

**A GENERAL MIXED-INTEGER NONLINEAR OPTIMIZATION  
MODEL FOR HUB NETWORK DESIGN**

by

Jean-Michel Guldmann  
Department of City and Regional Planning  
The Ohio State University  
Columbus, Ohio 43210, USA

and

Guoqiang Shen  
Department of Planning  
East Carolina University  
Greenville, North Carolina 27858, USA

Paper prepared for presentation at the 44th North American meeting of the Regional Science Association International, November 6-9, Buffalo, New York.

Correspondence to: Jean-Michel Guldmann, Department of City and Regional Planning,

The Ohio State University, 190 West 17th Avenue, Columbus, Ohio 43210. Fax: 614-292-7106.  
Phone: 614-292-2257. E-mail:Guldmann.1@osu.edu.

## **ABSTRACT**

A general discrete hub network model that accounts for fixed, capacity, and operating/congestion costs on links and at hubs, with both economies and diseconomies of scale, selects hubs and links, determines their capacities, and assigns O-D flows over paths, while minimizing all system costs. Initially formulated as a mixed-integer non-linear program, the model is transformed into a mixed-integer linear program through the linearization of the capacity and congestion cost functions. The methodology is illustrated by an application to a small-scale network with hypothetical data. Extensive sensitivity analyses are carried out to assess the trade-offs between the different link and hub costs.

## 1. INTRODUCTION

Hub networks, where hubs act as switching or transshipping points for flows between origins and destinations (O-D) and spokes connect O-Ds to the hubs, can generate large economies of scale, and therefore have been widely adopted by air and surface transportation systems, in which passengers, parcels, or cargoes carried by airplanes and vehicles are often transshipped at one or more major airports/stations, and computer and telecommunication systems, in which digital or analog information from and to individual computers or telephones is switched by satellites, computer servers, or telephone exchange stations.

Various optimization models have been developed for the design of such networks under various topological assumptions, including (1) single allocation, where each non-hub node is assigned to one hub only, (2) multiple allocation, where a non-hub node may be linked to several hubs, implying several paths between O-D pairs, (3) pure hub network, with no links between non-hub nodes, etc. Most of these models consider pure networks with single allocation, focus on link costs, which are taken proportional to link flows, while using a multiplicative discount factor for hub-hub links to reflect economies of scale due to flow concentration, and select hubs and links that minimize total system costs while allowing for all O-D flows to be carried over the network. Most of this research has primarily focused on the development of algorithms and heuristics to solve these NP-hard problems for networks of realistic sizes, and has given little attention to the following issues: (1) The determination of link and hub capacities (heretofore assumed unlimited), accounting for related fixed and variable costs and allowing for varying levels of economies of scale, must be part of the design of hub systems, and capacities should become endogenous decision variables; and (2) Congestion effects on links and hubs resulting from the interactions between capacities and flows, and the resulting costs incurred by the hub system operator and its users (time delays, failures, increased maintenance), must also be considered when designing hub systems.

The purpose of this paper is to develop a general hub network model that accounts for the above issues, and places as few as possible a priori restrictions on the design of the system, thus allowing for multiple hub allocations, non-hub direct links, and multiple paths between any O-D pair. For given sets of nodes and potential links, and given O-D flows between any pair of these nodes, the model selects the nodes to serve as hubs and the links to be built, determines their capacities, and allocates O-D flows over paths made of the selected hubs and links, while minimizing the sum of all hub and link fixed costs, variable capacity costs, and variable operating and congestion costs, subject to various flow conservation and capacity constraints. The model output may then range from a complete point-to-point network with no hub at all, to a network with any number of hubs of varying sizes and thus hierarchical importance. The resulting mixed-integer non-linear program is transformed into a mixed-integer linear program by piecewise linearizations of the capacity and congestion cost functions. In order to better understand the model properties and the influence of the various input parameters on

the optimal network design, the model is solved for a small-scale problem, and extensive sensitivity analyses are carried out.

The remainder of the paper is organized as follows. Section 2 consists in a brief literature review. The model assumptions, structure, and approximation are discussed in Section 3. The applications of the model are presented in Section 4. Conclusions and areas for further research are outlined in Section 5.

## **2. LITERATURE REVIEW**

Since the model developed in this paper is discrete, this review focuses on the discrete hub modeling literature. Hakimi (1964, 1965) modeled the location of a single switching center in a communication network, showing that its optimal location is always at a network node, and then extended this work to the case of multiple centers. Goldman (1969), analyzing multi-center location and multi-stage (origin-to-center, center-to-center, and center-to-destination) problems in a communication network, recognized the likely lower unit cost of hub-hub (H-H) links and the importance of scale economies. He developed a model to locate  $n$  centers in a network while minimizing the total multi-stage transportation cost. Marsten and Muller (1980) developed a mixed-integer program for hub-and-spoke (H-S) network design and fleet deployment. Their study was probably the first to recognize the nature and advantage of a H-S structure, discussing pure and mixed H-S networks, single and multiple hub allocations, interactions between hubs, and airplane assignments.

O'Kelly (1987) developed the first integer quadratic programming hub model. For given O-D flow and unit transportation cost matrices, this model minimizes the total transportation cost from origin-to-hub, hub-to-hub, and hub-to-destination. One distinct feature of this model is a discount rate associated to HH links to reflect scale economies due to flow concentration. The integer quadratic program is NP-hard, and has not been solved exactly. However, it has spurred the development of heuristics providing good, albeit sub-optimal, solutions (e.g., Klincewicz, 1991; Skorin-Kapov and Skorin-Kapov, 1994; O'Kelly et al., 1995; Campbell, 1996). A related quadratic programming model has been proposed by Helme and Magnanti (1989) to design satellite communication networks. However, the special structure of their model allowed for its efficient linearization and resolution.

Aykin (1994) developed a capacitated hub-and-spoke model allowing for non-hub to non-hub links. Various discount factors are used for different types of links. The problem is partitioned into hub location and routing sub-problems, and the approach combines heuristics and subgradient optimization. Campbell (1994) presented the hub location and network design problem as an uncapacitated mixed integer linear program. He also used a discount rate, and extends the model to  $p$ -hub center and hub covering problems.

