the goods, the costs and times incurred in loading/unloading activities, and the propensity for the use of some specific modes.

GIS Interface / Logical Network Generator

Since the analysis of alternative scenarios for multimodal transportation networks is obviously a spatial process, the MCMND model could strongly benefit from an interface with an interactive graphic environment capable of maintaining network data and displaying model results. Therefore, a simple interface was designed to link the MCMND model to one of the commercially available GIS-T[†] packages, consisting basically of a file interchange scheme between the model's solution algorithm and the GIS. This interface not only simplifies the data input-

[†]GIS-T defines the adoption and adaptation of GIS technology for transportation purposes.

output process, but also allows users to visually evaluate the distribution of freight flows over the network, the addition of new links and the location of transfer facilities, and, consequently, the impacts of alternative design options.

TransCAD (Caliper, 1988), a popular GIS-T software package appropriate for running on micro-computers, was chosen as the GIS platform. As a package designed primarily to be used for planning, management, operation, and analysis of transportation systems and facilities, TransCAD incorporates, in addition to basic GIS capabilities, specific procedures developed to solve transportation, logistics, and operations research problems. These analysis tools are very helpful in setting up and carrying out applications of the MCMND model, as is demonstrated later in Chapter VI by the analysis of the location of river ports along the Tietê-Paraná waterway in the Southeast region of Brazil.

In applying the MCMND model to the analysis of a certain regional network, the first requirement is the collection of the region's geographic data consisting of administrative boundaries, location and characteristics of cities, zone centroids, and other nodes in the network, in addition to the network links and their attributes. All this geographic information is then stored in TransCAD in the form of point, line, and polygon layers. A more detailed description of how to create the GIS database using TransCAD and how to take advantage of its graphic and analytical capabilities when applying the MCMND model is presented in the analysis of the Brazilian case, used as a prototype application of the model.

Once the network is represented in the GIS database, the MCMND model uses a program to transform the digital representation of the physical network (physical links) into a logical network containing both physical and logical links. As indicated in the last section, building the logical network involves the task of exploding the demand, supply, and transfer nodes by creating as many new nodes and arcs as necessary to model pickups, deliveries, and intermodal transfers.

The first step in generating the logical network is to transform each zone centroid into two logical nodes, one for supply and one for demand, to capture pickup and delivery movements. Next, for each mode that enters or leaves a node, a logical node for that type of mode is constructed. The last step involves, then, the addition of loading and unloading arcs linking the supply and demand logical nodes to each logical node-mode combination, together with the inclusion of intermodal transfer arcs. The logical network generator program used in the MCMND model reproduces most of the concepts behind the one incorporated in the BTMS (Ralston et al., 1994). To better explain the steps described above, Figure 2 depicts the logical structure resulting from the explosion of a centroid/transfer node served by road, rail, and water links.

Demand Modeling

As emphasized in the last two sections, the MCMND model assumes that the representation of the physical network is performed at a level of detail appropriate for strategic planning for a large region. Origins and destinations of

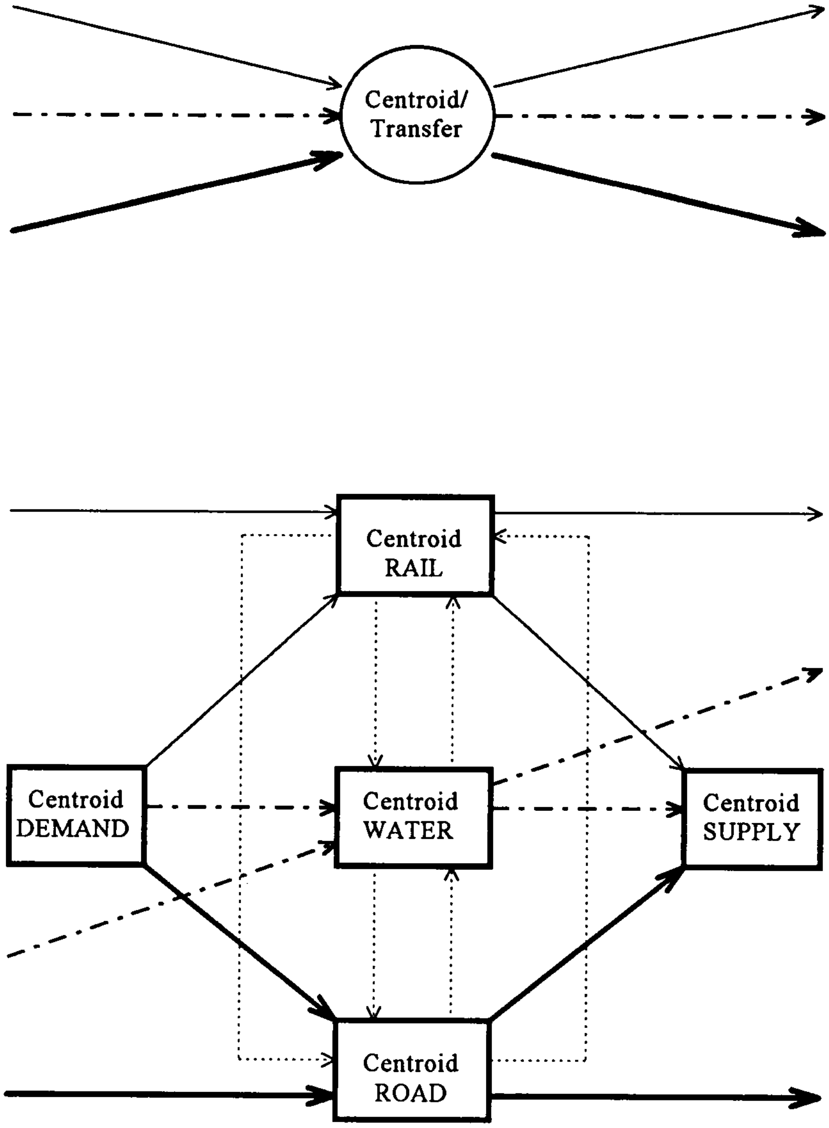

Figure 2

The Logical Structure of a Centroid/Transfer Node

commodity flows should, thus, correspond to traffic zones representing relatively large geographical areas, agglomerating, in some cases, several municipalities. Therefore, the model does not take into consideration goods movements within the urban environment.

The demand for transportation services is assumed to be fixed and exogenous to the model. In other words, the MCMND model is limited to deal with fixed intercity shipment matrices that do not vary as a result of the selected network improvements. Although this may seem a very restrictive assumption, the fixed demand case is sufficiently appropriate if one considers the types of products to be moved along multimodal routes (usually bulk commodities) and their low elasticity in shifting supply or demand markets. Moreover, it sounds inconsistent to assume that shipments between pairs of regions would vary due to network improvements without considering the effect of reduced transportation costs on the levels of production and consumption for certain regions. As discussed in Chapter II, the incorporation of a spatial price equilibrium model, although ideal, is not compatible with the level of effort appropriate for doing strategic planning analyses of freight transportation, especially considering the amount of data necessary to calibrate such a model.

As indicated before, the MCMND model can handle the simultaneous flow of several different commodities over a multimodal network. It takes as input O/D shipment matrices, expressed in tons per year, corresponding to the set of commodities under analysis. It is assumed that only commodities able to be shipped by all modes available in a specific application, and appropriate for intermodal transfers, will be included in the analysis.

Mode Split / Traffic Assignment

Mode choice in shipping freight is modeled in combination with traffic flow assignment, with the assumption that goods are shipped at minimum total generalized costs. Thus, the model considers explicitly the competition among modes, although not taking into account the micro-circumstances governing an individual shipper's decision.

Shipments are assigned stochasticly to the least disutility paths according to a multinomial logit model, formulated later on in this chapter. Path disutility is calculated as a function of link costs and travel times. The fixed cost of a physical link is dependent on its length and the per-ton-kilometer cost of the mode being used, whereas the free-flow time to traverse a link is a function of the length of the link and the average speed of the mode being used.

Link travel times are considered to be dependent on link flows when assigning shipments to rail and inland water links only. Roadways are assumed to operate below capacity, free of congestion effects. The approach adopted for modeling congestion on some modes does not explicitly consider the link or transfer capacities, but rather integrates them into volume/delay functions that affect travel time. These functions are calibrated as polynomial-shaped, nonlinear delay functions with volume and capacity values expressed in tons per unit of time. Examples of such functions are included in the analyses of the test network and the Brazilian pilot application, presented in Chapters V and VI, respectively.

A path-based stochastic user equilibrium assignment algorithm is used to distribute shipments over the multimodal network according to the above mentioned logit-type model. The assignment algorithm is described in detail in Chapter IV - Solution Procedure.

Infrastructure Investment Options

The MCMND model was designed to work with three main types of investment strategies:

- the addition of new physical links to the existing, base network (e.g., construction of new road sections, new rail lines, or new navigable river stretches).
- the improvement or upgrade of existing physical or logical links (e.g., paving gravel roads, doubling the number of tracks on a single-track rail line, or increasing handling capacity at river ports).
- the addition of new logical links, representing the siting of intermodal transshipment terminals.

Infrastructure investments are specified as a set of mutually exclusive link projects to be implemented during the planning horizon. Therefore, each link can possibly have one or several planned investment projects, but at most one of these projects can be selected. Each alternative investment project corresponds, then, to a certain level of improvement to the respective link. The improvement levels are characterized by new values of link impedance represented by the following measures: link capacity, transport rate, and free-flow speed. In addition, each level of improvement to a specific link has a corresponding investment cost incurred by the respective improvement project. The impedance values relating to unimproved arcs are defined at the level of "zero" improvement. For the case of non-existing arcs, the "zero" improvement level contains all three impedance values listed above equal to zero.

Finally, it is important to point out that each link addition and/or improvement has an associated cost which is independent of any other investment proposed to this specific link or to any other link in the network.

In the following section, we present the mathematical formulation for the MCMND model, reflecting the assumptions described above.

Mathematical Formulation

Define $G = \{N,A\}$ as a network consisting of $n \in N$ nodes and $a \in A$ directed arcs representing physical and logical links, with each one of these links involving a set of candidate investment options with their respective levels of improvement available. That is, each arc a, in addition to its natural attributes such as from-node, to-node, mode, and length, is also defined by the number of alternative investment projects available and the level of improvement that characterizes the current state of the arc.

Therefore, for each arc a, there is a corresponding set $L_a = \{l_a^0, l_a^1, l_a^2, \ldots, l_a^n\}$ consisting of the available levels of improvement relating to the candidate investment options, where l_a^0 corresponds to the state of the arc in the existing network, that is, a zero level of improvement with no investment being selected for this arc. Let T be a subset of A consisting only of arcs $t \in T$ for which the current level of improvement is different from zero. In other words, T is the set of improved arcs or, from a different perspective, the set of all candidate investment projects.

Let P be the set of commodities or classes of products p to be shipped over the network. Let R and S be subsets of N comprising nodes which correspond, respectively, to origins and destinations on the network (zone centroids). Note that the sets R and S are not mutually exclusive since nodes can serve simultaneously as origins and destinations for different shipments. Note also that not all nodes are required to be origins and destinations. Each distinct pair of nodes $r \in R, s \in S$, with $r \neq s$, is termed a *channel*.

Let K denote the set of all "efficient" paths in G as defined by Dial (1971), i.e., each link on an efficient path takes a shipment further from its origin and closer to its destination, k being an arbitrary element of K. Also, let $G^* = \{N, A^*\}$ be called the *base network* (unimproved network) where $A^* = A \setminus T$.

The following notation is adopted throughout this dissertation:

x_a	flow volume on link $a \in A$, in tons.
$t_a(x_a)$	travel impedance on link a as a function of link volume x_a, in dollars per ton.
K_{rs}	subset of K for channel rs.
K_t	subset of K in which each path $k \in K_t$ utilizes improved arc t.
q_p^{rs}	demand for commodity p between origin r and destination s (channel rs), in tons.
q_{rs}	total demand on channel rs, in tons, where $q_{rs} = \sum_{p \in P} q_p^{rs}$.
$f_{p,k}^{rs}$	flow of product p on path k of channel rs $(k \in K_{rs})$, in tons.
f_k^{rs}	total flow on path k of channel rs, in tons, where $f_k^{rs} = \sum_{p \in P} f_{p,k}^{rs}$.
c_k^{rs}	impedance on path k of channel rs $(k \in K_{rs})$, in dollars per ton.

h_k^{rs} distance traversed along highway links on path k of channel rs, in kms.

ϕ_h cost multiplier for penalizing the usage of highway links, in dollars per ton-km.

b_k^{rs} distance traversed along rail links on path k of channel rs, in kms.

ϕ_b cost multiplier for penalizing the usage of railway links, in dollars per ton-km.

$\delta_{a,k}^{rs}$ equals 1 if link a is part of path k on channel rs, 0 otherwise.

The objective function of the multicommodity, multimodal, mixed facility location/network design problem embedded in the MCMND model is defined as a minimization problem over elements of the sets R and S :

$$z_G^1 = Min \sum_{r\in R}\sum_{s\in S}\sum_{k\in K_{rs}} f_k^{rs} \cdot (c_k^{rs} + h_k^{rs} \cdot \phi_h + b_k^{rs} \cdot \phi_b) \qquad (3.1)$$

subject to the following set of constraints:

$$\sum_{t\in T} Y_t \cdot IC_t \leq B \qquad (3.2)$$

$$\sum_{k \in K_t} X_k \quad - \quad M \cdot Y_t \quad \leq \quad 0 \qquad \forall t \in T \tag{3.3}$$

f_k^{rs} satisfies the stochastic user equilibrium conditions described by Fisk (1980). (3.4)

$$X_k \in \{0,1\} \qquad \forall k \in K \tag{3.5}$$

$$Y_t \in \{0,1\} \qquad \forall t \in T \tag{3.6}$$

where:

X_k equals 1 if path k is selected, 0 otherwise.

Y_t equals 1 if improved arc t is part of one of the selected paths, 0 otherwise.

IC_t is the investment cost incurred by the improved arc t, in dollars.

M is some arbitrarily large number.

B is the available investment budget, in dollars.

and

$$c_k^{rs} = \sum_{a \in A} \delta_{a,k}^{rs} \cdot t_a(x_a) \qquad \forall k \in K_{rs}, \quad \forall rs \tag{3.7}$$

The objective function (3.1) minimizes the sum of the total shipping cost incurred by users (shippers) plus a measure of system impact costs that reflect the energy consumption, environmental, and safety impacts to the system of moving freight on less efficient modes of transportation. It is assumed that, regarding fuel consumption, emissions, and accident rates (the three standards used to assess the above listed impacts), the inland waterway mode is the most efficient, followed by rail and highway. Such impacts are translated into two cost multipliers, ϕ_h and ϕ_b, applied to the amount of ton-kms moved, respectively, on highway and railway links. The two cost multipliers are calculated and converted to dollar values according to the methodology proposed by Newstrand (1992), and reflect the relative differences in fuel use, emissions, and accident rates for trucks and rail against barges.

Equation (3.2) is an investment budget constraint which limits the total expenditure on implementing infrastructure projects onto the network. Inequality (3.3) ensures that any multimodal path used which requires the implementation of one or more projects can only be selected if the corresponding investment projects have been included in the solution. Constraints (3.5) and (3.6) impose integrality restrictions on the decision variables X_k and Y_t, respectively.

Constraint (3.4) states that the users of the transportation system (shippers) make their path and, consequently, mode choice decisions according to their individual perception of the route which will result in the lowest possible generalized shipping costs (shipping rate plus delay costs). At stochastic user equilibrium (SUE), no shipper can reduce his or her *perceived* shipping costs by unilaterally changing routes. This state of SUE results from the solution of an equivalent optimization problem that has the properties of an equilibrated traffic assignment in which flows are shared among a subset of paths $k^* \in K_{rs}$ (the ones for which $X_k = 1$) for each channel *rs*, according to an exponential (logit-type) formula, and in which path impedance values are determined as a function of the assigned flows. The definition of SUE was introduced by Daganzo and Sheffi (1977), who also showed that it was a generalization of the user equilibrium (UE) criterion established by Wardrop (1952).

Fisk (1980) formulated the above described optimization problem as a minimization of the following objective function:

$$z_G^2 = \frac{1}{\theta}\sum_{r\in R}\sum_{s\in S}\sum_{k\in K_{rs}} f_k^{rs}\cdot \ln f_k^{rs} + \sum_{a\in A}\int_0^{x_a} t_a(x)\cdot dx \tag{3.8}$$

subject to the constraints:

$$\sum_{k \in K_{rs}} f_k^{rs} = q_{rs} \qquad \forall rs \tag{3.9}$$

$$f_k^{rs} \geq 0 \qquad \forall k \in K_{rs}, \quad \forall rs \tag{3.10}$$

and the following link flow relationship:

$$x_a = \sum_{r \in R} \sum_{s \in S} \sum_{k \in K_{rs}} \delta_{a,k}^{rs} \cdot f_k^{rs} \qquad \forall a \in A \tag{3.11}$$

As demonstrated by Fisk (1980), the set of path flows f_k^{*rs} which solves the above problem can be defined as:

$$f_k^{*rs} = q_{rs} \cdot \left(\exp(-\theta \cdot c_k^{rs}) \Big/ \sum_{k^* \in K_{rs}} \exp(-\theta \cdot c_{k^*}^{rs}) \right) \tag{3.12}$$

It can be noticed that equation (3.12) reflects exactly a discrete choice logit model in which the total demand for channel *rs*, q_{rs}, is distributed among the set of selected paths, $k^* \in K_{rs}$ (the ones for which $X_k = 1$), proportionally to the exponential of their impedance values. As expected, these path flows also observe the two necessary conditions for a stochastic user equilibrium assignment, which are:

$$f_k^{*rs} > 0 \quad \text{implies} \quad \Omega_k^{rs} \leq \Omega_{k^*}^{rs} \qquad \forall k^* \neq k \tag{3.13a}$$

$$f_k^{*rs} = 0 \quad \text{implies} \quad \Omega_k^{rs} \geq \Omega_{k^*}^{rs} \quad \text{for at least one } k^* \neq k \tag{3.13b}$$

where Ω_k^{rs} is a random variable which represents the impedance value *perceived* by the users of path k of channel rs.

Finally, constraint (3.9) of the equivalent SUE problem guarantees the "conservation of flows" requirement for the mixed network design/facility location problem. It does this by stating that the sum of the flows on all paths k for each channel rs should be equal to the total demand on the channel. The conservation of flows is also expressed in the logit equation (3.12) that defines the flows that solve the equivalent problem.

Some Considerations on Bilevel Formulations

The formulation of the MCMND model presented above can be classified as a nonlinear bilevel multicommodity multimodal network design problem which utilizes a stochastic user equilibrium path choice problem at its lower level. LeBlanc and Boyce (1986) were one of the first to propose a nonlinear bilevel formulation for the network design problem. Their formulation, however, considered unimodal networks with user-optimal behavior and continuous link improvement variables.

Another interesting application of nonlinear bilevel programming to the design problem, one closer to the scope of the MCMND model, can be found in the work of Rothengatter (1986). He formulated a model for scheduling investments for interstate roads and railways in Germany as a mixed integer programming problem with a combined two-mode modal choice and travel assignment problem at the lower level. His emphasis was more on finding a good time schedule for investments rather than to determine the optimal investment program for a future time period, which is the case herein. His model also focused specifically on the effects on route and mode choice, ignoring the potential influences of investments on total demand (generation and attraction) or spatial distribution of activities.

After stating that the model formulation would result in such a large problem and, therefore, it would be hopeless to try to solve problems of practical relevance, Rothengatter decided to simplify his model by applying heuristic and exact procedures for pre-selecting a subset of the candidate projects and, then, solving a much smaller version of the original model. In pre-selecting projects before running the model, he considered the interdependencies of network improvements, both in time and space, suggesting a search approach consisting of four levels of analysis according to the scale of the problem and the corresponding desired degree of accuracy. The four levels of the project selection procedure were defined as: benefit/cost ratio, add-and-drop heuristic, branch-and-bound solution algorithm, and total enumeration.

It is important to notice that, although bearing a similar type of nonlinear bilevel formulation, the MCMND model differs considerably from Rothengatter's work. The differences arise not only in the sense that the former searches for a SUE state, while the latter takes the deterministic approach for the UE, but also due to the fact that the MCMND model is capable, by itself, of heuristically analyzing the whole set of candidate projects without the need for a pre-selection step as proposed by Rothengatter. Unquestionably, each one of the approaches has its advantages over the other depending on the type of analysis being carried out.