## **3. HUB NETWORK MODELING METHODOLOGY**

### **3.1 Assumptions**

Hub network modeling requires two network inputs: a set of nodes  $N$  and a set of links  $A$  connecting these nodes. The nodes correspond to the candidate hub locations and those origin and destination points where traffic originates and terminates. The flows (people, goods, information) between these nodes can be expressed in terms of an origin-destination (O-D) flow matrix. An O-D pair can be connected by any number of paths, which are sequences of connected links. Each link is characterized by a link performance function (LPF), which specifies the functional relationship between the variable travel cost on a link and its flow, capacity, and other parameters. A hub is also associated with a hub performance function (HPF), which relates hub transshipping costs to hub flow, capacity, and other parameters. Finally, we assume that the system is in a steady-state condition, that flows are conserved on links and at hubs, and that there are physical limits to the sizes of the links and hubs that can be selected.

### 3.2 Model Structure

Let  $r$  denote an origin (O),  $s$  a destination (D),  $j$  a path connecting the O-D pair  $rs$ ,  $km$  the link between nodes  $k$  and  $m$  (with direction  $k \rightarrow m$ ),  $X_j^{rs}$  the flow on path  $j$  between  $r$  and  $s$ ,  $Q^{rs}$  the total flow from  $r$  to  $s$ , and  $X_{km}$  the total directional flow on link  $km$ .  $X_{km}$  is the sum of all the flows on paths using link  $km$  and can be expressed as

$$X_{km} = \sum_{r \in R} \sum_{s \in S} \sum_{j \in J} d_{km}^{rsj} X_j^{rs} \quad (1)$$

where  $R$  is the set of origins,  $S$  the set of all destinations,  $J$  the set of available paths joining the O-D pair  $rs$ , and  $d_{km}^{rsj}$  the link-path incidence parameter, with  $d_{km}^{rsj} = 1$  if link  $km$  is on path  $j$ ,  $= 0$  otherwise.

To ensure O-D flow conservation, the sum of the flows on all the paths from  $r$  to  $s$  must equal the total flow from  $r$  to  $s$ , that is

$$\sum_{j \in J} X_j^{rs} = Q^{rs} \quad (2)$$

Let  $F_{km}$  and  $F_m^{Max}$  be the endogenous and maximum feasible flow capacities of link  $km$ , and  $Y_{km}$  a 0-1 integer variable, with  $Y_{km} = 1$  if  $F_{km} > 0$ ,  $Y_{km} = 0$  otherwise. The link capacity constraints are then

$$X_{km} \leq F_{km} \quad (3)$$

$$F_{km} \leq F_m^{Max} Y_{km} \quad (4)$$

Let  $CO_{km}(X_{km}, F_{km})$  be the variable operating cost for link  $km$ , reflecting travel time, congestion, fuel, labor, and maintenance costs. The total link operating cost is:  $\sum_{km \in A} CO_{km}(X_{km}, F_{km})$ .

Let  $cf_{km}$  be the unit fixed cost for building link  $km$ . The total fixed link cost is:  $\sum_{km \in A} cf_{km} Y_{km}$ .

Let  $CA_{km}(F_{km})$  be the variable capacity investment cost for link  $km$ . The total link capacity cost is:  $\sum_{km \in A} CA_{km}(F_{km})$ . The total link cost ( $TLC$ ) is then:

$$TLC = \sum_{km \in A} cf_{km} Y_{km} + \sum_{km \in A} CA_{km}(F_{km}) + \sum_{km \in A} CO_{km}(X_{km}, F_{km}).$$

The total flow originating from node  $m$  is  $\sum_s Q^{ms}$ , the total flow terminating at node  $m$  is  $\sum_r Q^{rm}$ , both exogenous to the model and calculated directly from the O-D flow matrix. The total exogenous flow at node  $m$ ,  $Q_m$ , is then

$$Q_m = \sum_r Q^{rm} + \sum_s Q^{ms}.$$

Let  $Y_m = 1$  if  $m$  is a hub ( $m \in N$ ) or  $Z_m > 0$ , and  $Y_m = 0$  if  $m$  is not a hub or  $Z_m = 0$ , where  $Z_m$  is the total flow transshipment at hub  $m$  and is endogenous to the model. The total endogenous inflow and outflow at a hub  $m$  minus the total exogenous flow  $Q_m$  is equal to twice the total transshipment at hub  $m$ , with

$$2Z_m + Q_m = \sum_{k \in O(m)} X_{km} + \sum_{k \in D(m)} X_{mk}. \quad (5)$$

where  $O(m)$  and  $D(m)$  are the sets of nodes  $k$  that send flows to and receive flows from  $m$ , and where the total endogenous flows to and from hub  $m$  are  $\sum_{k \in O(m)} X_{km}$  and  $\sum_{k \in D(m)} X_{mk}$ , respectively

Let  $F_m$  and  $F_m^{Max}$  be the endogenous and maximum feasible flow transshipment capacities at hub  $m$ . The hub capacity constraints are then:

$$Z_m \leq F_m, \quad (6)$$

$$F_m \leq F_m^{Max} Y_m. \quad (7)$$

Let  $cf_m$  be the fixed hub cost at node  $m$ . The total fixed hub cost is:  $\sum_{m \in N} cf_m Y_m$ . Let  $CA_m(F_m)$  be variable hub capacity cost at node  $m$ . The total hub capacity cost is:  $\sum_{m \in N} CA_m(F_m)$ . Let  $CO_m(Z_m, F_m)$  be the hub operating cost function at node  $m$ . The total operating cost at hub  $m$  is:  $\sum_{m \in N} CO_m(Z_m, F_m)$ . The total hub cost ( $THL$ ) is then:

$$THL = \sum_{m \in N} cf_m Y_m + \sum_{m \in N} CA_m(F_m) + \sum_{m \in N} CO_m(Z_m, F_m)..$$

The total transportation cost ( $TC$ ), which includes fixed, capacity, and operating costs incurred on all links and hubs, is then

$$TC = \sum_{km \in A} cf_{km} Y_{km} + \sum_{km \in A} CA_{km}(F_{km}) + \sum_{km \in A} CO_{km}(X_{km}, F_{km}) + \sum_{m \in N} cf_m Y_m + \sum_{m \in N} CA_m(F_m) + \sum_{m \in N} CO_m(Z_m, F_m). \quad (8)$$

### 3.3 Capacity and Operating Cost Functions

The objective function of the general model includes fixed, variable capacity, and variable operating costs for all links and hubs. The capacity and operating cost functions need to be further specified to make the general model operational.