A powerful feature common to the two models is that their bilevel formulations allow explicit consideration of the interaction between the public sector which supplies transportation facilities and the private sector which uses the given facilities. The bilevel approach seems to realistically model the transportation planning process since this can be viewed as a hierarchical process that includes public and private sector interactions. As mentioned before, this is typical in the case of developing countries with a mixed economic system in which the public sector plays the lead role in providing the transportation infrastructure. The government authorities usually make their decisions seeking to improve the overall performance of the transportation network while, at another level, users make individual shipping or traveling choices with regard to route, mode, destination, and so on. As LeBlanc and Boyce (1986) pointed out, although the responses of users to a given network improvement can be predicted, their decisions can not be dictated.

On the other hand, a very inconvenient property of a bilevel problem, resulting from this difficulty of dictating users' behavior, is that the objective function may not decrease monotonically as improvements are implemented to the network. This phenomenon, named "Braess's Paradox" (Braess, 1968), is inherently present in user equilibrium assignment problems. It can be better described by stating that, even in the case when the public decision maker and the individual users follow the same objective, the addition of a link to an existing network may increase the total travel time because the individuals do not take into account the external diseconomies that their decisions may produce on other users of the system. Rothengatter (1986) stressed the further impact of Braess's Paradox for the case in which the decision maker includes in the objective function aspects which are not considered by users, such as the case herein of considering environmental, safety, and energy consumption impacts of moving freight on highways and railways, as opposed to waterways. This property of a bilevel problem affects considerably the ability of developing exact solution algorithms for solving the formulation of the MCMND model presented above.

If this difficulty were not enough to prevent solving our model exactly, the computational complexity of the problem places another barrier for dealing with networks of realistic sizes. The discrete choice network design problem, by itself, is combinatorial in nature and NP-hard to solve (Johnson et al., 1978). Complete enumeration of all feasible and infeasible solutions to a combinatorial problem with P candidate projects requires the evaluation of 2^P solutions. When combined with our network optimization problem, it results in the solution of that same number of SUE assignment problems, and that translates into an enormous amount of computational time even for small networks with a restricted number of possible alternative investment options. These considerations led us to explore the alternative of developing a heuristic solution algorithm which could solve large scale problems in a reasonable time. Such an algorithm is presented and tested in the next two chapters.

At this point, it should be pointed out that there are some practical advantages of using heuristic procedures to solve large scale real problems as opposed to exact techniques. As stated by Janson and Husaini (1987), heuristics are frequently more straightforward to implement and can often be used to assess the quality of competitive solutions in order to judge a final solution's degree of superiority. Moreover, the search process of a heuristic may be more tractable and, thus, provide a greater insight into the trade-offs being made in the design process. This latter aspect gains in importance if we consider that modeling results serve as only one factor in the overall decision making process, making the development cost and computational burden of optimization models a deterrent to their implementation.

CHAPTER IV

SOLUTION PROCEDURE

Introduction

In the last chapter, the formulation of the MCMND model was presented. The model comprises a nonlinear bilevel multimodal network design problem with flows assigned according to a stochastic user equilibrium criterion. The upper level, network design problem is formulated as a mixed integer programming (MIP) problem with continuous flow variables and zero-one design variables. The solution of the network design problem is subject to a set of user optimal flows assigned stochastically to different paths serving each demand channel in the network. Such flows result from the solution of the lower level, traffic assignment problem which distributes each channel's commodity flows among the set of available paths according to a discrete choice logit model in which path disutilities are calculated by a combination of user costs and shipping delays.

Also discussed were the combinatorial nature of the upper level, network design problem and the difficulties involved in solving it exactly when inserted in a bilevel structure which required the solution of a very large number of assignment problems. It was, then, suggested that the only feasible approach to solve large scale problems would be the development and use of a heuristic solution procedure.

The following sections present a two-step heuristic algorithm, based on a column generation technique (Cullen et al., 1981), able to solve the upper level problem of the formulation for real-size networks in a reasonable time. Also described is the path-based stochastic user equilibrium assignment algorithm developed to solve the lower level problem. Chapter V presents a test problem designed to assess the efficiency of the proposed heuristic, both in terms of its ability to approximate the optimal solution as well as the required computational time.

Solution Approach

As emphasized above, the MCMND model, although formulated as an MIP problem, can not be solved directly by applying general purpose integer programming techniques (e.g., branch-and-bound or enumeration methods) due to its bilevel structure. The use of existing IP-solver computer programs becomes impossible due to the nonlinear formulation of the lower level problem. Further, integer programming techniques would require the enumeration of all paths in the network that comprise the sets, K, K_{rs}, and K_t. For medium to large size networks (over 20 nodes), this task would become prohibitively time consuming.

As demonstrated by Ray (1990), in the case of a fully connected network consisting of n vertices, the number of unique paths for each channel rs is given by:

$$|P_{rs}| = \binom{n-2}{1}(1)! + \binom{n-2}{2}(2)! + \ldots + \binom{n-2}{n-2}(n-2)! + 1 \qquad (4.1)$$

where the notation $|S|$ denotes the cardinality of S. Therefore, if there are q channels in the network, the total number of paths is:

$$|P| = q \cdot |P_{rs}| \qquad (4.2)$$

which is a large number for even a very small network with just a few demand channels.

As an illustration of how large the problem can become, consider a fully connected network containing the same number of nodes, 7, and channels, 4, as the test network to be presented in the next chapter. According to the calculations above, the total number of unique paths is equal to 684. For large scale networks, as the one used for the Brazilian application in Chapter VI, consisting of hundreds of arcs and dozens of demand channels, the size of P becomes extremely large.

These considerations, brought together with the combinatorial nature of the network design problem, led this research to explore the alternative of developing a heuristic solution procedure which would allow the solution of large scale problems in a reasonable time. Among the available alternatives, two already known types of heuristics appeared to fit the scope of the MCMND problem. First, a rank-add-delete-swap algorithm of the type proposed by Janson and Husaini (1987) and, secondly, a decomposition procedure based on column generation techniques such as the one applied by Ray (1990) to solve the FRFLP (both procedures previously reviewed in Chapter II). For this research, the second type of heuristic approach was considered to be more suitable for handling the bilevel formulation within the two-stage structure inherent to decomposition and column generation techniques. The design portion of the model is, thus, solved as the *restricted master problem* while having the network optimization (SUE) as part of its *subproblem*. The details and advantages of this approach are explained in the following section.

Solution Algorithm

The algorithm developed to solve the MCMND model consists of a heuristic two-step procedure based on column generation techniques with an embedded stochastic assignment algorithm. The adoption of such a procedure was influenced by the work of Ray (1990) who developed a similar methodology to solve the FRFLP, although his work did not take into account traffic congestion and several other assumptions inherent to the problem of this research, such as the multimodal environment and the consideration of multiple commodities.

Column Generation in Discrete Problems

The heuristic column generation technique, proposed herein, parallels the decomposition procedure commonly applied to solve large scale linear programs (LP's). As reviewed by Ray (1990), column generation has been successfully utilized not only for solving LP's, but also in a variety of discrete models such as routing and scheduling problems. Examples of such applications can be found in the works of Cullen, Jarvis, and Ratliff (1981), and Desrosiers, Soumis, and Desrochers (1984). As in these two other cases, the model in this research is a mixed integer rather than a pure linear programming problem.

Perhaps the first application of column generation techniques for solving discrete problems was carried out by Cullen, Jarvis, and Ratliff (1981). They adapted the original column generation procedure used for LP's into a heuristic procedure that could solve interactively a broad class of vehicle routing problems. The main idea behind their model, called a "set partitioning model", was that alternative vehicle routes could be enumerated as columns on a 0-1 node incidence matrix, having as its first row the costs incurred by each route (e.g., distance traveled). The problem, then, could be solved by selecting a set of columns (routes) such that every row (node) was represented in exactly one column and the sum of the costs of the selected columns was the smallest possible.

Since the number of possible routes (columns to be enumerated) would be very large even for small networks, they devised a mechanism to generate new "good" candidate columns based on the pricing information provided by a previous feasible solution. The assumption was that such mechanism would be able to generate optimum or near-optimum solutions by considering only a small subset of possible routes (i.e., by enumerating only a small number of candidate columns).

The fundamental concept underlying the efficiency of Cullen et al.'s solution procedure was the calculation of a set of "row prices" corresponding to a given feasible solution that could be used to drive the algorithm towards an improving direction. The row prices would also have to be able to indicate optimality when no improvement were possible. Therefore, these row prices could be viewed as heuristic counterparts of the dual values associated with an LP model.

Unfortunately, for the case of discrete (integer) problems, the row prices could not be determined exactly since a full description of the convex polytope associated with the problem was not always available. In other words, in any problem for which the integer and continuous solution differed, the dual values obtained using LP techniques were not guaranteed to lead the integer solution to its optimum direction. To overcome this problem, they heuristically calculated the row prices in proportion to the cost of serving the delivery points one-at-a-time, that is, as if each delivery trip would serve only one point. Thus, their solution algorithm was intrinsically heuristic and could not guarantee optimality. However, such heuristic procedure was proved to approximate the optimum solution in a satisfactory way for a variety of routing problems. Most importantly, it was able to solve large problems in a reasonable time and also to allow user interaction in defining new candidate routes to be analyzed by the model.

Column Generation in the MCMND Model

The adapted column generation heuristic discussed above, as well as the one applied by Ray (1990) to the FRFLP, served as the basis for the development of a similar procedure to solve the bilevel formulation of the multicommodity multimodal network design problem. As for the vehicle routing and the FRFLP applications, such a problem can also be stated as a set partitioning model in which columns correspond to paths or routes between pairs of O/D channels. The problem, then, becomes one of selecting the optimal subset of paths (columns) to carry commodity flows, and their corresponding investment projects, that result in the lowest total shipment cost. However, as opposed to Cullen et al.'s and Ray's models, the multimodal network design problem requires that path impedance values be dependent on link flows, leading to the search of an equilibrium state between shipping costs and path volumes.

In applying the adapted column generation heuristic to the MCMND model, the bilevel formulation is decomposed into a two-stage problem, following the decomposition principle commonly applied in linear programming. The multimodal network design problem is called the *restricted master problem*, with its *subproblem* being defined as a multi-criteria path finding optimization problem with two objectives: (a) minimize the total shipment cost in the network and (b) minimize the total investment cost of the improvements required by the selected paths.

The subproblem is solved by a "path construction submodel", representing the column generation step in the solution algorithm. The path submodel constructs new paths (columns) between origins and destinations such that these new paths are potential candidates to be selected by the restricted master problem. It consists basically of a specialized labeling algorithm that searches for the least cost multimodal path in each demand channel, taking into account the link impedance values that resulted from the last iteration of the restricted master problem (current best solution), in addition to the investment costs incurred by the selection of new network improvement projects. The heuristic dual values, called herein *surrogate costs*, associated with the selection of new paths requiring investment projects are calculated based on the two following premises:

- If investment projects have been selected for one or more links in the current best solution, then new paths utilizing those links can be added to the solution without incurring the corresponding investment costs.
- If investment projects have not yet been selected for some links, then adding a new path utilizing any of those links will require incurring the respective investment costs.

The restricted master problem, in turn, is solved using a specialized heuristic based on a path substitution (or swapping) algorithm of the type defined by Lin and Kernighan (1973). First, an initial feasible solution, containing a subset of all paths constructed so far by the subproblem (i.e., in-solution paths), is generated by a greedy algorithm. Then, the substitution algorithm iteratively calculates the marginal savings attributed to replacing each in-solution path with every out-of-solution path. The substitution that maximizes the potential improvement to the objective function is made and the flow configuration resulting from the new set of in-solution paths (current best solution) is sent back to the subproblem.

The whole process of generating an initial solution and improving this solution through the substitution algorithm is described in detail later in this chapter, after the definition of the path construction submodel and the statement of the embedded flow assignment algorithm used to achieve the SUE state on the network.

Path Construction Submodel

As explained above, the subproblem comprises a path construction (column generating) submodel responsible for generating new paths (columns) to be considered by the restricted master problem. This path submodel is solved by a specialized labeling algorithm, a modified form of Dijkstra's algorithm (Dijkstra, 1959), which seeks to find the shortest multimodal path from the origin to the destination of every channel, attempting to minimize shipping as well as investment costs.

The generalized cost of a path is defined as the sum of the link impedance values along the path multiplied by the total demand on the channel, plus the sum of the *surrogate costs, SC* (heuristic dual prices), of the investment projects required by that path. The cost of a path k on channel rs is, then, calculated according to the following equation:

$$c_k^{rs} = \sum_{a \in A} q^{rs} \cdot \delta_{a,k}^{rs} \cdot t_a(x_a^{n-1}) \quad + \quad \sum_{t \in T} \delta_{t,k}^{rs} \cdot SC_t \tag{4.3}$$

Link impedance is defined as a combination of the shipping cost on the arc and the time required to traverse it. Link travel times are assumed to vary with flow volume based on a polynomial delay function. Shipping times are converted to dollar values by multiplying the respective number of hours by a constant VT representing the average value of time, in dollars per hour, of the commodities being shipped over the network. The determination of an average value of time is necessary because the solution algorithm cannot handle a different set of shortest paths for each commodity class under analysis. The calculation of link impedance values is, then, performed according to the following equation:

$$t_a(x_a) \quad = \quad C_a \quad + \quad VT \cdot T_a \tag{4.4}$$

with

$$T_a = T_a^0 \cdot [1 + \alpha \cdot (x_a / Q_a)^{\beta}] \tag{4.5}$$

and where

$C_a = R_a \cdot D_a$ is the shipping cost on arc a, in US$/ton. (4.6)

$T_a^0 = D_a / S_a$ is the free-flow travel time on arc a, in hours. (4.7)

VT is the average value of time, in US$/h.

R_a is the base shipping rate for arc a, in US$/ton-km.

D_a is the length of arc a, in kms.

S_a is the average free-flow speed on arc a, in km/h.

Q_a is the practical capacity of arc a, in tons.

α, β are coefficients in the polynomial delay function.

As discussed in the last section, the surrogate costs are used to approximate the dual values, or row prices, indicating the potential advantages of submitting to the restricted master problem a new path which requires new investment projects. That is, they represent an estimate of the true dual cost associated with the selection of a new project required by a new path. The surrogate cost of some link investment project t, SC_t, is assigned heuristically by taking into account the current optimal solution in the following manner:

$SC_t = 0$ if link investment project t has already been selected by one of the paths included in the current optimal solution.

$SC_t = IC_t$ if link project t has not yet been selected (with no other project t' already selected for the same link) and selecting project t would not violate the budget constraint.

$SC_t = \gamma \,.\, IC_t$ if link project t has not been selected but some other project t' has already been chosen for the same link, or selecting investment project t would violate the budget constraint.

As mentioned before, the surrogate costs are an approximation of the true dual costs relating to the selection of new investment projects. This approximation is performed by multiplying the actual investment cost of a certain link project t, IC_t, by an extra-charge parameter, following the methodology used by Ray (1990). Such parameter can take on three different values (0, 1, or γ), according to the three conditions listed above.

The first condition presents an extra-charge parameter equal to zero and reflects the case in which a certain link project has already been selected, resulting in no costs at all to construct a new path going through that link. The second condition, with the extra-charge parameter equal to 1, relates to the situation in which selecting a new link project will not result in taking a previous selected project out of the current solution. That is, its addition does not exceed the budget constraint or does not conflict with a mutually exclusive in-solution project. The cost of selecting such a new project is, then, equal just to its corresponding investment cost, without any extra costs incurred in the construction of a new path using that specific link.

Finally, the last condition presents an extra-charge parameter equal to γ which can be defined as any value greater than 1 ($\gamma \geq 1$) and it is used to model the increase in the cost incurred by constructing a new path which requires removing an existing path from the current solution. This may be the case if either another investment project has already been selected for the same link (investments for the same link are mutually exclusive), or if the remaining budget is not sufficient to accommodate the new project, keeping the ones already selected. Therefore, as γ increases from 1, a higher charge is associated with creating paths that include projects not yet selected and, as a consequence, the path submodel is forced to generate alternative routes through the network.

The value of γ must, then, be carefully chosen for each specific application of the MCMND model. It is highly recommended that a sensitivity analysis be carried out to assess the impact of different values of γ on the results obtained from each run of the model. An example of the type of sensitivity analysis necessary to select an adequate γ is presented in the following chapter, for a fictitious test network.

One final comment becomes necessary at this point. Since the path construction submodel is stated as a bi-criteria optimization problem, we cannot guarantee that its resulting solutions (i.e., the shortest paths found) are optimal. The issue of sub-optimality in multicriteria path problems has been recently addressed in the literature and some approaches have been suggested to fix this problem (see, for instance, Berman et al., 1990). However, the available approaches to guarantee optimality for the shortest path algorithm embedded in the MCMND model significantly increase the amount of computational time required to process networks of large sizes. Considering the fact that the overall solution procedure used in the MCMND model is already heuristic due to the approximate dual (surrogate) costs, the efficiency of the model may not be substantially improved by incorporating an exact procedure for determining shortest paths. Therefore, it was decided to keep the adapted labeling algorithm, taking advantage of its efficiency in generating new paths with little computational effort.

Path-based Stochastic User Equilibrium Assignment Algorithm

As noted in equation (4.3), in seeking for the least cost path on channel *rs*, the path submodel considers link impedance values calculated based on the volumes that resulted from the last (n-1) major iteration, i.e., the last solution of the restricted master problem. This procedure requires, therefore, that a flow assignment be performed at each iteration of the solution algorithm in order to the path submodel to generate new paths according to the last performed assignment. This section presents, thus, the path-based logit-type stochastic user equilibrium algorithm for performing flow assignment within the MCMND model.

The path-based SUE assignment algorithm is based on the well known *method of successive averages* (MSA). The MSA procedure has been extensively used for solving the network user equilibrium problem (UE) as an incremental-assignment-type heuristic. It was first suggested as an alternative for solving the logit-based stochastic equilibrium problem by Fisk (1980). Powell and Sheffi (1982) provided a formal proof of convergence for the MSA in its applications to network equilibrium problems, including the UE and SUE formulations.

The version of the MSA method applied by this assignment algorithm is based on a predetermined move size, $\mu_i = 1/i$, along the descent direction, where i is the iteration counter. The algorithm takes a set of selected paths on each channel and performs a path-based stochastic network loading, assigning volumes to paths according to a multinomial logit function of the type defined by equation (3.12). At each iteration i of the algorithm, the new volume on each selected path is determined as the average of the volumes assigned in all preceding iterations.

A key aspect that makes this assignment algorithm different from past developments of SUE models is its unique procedure for performing a multicommodity logit-based network loading. Instead of the usual link-based methods, among which the most known is the Dial's algorithm (Dial, 1971), the network is loaded on a path-based approach that distributes commodity flows among the set of pre-selected paths according to their travel disutility values, i.e., considering only path variables. The path-based procedure is necessary because the assignment algorithm can only assign flows to paths belonging to the in-solution set defined by the restricted master problem. Therefore, the same set of used paths must be kept throughout the execution of the assignment algorithm, what cannot be guaranteed while applying a link-based procedure which assigns flows using only link and node variables.

As mentioned above, flows corresponding to each demand channel are distributed among the selected paths according to their disutility (impedance) values. The disutility of a path is calculated as the sum of the link impedance values along that path. However, as opposed to the path construction submodel, path impedance may have a distinct value for each commodity or class of products being considered. The logit-based assignment algorithm assumes that different commodities have different values of time when combining the shipping cost and time values on a route into a single measure of path impedance. The disutility of a path k in moving commodity p on channel rs is, then, defined by the following commodity-specific function:

$$U_{p,k}^{rs} = \theta_{1p} \cdot C_k^{rs} + \theta_{2p} \cdot T_k^{rs} \tag{4.8}$$

where

$$C_k^{rs} = \sum_{a \in A} \delta_{a,k}^{rs} \cdot C_a \tag{4.9}$$

$$T_k^{rs} = \sum_{a \in A} \delta_{a,k}^{rs} \cdot T_a \tag{4.10}$$

The coefficients θ_1 and θ_2 in equation (4.8) are assumed to vary by commodity type. They act as weights in determining the shipping disutility of a path. Both coefficients should be non-positive since shippers are expected to avoid routes with higher costs and/or longer travel times. The ratio $\theta_{1p} / \theta_{2p}$ represents the value of time for commodity p. The values of both coefficients are crucial to flow assignment in the MCMND model and, therefore, should be carefully calibrated for the different commodities under consideration. An example of how to calibrate such values is presented in Chapter VI for the Tietê-Paraná pilot application.