#### 3.3.1 Link and Hub Capacity Cost Functions

Since it is precisely economies of scale due to traffic concentration on links and at hubs that make a hub-and-spoke network attractive economically, it is reasonable to specify power cost functions that yield variable costs to scale, with

$$CA_{km}(F_{km}) = ca_{km}(F_{km})^{b_0} \quad (9)$$

$$CA_m(F_m) = ca_m(F_m)^{b_1} \quad (10)$$

where  $b_0$  and  $b_1$  are exogenous exponents,  $ca_{km}$  and  $ca_m$  exogenous unit capacity costs for link  $km$  and hub  $m$ , and  $F_{km}$  and  $F_m$  the endogenous capacities of link  $km$  and hub  $m$ . Procedures for piecewise linear approximation of these functions are well known. When these functions are concave ( $0 \leq b_0 \leq 1$ ,  $0 \leq b_1 \leq 1$ ), additional zero-one variables must be introduced for each segment of the piecewise approximation.

Note, however, that the only consideration of fixed costs,  $cf_{km} Y_{km}$  and  $cf_m Y_m$ , and linear capacity costs, would also account for economics of scale via decreasing average costs and would be computationally more convenient. A related linear approximation of the power capacity functions is proposed further on.

#### 3.3.2 Link Operating Cost Functions

Although various factors may affect the level of service (LOS) on links, the primary component of LOS, however, is travel time. Because of congestion, the travel time on a link is an increasing function of the flow on this link. Several general functional forms have been used to approximate link performance functions (LPF) in surface transportation systems. However, the use of LPFs has not been prevalent in air transportation and telecommunications. In this study, we use Davidson's LPF (Sheffi, 1972):

$$t_{km} = t_{km}^0 [1 + vX_{km} / (F_{km} - X_{km})]$$

The operating cost for link  $km$  is  $X_{km}t_{km}$ , and the total operating cost for all links is then:

$$\sum_{km \in A} CO_{km}(F_{km}, X_{km}) = \sum_{m \in N} co_{km} X_{km} t_{km}^0 [1 + vX_{km} / (F_{km} - X_{km})] \quad (11)$$

where  $co_{km}$  is the time-cost conversion factor for link  $km$ .

### 3.3.3 Hub Operating Cost Function

A hub needs a certain amount of time (waiting time, transshipping time, etc.) to process all the arrivals, storage, and departures. Following conventional assumptions and notions of  $M/M/1$  queues in queuing theory, we regard each hub as a single server with a queuing performance function. Two input elements are critical for queuing analysis at a hub: mean arrival flow rate and mean service rate (i.e., vehicles per hour, bytes per second).

The flow transshipping service capacity at hub  $m$ ,  $F_m$ , is analogous to mean service rate, and  $Z_m$ , the total flow transshipment at hub  $m$ , is analogous to the mean arrival flow rate. The mean transshipping time is then  $t_m = Z_m / (F_m - Z_m)$ . The corresponding hub operating cost is proportional to the product of transshipping flow and time, with

$$\sum_{m \in N} CO_m(F_m, Z_m) = \sum_{m \in N} co_m Z_m^2 / (F_m - Z_m), \quad (12)$$

where  $co_m$  is the time-cost conversion factor for hub  $m$ .

## 3.4 General Model Summary

The general model is restated below (a complete listing and the definitions of the indices, variables, and parameters are provided in the Appendix):

$$\begin{aligned} \text{Min TSC} = & \sum_{km \in A} \{ cf_{km} Y_{km} + ca_m (F_{km})^{b_0} + co_{km} X_{km} t_{km}^0 [ (1 + vX_{km} / (F_{km} - X_{km})) ] \} + \\ & \sum_{m \in N} [ cf_m Y_m + ca_m (F_m)^{b_1} + co_m Z_m^2 / (F_m - Z_m) ] \end{aligned} \quad (13)$$

$$\text{S.t.} \quad X_{km} = \sum_{r \in R} \sum_{s \in S} \sum_{j \in J^{rs}} d_{km}^{rsj} X_j^{rs} \quad (14)$$

$$\sum_{j \in J^{rs}} X_j^{rs} = Q^{rs} \quad (15)$$

$$X_{km} \leq F_{km} \quad (16)$$

$$F_{km} \leq F_{km}^{Max} Y_{km} \quad (17)$$

$$Z_m \leq F_m \quad (18)$$

$$F_m \leq F_m^{Max} Y_m$$

(19)

$$2Z_m + Q_m = [ \sum_{k \in O(m)} X_{km} + \sum_{k \in D(m)} X_{mk} ] \quad (20)$$

$$Y_{km}, Y_m = [0,1]$$

$$Z_m, F_{km}, F_m, X_{km}, X_j^{rs} \geq 0,$$

$$k, m \in N, r \in R, s \in S, km \in A, j \in J^{rs}$$

The objective function to minimize is the total system cost (*TSC*) incurred on links and at hubs.

Solving the general model yields the optimal hub locations ( $Y_m = 1$ ), the number of hubs ( $\sum_m Y_m$ ), and other information about node utilization. For instance, if  $Q_m > 0$  and  $Z_m = 0$ , then node  $m$  is a pure origin/destination node; if  $Q_m > 0$  and  $Z_m > 0$ , then node  $m$  acts as both hub and origin/destination node. If  $Q_m = 0$  and  $Z_m = 0$ , then node  $m$  is not used at all; if  $Q_m = 0$  and  $Z_m > 0$ , then node  $m$  is a pure transshipment hub. In addition, the model also yields endogenous link and hub capacities, link flow patterns, and link and hub individual costs. Unfortunately, the model is a mixed integer non-linear program with four non-linear terms in the objective function: two for the hub/link flow congestion costs ( $Z_m^2 / (F_m - Z_m), X_{km}^2 / (F_{km} - X_{km})$ ) and two for the hub/link capacity costs ( $(F_m)^{b_1}, (F_{km})^{b_0}$ ). As the objective function is not convex, a global optimal solution is difficult to obtain. In the following, we propose procedures to obtain near optimal solutions.

### 3.5 Piecewise Linearization of the Congestion Cost Functions

Consider the link congestion function  $\mathcal{G}(X_{km}, F_{km}) = X_{km}^2 / (F_{km} - X_{km})$ . Since  $F_{km} \geq X_{km}$ , we have:  $\mathcal{G}'(X_{km}, F_{km}) / \mathcal{G}'(X_{km}) = (2X_{km}F_{km} - X_{km}^2) / (F_{km} - X_{km})^2 > 0$  and  $\mathcal{G}''(X_{km}, F_{km}) / \mathcal{G}''(X_{km}) = 2(F_{km} - X_{km}) / (F_{km} - X_{km})^4 > 0$ . Therefore, the function  $\mathcal{G}(X_{km}, F_{km})$  is strictly convex with regard to  $X_{km}$ , for a fixed  $F_{km}$ , as shown in Figure 1.