As emphasized before, the assignment algorithm calculates each path's share of a certain commodity flow according to a multinomial logit model. Path shares are determined by multiplying the total flow on a channel by the probability of a shipment taking a specific path. The probability of shipping a certain commodity p on a given path k' of channel rs is determined based on the following logit-formula:

$$P^{rs}_{p,k'} = \frac{\exp(U^{rs}_{p,k'})}{\sum_{k \in K_{rs}} \exp(U^{rs}_{p,k})} \tag{4.11}$$

The amount of commodity p moving on path k' of channel rs is then determined by :

$$f^{rs}_{p,k'} = q^{rs}_{p} \cdot P^{rs}_{p,k} \tag{4.12}$$

At each iteration of the assignment algorithm, path shares are calculated separately for each commodity based on the values of path impedance resulting from the previous iteration. The corresponding commodity flows are added up for each channel and the new path volumes are determined as the average of the volumes assigned in the preceding iterations. Therefore, the network is loaded channel by channel, commodity by commodity, at each iteration of the assignment algorithm.

The overall process of loading the network and determining link volumes can be summarized by the following steps:

Step 0: *Initialization.* Perform a path-based stochastic network loading for all commodities $p \in P$, according to equation (4.12), based on the set of free-flow link impedance values $\{t_a^0\}$. This generates a set of path flows $\{f_k^{1,rs}\}$ which corresponds to a set of link volumes $\{x_a^1\}$. Set $i = 1$.

Step 1: *Update.* Set $t_a^i = t_a(x_a^i), \quad \forall a$.

Step 2: *Direction finding.* Perform another stochastic network loading procedure for all commodities $p \in P$, based on the current set of link impedance values, $\{t_a^i\}$. This yields an auxiliary path flow pattern $\{g_k^{i,rs}\}$.

Step 3: *Move.* Find the new flow pattern by setting:

$$f_k^{i+1,rs} = f_k^{i,rs} + \mu_i \cdot (g_k^{i,rs} - f_k^{i,rs}), \quad \mu_i = 1/i$$

which will lead to a new set of link flow volumes $\{x_a^{i+1}\}$.

Step 4: *Stopping criterion.* If $i = 10$, stop. If not, set $i = i + 1$ and go to step 1.

The stopping criterion in the MCMND model is defined as a predetermined number of assignment iterations equal to 10. Since the final volumes are calculated as averages of the volumes assigned on all preceding iterations, the contribution of the 10th iteration will be equal to only 10% of the final volumes. If the assignment algorithm was left running for more than 10 iterations, the differences in the assigned volumes would become insignificant as the number of iterations increased. As will be shown in the following sections, the solution of the MCMND model requires a large number of assignments. Therefore, a predetermined number of assignment iterations helps considerably to reduce the overall computational time of the solution algorithm, without compromising on the quality of the final solution.

Starting the Two-Step Solution Algorithm

As indicated before, the restricted master problem, due to its size, is solved using a specialized heuristic consisting of a substitution algorithm. This algorithm is an improvement heuristic which, given an initial feasible solution, attempts to improve the objective function by replacing in-solution paths by out-of-solution paths. Therefore, the algorithm requires an initial step involving the generation of a feasible solution, before the improvement step can be applied. As will be detailed in the next section, a greedy-type algorithm is incorporated into the MCMND model to find an initial feasible, and good, solution by adding paths to the solution set in a sequential manner.

Before the greedy algorithm is activated on the very first iteration of the two-step solution algorithm, a set of base paths is generated for each channel, comprising one shortest path on each available mode and one shortest multimodal path. A multimodal path can consist of unimodal or intermodal shipments, whichever leads to the least cost route. These initial base paths are generated considering only the base network G^*, without the inclusion of any investment projects. These paths are, then, attached to the so far empty in-solution set. The shortest multimodal path is included only if it is different from the unimodal paths already considered. Shipments are finally assigned to the base network G^*, by running the path-based stochastic assignment algorithm described in the last section, considering the set of in-solution paths comprising only the base paths selected for each channel.

This solution is, then, sent to the subproblem (path construction submodel) which will try to generate one new shortest path for each channel based on the link impedance values that resulted from the last performed assignment. At this stage, the path submodel is free to pick any unimodal or multimodal path that requires the selection of link investment projects, according to the surrogate costs incurred by them. If not even one single shortest path, different from all the ones already selected, could be generated for any of the channels, then the algorithm stops. Otherwise, the new generated paths are attached to the out-of-solution set, together with the base paths that were part of the in-solution set, which will become empty once again.

With all generated paths as part of the out-of-solution set, the greedy algorithm is called to sequentially add paths to the in-solution set until a good feasible solution is found. The overall two-step, greedy-swapping, algorithm is described in detail in the following two sections.

STEP 1: Generating an Initial Solution (Greedy Algorithm)

As explained above, the greedy algorithm is necessary as part of the overall solution procedure because it generates an initial feasible solution to be taken as a starting point by the substitution (improvement) step. This initial feasible solution should, thus, observe some characteristics required by the restricted master problem, which are:

- The total allocated cost of implementing the selected investment projects cannot exceed the available budget.
- Each path can be in the solution if, and only if, every link project required by the path has been selected.
- No more than one investment project can be selected for each link.

Considering the requirements listed above, the greedy algorithm strives to generate an initial good feasible solution by sequentially adding paths to the in-solution set, one at a time for each channel. In this process, besides searching for just a feasible solution, the algorithm also tries to generate a good one to speed up the improvement step. Therefore, in picking the paths to be added to the in-solution set, the greedy algorithm selects the ones with the highest marginal benefit in respect to reducing the objective function value. The computation of marginal benefits as well as all the other concepts behind the development of the greedy algorithm are clarified in the mathematical statement to follow.

First, define K^n as the set of all paths generated by the path submodel up to the major iteration n (iteration counter for the restricted master problem). The set K^n can be partitioned into a set of in-solution paths, K_B^n, and a set of out-of-solution paths, K_{NB}^n, where $K_B^n \cup K_{NB}^n = K^n$ and $K_B^n \cap K_{NB}^n = \varnothing$. Then, also define the following variables:

r	=	Iteration counter for the greedy algorithm.
Y_t^r	=	1 if project t is selected in iteration r, 0 otherwise.
B^r	=	Total allocated budget in iteration r.
$b(k,r)$	=	Marginal benefit of path k at iteration r.

The greedy algorithm used in the MCMND model can be stated, then, as an iterative procedure composed of the following steps:

Step 0: *Initialization.* Set $r = 1, K_{NB}^r = K^r, K_B^r = \varnothing, B^r = 0$, and set $Y_t^r = 0$ for all $t \in T$.

Step 1: Find, for each channel, the least generalized cost path $k \in K_{NB}^r$ based on the free-flow link impedance values. Add these paths to the set of in-solution paths, K_B^r. Update the investment vector Y_t^r and the used budget B^r.

Step 2: Assign flows to the set of in-solution paths according to the stochastic assignment algorithm and recalculate path impedance values based on the new link volumes.

Step 3: For each channel, find the path $k \in K^r_{NB}$ with the highest marginal positive benefit $b(k,r)$. If a single path could not be found for any one of the channels, stop. Else, go to step 4.

Step 4: Update the path sets, $K^{r+1}_B = K^r_B \cup \{k\}$ and $K^{r+1}_{NB} = K^r_{NB} \setminus \{k\}$. Also, update the investment vector Y^r_t and the used budget B^r. Set $r=r+1$. Go to step 2.

The above stated algorithm, following in several aspects the greedy algorithm developed by Ray (1990), starts with an empty set of in-solution paths and an out-of-solution set containing all the paths generated so far by the path construction submodel. Initially, the investment vector is set to null, with no link projects being selected. At step 1, the algorithm identifies, for each channel, the path with the lowest free-flow impedance among the ones belonging to the out-of-solution set, and adds it to the set of in-solution paths. Then, at step 2, the greedy algorithm performs an initial assignment and recalculates in-solution, as well as out-of-solution, path impedance values. At this point, each channel is being served solely by one single path.

The algorithm proceeds, at step 3, by searching for the path on each channel with the highest positive marginal benefit to be included in the in-solution set. The model defines marginal benefit of a path k' at iteration r, $b(k',r)$, as the reduction in the objective function value due to the inclusion of such a path, minus the sum of the investment costs incurred by that path. In calculating the total generalized cost (shipping plus system impact costs) related to path k', an <u>estimated</u> amount of cargo that would use this path is determined based on the ratio of the exponential of the disutility of path k' and the sum of the exponential of all paths' disutilities, including k', according to equations (4.11) and (4.12). The resulting reduction in the objective function value is determined by comparing the total generalized cost incurred on path k' to the average value for the paths being used by the specific channel. The average generalized cost on a channel is calculated as a weighted average of path impedance and system impact costs, with each path's values being weighted by their assigned volumes. The reduction in the value of the objective function is estimated according to the following expression:

$$\Delta Z_G = f_{k'}^{rs} \cdot [(\overline{c}_k^{rs} + \overline{h}_k^{rs}\phi_h + \overline{b}_k^{rs}\phi_b) - (c_{k'}^{rs} + h_{k'}^{rs}\phi_h + b_{k'}^{rs}\phi_b)] \tag{4.13}$$

The marginal benefit of path k' is then expressed by:

$$b(k',r) = \Delta Z_G - \sum_{t \in T} \lambda_t^r \cdot (\delta_{t,k'}^{rs} \cdot IC_t) \tag{4.14}$$

where

$$\lambda_t^r = \begin{cases} 1 & if \quad Y_t^r = 0, B^r \leq (B - IC_t), and\ Y_{t'}^r = 0, \forall t' \neq t \in L_a \\ \infty & if \quad Y_t^r = 0\ and\ B^r \geq (B - IC_t), or\ Y_{t'}^r = 1\ for\ any\ t' \neq t \in L_a \\ 0 & if \quad Y_t^r = 1 \end{cases}$$

The λ_t^r values, present in equation (4.14), parallel the concept of the surrogate costs computed by the path construction submodel. As in the case of the surrogate costs, they are also used to estimate the impact of bringing a path into the solution which requires the selection of some link investment projects. If the required projects have not yet been selected (with no other projects already chosen for the same links) and the remaining budget is greater than the corresponding investment cost, λ is made equal to 1. On the other hand, if the remaining budget is not enough to afford the investments required by the path, or if some other projects have already been selected for the same links along this path, then λ equals infinity. Finally, if the required projects have already been selected by previous paths included in the in-solution set, λ is zero. Therefore, the computation of each path's marginal benefit ensures the feasibility of the solution, since a λ equal to infinity will avoid the selection of a path with investment projects exceeding the available budget or the consideration of two mutually exclusive projects for the same link.

After determining, for each channel, the most beneficial path, the algorithm moves it from the out-of-solution set to the in-solution set, as stated in step 4, and reassigns flows considering now the new set of in-solution paths. For every subsequent iteration, a feasible solution is maintained through the computation of the new benefit values which are defined with respect to the last performed flow assignment. When not even one single path with a positive marginal benefit value can be found for any of the channels, the greedy algorithm stops. Upon termination, the resulting in-solution and out-of-solution sets, as well as the investment vector, serve as inputs to the improvement step which is described in the next section.

STEP 2: Improving the Initial Solution (Substitution Algorithm)

Departing from the initial solution generated by the greedy algorithm, the substitution (improvement) algorithm starts by selecting the first path in the list of out-of-solution paths. It then calculates the change in the objective function that would result from swapping this out-of-solution path for every in-solution path. The calculation of the improvement to the objective function resulting from each swap is performed by temporarily replacing the in-solution by the out-of-solution path and, then, reassigning shipments using the stochastic user equilibrium assignment algorithm. The new objective function value is calculated and the corresponding improvement is determined by the difference between the value before and after the substitution took place. The swap resulting in the highest positive improvement to the objective function is kept eventually.

It is important to point out that swaps are not restricted to pairs of paths belonging to the same channel. The algorithm allows the substitution of paths between different channels, making possible a variation on the number of paths serving each channel. However, the total number of paths in the in-solution set cannot change during the execution of the algorithm. Another important characteristic of the substitution algorithm is that only one path can be swapped at a time.

After testing the first path in the out-of-solution set against all the in-solution paths, if none of the candidate swaps results in a reduction on the value of the objective function, the algorithm repeats the same procedure for the next out-of-solution path. It then proceeds in the same manner, testing every out-of-solution path against every path already in the solution. When, after going through the whole set of out-of-solution paths, no swaps are performed, the algorithm terminates. The resulting solution is submitted back to the path submodel, which will try to generate new shortest paths based on the new configuration of the network, i.e., new link impedance values and selected investment projects.

The overall substitution algorithm, devised as part of the solution procedure to the MCMND model, can be better summarized by the steps listed below. First, let n denote the current major iteration of the overall solution algorithm and s denote the number of substitutions made during n. Define also the following operation on paths k_i and k_j:

$k_i \oplus k_j$ = the marginal reduction in the value of the objective function resulting from the substitution of path k_i for path k_j in the current solution.

The mathematical statement of the algorithm is, then:

Step 0: Set $s = 0$. Select the first path $k_i \in K_{NB}$. Go to step 2.

Step 1: Select the next path $k_i \in K_{NB}$. If all paths in K_{NB} have already been processed and $s = 0$, stop. If all paths in K_{NB} have been processed and $s > 0$, go to step 0. Otherwise, go to step 2.

Step 2: Find path $k_j \in K_B^n$ such that $k_j = \underset{k \in K_B^n}{Arg\,Max}\{k_i \oplus k\}$. If $k_i \oplus k_j \leq 0$, go to step 1. Otherwise, go to step 3.

Step 3: Set $K_B^n = (K_B^n \setminus \{k_j\}) \cup \{k_i\}$ and set $K_{NB}^n = (K_{NB}^n \cup \{k_j\}) \setminus \{k_i\}$. Update the investment vector Y^*. Set $s = s + 1$ and go to step 1.

The above described substitution algorithm maintains most of the characteristics of the algorithm proposed by Ray (1990) for handling the FRFLP.

The main distinction between the two algorithms is that, for the case herein, flows are assigned to the network when testing every candidate swap and also after every substitution takes place. Although this task may appear to be too time consuming, the path-based stochastic algorithm used for assigning flows turns out to be quite fast, as will be demonstrated in the next two chapters when solving the test problem and the Brazilian multimodal network.

Summarizing the Overall Solution Procedure

The flow diagram, depicted on Figure 3, summarizes the overall solution procedure developed to solve the MCMND model. It illustrates the process of breaking the design problem into a restricted master problem, solved using the two-step improvement heuristic, and a subproblem in charge of generating new paths to be considered by the master problem.

The algorithm starts at iteration 0 with an initial feasible solution containing the base paths over the base network, with no investment projects being considered. It then proceeds at iteration n with the submodel being used to generate a new path for each channel. If no new paths are found, the algorithm terminates since the value of the objective function cannot be reduced as the set of available paths is the same as in the last iteration. If one or more new paths are found, then the restricted master problem is solved using the greedy/substitution two-step procedure, resulting in a new flow assignment and a new set of investment projects being selected. If after resolving the master problem, no

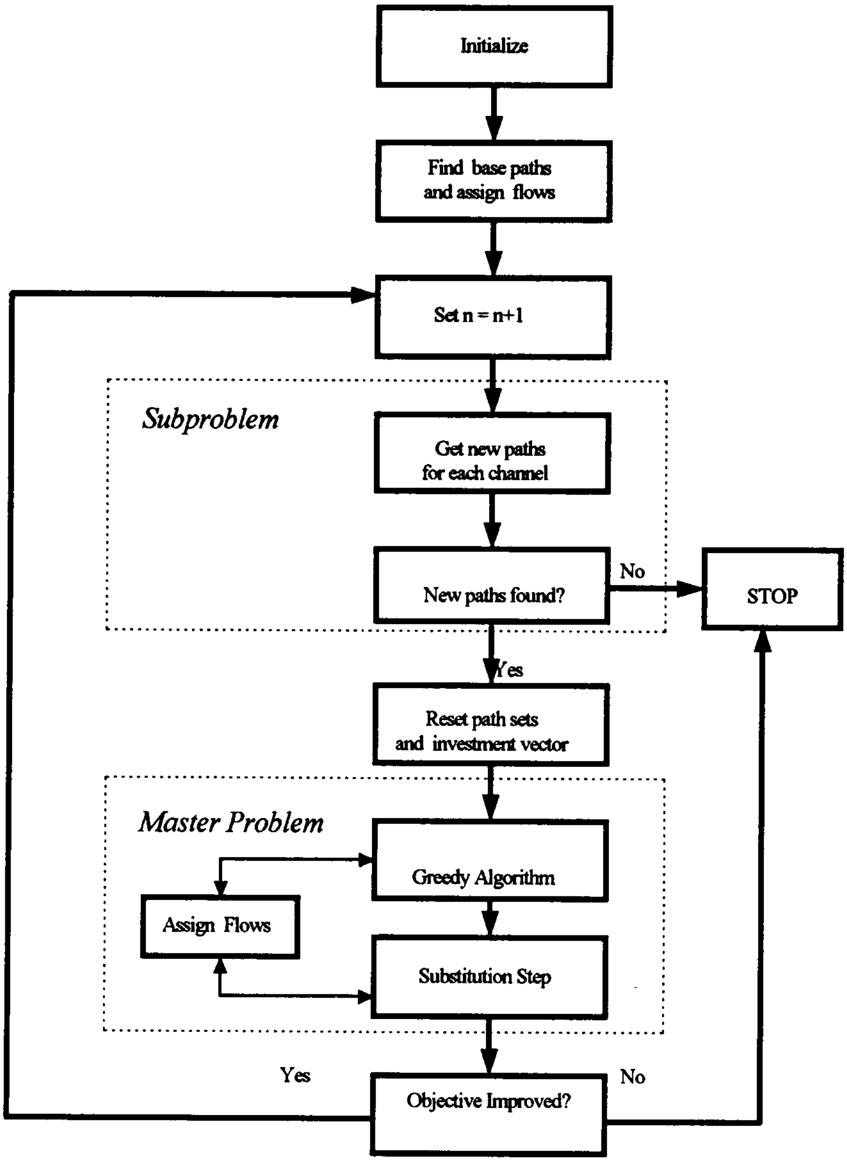

Figure 3

Flow Diagram for the Solution Algorithm

considerable reduction in the value of the objective function is found, the algorithm terminates. Otherwise, the iteration counter is incremented and the path construction submodel is resolved using the surrogate costs based on the solution to iteration $n - 1$.

In the next chapter, a test network is used to exemplify the application of the MCMND model and to evaluate the efficiency of the above proposed solution procedure in approximating the exact solution of the problem. In addition to this test problem, the model will be also applied, in Chapter VI, to a real size network consisting of approximately 170 zones, 1000 nodes, and 1500 physical arcs, to prove its applicability in solving large scale problems.

CHAPTER V

TEST PROBLEM

Introduction

As stressed in the last chapter, the algorithm developed to solve the MCMND model is intrinsically heuristic and, therefore, is not guaranteed to provide the optimal solution to the problem. The option for a heuristic approach resulted from the inability of solving the embedded multicommodity multimodal bilevel network design problem using the available general purpose integer programming techniques. Moreover, even if we were able to solve the problem exactly, its combinatorial nature, coupled with the requirement for path enumeration, would still make the solution of medium to large size problems very computationally intensive and prohibitively time consuming.

Prior to applying the proposed heuristic algorithm to a large scale and real-life network, it was decided to first design a small multimodal test network with a limited number of candidate investment projects, and assess the effectiveness of the MCMND model using this test network. The results of the model were compared to the results of stochastic equilibrium assignments on the same test network for all possible combinations of investment projects. The sections to follow describe the test procedure and its results in detail.

Multimodal Test Network

To illustrate the application of the MCMND model and to evaluate the effectiveness of its solution algorithm, a test network, depicted in Figure 4, was developed comprising 7 nodes, 18 directed physical arcs, and 68 logical arcs. As explained in Chapter III, the logical arcs are created by the logical network generator program embedded in the MCMND model to represent pickups, deliveries, and transshipment movements. Among these logical arcs, 16 represent possible locations for intermodal transfer facilities between pairs of the three available modes - road, rail, and water. It is assumed that no transfer facilities are present in the base (unimproved) network. The problem, then, becomes one of choosing the optimal number and location of transfer facilities so as to minimize the generalized cost of moving freight on this network.

The test problem to be solved by the MCMND model assumes the analysis of a unique general commodity "cargo", with shipments expressed in thousands of tons per year. The origin-destination distribution of the "cargo" is shown in Table 1, with nodes 0, 1, and 2 representing the flow origins, and nodes 3 and 6 representing destination points. Table 2 summarizes the modal attributes on the test network. Transfer movements are allowed only at nodes 0, 3, 4, and 5. The numbers close to each arc in Figure 4 represent the arc's practical capacity, in thousands of tons per year, while the numbers in parentheses are the lengths of the arcs in kms. It can be noticed that the capacity of all roadway links is set to a very high value, equal to 5,000,000 tons per year, to represent the fact that those links

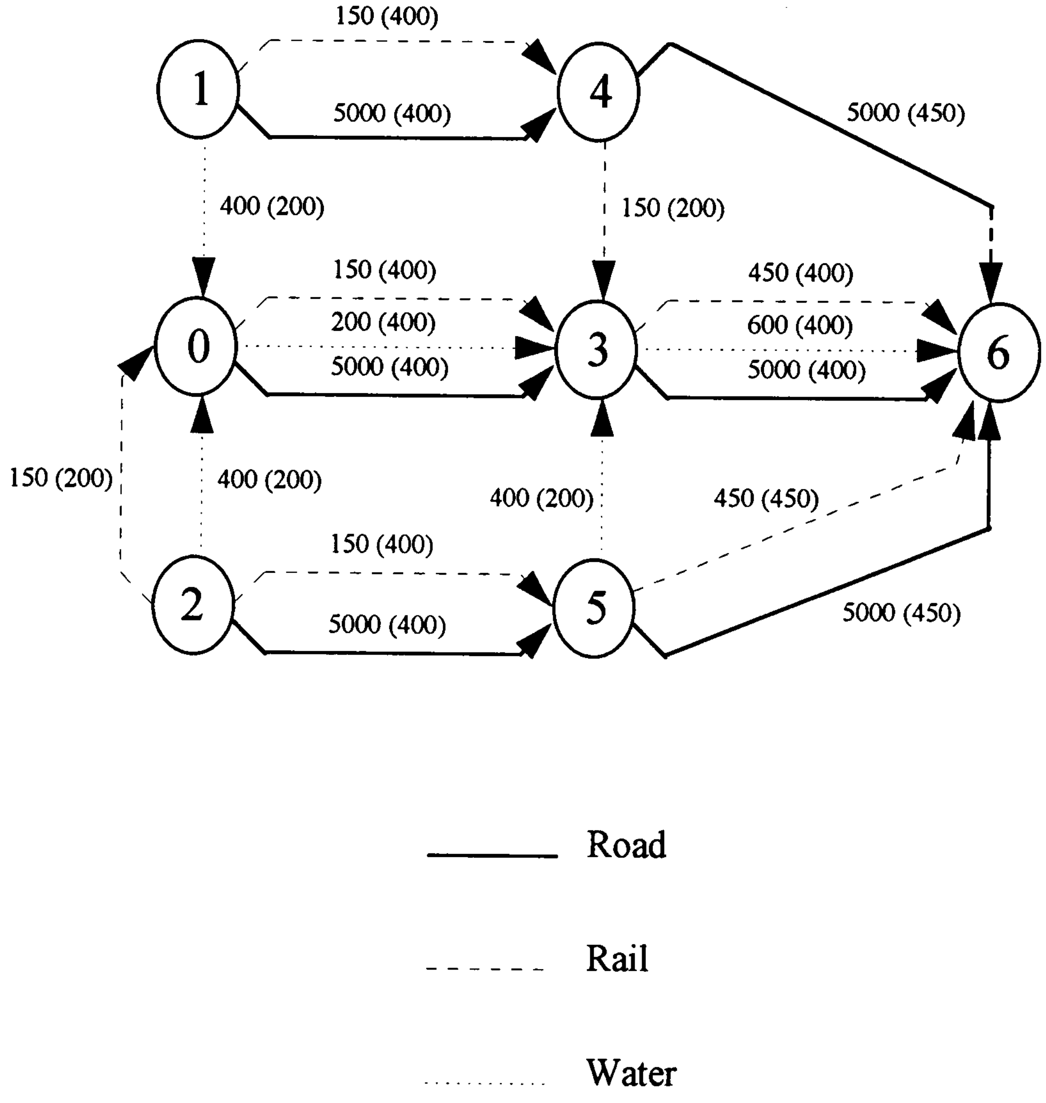

Figure 4

Multimodal Test Network

Table 1

Test Network - O/D Table (1000 tons/year)

Node	3	6
0	600	2000
1	-	800
2	-	600

Table 2

Test Network - Modal Attributes

Mode	Rate (US$/ton-km)	Speed (km/h)	System Impact Cost (US$/ton-km)	Transfer Cost (US$/ton)
Road(0)	0.04	60	0.01	1.50
Rail(1)	0.02	30	0.005	1.50
Water(2)	0.01	10	-	1.50

operate below capacity, free of congestion effects. For the case of rail and water links, shipping times are calculated according to a polynomial delay function of the type known in the literature as the "BPR function", with coefficients α and β equal, respectively, to 0.15 and 4.0 (refer to equation 4.5).

The shipping rates relating to the commodity "cargo" are equal to US$ 0.04 per ton-km on road, US$ 0.02 per ton-km on rail, and US$ 0.01 per ton-km on barges. A fixed unit transfer cost of US$ 1.50 per ton of "cargo" is assigned to each candidate transfer terminal. The problem assumes that all terminals will operate below capacity, with no increases in shipping delays due to transshipment operations. Loading and unloading movements at origins and destinations are disregarded in this test problem. The other characteristics of each mode affecting flow assignment are displayed on Table 2.

The test problem assumes that the value of time (VT) for commodity "cargo" is equal to US$ 0.10 per hour, what translates into values of coefficients θ_1 and θ_2 equal, respectively, to -0.1 and -0.01 (refer to equation 4.8). These logit coefficients were selected just for the purpose of this hypothetical exercise, without any calibration with real data.

As can be seen in Table 2, when calculating the value of the objective function, the cost of the environmental, safety, and energy impacts (system impact costs) of moving freight on roadways and railways, as opposed to inland waterways, is estimated as a fixed value equal to US$ 0.01 per ton-km for roads and US$ 0.005 per ton-km for rail.

Investment costs for locating transfer facilities differ by pairs of modes involved, but do not vary among nodes in the network. The cost of locating one transfer terminal between road and rail is assumed to be equal to US$ 1,000,000.00, whereas the investments on transfers between water and road, and water and rail are estimated as US$ 1,500,000.00.

MCMND Results

The test problem described above was solved by the MCMND model for several different investment scenarios. A series of runs was performed considering ten different budget levels, for which the results are presented in Table 3. The model was solved on a 486-66 DX2 IBM-compatible microcomputer. The corresponding running time for each case is also displayed on Table 3.

All the ten cases listed on Table 3 were solved using a value of γ equal to 9. Before deciding on which parameter value to use in the calculation of the surrogate costs, several different integer values of γ, ranging from 1 to 100, were tested with all ten budget levels. The analysis indicated no appreciable changes from the results displayed on Table 3. For values of γ close to 1, the path construction submodel generated more alternative paths requiring the location of new transfer facilities. However, the values of the objective function for all cases

Table 3

MCMND Results of the Test Problem

Case	Budget (US$ 10^3)	# Iters	# Subs	# Paths Total	# Paths Used	Obj.Func. Value (US$$10^3$)	Transfer Facilities Located	Invest Cost (US$ 10^3)	Run Time (sec)
Base	-	-	-	12	12	190,029.49	-	-	-
I	0	3	2	14	12	185,462.53	-	0	3.9
II	1000	3	3	15	13	179,293.23	h5r	1000	5.6
III	1500	3	6	16	13	179,293.23	h5r	1000	6.8
IV	2000	3	6	17	13	179,293.23	h5r	1000	6.8
V	2500	2	1	15	13	173,847.95	h5r, h3w	2500	2.9
VI	3000	2	2	15	13	173,847.95	h5r, h3w	2500	3.1
VII	3500	3	5	16	13	166,595.47	h5r, h3w, h3r	3500	7.9
VIII	4000	3	3	16	14	169,774.11	h5r, h3w, h3r	3500	5.7
IX	4500	4	6	17	14	169,774.11	h5r, h3w, h3r	3500	9.9
X	5000	3	2	16	15	169,995.98	h5r, h3w, h3r, w0h	5000	5.1

were exactly the same as the ones in Table 3. On the other hand, for values of γ equal to or greater than 10, the number of alternative paths generated by the path submodel was somewhat less, sometimes causing the model to reach its final solution faster. Again, except for case I, no differences whatsoever were found in the final values of the objective function due different values of γ. Eventually, the value 9 was selected as a threshold in the trend of the path submodel for generating a significant amount of candidate paths to enter into the final solution.

The analysis of the results presented in Table 3 indicates that, for budget levels up to US$ 2,000,000.00, the best solution found, by the MCMND model, to the test problem is the location of a single intermodal transfer facility at node 5, between road and rail (h5r). If the budget is increased to US$ 3,000,000.00, the model then locates one additional facility at node 3, between road and water (h3w). For budget levels equal to or greater than US$ 3,500,000.00 and less than US$ 5,000,000.00, a third facility is selected to be located at node 3, between road and rail (h3r). Finally, for budget values equal to or over US$ 5,000,000.00, the model assigns a fourth transfer facility, this time at node 0, between water and road (w0h), with the same objective function values and corresponding flow assignments as in case X.

It is interesting to note that the network configuration which led to the lowest objective function value (US$ 166,595,470.00), represented by case VII, included the location of only 3 transfer facilities. Although it may seem inconsistent that by increasing the budget available, and locating an additional facility, the value of the objective function increased, it is important to remember the stochastic nature of the path-based flow assignment. Since flows are assigned to paths based on probabilistic values calculated as a function of path impedance, an increase in the number of paths belonging to the in-solution set may cause a greater amount of flow to be assigned to paths with higher values of travel impedance. This may, ultimately, result in an increase in the final value of the objective function.

Due to the intrinsic characteristics of the two-step improvement algorithm, once an initial solution, containing a certain number of paths, is generated by the greedy step, the substitution step is not able to change the cardinality of the in-solution set. Therefore, it is perfectly possible that, by increasing the budget, the number of paths included by the greedy procedure in the initial feasible solution may also increase, resulting in a higher value for the objective function. A good example of this situation can be observed by comparing the results of cases VII and VIII. In both cases, the final solution indicated the selection of the same three facilities; however, the values of the objective function were different. The disparity was caused by the addition, in case VIII, of one extra path to the initial feasible solution, and that prevented the substitution algorithm from reaching the same objective function value as in case VII.

Another important aspect that deserves attention is the possible effect of the Braess's paradox to the problem solved herein. As discussed in the last section of Chapter III, since the bilevel formulation involves an objective function that includes cost factors not considered by the users of the system, the value of the objective function may not always decrease monotonically as new links (transfer facilities) are added to the network.

Another aspect that should be pointed out involves the computer execution speed of the overall solution algorithm. It is interesting to note that case IX, the one with the highest number of iterations and substitutions, consequently presenting the longest running time, was solved in less than 10 seconds. The solution of this case required the execution of the assignment algorithm for more than 100 times, and the speed of execution proves the efficiency of the path-based logit-type approach used by the model to achieve a stochastic state of user equilibrium.

BTMS Results

To assess how good was the solution given by the MCMND model to the test problem, another transportation model, called the Bangladesh Transportation Modeling System - BTMS (Ralston et al., 1994), was used to run a series of stochastic equilibrium assignments on the test network using different combinations of locations for intermodal transfer facilities. The idea was to compare the results of the ten cases analyzed with the MCMND model to those for similar scenarios analyzed with the BTMS.

Since it was practically impossible to run, one by one, all possible combinations of facilities even for this very small test network (2^{16} = 65,536 different network configurations), it was decided to group facilities by candidate transfer node and, then, run all possible combinations on those nodes (2^4 = 16). The results of these 16 runs, including the transfers allowed and the ones actually used in the assignment, are presented in Table A-1 in the appendix. The analysis included the 5 transfer facilities for which the BTMS assigned flow when considering a network with all 16 transfer terminals (w0h, h3r, h3w, h5r, h5w) and the 32 scenarios (2^5 = 32) consisting of all possible combinations among those 5 facilities were tested. The results of these 32 runs are summarized in Table A-2 in the appendix. A total of 48 different network scenarios were run in the BTMS. Table 4 displays a list of the 15 different values of the objective function found among the 48 runs of the BTMS, with the respective transfer facilities used in the assignment of each case. It can be noticed that several different network configurations resulted in the same flow assignment and, consequently, in the same objective function value.

Before comparing the results of the two models, a very important point deserves to be emphasized. The BTMS assigns flows according to a stochastic user equilibrium criterion, not taking into account the system impact costs present in the MCMND model. Therefore, the two models should not produce the same final assignments and the same resulting objective function values to the test problem. Even though the values of the objective function displayed on Table 4

Table 4

BTMS Results of the Test Problem

Case	Transfers Used	Objective Function (US$ 10^3)
1	w0h, h3r, h3w, h5r, h5w	169,561.21
2	h3w, h5r, h5w	170,200.15
3	h3r, h3w, r3w, h5r, h5w	170,302.30
4	h3w, h5r	170,854.54
5	h3w, h5w	178,629.10
6	w0h, h3r, h3w, h5w	178,635.22
7	w0h, h3w	181,640.73
8	w0h, h3r, h3w	181,645.97
9	h3r, h5r	182,736.45
10	h3w	182,794.18
11	w0h, h3r, h5r	183,671.28
12	r3w, h3w	185,373.10
13	h5r	185,638.70
14	h3r	193,312.06
15	No transfers used	209,264.91

were adapted to reflect the same procedure of the MCMND model, we still cannot expect that similar network scenarios (same combination of facilities) running on the two models will result in the same objective function value. This is true even if the BTMS was modified to optimize system costs as part of its objective function. The reason for this apparent inconsistency is that, although both models assign flows according to a SUE criterion and using the same logit function, the BTMS applies Dial's algorithm (Dial, 1971) as opposed to the MCMND's path-based algorithm. That is, by following Dial's algorithm, the BTMS assigns flows according to a link-based stochastic network loading procedure, which differs from our path-based approach. As discussed before, the link-based method has the advantage of precluding path enumeration. However, this advantage turns out to be inappropriate for the MCMND formulation since the link-based procedure does not maintain the same set of in-solution paths throughout its execution.

Despite the different methodologies for performing flow assignment, the comparison of the results obtained by running the two models can still provide an indication of the efficiency of the MCMND model, as it will be showed in the following section. As a final comment, the use of the BTMS as a means of comparison to the MCMND model was the only available alternative, since no other path-based stochastic assignment algorithm was found in the literature.

Comparing the Results of the Two Models

By comparing Table 3 to Table 4, it is clear that the two models provided different best solutions to the test problem. The lowest value of the objective function found by the BTMS required the location of 5 transfer facilities, represented by case 1 in Table 4. The MCMND model, on the other hand, suggested the location of only 4 transfer terminals (case X, Table 3), with the total investment cost never exceeding US$ 5,000,000.00. However, all 4 facilities selected by the MCMND model are also part of the best solution provided by the BTMS, which is an indication that both models progressed in the same optimal direction.

Another striking conclusion is that the different best solutions found by the two models resulted in very close values of the objective function, indicating a convergence trend to an optimal value. This similarity in the objective function value was somewhat unexpected due to the distinct assignment techniques used by the two models. The effect of the two different assignment procedures to the value of the objective function becomes clear when we compare case I in Table 3 to case 15 in Table 4. Both cases dealt exclusively with the base network, with no transfer facilities being located. The assignment algorithm embedded in the MCMND model was able to come up with an objective function value over 11% lower than the one provided by the BTMS.

One inconsistency found in the results given by the BTMS is that the presence, or absence, of some unused transfer terminals sometimes result in a different flow assignment. This fact can be illustrated by comparing, for instance, cases 3, 4, and 9 in Table A-1, in the appendix. When running case 3, which included only facilities at node 3, the model assigned flows to facilities h3w and r3w, resulting in an objective function value equal to US$ 185,373,100.00. For case 4, with only facilities at node 4, no transfers were used and the objective function totaled US$ 209,264,910.00. On the other hand, when all facilities were located at both nodes 3 and 4, in case 9, only one facility was used (h3w), reducing the value of the objective function to US$ 182,794,180.00.

A final observation regards the fact that, in none of the assignments performed by the BTMS, were the same sets of transfer terminals selected by the MCMND model in cases VII through X used. This is even more striking if we consider that case VII resulted in an objective function value that could not be achieved by the BTMS assignment algorithm. Again, this disparity may have been caused by the distinct assignment techniques applied in the two models.

This comparative analysis is concluded by stating that, although it was not possible to prove that the solution algorithm embedded in the MCMND model reached the optimal solution for the test problem, the algorithm was definitely able to find a good solution to such a problem, generating a very small number of paths in a short period of time.

The following chapter will present the application of the MCMND model to a similar problem of locating intermodal transfer facilities on a multimodal transportation network, however this time considering a very large, real-life network consisting of approximately 1,000 nodes and 1,500 physical arcs. This pilot application of the MCMND model will serve to test its applicability for the case of large network problems.

CHAPTER VI

THE TIETÊ-PARANÁ VALLEY APPLICATION

Introduction

In order to illustrate the applicability of the MCMND model to a real-life large network problem, the Tietê-Paraná valley area, located in the Southeast region of Brazil (refer to the maps in Figure 5), was selected for a pilot application of the model. Since the appropriate data needed to carry out a detailed analysis of the region's freight transportation system could not be gathered, this pilot application was restricted to the hypothetical problem of locating intermodal transshipment facilities along the Tietê and Paraná rivers. The problem was defined as hypothetical because it did not consider the existing river ports and terminals, as well as the current movements of commodities in that waterway system. The main purpose of applying the MCMND model to such a problem is to prove that the model can be solved for a large network in a reasonable amount of time.

After a brief description of the study area and its multimodal transportation system, the following sections present the steps involved in setting up such an application. The emphasis will be mainly on the representation of the network, the collection of the required input data, and the calibration of the model's

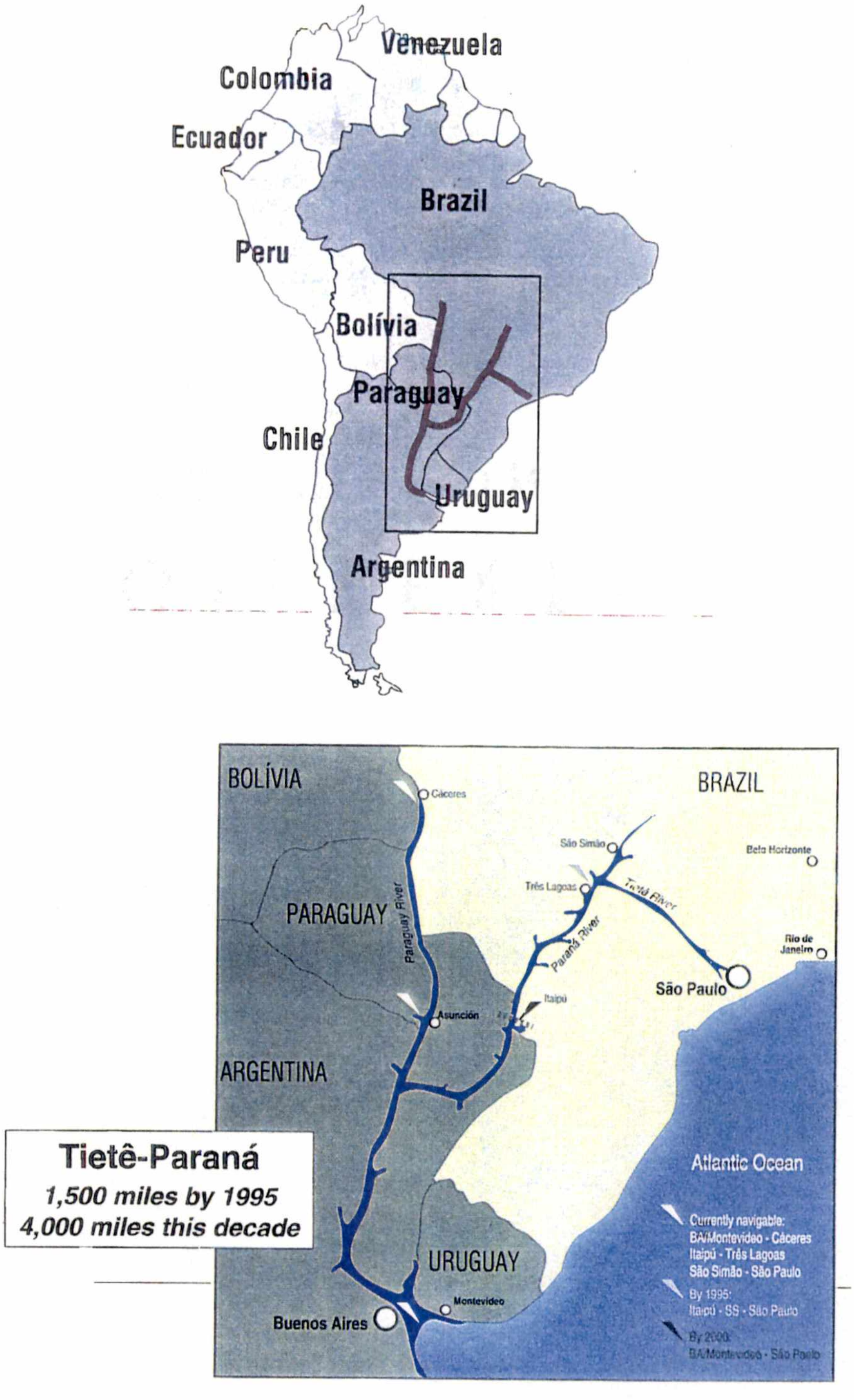

Figure 5

Tietê-Paraná-Paraguay Waterway System

parameters. As stated in the introductory chapter, the development of this pilot study also serves to demonstrate the importance of implementing the MCMND model within a GIS framework, especially regarding the three step listed above. The last section of this chapter summarizes the findings and results of this analysis. In addition, it indicates the data requirements for a future in-depth, realistic, application of the model to the region in question.