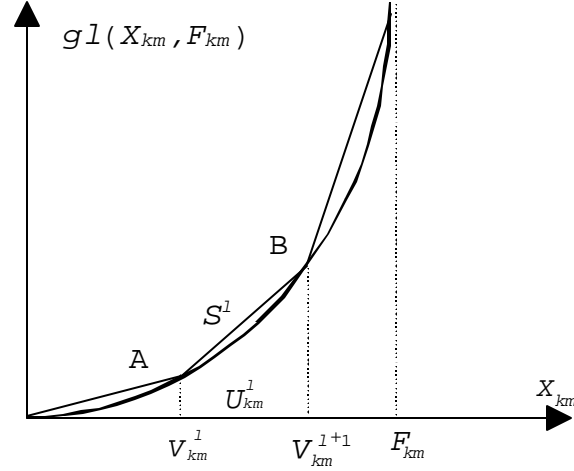


Figure 1 A Typical Curve of  $g_l(X_{km}, F_{km})$

Let the interval  $[0, F_{km}]$  be divided into  $L$  intervals ( $l=1,2,3,\dots,L$ ). The two end points of interval  $l$  are  $V_{km}^l, V_{km}^{l+1}$ , the slope of the line segment AB that approximates the curve between A and B is  $S^l$ , and the interval length is  $F_{km} / L$ . We obtain:

$$\begin{aligned} S^l &= [g_l(V_{km}^{l+1}, F_{km}) - g_l(V_{km}^l, F_{km})] / (V_{km}^{l+1} - V_{km}^l) \\ &= \frac{2Ll + L - l^2 - 1}{(L - l)(L - l - 1)}. \end{aligned}$$

It is important to note that the slope equation does not include the variables  $X_{km}$  and  $F_{km}$ . We assign new continuous variables  $U_{km}^l$  to all intervals  $l$ . The following constraints must then be satisfied:

$$U_{km}^l \leq \frac{1}{L} F_{km} \quad (21)$$

$$X_{km} = \sum_{l=0}^{L-1} U_{km}^l \quad (22)$$

The function  $g_l(X_{km}, F_{km})$  is approximated as follows:

$$g_l(X_{km}, F_{km}) \approx \sum_1 S^l U_{km}^l = \sum_{l=0}^{L-2} \left[ \frac{2Ll + L - l^2 - 1}{(L - l)(L - l - 1)} U_{km}^l \right] + S^{L-1} U_{km}^{L-1}.$$

where  $S^{L-1}$ , the slope of the last interval, cannot be computed using the basic formula for  $S^l$  (since  $S^l = \infty$  when  $l = L - 1$ ) and is approximated as the slope of a line with an inclination of  $89^\circ$ , or  $S^{L-1} = 57.3$ .

The total link congestion cost can then be written as:

$$\sum_{km \in A} c_{o_{km}} t_{km}^0 X_{km} + \mathbf{a} \sum_{km \in A} c_{o_{km}} t_{km}^0 \left\{ \sum_{l=0}^{L-2} \left[ \frac{2Ll + L - l^2 - 1}{(L-1)(L-1-1)} U_{km}^L \right] + 57.3 U_{km}^{L-1} \right\} \quad (23)$$

Since the hub flow congestion cost function has the same mathematical structure, the same procedure is used. If the interval  $[0, F_m]$  is divided into  $W$  intervals indexed by  $w = 1, 2, 3, \dots, W$ , and the continuous variables  $U_m^w$  are defined for all intervals  $w$ , then the total hub congestion cost function is approximated by

$$\sum_{m \in N} c_{o_m} \left\{ \sum_{w=0}^{W-2} \left[ \frac{2Ww + W - w^2 - w}{(W-w)(W-w-1)} U_m^w \right] + 57.3 U_m^{W-1} \right\}, \quad (24)$$

and the following constraints apply:

$$U_m^w \leq \frac{1}{W} F_m, \quad (25)$$

$$Z_m = \sum_{w=0}^{W-1} U_m^w. \quad (26)$$

### 3.6 Approximation of the Capacity Cost Functions

Consider the case of link  $km$  capacity cost function, as illustrated in Figure 2, where the exact function curve  $OCA = (F_{km})^{b_0}$  is always above the chord  $OA$  (with slope  $S^{OA}$ ) and below the two-segment curve  $OBA$ , where  $BA$  is tangent to the exact curve at  $A$ . Thus,  $OBA$  and  $OA$  can be used as upper and lower bounds for the function  $(F_{km})^{b_0}$ . In the following, we use the upper bound approximation, which can be expressed as a linear function of the variables  $(Y_{km}, F_{km})$ , with:

$$f(F_{km}) = (F_{km})^{b_0} \approx (1 - b_0) (F_{km}^{Max})^{b_0} Y_{km} + b_0 (F_{km}^{Max})^{b_0 - 1} F_{km}.$$

The same approximation is applied to the hub capacity cost function.

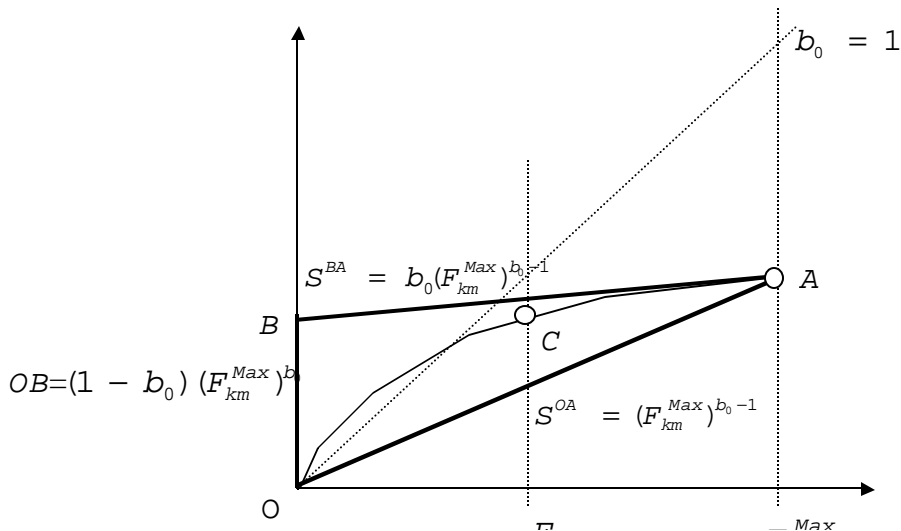


Figure 2 Upper and Lower Bounds for  $(F_{km})^{b_0}$

### 3.7 The Linearized Total System Cost Function

The total system cost function TSC is approximated by:

$$\begin{aligned}
 \text{TSC} = & \sum_{km \in A} c f_{km} Y_{km} + \sum_m c f_m Y_m \\
 & + \sum_{km \in A} c a_{km} [ (1 - b_0) (F_{km}^{Max})^{b_0} Y_{km} + b_0 (F_{km})^{b_0-1} F_{km} ] \\
 & + \sum_{km \in A} c o_{km} t_{km}^0 \{ X_{km} + a \sum_{l=0}^{L-2} [ \frac{2Ll + L - l^2 - l}{(L-l)(L-l-1)} U_{km}^L ] + 57.3 U_{km}^{L-1} \} \\
 & + \sum_{m \in N} c a_m [ (1 - b_1) (F_m^{Max})^{b_1} Y_m + b_1 (F_m^{Max})^{b_1-1} F_m ] \\
 & + \sum_{m \in N} c o_m \{ \sum_{w=0}^{W-2} [ \frac{2Ww + W - w^2 - w}{(W-w)(W-w-1)} U_m^W ] + 57.3 U_m^{W-1} \} \tag{27}
 \end{aligned}$$

The final model involves minimizing (27) subject to constraints (14)-(22), and (25)-(26). This is now a mixed-integer linear program (MILP), with the zero-one integer variables  $Y_{km}, Y_m$  ( $NV2=N^2$ ) and the continuous variables  $Z_m, F_{km}, F_m, X_{km}, X_j^{rs}, U_m^w, U_{km}^1$  ( $NV1=N^2(J+L+2) - N(J+L-W)$ ).

## 4. MODEL APPLICATIONS

The MILP developed in the previous section is solved using the OSL solver of GAMS for a small-scale network and with synthetic data input. The focus is on analyzing model properties and behavior, in particular the trade-offs between the different costs and the impacts of the input parameters on the optimal solution. Extensive sensitivity analyses are carried by varying each input parameter separately, as well as over a grid of values, leading to the estimation and analysis of response surfaces.

### 4.1 Data Input

We consider the 5-node rectangular network ( $N=5$ ) presented in Figure 3. The nodes are the 4 vertices and the rectangle center. Each node may be potentially linked to any other node. Link distances are also indicated in Figure 3. We assume that each O-D pair can be linked by a path with at most two intermediate hubs, leading to 10 possible paths (one path without any hub, three paths with one hub, and six paths with two hubs).

The input parameters include unit costs, distances, free-flow times, maximum feasible capacities, function exponents, and O-D flows. First, we assume that all hubs have the same reference values for fixed cost  $cf_0 = 50$ , capacity cost  $ca_0 = 1$ , time-cost conversion factor  $co_0 = 1$ , and maximum feasible capacity  $f_0 = 500$ . Then, we scale these values with the multipliers  $(u_1, u_2, u_3, u_7)$ , which vary within controlled value ranges, as follows:  $cf_m = u_1 * cf_0$ ,  $ca_m = u_2 * ca_0$ ,  $co_m = u_3 * co_0$ , and  $F_m^{Max} = u_7 * f_0$ . We assume that the link parameters are proportional to link length,  $d_{km}$ . Again, we use reference values for fixed cost  $cf_1 = 1$ , capacity cost  $ca_1 = 1$ , time-cost conversion factor  $co_1 = 0.5$ , free-flow time  $t_1 = 1$ , feasible capacity  $f_1 = 400$ , the multipliers  $(u_4, u_5, u_6, u_8, u_9)$ , and the link lengths  $d_{km}$  as follows:  $cf_{km} = u_4 * d_{km} * cf_1$ ,  $ca_{km} = u_5 * d_{km} * ca_1$ ,  $co_{km} = u_6 * co_1$ ,  $t_{km} = u_8 * d_{km} * t_1$ , and  $F_{km}^{Max} = u_9 * f_1$ . Note that the link time-cost factors  $co_{km}$  are set equal for all links, and so are the same factors  $co_m$  for all hubs. However, we allow for  $co_{km} \neq co_m$  to account for users' perceived differences in the value of time at hubs and over links. All O-D flows are assumed equal and defined by the reference value  $Q_1 = 40$  and the multiplier  $u_{10}$ , with:  $Q^{rs} = u_{10} * Q_1$ . The remaining inputs are the exponents  $(b_0, b_1)$  of the link and hub capacity cost functions and the multiplier  $v$  in the Davidson's LPF.

We assign to each of the 13 parameters  $(u_1-u_{10}, b_0, b_1, v)$  10 values within the following intervals:  $u_1 \in [4-44]$ ,  $u_2 \in [6-66]$ ,  $u_3 \in [1-11]$ ,  $u_4 \in [84-924]$ ,  $u_5 \in [0.8-8.8]$ ,  $u_6 \in [0.4-4.4]$ ,  $u_7 \in [0.4-4.4]$ ,  $u_8 \in [0.2-2.2]$ ,  $u_9 \in [0.4-4.4]$ ,  $u_{10} \in [0.3-3.3]$ ,  $b_0 \in [0.1-1.0]$ ,  $b_1 \in [0.1-1.0]$ , and  $v \in [0.1-1.0]$ .

## 4.2 Sensitivity Analyses Over Individual Parameters

The model has been solved while varying each of the 13 multipliers over its range, while keeping the other multipliers at their mid-range values. The absolute and relative changes in the total cost resulting from these variations are summarized in Table 1.

Table 1 Changes in the Total Cost (TC)

Parameter	TC Absolute Changes	TC Relative Changes (%)
u1	1800	12.78
u2	1708	12.09
u3	4159	34.25
u4	14496	216.15
u5	2168	15.63
u6	4593	36.03
u7	5417	42.61
u8	4593	36.03
u9	2829	19.01
u10	13479	184.03
b0	11699	85.30

b1	4136	29.56
v	759	5.25

The following general results are observed:

(a) The total cost increases with the parameters  $(u_1-u_6, u_8, b_0, b_1, v)$ , and, as expected, decreases with  $u_7$  and  $u_9$ , the maximum hub and link capacity multipliers.