Characterization of the Study Area

The selection of the region under the influence of the Tietê-Paraná watershed (refer to Figure 5) as the study area for the pilot application of the MCMND model was motivated mainly by the multimodal characteristics of its freight transportation system and its socio-economic relevance within Brazil and the continent of South America. The region encompasses five Brazilian states and includes the São Paulo metropolitan area, with a total population of approximately fifty million people. It is South America's leading industrial and agricultural pole, accounting for over 75% of Brazil's gross national product.

With the advent of the Southern Cone Common Market (Mercosul), a free trade agreement involving Brazil, Argentina, Uruguay, and Paraguay scheduled to take effect on January 1st, 1995, this region is expected to experience a boost in economic growth. The anticipated increase in commodity flows resulting from the estimated growth in production and consumption will require major infrastructure investments in the region's transportation system. Both state and federal governments in Brazil have already expressed interest in promoting infrastructure improvements to the regional multimodal network. Priority areas of investments have been defined with the main objectives of improving the region's waterway system through the construction of new locks and transshipment terminals, upgrading existing railroad facilities, and allowing concession and privatization opportunities for maintaining and expanding the highway system.

Among the investment strategies already being planned for the region, the development of the Tietê-Paraná waterway system is foreseen as one of the key factors in fostering economic growth within the future free trade zone. The river system, which comprises the Tietê and Paraná rivers and some minor tributaries, drains an area 50% larger than that drained by the Mississippi river. It economically influences about 70 million hectares reaching the states of São Paulo, Paraná, Minas Gerais, Goiás, and Mato Grosso do Sul (refer to the political map of Brazil in Figure 6). Presently, 1,040 km of navigable ways connect the metropolitan region of the city of São Paulo all the way north to the states of Goiás and Minas Gerais. Waterborne traffic, estimated as approximately 3.3 million tons in 1993, is already flowing in the total extension of the Tietê river and the north stretch of the Paraná river.

It is expected that by 1995, with the completion of two new lockage facilities, the Tietê waterway will be also interconnected to the already navigable south stretch of the Paraná river, reaching the Itaipu hydroelectric dam and the town of Foz do Iguaçu at the border of Brazil and Paraguay (see Figure 5). By

Figure 6

Administrative and Political Map of Brazil

then, there will be a total of 2,400 km of navigable ways linking the Brazilian waterway system to the Paraguay river, allowing cargo convoys to move directly from São Paulo to the capitals of the other three countries part of Mercosul - Asuncion, Buenos Aires, and Montevideo. Therefore, it is not without reasons that the Tietê-Paraná-Paraguay waterway complex is already being called "the highway through the heart of Mercosul".

The concept of the waterway area of influence gets an even broader dimension if one considers its interconnections with the existing railway and highway networks. The Southeast region of Brazil incorporates a comprehensive road network comprising a total of 42,000 km of paved and 428,000 km of unpaved federal, state, and municipal roads. Its railroad network, in turn, consists of approximately 15,000 km of metric- and broad-gauge tracks managed by several public and private owned companies. Figure 7 depicts a schematic representation of the multimodal network in the state of São Paulo and surroundings.

Currently, around 80% of the freight traffic moves on trucks, with rail and water accounting approximately for only 18% and 2%, respectively. Although planners estimate a fivefold increase in the amount of freight moving on barges by the next decade, the waterway system will be increasingly dependent on the two other modes - highway and rail. The main reason for this interdependency arises from a physical barrier imposed by nature. The Tietê river rises just 22 km from the Atlantic ocean (with its first navigable section being located about 100 km

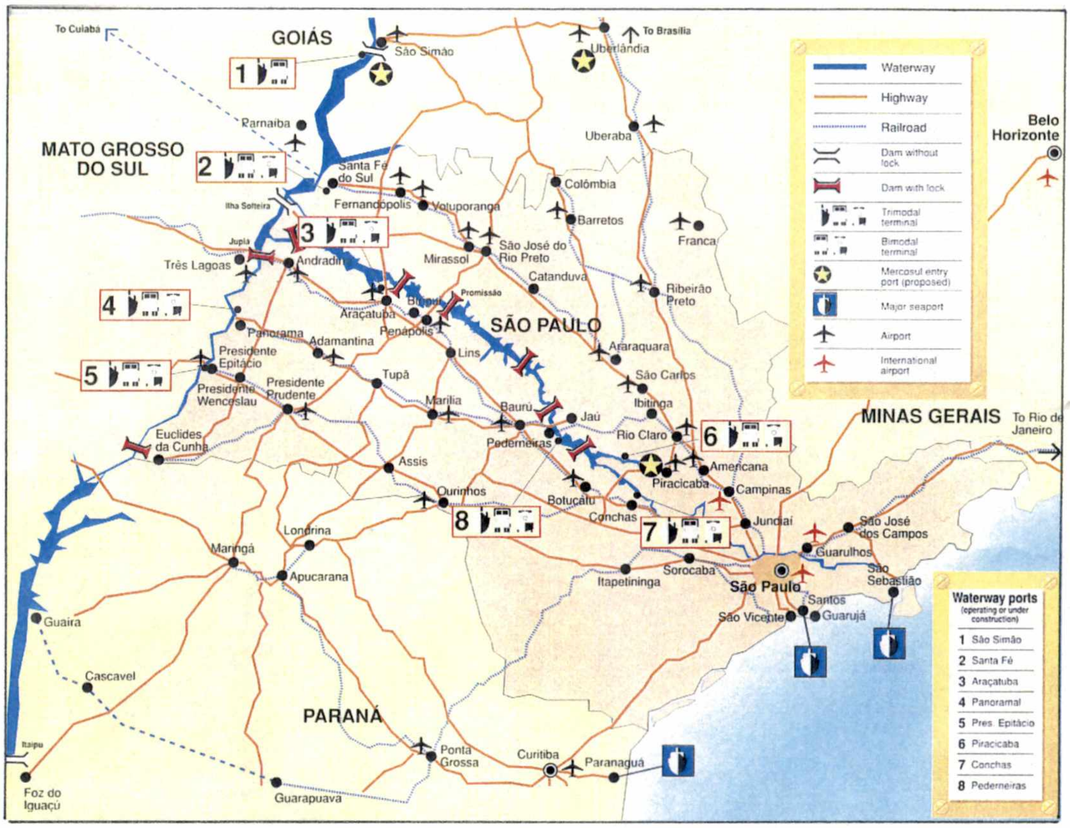

Figure 7

São Paulo's Multimodal Network

west of the city of São Paulo), but its waters must flow almost 4,800 km west and south before reaching the sea in the River Plate estuary (refer to the maps in Figure 5). The 790-meter Serra do Mar mountain encarpment makes it impossible for goods moving east on the waterway system to reach Brazil's Atlantic ports. This natural barrier means that barge traffic from the interior of Brazil must either take the long route south through the Mercosul zone, or head east and join up with the road or rail systems for the final leg down the mountain to the major sea ports such as Santos and Paranaguá. Thus, the region presents a strong potential for intermodal activities, with mixed-mode shipments expected to play an important role in intercity goods movements. Therefore, the transportation economic benefits of the waterway system will be increasingly higher when both the end to end and supplementary transportation are time and cost efficient.

The products shipped through multimodal routes in the Southeast region of Brazil include mainly grains and brans (e.g., soy beans, wheat, corn, and rice), iron ore, steel products, coal, alcohol and oil by-products, sugar cane, cement, building materials, fertilizers and agricultural limestone, charcoal, timber, and cattle. Other types of commodities are also transported in smaller quantities. The movement of freight within the region is estimated to be approximately 300 million tons per year. Presently, some of those commodities are already being moved on barges. Once the Tietê-Paraná waterway gets fully operational, the traffic on the two rivers is expected to consist primarily of agricultural products and fertilizers, as well as petroleum by-products and manufactured goods, according to the chart depicted in Figure 8.

In a recent study performed by the University of Montreal in cooperation with the Brazilian government (Florian and Crainic, 1989), this same region was also used as a pilot study area for the development of a multimode multiproduct network assignment model (Guelat et al., 1990). In their study, the Southeast region of Brazil was subdivided in 165 internal traffic zones and some other external zones, to represent shipments to other regions of the country. The respective origin-destination data were gathered for eight major products - iron ore, steel products, coal, cement, fertilizers, Soya oil, Soya grain, and Soya meal.

The study also required data describing the components of the multimodal transportation network (modes, nodes, links, and transfers) and their attributes (vehicles, lengths, capacities, unit costs, associated functions, etc.). Although their analytical model, STAN (reviewed in Chapter II), was able to handle a variety of modes of transportation, including the micro-simulation of ocean and river ports operation, their pilot study did not consider inland navigation as one of the options for freight movements in the region. Nevertheless, due to its comprehensiveness both in terms of modeling techniques and data collection, that study served as a main source of reference for this pilot application of the MCMND model, as will be noticed in the sections to follow.

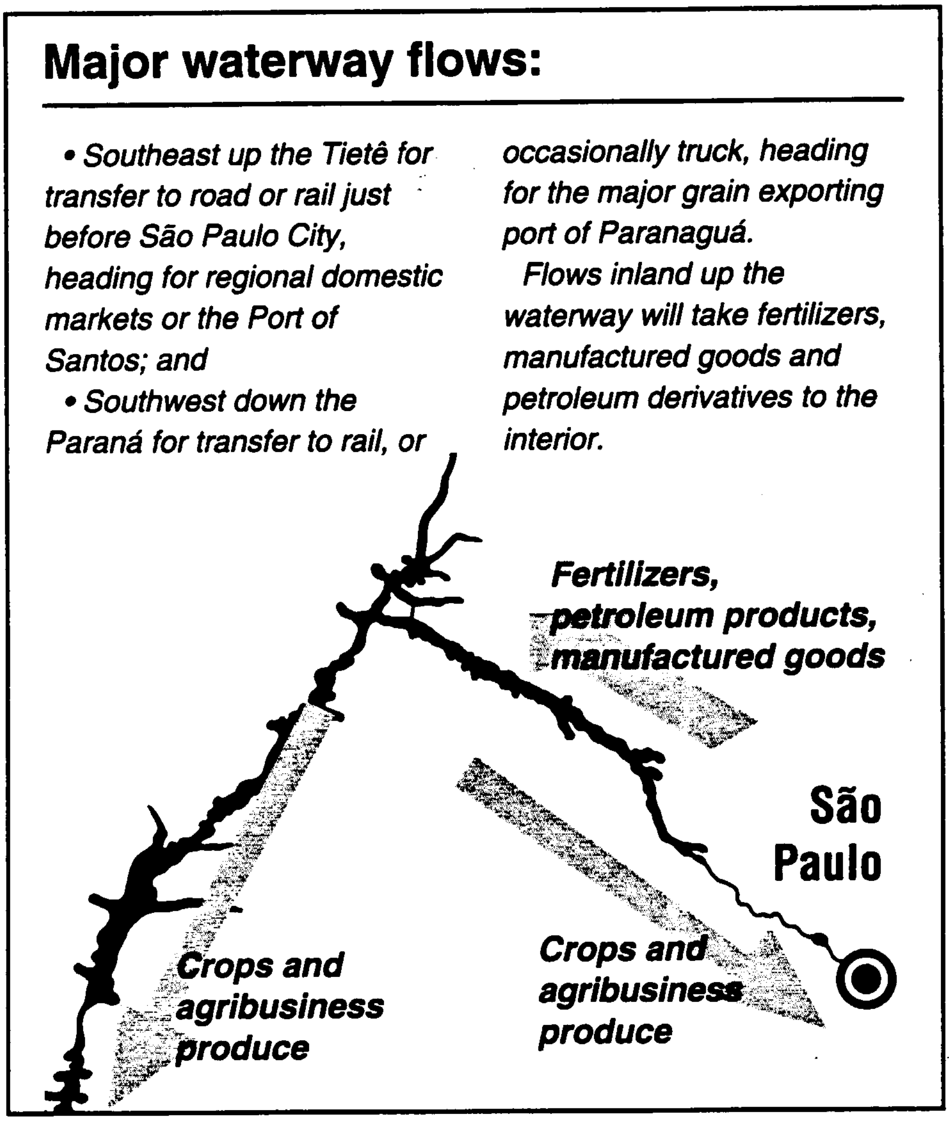

Figure 8

Projected Major Flows on the Tietê and Paraná Rivers

Data Requirements

This section describes the steps involved in collecting the data required for the Tietê-Paraná valley application, emphasizing on the suitability of using a GIS platform for storing, retrieving, and analyzing the information on network attributes and commodity flows. The following four subsections describe individually the tasks of representing the multimodal network, defining mode and link specific costs and functions, compiling O/D data, and establishing the set of candidate investment options to be considered by the MCMND model.

Network Representation

In applying the MCMND model for the location of transfer terminals along the Tietê and Paraná rivers, the first step was the representation of the physical multimodal transport network in a level of detail appropriate for strategic planning. The study area encompassed the whole region influenced by the Tietê-Paraná valley, taking into account the origin and destination of commodity flows likely, at some point, to use the waterway system. The network was represented for seven Brazilian states including the totality of the Southeast region and parts of the West-Central and South regions of Brazil. The digital representation of the road, rail, and inland water links for the states of São Paulo (SP), Rio de Janeiro (RJ), Minas Gerais (MG), Espírito Santo(ES), Paraná (PR), Goiás (GO), and Mato Grosso do Sul (MS) was provided by the Technical Research Institute (IPT-SP) of the São Paulo state government. This database was digitized in Brazil using the GIS-Plus package, a simpler version of TransCAD, and consisted initially of 5,170 nodes and 5,772 links.

The digital database generated by IPT-SP also included some link attributes to roads and rail lines, as well as the location and name of the existing ports along the Tietê and Paraná rivers. For roadways, in addition to the from- and to-nodes, and the length of the link, the database also contained information on the designated name of the road and its functional type (primary or secondary road). For railroads, in turn, it included the operating company for each line (RFFSA, federal-owned company, or FEPASA, São Paulo's state-owned company), its regional administrative office for the case of RFFSA lines, type of gauge (metric or broad), type of link (single or double tracks), and whether the link was electrified or not.

Based on the IPT data and several other sources, referenced later on this chapter, a TransCAD application was developed consisting of two point and two line databases representing, respectively, zone centroids, existing river ports, political boundaries, and the multimodal regional network. The network line database comprised two separate layers containing nodes and physical links.

In order to adapt the network to the purposes of this pilot application, a link-chaining process was carried out to reduce the number of nodes and links in the database, maintaining only those nodes which represented zone centroids, rail stations, transfer terminals, and intersections between three and more links. The multimodal network that resulted from this editing process comprised, then, 895 nodes and 1,465 physical links corresponding to 41,760 km of state and federal roads, 14,485 km of metric (1.0 m) and broad (1.6 m) gauge railroads, and 1,640 km of inland waterway routes. This digital representation of the physical network consisted almost of the totality of all paved highways, railroads, and waterways in the region under study. Figures 9, 10, and 11 show maps generated out of TransCAD depicting, respectively, the inland waterway, railway, and roadway links stored in the network database.

Unit Costs and Functions

As specified in Chapter III, the MCMND model assigns freight flows based on a combination of two impedance criteria: shipping cost and time delay. Therefore, it was necessary to collect cost and time information for each individual mode and each different group of commodity to be considered in this pilot application.

The Transportation and Logistics Division of IPT-SP was once again the main source for data regarding fixed costs and times of loading/unloading activities, transport rates charged by carriers, average speed of typical vehicles used by each mode, and transshipment costs and times at transfer terminals. Additional information was also obtained from the report entitled "Strategic Planning of Freight Transportation in Brazil: Methodology and Applications" (Florian and Crainic, 1989) which summarized the development and application of the STAN model to the Brazilian transportation system.

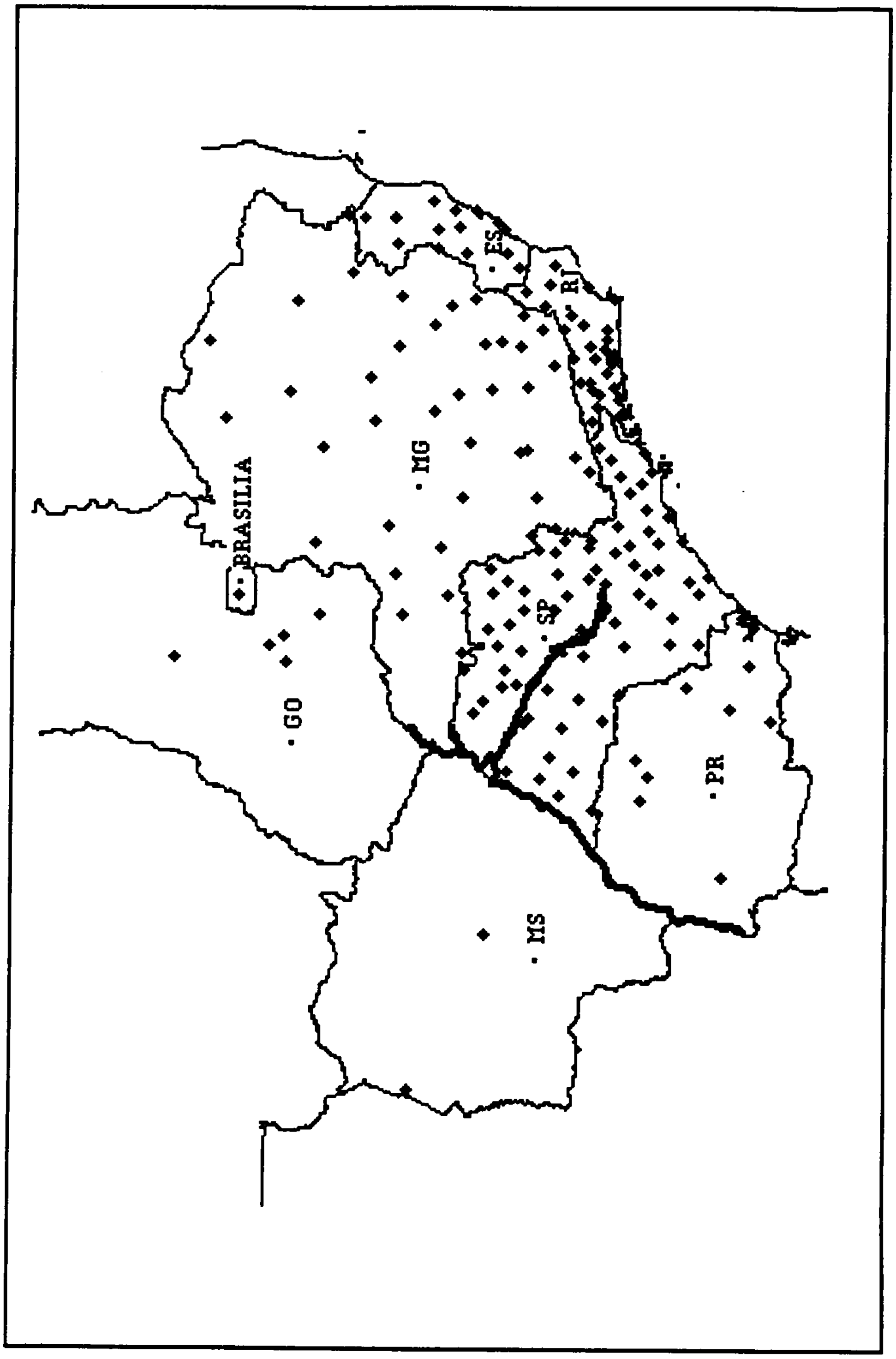

Figure 9

Digital Representation of the Inland Water Network and Zone Centroids

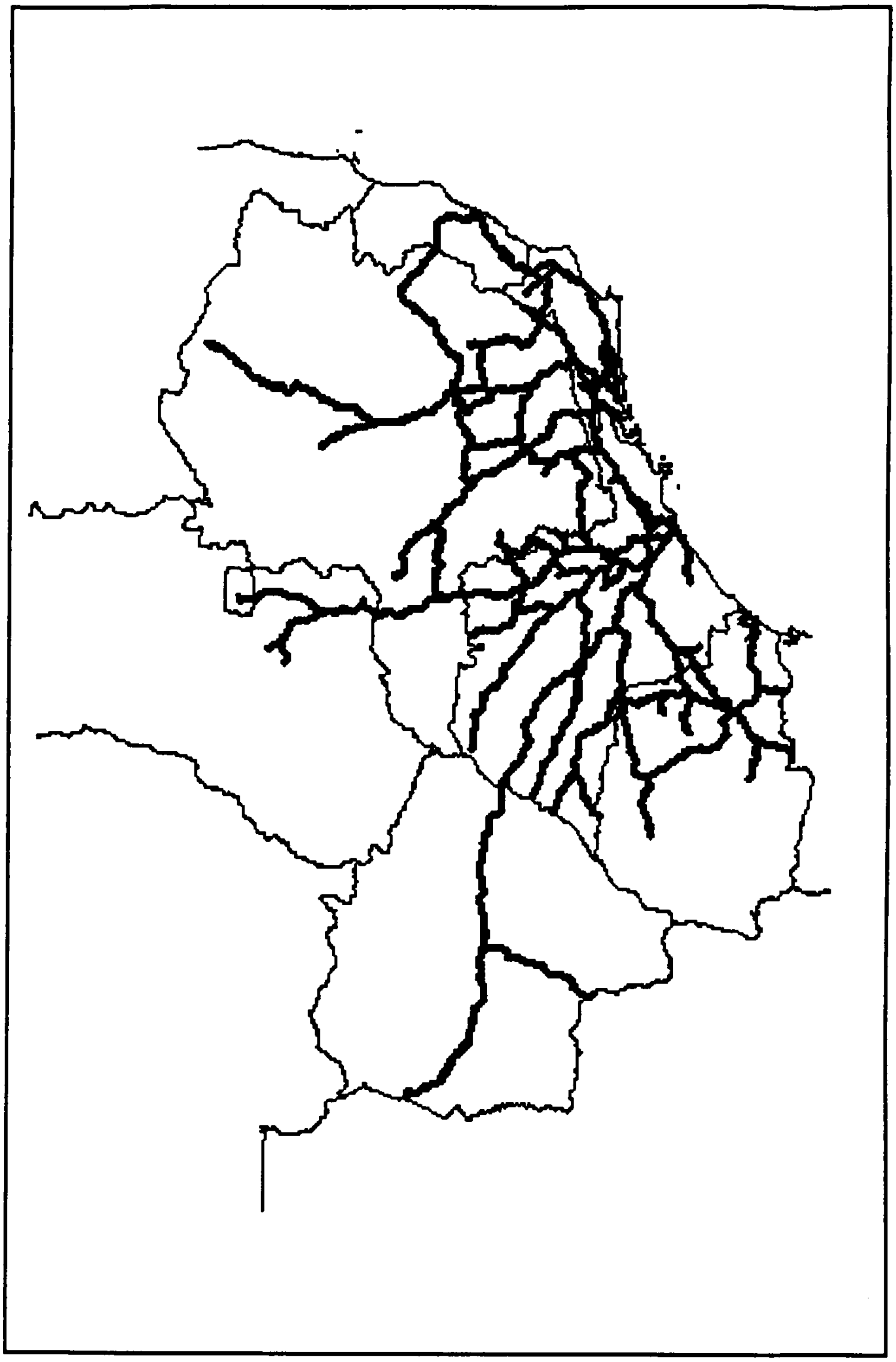

Figure 10

Digital Representation of the Rail Network

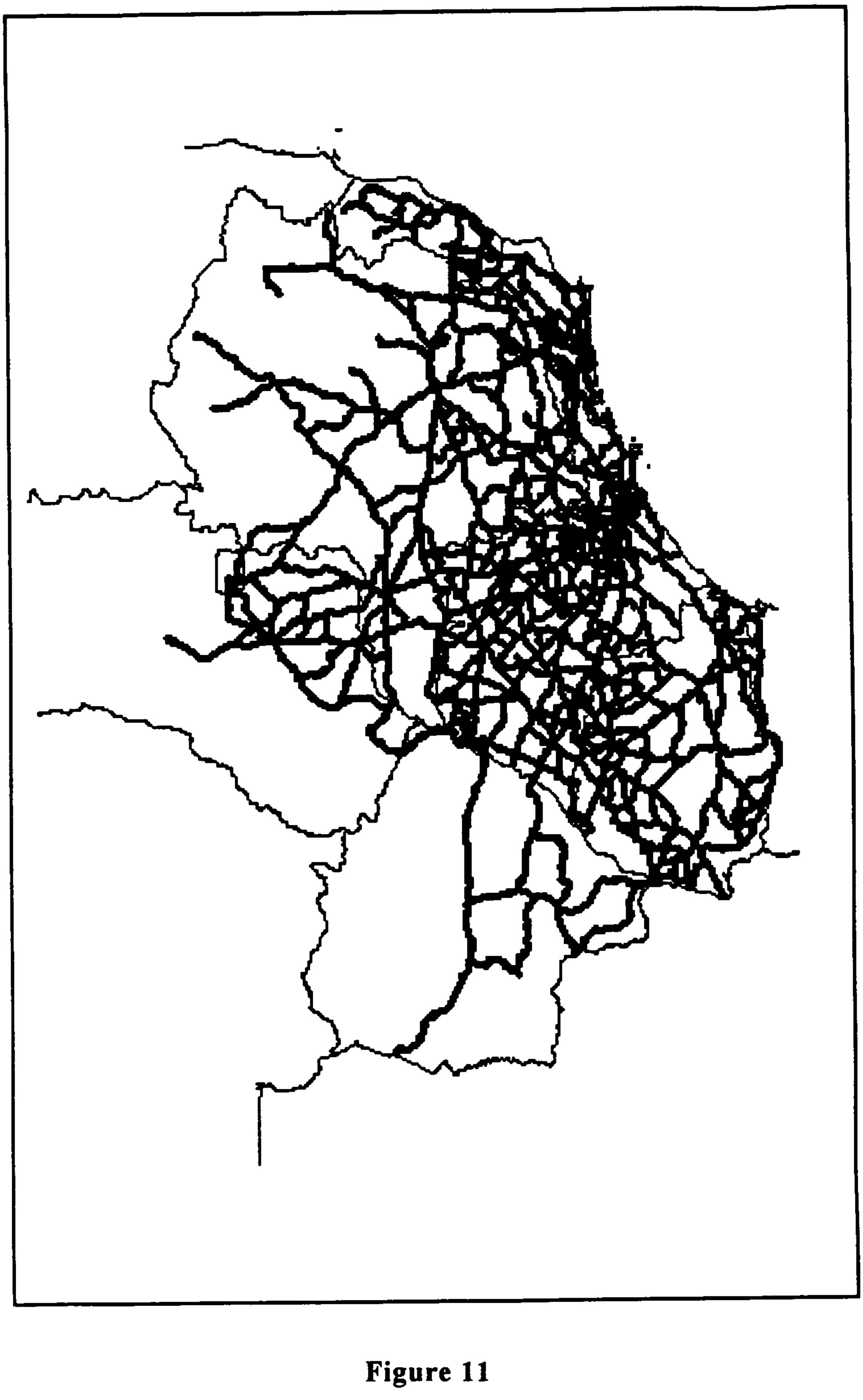

Figure 11

Digital Representation of the Road Network

Even though unit costs and times usually vary by commodity type, the MCMND model requires the use of modal average values when calculating the total generalized cost of moving freight on the network. Therefore, mode representative values for the above listed attributes were estimated based on the information gathered from the two sources mentioned in the last paragraph. Table 5 displays the unit costs and times by mode used in this application. The commodity-specific data were considered, however, during the calibration of the logit function coefficients, as will be discussed later in the section describing the model calibration steps.

This application of the MCMND model assumed that congestion would only be a factor when assigning flows to single track rail links. Thus, shipping costs and delays on roads, waterways, and double track railroads were determined by linear functions of the distance traveled according to equations (4.6) and (4.7). Such an assumption followed the modeling approach implemented by Florian and Crainic (1989) in their analysis of the São Francisco river transportation corridor, in the Northeast and Southeast regions of Brazil. It was thus considered adequate for the purposes of this pilot application.

The calculation of over-the-line delay on single track rail links was performed based on a simplified version of the polynomial function used in the STAN model and calibrated specially for the Brazilian railway system. The original delay function proposed by Florian and Crainic (1989) was a polynomial approximation of the form:

Table 5

Tietê-Paraná Application - Modal Attributes

Mode	Shipping Rate (US$/ton-km)	Average Speed (km/h)	Loading or Unloading Fixed Charges (US$/ton)	System Impact Cost (US$/ton-km)	Transfer Cost (US$/ton)
Road(0)	0.050	55	3.20	0.01	1.50
Rail(1)	0.025	30	1.20	0.002	1.50
Water(2)	0.015	10	0.60	-	1.50

$$\overline{D}_a = FRT_a \cdot (1 + \phi_1 \cdot T_a + \phi_2 \cdot (T_a / CAP'_a)^{\exp})$$

where

$\overline{D}_a$ estimate of delay on link a.

FRT_a free running time on link a.

T_a total train flow on the link (both directions).

CAP'_a value of the total train flow where the line begins to be considered congested.

ϕ_1 parameter reflecting the linear part of the function.

ϕ_2, exp congestion parameters reflecting the nonlinear portion of the delay curve.

Their study suggested values for ϕ_1, ϕ_2, and exp equal respectively to 1.1/ CAP'_a , 4.12, and 6.0. For this application, a simplified polynomial function of the type defined by equation (4.5) was adopted to approximate the delay function above, however with link volume and capacity expressed in thousands of tons per year instead of number of trains on the link. The same values of ϕ_2 = 4.12 and exp = 6.0 were used for the parameters α and β in equation 4.5. The practical capacity of a link was estimated as 80% of its nominal capacity, assuming a maximum of 1,500 trains per year (equivalent approximately to 30 trains per week), an average train convoy of 40 wagons, and a maximum net loading capacity of 30 tons per wagon. These figures were also obtained from the STAN report. The value of 1,440 thousands of tons per year was then considered as the practical capacity for all single track links.

Regarding the operation of transshipment terminals, this application assumed fixed values of cost and time incurred by each land-land or land-water transfer activity. Since transfer facilities are represented in the MCMND model as logical arcs, the corresponding costs to traverse these arcs were determined assuming a US$ 1.50-rate per ton-km and a 1.0-km arc length for both land-land and land-water facilities. In calculating the time required to traverse the intermodal transfer arcs, a symbolic value of speed expressed in km per hour was used to represent a fixed delay value of 24 hours per transfer. For the case of loading and unloading arcs, the same procedure was adopted with fixed costs equal to US$ 3.20, US$ 1.20, and US$ 0.60 per ton for road, rail, and water shipments, respectively; whereas, the fixed delay period was defined as equal to 2 hours for either loading or unloading trucks, and 24 hours per operation for the case of rail and water movements.

In estimating the system impact costs accruing from moving freight in trucks and trains, as opposed to barges, only fuel consumption was considered as a measure of the net system benefit resulting from the use of more efficient modes of transportation. Emissions and accident rates were disregarded in this pilot application due to the difficulties and controversies involved in converting these two measures of efficiency into monetary values. Fuel consumption rates for the three modes being considered in this application were extracted from the study developed by Newstrand (1992) on the environmental impacts caused by modal shifts. As indicated in Chapter III, that study provided the methodology used in the MCMND model to estimate system impact costs. The fuel consumption rates suggested by Newstrand were converted from ton-mile/gallon to ton-km/liter, resulting in rates equal to 25, 86, and 216 ton-km/liter for trucks, rail, and barges, respectively. The cost of one liter of fuel oil in Brazil was obtained from the STAN report and converted to the present value of US$ 0.30. The ϕ_h and ϕ_b cost multipliers in equation (3.1) were then estimated, respectively, as US$ 0.01/ton-km and US$ 0.002/ton-km. These values are displayed in Table 5.

Transport Demand

One of the first steps in setting up an application of the MCMND model is to divide the region under study into homogeneous traffic zones and to associate a centroid to each zone. For this pilot application, it was adopted the zoning system defined by GEIPOT, the transportation planning agency of the Ministry of Transportation of Brazil. As mentioned before, a TransCAD point database was created containing the name and location of the GEIPOT's zone centroids situated inside the study area. A total of 170 zones were designated for the seven Brazilian states, all of them having direct access to the road sub-network, 120 being accessed by road and rail, and two of them being directly connected to the inland waterway network.

The demand for transportation was, then, represented in the MCMND model by a set of origin-destination (O/D) matrices containing the quantity (in thousands of tons) of each product category that had to be moved between the centroids of the network. Five different homogeneous classes of commodities were defined for this application, considering aspects such as transportation mode usage, cargo and storage handling. These product classes are:

- general cargo: cement, building materials, wood, etc.
- mineral solid cargo: iron ore, bauxite, manganese, zinc, aluminum, phosphatic rock, lime, gypsum, mineral coal and charcoal.
- steel products.
- agricultural solid cargo: Soya bean and Soya meal, rice, corn, wheat, sugar, salt, and fertilizers.
- liquid bulk cargo: fuel oils, brut petroleum, petroleum by-products and alcohol.

The same O/D matrices used during the development of the STAN system were obtained from the Department of Industrial Engineering of the Pontificia Universidade Católica of Rio de Janeiro (PUC-RJ). The tables contained 1986 observed flows of the commodity classes listed above for roads and railways.

Since more recent demand data could not be gathered, 1986 was defined as the base year for this application. The MCMND model, then, tried to select the transfer terminals to be located along the rivers according to their impact in reducing the transportation costs observed for that base year, through the reassignment of flows to the improved network.

The 1986 O/D data are summarized in Tables 6, 7, and 8 which display the total amount and percentage of tons by product category moved on each mode and the intra and interstate shipment totals. It is important to notice that the total number of tons moved on the network adds up to over 1 billion tons, while the actual value for the region under analysis in 1986 was estimated in approximately 200 million tons. This discrepancy results from the fact that many shipments were broken down into several intermediate movements, causing a double-counting effect on the amount of tons transported over the network for that year.

Investment Options

As stated earlier in this chapter, this pilot application comprised the solution of a hypothetical transfer terminal location problem along the Tietê and Paraná rivers. Such a problem assumed that, in the base year, no freight cargo was transported by the waterway system due to the lack of water-road and water-rail transshipment facilities. Then, 25 possible locations for connecting the waterway system to the road and rail sub-networks were identified, as depicted in Figure 12. Table 8 presents a list of those candidate port locations.

Table 6

Commodity Flows by Mode - 1986 O/D Data

(in 10^3 tons)

	General Cargo	Minerals	Steel Products	Agricultural Solid	Liquid Bulk	**Total**
Road	52,899	433,193	33,255	78,400	152,110	**749,857**
Rail	23,337	184,963	34,732	25,831	40,933	**309,796**
Total	**76,236**	**618,156**	**67,988**	**104,231**	**193,043**	**1,059,653**

Table 7

Commodity Flows by Mode - 1986 O/D Data

(in %)

	General Cargo	Minerals	Steel Products	Agricultural Solid	Liquid Bulk	**Total**
Road	69.4	70.1	48.9	75.2	78.8	**70.8**
Rail	30.6	29.9	51.1	24.8	21.2	**29.2**
Total	**7.2**	**58.4**	**6.4**	**9.8**	**18.2**	**100.0**

Table 8

Interstate O/D Movements in 1986

(in 10^3 tons)

State	MG	ES	RJ	SP	PR	MS	GO	Total
MG	103,125	284,031	169,657	66,747	405	1	8,657	632,624
ES	20,451	5,466	1,155	2,682	4	0	405	30,162
RJ	7,180	3,597	62,225	10,802	309	0	614	84,726
SP	13,008	2,137	5,724	165,678	9,029	4,159	6,012	205,746
PR	767	1,584	193	11,388	75,931	21	47	89,930
MS	21	0	0	1,702	1,721	701	103	4,249
GO	2161	7	164	5,371	127	629	3,759	12,218
Total	146,714	296,822	239,117	264,370	87,526	5,511	19,597	1,059,657

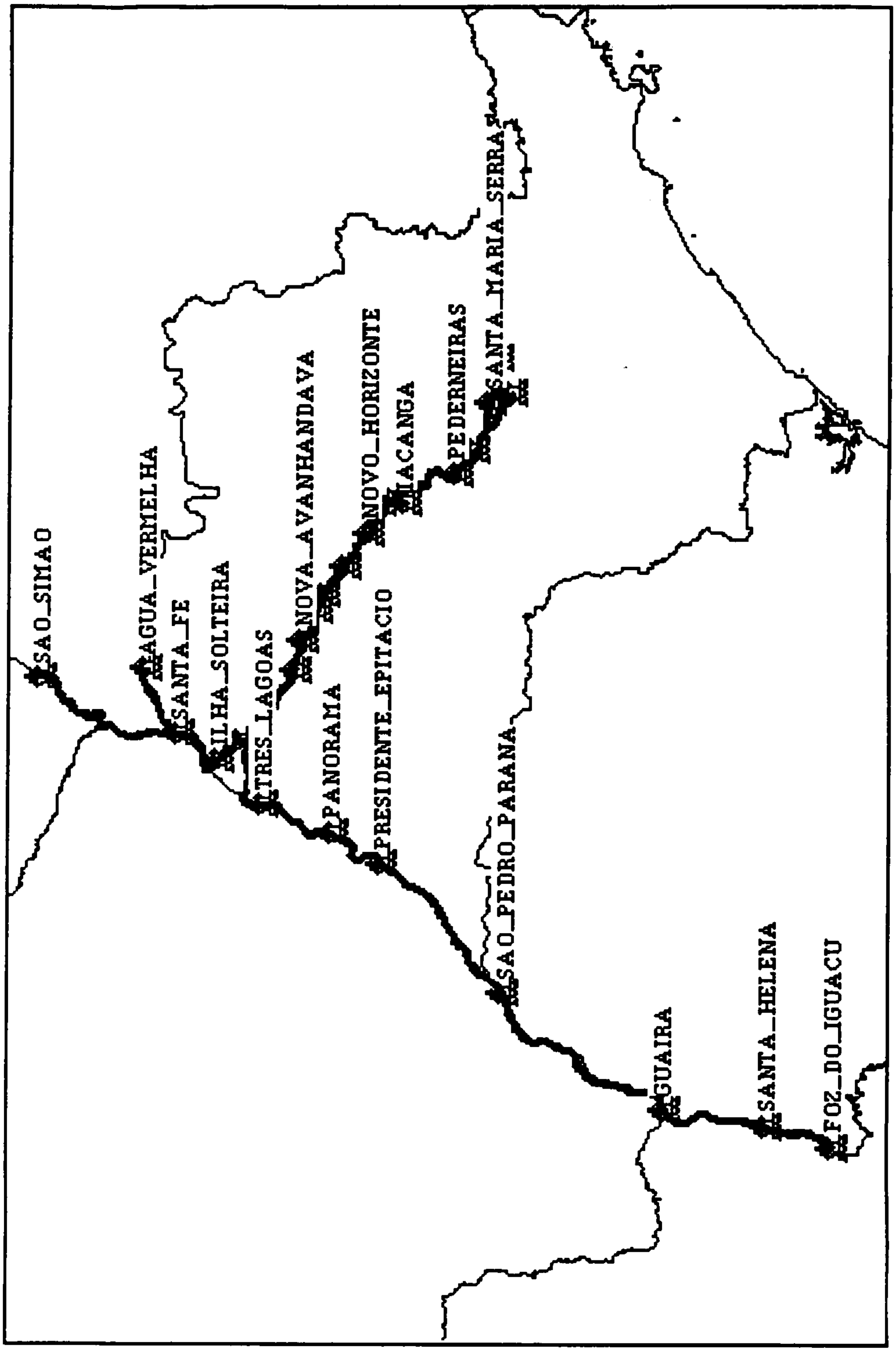

Figure 12

Candidate Locations for Transfer Terminals

Table 9

List of Candidate Locations for Transfer Terminals

along the Tietê and Paraná Rivers

Node-ID	Node Name
.034107	Porto Anhumas (Conchas)
.034112	Porto Artêmis (Piracicaba)
.034323	Porto Santa Maria da Serra
.034080	Porto Igaraçu do Tietê
.034500	Porto Pederneiras
.034087	Porto Iacanga
.034093	Ibitinga
.034322	Novo Horizonte
.034060	Sabino
.034321	Promissão
.034320	Barbosa
.034319	Nova Avanhandava
.034027	Porto Araçatuba
+003464	Pereira Barreto (Barragem Três Irmãos)
.034317	Barragem Ilha Solteira
.034318	Porto Santa Fé
.034040	Água Vermelha
.052014	Porto São Simão
+005006	Três Lagoas
.034006	Porto Panorama
.034005	Presidente Epitácio
.041055	São Pedro do Paraná
.004132	Porto Guaíra
.041018	Porto Santa Helena
.041000	Foz do Iguaçu

The cost of building a bi-modal intermodal terminal was defined as US$ 5 million, whereas the location of a tri-modal terminal would require a US$ 10 million investment. These figures are comparable to the estimates released by the state government of São Paulo on the future investment opportunities for the Tietê-Paraná valley. As will be seen on the section presenting the results of the MCMND model, the problem was solved for a US$ 60 million budget level, corresponding to the twelve transfer terminals currently present in the real network.