(b) Varying the multipliers  $u_1, u_2, u_5, u_9$ , and  $v$  produces only small changes (less than 2,830 or 19.01%) in the total cost. In the cases of  $u_1$  and  $u_2$ , the changes in the total cost (12.78% and 12.09%) correspond exclusively to changes in the costs associated with these multipliers (hub fixed and variable capacity costs). However, in the cases of  $u_5, u_9$ , and  $v$ , total cost variations (15.63% , 18.08%, and 5.25%) result from trade-offs between all the three link costs and hub congestion costs.

(c) Varying the multipliers  $u_3, u_4, u_6, u_7, u_8, u_{10}, b_0$ , and  $b_1$  generates steeper changes (a minimum of 4,000 or 29.56%) in the total cost, with trade-offs among individual costs. The link fixed cost multiplier ( $u_4$ ) and the O-D flow multiplier ( $u_{10}$ ) produce the largest total cost changes (216.15% and 184.03%).

(d) When varying  $u_1, u_2, u_5, u_6, u_8$ , and  $v$ , we always obtain one-hub networks the configurations of which are virtually the same as in Figure 4 (i.e., same hub and same links), with only slight variations in link flows, hub transshipment, and link and hub capacities. This result suggests that the structure of the optimal networks is not sensitive to these multipliers. This is further confirmed by the small changes in the link and hub flow concentration indicators FCL and FCH, which are computed as follows:

$$FCL = \left( \sum_{km \in A} X_{km} \right) / \left( \sum_{km \in A} F_{km}^{Max} Y_{km} \right) \quad (28)$$

$$FCH = \left[ \sum_{m \in N} (Z_m + Q_m Y_m) \right] / \sum_{m \in N} (Q_m Y_m) \quad (29)$$

The results obtained when varying the multipliers of hub congestion cost ( $u_3$ ), link fixed cost ( $u_4$ ), maximum feasible hub capacity ( $u_7$ ), maximum feasible link capacity ( $u_9$ ), O-D flow ( $u_{10}$ ), and capacity function exponents ( $b_0$  and  $b_1$ ) are discussed in more details below.

$u_3$  (hub congestion cost multiplier): As expected, when  $u_3$  increases, the number of selected links  $(\sum_{km} Y_{km})$  increases from 10 to 14, and the total hub flow transshipment  $(\sum_m Z_m)$  decreases from 600 to 360, leading to decreasing hub flow concentration FCH (2.87 to 2.12) and link flow concentration FCL (0.45 to 0.28). Interestingly, despite the above changes, the same one hub is selected throughout. While the hub fixed and capacity costs display no changes, hub congestion costs fluctuate up and down. Link capacity costs, increasing with growing link capacity, induce a decline in link congestion costs.

$u_4$  (link fixed cost multiplier): Most cost and performance indicators increase with  $u_4$ , except the number of selected links, which drops from 20 to 10, and link capacity costs, which also decrease. The

increase in link fixed costs is slightly compensated by decreasing link capacity costs, due to the smaller number of links. Of all the multipliers, the largest total cost increase (from 6707 to 21203) is due to  $u_4$ .

$u_7$  (hub capacity multiplier): As with  $u_4$ , changes in  $u_7$  also induce significant changes in all cost components and performance indicators. The number of selected links decreases from 20 to 8, corresponding to the change from a point-to-point network to a minimum spanning tree. The decreasing total cost is associated with decreasing link fixed and capacity costs, and with increasing link congestion and hub costs. The dominant factor, however, is the decline in link fixed costs due to the declining number of links made possible by the increasing hub capacity. Hub and link flow concentrations increase from 0 to 3.25 and from 0.15 to 0.6, respectively, and so does total hub transshipment (0 to 720).

$u_9$  (link capacity multiplier): Increasing  $u_9$  induces a strong decrease in link fixed and congestion costs, a slight increase in link capacity and hub congestion costs, and no changes in the hub fixed and capacity costs, with, of course, total costs decreasing.

$u_{10}$  (O-D flow multiplier): All costs increase with  $u_{10}$ , except hub fixed costs, which remains constant, and hub congestion costs, which fluctuate up and down. Hub capacity costs remains constant and equal to 949 when  $u_{10} \geq 0.6$ . The largest contributors to the total cost increase are link fixed and congestion costs. The total number of links increases with  $u_{10}$  from 8 to 17: as O-D flows increase, it becomes necessary to use more of the available link capacity. As the maximum feasible capacity of each link is fixed, it is then necessary to put more links into service to carry all flows. At the limit, all 20 links may be selected, but not necessarily as a point-to-point network (using a hub may still be optimal). This pattern is clearly confirmed by the increasing link and hub flow concentration indicators.

$b_0$  (power of the link capacity cost function): Increasing  $b_0$  results in a large total cost increase (from 13716 to 25415). Link capacity costs become larger than link congestion costs when  $b_0 \geq 0.7$ , and become equal to link fixed costs when  $b_0 = 1.0$ . When  $b_0$  shifts from 0.9 to 1.0, hub congestion costs decrease, which is compensated primarily by increases in link congestion and capacity costs.

$b_1$  (power of the hub capacity cost function): The value of  $b_1 = 0.7$  splits the results into two groups: one with 13 selected links and a hub, the other a point-to-point network with no hub, and thus zero hub capacity cost. For  $0.1 \leq b_1 \leq 0.7$ , most costs and performance indicators remain constant, except hub capacity costs, which are the only contributor to the total cost increase. However, for  $0.7 < b_1 \leq 0.8$ , there are drops in hub congestion and capacity costs, which are offset by increases in link fixed and capacity costs. The sudden change in the optimal network structure with the change in  $b_1$  within the small interval  $[0.7, 0.8]$  underscores the discrete nature of the model solution space.

### 4.3 Optimal Network Analysis

We have obtained 130 optimal networks corresponding to the 13 multipliers, each with 10 values. A total of 10 distinct optimal networks (DON) in terms of unique hub and link selections are presented in Figures 4 through 13. The remaining 120 optimal networks are either similar or identical to

the 10 DONs. A similar network (SN) is defined as having the same network structure as one of the DONs, but with different link and hub flows and capacities. An identical network (IN) has exactly the same hub and link flows and capacities as one of the DONs. We observe the following:

(a) The optimal networks range from a point-to-point (P-P) configuration without hubs, as presented in Figure 8, to a minimum spanning tree (MST) structure, as presented in Figure 12. While the P-P network has the lowest levels of link and hub flow concentrations (FCH=0, FCL=0.15), the MST network has the highest ones, with FCH=3.25 and FCL=0.6.

(b) In all the optimal hub networks (i.e., not P-P), only one hub, located at node 5, is selected. The hub flow concentration FCH ranges from 1.45 to 3.25, and the total hub flow transshipment ranges from 240 to 720.

(d) All optimal networks include the shortest links between node 5 and nodes 1, 2, 3, and 4. Some networks also select the next shortest links, such as the links between nodes 1 and 4, nodes 1 and 2, and nodes 2 and 3. The longest links, such as the links between nodes 1 and 3 and nodes 2 and 4, are never selected in any optimal network with a hub. Only the optimal point-to-point network (Figure 8) selects the longest links. This is not surprising since the link costs are distances-related.

#### 4.4 Response Surface Analysis

In order to assess more precisely the cost trade-offs in the design of optimal hub networks, the model has been solved over a grid of values for the six cost multipliers ( $u_1 - u_6$ ), which are each assigned the lowest, mid-range, and highest values in the intervals specified in Section 4.1. A total of  $3^6 = 729$  distinct combinations of multiplier values was considered. All the other multipliers were kept at their mid-range values. The resulting optimal costs, in total and for each cost category, were then regressed over the input multipliers. Both linear and log-linear functional forms were considered. Higher  $R^2$  were obtained in the log-linear cases, the results of which are retained for further analysis and reported in Table 2. Total (TC) and link-specific (TLC, FIXLC, CAPLC, COMLC) costs are in logarithmic form ( $\ln X$ ). However, as hub-specific costs (THC, FIXHC, CAPHC, COMHC) do take a value of zero (i.e., there is no hub) over the grid, we cannot use the logarithmic transformation for these costs, and use, instead, the Box-Cox transformation  $X(\mathbf{I}) = (X^{\mathbf{I}} - 1) / \mathbf{I}$ , with  $\mathbf{I} = 0.1$ . The general form of the regression is:

$$\ln X \text{ or } X(\mathbf{I}) = a_0 + \sum_{i=1}^6 a_i \ln(u_i) \quad (30)$$

Table 2 Optimal Costs Regression Models

	Constant	$\ln u_1$	$\ln u_2$	$\ln u_3$	$\ln u_4$	$\ln u_5$	$\ln u_6$	R-Sq
TC	6.088	0.041	0.039	0.096	0.505	0.067	0.143	0.938
t-values	154.50	7.69	7.42	18.19	95.39	16.40	34.71	
TLC	6.781	0.026	0.021	0.102	0.359	0.069	0.188	0.906

t-values	167.42	4.76	3.94	18.78	65.93	16.32	44.35	
FIXLC	4.809	0.043	0.036	0.153	0.597	-0.013	0.028	0.948
t-values	120.15	8.02	6.78	28.52	110.98	-3.15	6.74	
CAPLC	6.460	0.024	0.021	0.079	-0.178	0.950	0.076	0.994
t-values	235.67	6.49	5.77	21.51	48.30	331.79	26.59	
COMLC	6.931	-0.031	-0.029	-0.032	0.092	0.055	0.897	0.992
t-values	238.18	7.95	7.44	8.31	23.63	17.97	295.02	
THC	-49.62	-0.391	-0.373	-0.324	10.012	0.139	-0.321	0.870
t-values	46.07	2.70	2.58	2.24	69.17	1.23	2.86	
FIXHC	-44.14	0.531	-0.756	-0.360	8.434	0.085	-0.302	0.852
t-values	45.01	4.03	5.73	2.73	64.00	0.83	2.95	
CAPHC	-44.50	-0.759	0.555	-0.365	8.383	0.087	-0.301	0.852
t-values	45.61	5.79	4.23	2.79	63.93	0.85	2.95	
COMHC	-46.76	-0.511	-0.466	-0.191	9.419	0.163	-0.295	0.876
t-values	47.57	3.87	3.53	1.45	71.30	1.59	2.87	

R-sq: R square values

TC: total cost, TLC: total link cost, FIXLC: link fixed cost, CAPLC: link capacity cost, COMLC: link operating cost; THC: total hub cost; FIXHC: hub fixed cost, CAPHC: hub capacity cost, COMHC: hub congestion cost.

When  $\ln X$  is used, the coefficient  $a_i$  represents the constant elasticity  $e_i$  of  $X$  with respect to  $u_i$ . However, when using the Box-Cox transformation, this elasticity is no longer constant, but varies with the levels of the variables, with:  $e_i = a_i / X^l$ . We discuss below in detail the regression results for each of the dependent cost variables.

(a) Total Cost (TC): The coefficients have the expected positive signs. The elasticity of the link fixed cost multiplier is the highest (0.505), followed by the link and hub congestion cost elasticities (0.143 and 0.096). The hub fixed and capacity costs, and the link capacity costs have the least effects on the total cost (0.041, 0.039, 0.067).

(b) Link Fixed Cost (FIXLC): The coefficients are all positive, except for the link capacity cost multiplier ( $u_5$ ), whose negative effect is very small (-0.013). The elasticity of the link fixed cost multiplier  $u_4$  (0.597) is the highest. To compensate for the increasing link capacity costs resulting from an increase in  $u_5$ , the number of links is reduced, thus reducing link fixed costs.

(c) Link Capacity Cost (CAPLC): Consistent with the discussion in (b), the coefficients have all positive signs, except for the link fixed cost multiplier  $u_4$  (-0.178). When  $u_4$  increases, the network optimization process favors a smaller number of links, and thus smaller aggregate link capacity and cost. The elasticity of the link capacity cost multiplier is high and close to unity (0.95).

(d) Link Congestion Cost (CONLC): The coefficients of the hub multipliers ( $u_1, u_2, u_3$ ) are negative and relatively small. However, the coefficients of the link multipliers ( $u_4, u_5, u_6$ ) are positive, and the coefficient of the link congestion multiplier  $u_6$  is large (0.897). When the hub multipliers increase, the network optimization process favors selecting links over hubs, leading to a lower link congestion cost.

(e) Total Link Cost (TLC): All coefficients are positive, and the largest elasticities are those for  $u_4$  (0.359),  $u_6$  (0.188), and  $u_3$  (0.102). The trade-offs between the three link costs, as uncovered in the analysis of the individual cost equations, are hidden when considering aggregate link costs.