Logit Model Calibration

In order to calibrate the commodity-specific logit functions used by the stochastic UE assignment algorithm embedded in the MCMND model, the same O/D tables containing observed flows for 1986 were utilized for estimating the combined modal split / route assignment parameters. The tables were, first, stratified by mode and commodity, considering only those O/D observations for which the two zone centroids had direct access to both road and rail links. That is, 1,838 observations for the five groups of products were identified for which the respective O/D pair faced both road and rail choices, covering over 60% of the total movement of goods in the region.

The next step was to determine the shortest distances on both road and rail sub-networks between all pairs of the selected 120 centroids, what was done using the shortest path routine available in TransCAD. These distances were, then, used to calculate the shipping costs and times for each O/D pair on both modes, assuming the same values of average speed by mode displayed on Table 5 and commodity-specific values for transport rates and loading/unloading costs and delays. Table 10 presents these fixed unit costs and times by commodity and mode.

The modal split calibration package available in the BTMS suite of programs was used for estimating the utility function coefficients θ_{1p} and θ_{2p}, as specified in equation (4.8). The BTMS calibration program applies a maximum likelihood method in fitting the modal split utility functions to the demand data relative to each commodity class. It requires two input data files containing the observed modal shares for each O/D pair and the corresponding shipping cost and time values on each mode. The results of the calibration program include estimates of the coefficients and a set of summary statistics consisting, among others, of asymptotic standard errors and t-values for each parameter, a likelihood ratio test, and a rho squared measure of the goodness of fit. In addition to the cost and time parameters required by the MCMND model, the BTMS calibration routine is also able to estimate mode-specific constants to reflect attributes not considered in the model, such as reliability of certain modes or probability of loss or damage to the cargo. However, these modal constants are not allowed in the utility functions embedded in the MCMND model formulation.

Table 10

Modal Unit Costs and Times by Commodity Group

Product Group	Mode	Shipping Rate (US$/ton-km)	Loading/ Unloading Fixed Charge (US$/ton)	Loading/ Unloading Fixed Delay (hours)
General Cargo	Road	0.050	3.20	2
	Rail	0.025	1.50	24
Minerals	Road	0.050	3.20	2
	Rail	0.013	0.80	24
Steel Products	Road	0.050	3.20	2
	Rail	0.040	1.50	24
Agricultural	Road	0.050	3.20	2
	Rail	0.012	1.20	24
Liquid Bulk	Road	0.050	3.20	2
	Rail	0.028	-	36

In the first attempt to calibrate the logit coefficients for the five groups of commodities considered in this application, it was observed that, for some product classes, the calibration program produced values with signs opposite to what was expected. Moreover, the corresponding rho squared values for all classes were under 10%. Groups such as agricultural solid cargo and liquid bulk were assigned positive values for both cost and time coefficients, meaning that as the combined shipping impedance measure of a given mode increased, the greater was the probability of the shipment being assigned to that mode.

It was then concluded that the demand data were not in a level of aggregation appropriate for calibrating a logit model of the type required by the MCMND model. Assuming that the unit costs and times used to estimate modal impedance were representative of the values observed in the region, there were two possible reasons for the inconsistency in the parameter estimates. Either the products were not grouped appropriately in homogeneous sets, or some of the O/D observations were not actually facing similar alternative choices. The latter problem is more likely to have happened in the context of this application, since a choice scenario between road and rail was assumed only based on the direct access from the zone centroid to both sub-networks. There were no considerations, for instance, on the existence of the appropriate equipment for handling a specific commodity during loading/unloading activities, or on the proximity of plants and farms to rail stations and storage facilities.

In order to fix the "wrong-sign" problem observed for steel products, agricultural solid, and liquid bulk, it was decided to consider for these three classes only shipments originating at zones for which the total supply of the commodity group in question exceeded 1 million tons in 1986. In addition, zones having access to the rail sub-network but presenting only trucking shipments were disregarded. These restrictions somewhat reduced the number of observations, although not causing a significant drop in the total amount of tons under analysis. Even though the resulting rho squared values were still very low, the results of the calibration procedure using the limited sets of observations (O/D-product triplets), summarized in Table 11, were considered satisfactory within the scope of this pilot application. It is important to emphasize, however, that for real applications of the MCMND model, the logit coefficients should be calibrated with time and cost data collected in the field for a sufficiently large number of O/D pairs, instead of using aggregated demand matrices.

The cost and time coefficients estimated by the BTMS calibration program served also to determine the value of time for each product class. These values were calculated as the ratio of the respective time and cost coefficients, expressed in US$ per hour. They were, then, used to estimate the average value of time on the network to be used by the path construction submodel in the calculation of the combined impedance value of a path, according to equation (4.4). The value of time on the network was defined as the average of each commodity's value of time weighted by the corresponding amount of tons being assigned to the

Table 11

Logit Model Parameter Estimates

Parameter	General Cargo	Minerals	Steel Products	Agricultural Products	Liquid Bulk
COST (θ_{1p})	-0.06756	-0.10705	-0.03060	-0.11053	-0.06459
Cost |t-value|	40.9	142.8	16.8	80.3	36.4
TIME (θ_{2p})	-0.01504	-0.03639	-	-0.02358	-0.00173
Time |t-value|	42.0	137.6	-	52.8	7.2
VT (θ_{2p}/θ_{1p})	0.22	0.34	0.00	0.21	0.03
# obs. (O/D-p)	351	325	220	206	269
Total Tons. (10^3)	61,510	302,254	54,595	77,243	159,849
rho squared	0.02	0.05	0.01	0.15	0.06
-2 log likelihood	1882.1	22294.2	284.5	15481.6	12545.5

network. Therefore, the total amount of tons considered in the calculation of the value of time related to the volumes actually being assigned to the network, which were different from the ones used during the calibration procedure. A final value of time equal to US$ 0.12 per hour was adopted for this application.

Running the MCMND Model

After defining the input parameters described in the last two sections, the multimodal physical network comprising 895 nodes and 1,465 physical links was expanded into a logical network consisting of 1,429 logical nodes and 4,362 logical and physical arcs. As indicated before, the 1986 O/D matrices obtained for the region under analysis contained a total of 170 zones, with a corresponding number of demand channels equal to 1,662. In order to simplify the solution of the hypothetical transfer terminal location problem, the number of channels was significantly reduced to embrace only the flows that had a real potential for using the waterway system. Moreover, only demand channels with flows amounting to more than 500,000 tons were considered. In the end, a total of 40 channels with flows corresponding to 75% of the tonnage capable of being shipped on barges, and around 7% of the total movement of goods in the region, was selected. The majority of these flows related to liquid bulk (62%) and agricultural solid products (20%), matching the projected major flows of commodities to use the waterway system anticipated by the planning authorities in Brazil (see Figure 8).

The solution algorithm was coded in C and compiled using the Sun C compiler. Due to insufficient memory requirements, this application could not be solved using a PC micro-computer. Instead, a Sun SPARCstation 10 was used to solve the problem. As mentioned before, a US$ 60 million investment budget was selected, corresponding to the location of a maximum of 12 facilities. This is the actual number of existing transfer terminals along the two rivers.

The maximum number of major iterations was set to 10. The parameter γ was set to 1, in order to allow the construction of as many multimodal paths as possible requiring the location of transfer facilities. Again, the average value of time (VT) for the five groups of commodities equaled US$ 0.12/hour and it was determined based only on the flows actually being assigned to the network. The convergence criterion for the value of the objective function was defined as equal to or less than US$ 5 million, establishing the threshold where the cost of locating an additional terminal would offset the corresponding reduction in the total cost of moving freight on the network.

Model Results

Table 12 shows the results of the MCMND model, iteration by iteration. The best solution to the problem was found after 6 major iterations, which took approximately 48 hours of CPU time in the Sun SPARC 10 workstation. The model selected only 3 logical arcs representing water-land intermodal transfer

Table 12

MCMND Results for the Pilot Application

Major Iter.	# of Paths (Total)	# of Paths Used	Obj. Value before Subs. (US$ 10^3)	Obj. Value after Subs. (US$ 10^3)	# of Subs.	# of Minor Iter.	# of River Ports Located
0	78	0	14,990,498.89	-	-	-	-
1	116	114	3,969,101.45	3,798,273.87	2	5	-
2	149	148	3,648,823.58	3,517,034.52	1	2	-
3	167	164	3,449,692.18	3,185,916.23	3	9	-
4	182	178	3,382,521.90	3,139,505.92	5	15	2
5	194	183	3,264,486.88	2,890,024.22	17	113	3
6	205	186	3,296,445.83	2,885,620.58	24	259	3

terminals. A bi-modal terminal was located at the zone centroid representing the town of Três Lagoas (MS), between water and highway. In addition, a tri-modal terminal consisting of transfer movements between water and rail, and water and highway, was located at the node representing the town of Piracicaba (SP). It is interesting to note that a tri-modal terminal is currently under construction in Piracicaba (Artêmis Terminal), which is expected to handle the largest amount of commodities among all terminals in the Tietê river. Três Lagoas, in turn, is also foreseen as one of the future major transshipment points in the waterway system, once the lockage facility in the Jupiá dam is completed by 1995. As discussed in the section describing the study area, the Jupiá lock will allow the connection between the Tietê river and the south stretch of the Paraná river, linking the Brazilian waterway system to the Paraguay river.

The selection of only 3 land-water movements, leaving a US$ 45 million slack in the investment budget, was mainly a consequence of the high delays experienced at river ports in Brazil. Each loading or unloading activity between trucks and barges, or trains and barges, is currently taking around 24 hours to be processed. Since the main production and consumption centers in the region under analysis are not located along the waterway system, at least a 48-hour delay is experienced for multimodal shipments using the waterway. Considering an average shipping distance of approximately 400 km, transfer delays offset the lower shipping rate advantage of barges against rail and road. This is actually one of the main reasons why waterway traffic accounts for only 2% of the freight movements in that region.

The flow assignment corresponding to the best solution found by the MCMND model (iteration 6) was taken back to TransCAD for an analysis of the resulting movement of goods over the network. The statistical routines available in TransCAD allowed an easy assessment of the total amount of goods shipped by each mode, the corresponding modal shares, and the total number of ton-km moved on the network. These and other figures are summarized in Table 13.

Those same summary statistics are also provided in Table 14 for the flow assignment resulting from iteration 3. The solution at the end of iteration 3 corresponded to the lowest value of the objective function without the location of transfer terminals along the waterway. In other words, it represented the best flow assignment considering only the road and rail sub-networks. This network configuration, thus, can be seen as the situation in the base year for the unimproved (base) network.

The comparison of the network flow scenarios at the end of iterations 3 and 6 provides a clear indication of the benefits accruing from the construction of land-water transshipment facilities and the consequent use of the inland waterway system. The inclusion of water shipments reduced the total generalized cost of moving freight on the network by almost 10%. This reduction reflected a total savings to shippers of approximately US$ 300 million. The respective savings in fuel consumption were around US$ 1 million, expressed by the difference in the

Table 13

Summary Statistics of the Flow Assignment from Iteration 6

Variables	Highway Sub-network	Railway Sub-network	Waterway Sub-network	Multimodal Network
Total Flow (1000 tons)	878,877.83	187,482.87	18,543.84	1,084,904.54
Avg. Flow per Link (1000 tons)	3,739.91	1,575.49	1,545.32	2,964.22
Total Length (km)	7,225.61	6,776.17	463.20	14,464.98
Avg. Length per Link (km)	30.75	56.94	38.60	39.52
# of Loaded Links	235	119	12	366
Total Ton-km (1000)	20,749,778.36	9,864,533.98	715,792.20	31,330,104.54
Modal Split (%)	66.2	31.5	2.3	100.0
System Impact Cost (US$ 1000)	207,497.78	19,729.07	-	227,226.85

- Objective Function Value = US$ 2,885,620,580.00
- Average Shipping Cost on the Network = US$ 0.092/ton-km

Table 14

Summary Statistics of the Flow Assignment from Iteration 3

Variables	Highway Sub-network	Railway Sub-network	Multimodal Network
Total Flow (1000 tons)	879,563.89	189,524.96	1,069,088.85
Avg. Flow per Link (1000 tons)	3,807.64	1,633.84	3,080.95
Total Length (km)	7,316.84	6,617.75	13,934.59
Avg. Length per Link (km)	31.68	57.05	40.16
# of Loaded Links	231	116	347
Total Ton-km	20,845,517.62	9,856,487.19	30,702,004.82
Modal Split (%)	67.9	32.1	100.0
System Impact Cost (US$ 1000)	208,455.18	19,712.97	228,168.15

- Objective Function Value = US$ 3,185,916,230.00
- Average Shipping Cost on the Network = US$ 0.104/ton-km

values of total system impact costs for the two cases. When compared to the US$ 15 million invested on the construction of the two transshipment terminals along the waterway, the benefit/cost ratio of the selected investments is clearly in favor of their implementation.

One final aspect that deserves to be emphasized relates to the fact that the MCMND model was able to provide percentages of modal split between road and rail somewhat close to the values observed in the study area for the commodities being considered. As it can be seen in Table 14, the flow assignment at the end of iteration 3 resulted in modal shares that approximate the values of 76% and 24% for road and rail, respectively, which were calculated based on the portion of the 1986 O/D data used in this application. Considering that the logit function coefficients were estimated based on very aggregated data, as discussed in the calibration section, these results indicate that the MCMND model was able to accurately simulate the flow of commodities over the Brazilian network. Another indication of the simulating capability of the model is the percentage of ton-km assigned to the waterway routes in the best solution (Table 13), which was equal to 2.3%. This value corresponds exactly to the current share of commodity flows being shipped on barges in the Tietê-Paraná waterway system.

Final Comments

As emphasized in the introductory section of this chapter, the purpose of developing this pilot application of the MCMND model was mainly to test its efficiency in solving large network problems. The hypothetical problem of locating intermodal transfer terminals along the Tietê and Paraná rivers should be seen as just an illustration of the types of problems that can be addressed by the MCMND model. The results presented in the last section must be evaluated considering the limitations of the data gathered for running the model with the Brazilian network. Despite the academic nature of this prototype application, there is a clear indication that the model is able to accurately simulate the flow of commodities on a real-life large network.

The development of the Tietê-Paraná valley application also served to illustrate the steps involved in collecting the necessary data for applying the MCMND model and calibrating the model's parameters. In addition, this application proved the advantages of using a GIS package in connection with the model to store and modify the network information, as well as to analyze and display the obtained results.

CHAPTER VII

SUMMARY AND CONCLUSIONS

Summary

The primary purpose of this research was the development of a mathematical model to be used in the selection of investment strategies for a regional multimodal freight transportation network. The model was intended to provide useful information for infrastructure planning at a strategic level. It would be especially useful for the cases in which the number of investment options are large enough to prevent the use of classical economic analysis for comparing individual investment alternatives.

This research achieved its objectives through the formulation and solution of the *Multicommodity Multimodal Network Design (MCMND)* model presented in this dissertation. As described in Chapter III, the MCMND model is formulated as a nonlinear bilevel mixed-integer programming problem in which a combined network design / facility location problem is subject to a set of user optimal flows assigned stochastically to the network. In other words, the model's mathematical formulation consists of an optimization problem to search for the network configuration that would result in the lowest total transportation costs, assuming that shipments are assigned to the network according to a stochastic user equilibrium criterion. The changes to a network that can be handled by the MCMND model include link improvements or additions, and the location of intermodal transshipment terminals. The model also considers explicitly the effect on shippers' decisions of the delays caused by congestion on links and transfer terminals. Total transportation costs are defined as a combination of the expenses incurred by shippers, both in terms or shipping rates and delays, plus the monetary value of the environmental impacts caused to the system by moving freight on less efficient modes of transportation. These environmental impacts are characterized as the net increase in fuel consumption, emissions, and accident rates resulting from the use of trucks and rail as opposed to barges.

As discussed in Chapter IV, due to the combinatorial nature of the design problem and its bilevel structure, the exact solution of the MCMND model is possible only for very small, trivial, networks. Therefore, this research also proposed the adoption of a heuristic procedure to search for a near-optimal solution for real-world applications of the model. This procedure is based on a decomposition / column generation technique in which new paths, requiring the selection of some investment projects, are generated by a path construction submodel and submitted to the restricted master problem to be considered for inclusion in the final solution. The restricted master problem is then solved by a two-step greedy-swapping algorithm that, departing from an initial feasible solution consisting of a set of in-solution paths, would swap paths between the out-of-solution and the in-solution sets until no further improvement to the value of the objective function can be attained. The advantages of this heuristic algorithm depend on three main assumptions: first, that the path construction submodel will be able to generate new paths quickly and efficiently; secondly, that the submodel will construct only the paths which are potential candidates to enter into the solution; and, finally, that the algorithm will reach its best solution after the generation of only a small subset of all possible paths on the network.

In Chapter V, the heuristic solution algorithm embedded in the MCMND model was applied to a test problem comprising the location of transfer terminals in a small transportation network. The results of the MCMND model for this problem were, then, compared against those obtained by using the BTMS model for different configurations of the test network. It was found that the solution provided by the MCMND model resulted in an objective function value surprisingly close to the lowest value observed among the scenarios considered by the BTMS. Although the number and locations of facilities selected by the MCMND model did not match exactly with those suggested by the BTMS, it was clear that both models progressed in the same optimal direction. Most importantly, the MCMND model was able to provide a good solution to the test problem in a short period of time, generating only a small number of candidate paths.

The MCMND model was also applied to a real-world problem of locating river ports along the Tietê-Paraná waterway in the Southeast region of Brazil, as described in Chapter VI. The purpose of this application was mainly to prove that the model could be used to solve real-life large multimodal transportation networks in a reasonable amount of time. There were no intentions of using the results of the model for suggesting policies in the Brazilian case, since the appropriate data for this type of analysis were not available. The analysis of the Brazilian case was also carried out to illustrate the data collection and calibration steps involved in a real application of the MCMND model. In addition, it served to emphasize the importance of a GIS interface to enhance the analytical capability of the model.

Major Contributions

This research has made several important contributions to the broad field of network optimization. The formulation of the MCMND model is a contribution as it will enable planners/analysts to address a transshipment facility location problem within a multimodal network design structure. The model enables planners to examine the trade-offs between building new intermodal transfer facilities and investing on some of the physical links of the multimodal network. Furthermore, the MCMND model explicitly considers the real phenomenon of congestion on links and terminals and its impact on the freight flow pattern over a multimodal network. It is possible, thus, to identify capacity bottlenecks at some points on the network and to choose the most effective infrastructure investment strategies to increase the overall performance of the system.

The MCMND model also introduces a new approach for developing and comparing investment alternatives on multimodal networks. As discussed in the literature review chapter, previously developed network models from the perspective of applications to real world problems fall into two distinct categories. They are either network analysis models developed to simulate the flow of commodities over a given network, or they can be in the category of network synthesis (design) models which seek for an optimal network configuration to satisfy a given set of requirements. The models of the first category, such as the BTMS and the STAN system, have to rely on the knowledge and experience of the transportation analyst who defines a small set of alternative investment scenarios and, then, carries out an economic evaluation analysis based on the performance measures resulting from the simulation of these selected alternative network configurations. Therefore, these models are restricted to the small set of investments initially defined by the planner, and the effectiveness of the approach relies on the experience of the planner.

The second type of models, on the other hand, is able to handle a much larger set of candidate investment options, including both discrete and continuous investment choices. These models, formulated as conventional network design problems, are very computationally intensive and cannot be solved completely for large scale networks. Consequently, an initial step of preselecting good candidate projects has to be performed even before a heuristic solution procedure can take place. Interesting examples of such models can be found in the works of Rothengatter (1986) and Janson, Buckels, and Peterson (1991).

The MCMND model falls into this second category. However, in contrast to the existing multimodal network design models, its solution algorithm is able to handle all possible investment options, without the need for a preselection step. The heuristic procedure decides by itself which candidate investment projects should be part of the optimal flow scenario while searching for new shortest paths in the network. Therefore, the decisions on selecting certain investment options are made considering simultaneously their effect on the shipping behavior of the users of the system (shippers/carriers).

This interaction between the central planning entity and the users of the transportation system, reflected in the bilevel formulation of the MCMND model, appears to be one of its major strengths. Moreover, the formulation also assumes that users do not have a full knowledge of all aspects influencing their shipping decisions and, thus, encompasses a probabilistic approach for modeling route/mode choices. The implementation of a logit choice model within the network design problem, while searching for a flow equilibrium state in a capacitated network, is believed to be a step forward towards more realistic solutions.

Another important innovation introduced by the MCMND model is the adaptation of the column generating technique to the solution of the multimodal network design problem. As reviewed in previous chapters, such a technique was initially developed to solve linear programming problems. It was then adapted by Cullen, Jarvis, and Ratliff (1981) into a heuristic procedure to address a specific class of discrete problems, that is, vehicle routing problems on an uncapacitated network. This research, thus, was innovative by proposing a specialized version of this heuristic procedure able to solve a discrete choice network design problem coupled with a stochastic user equilibrium assignment problem.

Although the heuristic procedure presented in Chapter IV is not guaranteed to find the optimal solution to the model, it does allow the solution of large scale problems in a reasonable time. Since the MCMND model was not designed to serve as the unique analytical tool during the decision-making process, the option for a heuristic approach should not be seen as a compromise. The scope of the model makes possible a quick analysis of the types of solutions resulting from different budget levels. Thus, analysts have the opportunity of comparing competitive solutions in a marginal benefit type of analysis, considering the marginal impact of each additional amount of money on network efficiency. In addition to this sensitivity analysis capability, the results suggested by the MCMND model may also be considered in more detailed studies using individual project evaluation techniques. After the selection by the MCMND model of the very few good candidate projects, these top-ranked projects may, for instance, be used for defining several investment scenarios to be considered by a transportation analysis model such as the BTMS. Of course, before setting their final recommendations, analysts must also submit the results of any quantitative deterministic model to a broader qualitative set of social-political criteria. Therefore, in reality, optimal solutions resulting from exact mathematical models may lose a great deal of importance within the overall decision-making process.

Future Extensions

The formulation and solution procedure of the MCMND model have several features that deserve to be extended in the future towards a more realistic modeling approach. Perhaps the most critical limitation of this model is its static nature reflected in a single-period planning assumption. Investment strategies are defined for a fixed planning horizon and analyzed assuming the same traffic patterns observed in the base year.

There are two problems with this type of formulation. First, it considers neither the temporal nor the spatial interdependencies among investment projects. That is, the net benefit of selecting a project is not dependent on the inclusion or not of another project planned for another link or for the same link at a different point in time. Secondly, the impacts of adding or deleting links and facilities on the generation and distribution of freight flows over the network, although clearly influential, are disregarded by the model. Production and consumption levels are assumed to vary throughout the planning period according to an estimated fixed rate of economic expansion or contraction. Origins and destinations of commodity flows, in turn, remain the same as those observed in the base year. Undoubtedly, a multi-period, dynamic approach, with variable O/D tables, would considerably enhance the analytical capability of the model through a better representation of reality. It is important to remember, however, that the model was designed for strategic planning purposes. Consequently, the level of effort required to carry out an application of the model should be kept compatible to the decisions made at the strategic stage of planning. Marginal increases in accuracy may not be worthwhile when preceded by a much longer period of data collection, parameter calibration and model implementation.

Another limitation imposed by the formulation of the MCMND model involves the calculation of shipment delays due to congestion on links and transfer terminals. The model assumes similar delay functions for all links of the network. The consideration of link-specific congestion functions would certainly add more realism to the model, although significantly increasing the burden on data collection and calibration.

One simplistic assumption embedded in the model that can be relaxed easily is the assignment of freight flows in tons per unit of time. To better represent the flow of commodities on the physical network, tonnages should be converted to the corresponding number of vehicles using the most representative types of trucks, train cars, and barges being utilized in the system under analysis. The use of vehicles, instead of tons, would also allow the development of more adequate link impedance functions. In addition, it would be possible to consider the movement of empty vehicles, known as "dead-heading", within the model's structure.

Passenger movements could also be converted into vehicles and taken into consideration by an expanded version of the MCMND model.

Another aspect of the model that should be considered in future works relates to the calculation of the environmental impacts resulting from moving freight on trucks and rail as opposed to barges. Currently, a very simplistic approach has been adopted which estimates the monetary value, per ton-km, of the extra liters of fuel consumed, the increase in pollutant emissions, and the additional number of accidents and fatalities occurring as a consequence of not shifting flows to the most efficient modes of transportation. This research did not go further in developing a more detailed methodology for estimating these impacts especially because the main purpose of introducing these costs in the objective function was to call the attention to the importance of considering such factors in the search for the optimal system configuration. Certainly, other factors also should be taken into consideration when calculating the system impact costs in the network. A good example is the net increase in maintenance costs accruing from a more intensive use of the rail and, especially, the road network. As in the case of Brazil, most developing countries have been unable to properly maintain their road networks due to rapidly deteriorating pavement conditions. Therefore, diverting freight traffic out of the roadway system to reduce its maintenance costs may be a crucial factor when planning investments to a multimodal network. This and other types of system costs can be incorporated into the model's framework easily.

A final suggestion for future extensions of the MCMND model refers to the development of a more comprehensive graphic interface between the model's solution algorithm and a GIS software package. A user-friendly interface allowing users to graphically input the parameters necessary to run the model, such as the set of candidate investment projects, would certainly make the model more attractive to planning agencies in charge of defining investment priorities. In fact, it would be ideal to incorporate the model into a true spatial decision support system (SDSS) in which an assortment of different transportation analysis tools would be available to help planners and analysts at the various stages of the planning process.

BIBLIOGRAPHY

BIBLIOGRAPHY

Berman, O., Einav, D., and Handler, G. (1990). "The constrained bottleneck problem in networks", *Operations Research*, 38, 178-181.

Boyce, D.E. and Janson, B.N. (1980). "A discrete transportation network design problem with combined trip distribution and assignment". *Transportation Research*, 14(B), 147-154.

Boyce, D.E. and Soberanes, J. (1979). "Solutions to the optimal network design problem with shipments relating to transportation cost", *Transportation Research*, 13(B), 65-80.

Boyce, D.E., Farhi, A., and Weischedel, R. (1973). "Optimal network problem: a branch and bound algorithm", *Environment and Planning*, 5, 519-533.

Braess, D. (1968). "Uber ein Paradox der Verkehrsplanung", *Unternehmensforschung*, 12, 258-268.

Bronzini, M.S. (1980). "Evolution of a multimodal freight transportation network model", in *Proceedings of the 21st Annual Meeting, Transportation Research Forum*, 475-485.

Caliper Corporation (1988). *Transcad - Transportation Workstation Software - Reference Manual.*

Crainic, T.G. (1987). "Operations research models of intercity freight transportation: the current state and future research issues", *Logistics Transportation Review*, 23, 189-206.

Cullen, F.H., Jarvis, J.J., and Ratliff, H.D. (1981). "Set partitioning based heuristics for interactive routing", *Networks*, 11, 125-143.

Current, J.R. (1988). "The design of a hierarchical transportation network with transshipment facilities", *Transportation Science*, 22(4), 270-277.

Current, J.R. and Min, H. (1986). "Multiobjective design of transportation networks: taxonomy and annotation", *European Journal of Operational Research*, 26, 187-201.

Current, J.R., ReVelle, C., and Cohon, J. (1985). "The application of location models to the multiobjective design of transportation networks", *Regional Science Review*, 14.

Current, J.R., ReVelle, C.S, and Cohon, J.L. (1986). "The hierarchical network design problem", *European Journal of Operational Research*, 27, 57-66.

Daganzo, C.F. and Sheffi, Y. (1977). "On stochastic models of traffic assignment", *Transportation Science*, 11(3), 253-274.

Dantzig, G.B., Harvey, R., Robinson, D., Lansdowne, Z. and Maier, S.F. (1979). "Formulating and solving the network design problem by decomposition", *Transportation Research B*, 13(B), 5-17.

Densham, P.J. and Rushton, G. (1988). "Decision support systems for locational planning", in R. Golledge and H. Timmermans eds., *Behavioural Modeling in Geography and Planning*, Croom-Helm, London, 56-90.

Desrosiers, J., Soumis, F., and Desrochers, M. (1984). "Routing with time windows by column generation", *Networks*, 14, 545-565.

Dial, R.B. (1971). “A probabilistic multipath traffic assignment model which obviates path enumeration”, *Transportation Research*, 5, 83-111.

Dijkstra, E.W. (1959). "A note on two problems in connection with graphs", *Numerische Mathematik*, 1, 269-271.

Dionne, R. and Florian, M. (1979). "Exact and approximate algorithms for optimal network design". *Networks*, 9(1), 37-59.

Elshaygi, G.H. (1983). *A Road Investment Model for the Sudan*, Ph.D. Dissertation, Department of Geography, University of Tennessee, Knoxville, Tennessee.

Fisk, C. (1980). “Some developments in equilibrium traffic assignment”, *Transportation Research*, 14(B), 243-255.

Florian, M. and Crainic, T.G. (1989), "Strategic Planning of Freight Transportation in Brazil: Methodology and Applications", Final Report, 5 volumes, Publications #638-642, Centre de recherche sur les transports, Universite de Montreal.

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

Friesz, T.L., Gottfried, J.A., and Morlok, E.K. (1986). "A sequential shipper-carrier network model for predicting freight flows", *Transportation Science*, 20, 80-91.

Friesz, T.L., Tobin, R.L., and Harker, P.T. (1983). "Predictive intercity freight network models", *Transportation Research*, 17(A), 409-417.

Goldman, A.J. (1969). "Optimum location for centers in a network", *Transportation Science*, 3, 352-360.

Guelat, J., Florian, M., and Crainic, T.G. (1990). "A multimode multiproduct network assignment model for strategic planning of freight flows", *Transportation Science*, 24, 25-39.

Hakimi, S.L. (1964). "Optimal locations of switching centers and the absolute centers and medians of a graph", *Operations Research*, 12, 450-459.

Hakimi, S.L. and Maheshwari, S.N. (1972). "Optimum location of centers in networks", *Operations Research*, 20, 967-973.

Harker, P.T. and Friesz, T.L. (1986). "Prediction of intercity freight flows I & II: Theory and Mathematical formulation", *Transportation Research*, 20(B), 139-174.

Hoang, H.H. (1973). "A computational approach to the solution of an optimal network", *Management Science*, 19(5), 488-498.

Hu, J. (1992). *SDSSLF: A Spatial Decision Support System for Locating Landfills*, Ph.D. Dissertation, Department of Geography, University of Tennessee, Knoxville, Tennessee.

Janson, B.N. and Husaini, A. (1987). "Heuristic ranking and selection procedures for network design problems", *Journal of Advanced Transportation*, 21(1), 17-46.

Janson, B.N., Buckels, L.S., and Peterson, B.E. (1991). "Network design programming of U.S. highway improvements", *Journal of Transportation Engineering*, 117(4), 457-479.

Johnson, D.S., Lenstra, J.K., and Rinnooy Kan, A.H.G. (1978). "The complexity of the network design problem", *Networks*, 8, 279-285.

Jones, P.S. and Sharp, G.P. (1977). "Multi-mode intercity freight transportation planning for underdeveloped regions", *in Proceedings of the 18th meeting of the Transportation Research Forum*, 523-531.

Kim, T.J., Park, C., and Rho, J.H. (1985). "Modeling investment priorities for national road improvements: a case study of Korea", *Papers of the Regional Science Association*, 57, 91-105.

Kresge, D.T. and Roberts, P.O. (1971). *Systems Analysis and Simulation Models*, in *Techniques of Transport Planning*, vol.2, John R. Meyer (ed.), The Brookings Institution, Washington, D.C..

LeBlanc, L.J. (1975). "An algorithm for the discrete network design problem", *Transportation Science*, 9(2), 183-199.

LeBlanc, L.J. and Boyce, D.E. (1986). "A bi-level programming algorithm for exact solution to the network design problem with user optimal flows", *Transportation Research B*, 20(3), 259-265.

Lewis, S., Cook, P., and Minc, M. (1990). "Comprehensive transportation models: past, present and future", *Transportation Quarterly*, 44, 249-265.

Lin, S. and Kernighan, B. (1973). "An effective heuristic algorithm for the Traveling Salesman Problem", *Operations Research*, 21, 498-516.

Los, M. and Lardinois, C. (1980). "Combinatorial programming, statistical optimization and the optimal transportation network problem", *Transportation Research*, 16(B), 89-124.

Magnanti, T.L. and Wong, R.T. (1984). "Network design and transportation planning models and algorithms". *Transportation Science*, 18(1), 1-56.

Minoux, M. (1989). "Network synthesis and optimum network design problems: models, solution methods and applications", *Networks*, 19, 313-360.

Mirchandani, P.B. and Odoni, A.R. (1979). "Locating new passenger facilities on a transportation network", *Transportation Research*, 13(B), 113-122.

Newstrand, M.W. (1992). “Environmental Impacts of a Modal Shift”, *Transportation Research Record*, 1333, 9-12.

Ogryczak, W., Studzinski, K., and Zorychta, K. (1992). "Computer-assisted analysis system for multiobjective transshipment problems with facility location", *Computers & Operations Research*, 19(7), 637-647.

Pirkul, H., Current, J.R., and Nagarajan, V. (1991). "The hierarchical network design problem: a new formulation and solution procedures", *Transportation Science*, 25(3), 175-182.

Poorzahedy, H. and Turnquist, M.A. (1982). "Approximate algorithms for the discrete network design problem", *Transportation Research B*, 16(1), 45-55.

Powell, W.B. and Sheffi, Y. (1982). "The convergence of equilibrium algorithms with predetermined step sizes", *Transportation Science*, 16(1), 45-55.

Ralston, B.A. and Liu, C. (1992). *The Bangladesh Transportation Modeling System*, World Bank, Washington, D.C.

Ralston, B.A., Tharakan, G., and Liu, C. (1994). "A spatial decision support system for transportation policy analysis", *Journal of Transport Geography*, 2(2), 101-110.

Ray, J.J. (1990). *The Flow-resource Facility Location Problem*, Ph.D. Dissertation, Department of Geography, University of Tennessee, Knoxville, Tennessee.

Revelle, C.S. and Swain, R.W. (1970). "Central Facilities Location", *Geographical Analysis*, 2(1), 30-42.

Rothengatter, W. (1979). "Application of optimal subset selection to problems of design and scheduling on urban transportation networks", *Transportation Research*, 13(B), 49-63.

Rothengatter, W. (1986). “Scheduling of interstate road and railway investments”, *Environment and Planning A*, 18, 465-483.

Samuelson, P.A. (1952). “Spatial price equilibrium and linear programming”, *Amer.Econ.Rev.*, 42, 283-303.

Sharp, G.P. (1979). "A multi-commodity intermodal transportation model", in *Proceedings of the 20th Annual Meeting, Transportation Research Forum*.

Tansel, B.C., Francis, R.L., and Lowe, T.L. (1983). "Location on networks: a survey; part I. The p-center and p-median problems; part II. Exploiting tree network structure". *Management Science*, 29, 482-511.

Teitz, M.B. and Bart, P. (1968). "Heuristic methods for estimating the generalized vertex median of a weighted graph", *Operations Research*, 16, 955-961.

Wardrop, J.G. (1952). “Some theoretical aspects of road traffic research”, *Proc. Inst. Civil Engineers*, Part II, 1, 325-378.

Wendell, R.E. and Hurter, A.P. (1973). "Optimal locations on a network", *Transportation Science*, 7, 18-23.

Xiong, Y. and Schneider, J.B. (1992). "Transportation network design using a cumulative genetic algorithm and a neural network", *paper presented at the 71st Annual meeting of the Transportation Research Board*, paper no. 920685.

APPENDIX

Table A-1

Tietê-Paraná Application - BTMS Results

(Combinations of Nodes)

Case	Description	Transfer Facilities Allowed	Obj.Function (US$ 10^3)
1	no transfers	-	209,264.91
2	transfers at 0	r0h, r0w, w0h, w0r	209,264.91
3	transfers at 3	h3r, (h3w), r3h, (r3w), w3h, w3r	185,373.10
4	transfers at 4	h4r, r4h	209,264.91
5	transfers at 5	(h5r), h5w, r5h, r5w	185,638.70
6	0 & 3	r0h, r0w, (w0h), w0r, (h3r), (h3w), r3h, r3w, w3h, w3r	181,645.97
7	0 & 4	r0h, r0w, w0h, w0r, h4r, r4h	209,264.91
8	0 & 5	r0h, r0w, w0h, w0r,(h5r), h5w, r5h, r5w	185,638.70
9	3 & 4	h3r, (h3w), r3h, r3w, w3h, w3r, h4r, r4h	182,794.18
10	3 & 5	(h3r), (h3w), r3h, (r3w), w3h, w3r, (h5r), (h5w), r5h, r5w	170,302.30
11	4 & 5	h4r, r4h, (h5r), h5w, r5h, r5w	185,638.70
12	0, 3 & 4	r0h, r0w, (w0h), w0r, (h3r), (h3w), r3h, r3w, w3h, w3r, h4r, r4h	181,645.97
13	0, 3 & 5	r0h, r0w, w0h, w0r, h3r, (h3w), r3h, r3w, w3h, w3r, (h5r), (h5w), r5h, r5w	170,200.15
14	0, 4 & 5	r0h, r0w, w0h, w0r, h4r, r4h, (h5r), h5w, r5h, r5w	185,638.70
15	3, 4 & 5	h3r, (h3w), r3h, r3w, w3h, w3r, h4r, r4h, (h5r), (h5w), r5h, r5w	170,200.15
16	0, 3, 4 & 5	r0h, r0w, (w0h), w0r, (h3r), (h3w), r3h, r3w, w3h, w3r, h4r, r4h, (h5r), (h5w), r5h, r5w	169,561.21

obs.: The facilities in parentheses represent the ones which had flows assigned to them.

Table A-2

Tietê-Paraná Application - BTMS Results

(Combinations of the 5 Selected Facilities)

Case	Transfer Facilities Allowed	Obj.Function (US$ 10^3)
1	-	209,264.91
2	(h3r)	193,312.06
3	(h3w)	182,794.18
4	(h5r)	185,638.70
5	h5w	209,264.91
6	w0h	209,267.91
7	h3r, (h3w)	182,794.18
8	(h3r), (h5r)	182,736.45
9	(h3r), h5w	193,312.06
10	(h3r), w0h	193,312.06
11	(h3w), (h5r)	170,854.54
12	(h3w), (h5w)	178,629.10
13	(h3w), (w0h)	181,640.73
14	(h5r), h5w	185,638.70
15	(h5r), w0h	185,638.70
16	h5w, w0h	209,264.91
17	h3r, (h3w), (h5r)	170,854.54
18	h3r, (h3w), (h5w)	178,629.10
19	(h3r), (h3w), (w0h)	181,645.97
20	(h3w), (h5r), (h5w)	170,200.15
21	(h3w), (h5r), w0h	170,854.54
22	(h5r), h5w, w0h	185,638.70
23	(h3r), (h5r), h5w	182,736.45
24	(h3r), (h5r), w0h	182,736.45
25	(h3w), (h5w), w0h	178,629.10
26	(h3r), h5w, w0h	193,312.06
27	h3r, (h3w), (h5r), (h5w)	170,200.15
28	h3r, (h3w), (h5r), w0h	170,854.54
29	(h3r), (h5r), h5w, (w0h)	183,671.28
30	(h3r), (h3w), (h5w), (w0h)	178,635.22
31	(h3w), (h5r), (h5w), w0h	170,200.15
32	h3r, (h3w), (h5r), (h5w), w0h	170,200.15

obs.: The facilities in parentheses represent the ones which had flows assigned to them.

VITA

Carlos Felipe Grangeiro Loureiro was born in Fortaleza, Brazil on March 28, 1968. He attended elementary and high schools in his hometown and graduated from high school in December 1984. The following March, he entered the Universidade Federal do Ceará, Brazil and in February 1990 received the B.S. degree in Civil Engineering.

In August 1990, he came to Knoxville, TN to start his graduate studies at The University of Tennessee. He received a M.S. in Civil Engineering in December 1991 and the Doctor of Philosophy degree in Civil Engineering in August 1994, both with concentration in Transportation Engineering. During his Master's program, Mr. Loureiro was sponsored by a scholarship awarded by The Rotary Foundation. For his Ph.D. program, he was awarded a three-year grant by the Federal Government of Brazil, through CAPES - Federal Agency for Post-Graduate Education.

Mr. Loureiro is married to Lídia Amélia de Barros Cardoso. Following graduation, he will go back to his hometown in Brazil where he intends to become a professor at the same university from which he received his B.S..