(f) Hub Fixed Cost (FIXHC): The strongest positive effects are due to the fixed cost multipliers  $u_1$  (hub) and  $u_4$  (link). When  $u_4$  increases, the model moves from a zero-hub to a one-hub system to compensate for the increasing link costs. The strongest negative effects are due to the hub capacity and operating cost multipliers,  $u_2$  and  $u_3$ . When these costs increase, the model tends to reduce the number of hubs, and thus their fixed costs. The link operating cost multiplier  $u_6$  has a negative effect (-0.302): to compensate for these link costs, the model tends to increase the number of links and their capacities, thus decreasing the number of hubs and their costs.

(g) Hub Capacity Cost (CAPHC): The effects of  $u_1$  (-0.759),  $u_3$  (-0.365), and  $u_6$  (-0.301) are negative because the model, to compensate for these costs, increases the number of links and their capacities, and thus reduces the number of hubs and/or their capacities, hence the hub capacity costs. The strongest positive effects are due to  $u_4$  and  $u_2$ . When link fixed costs increase, the model substitutes hubs and hub capacity for links.

(h) Hub Congestion Cost (CONHC): The only positive effects are due to the link multipliers  $u_4$  (9.419) and  $u_5$  (0.163): when these link costs increase, the model will attempt to reduce both the number of links and their capacities, thus increasing the flow through the hub and the resulting congestion. However, the effect of  $u_6$  (link congestion cost multiplier) is negative, which suggests that the model, in order to compensate for such costs, increases the number of links and their capacities, leading to a decrease in hub flow and congestion cost. Finally, the impacts of the hub multipliers ( $u_1 - u_3$ ) are all negative: an increase in these costs is compensated by selecting more links with larger capacities, ultimately leading to lesser hub flows and congestion.

(i) Total Hub Cost (THC): The effects of the multipliers are similar to those in the hub congestion case (h), because these costs generally dominate the two other hub costs.

## 6. CONCLUSIONS

We have presented a general hub network model that considers all fixed, capacity, and operating/congestion costs on links and at hubs, accounting for both economies and diseconomies of

scale. The model selects hubs and links, determines their capacities, and assigns O-D flows over paths, while minimizing all system costs. The model, initially formulated as a mixed-integer non-linear program, is transformed into a mixed-integer linear program through the linearization of the capacity and congestion cost functions. The methodology is illustrated by an application to a small-scale network with hypothetical data. Extensive sensitivity analyses have been carried out to assess the trade-offs between the different link and hub costs.

Several issues call for further research. First, while the MILP model has been easily solved for the small-scale problem considered here, it is clear that its application to much larger problems might be more difficult or possibly unfeasible in terms of computational requirements, thus calling for the development of heuristic procedures providing very good, although sub-optimal, solutions. The present model would provide benchmark solutions against which such heuristics could be evaluated. Second, the model should be applied to a real-world setting, which would require the gathering of data and the calibration of the capacity and congestion cost functions used in the model. Finally, the model could be extended to account for dynamic factors and the time dimension, as well as stochastic factors and reliability.

## REFERENCES

Aykin, T. "Lagrangian Relaxation Based Approaches to Capacitated Hub-and-Spoke Network Design Problem." *European Journal of Operational Research* 79 (1994): 501-523.

Campbell, J. F. "Integer Programming Formulations of Discrete Hub Location Problems," *European Journal of Operational Research* 72 (1994): 387-405.

Campbell, J. F. "Hub Location and the p-Hub Median Problem." *Operations Research* 44, 6 (1996): 923-935.

Goldman, A. J. "Optimal Locations for Centers in a Network." *Transportation Science* 3 (1969): 352-360.

Hakimi, S. L. "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph." *Operations Research* 12 (1964): 450-459.

Hakimi, S. L. "Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems." *Operations Research* 13 (1965): 462-475.

Helme, M. P. and Magnanti, T. L. "Designing Satellite Communication Networks by Zero-One Quadratic Programming." *Networks* 19 (1989): 427-450.

Klincewicz, J. G. "Heuristics for the  $p$ -hub Location Problem." *European Journal of Operational Research* 53 (1991): 25-37.

Marsten, R E., and Muller, M R. "A Mixed-Integer Programming Approach to Air Cargo Fleet Planning." *Management Science* 26 (11 1980): 1096-1107.

O'Kelly, M. E. "A Quadratic Integer Program for the Location of Interacting Hub Facilities." *European Journal of Operational Research* 32 (1987): 393-404.

O'Kelly, M. E., Skorin-Kapov, D., and Skorin-Kapov, J. "Lower Bounds for the Hub Location Problem." *Management Science* 41 ,4 (1995): 713-721.

Sheffi, Yosef. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. New Jersey: Prentice-Hall, 1985.

Skorin-Kapov, D., and Skorin-Kapov, J. "On Tabu Search for the Location of Interacting Hub Facilities." *European Journal of Operational Research* 73 (1994): 502-509.

## APPENDIX

### General Model Notations

#### Indices

$N$ : set of nodes  $(k, m) \in N$

$R$ : set of origins,  $r \in R$

$S$ : set of destination,  $s \in S$

$A$ : set of links  $km \in A$

$J^{rs}$ : set of paths from  $r$  to  $s$ ,  $j \in J^{rs}$

$O(m)$ : set of origin nodes  $k$  linked to destination node  $m$  on link  $k \rightarrow m$

$D(m)$ : set of destination nodes  $k$  linked to origin node  $m$  on link  $m \rightarrow k$

#### Decision Variables

$X_j^{rs}$  : flow on path  $j$  from  $r$  to  $s$

$X_{km}$  : endogenous total flow on link  $k \rightarrow m$

$F_{km}$  : endogenous flow capacity of link  $k \rightarrow m$

$Z_m$  : total endogenous flow transshipment at  $m$

$F_m$  : endogenous transshipment capacity of hub  $m$

$Y_{km}$  : decision variable,  $Y_{km} = 1$  if link  $k \rightarrow m$  is selected,  $Y_{km} = 0$  otherwise

$Y_m$  : decision variable,  $Y_m = 1$  if  $m$  is a hub or  $Z_m > 0$ ,  $Y_m = 0$  if  $m$  is not a hub or  $Z_m = 0$

$t_{km}$  : endogenous travel time on link  $k \rightarrow m$

## Parameters

$Q_j^{rs}$  : total flow from  $r$  to  $s$

$F_{km}^{Max}$  : exogenous maximum feasible capacity of link  $k \rightarrow m$

$cF_{km}$  : fixed unit length cost of link  $k \rightarrow m$

$ca_{km}$  : capacity unit cost of link  $k \rightarrow m$

$CO_{km}$  : time-cost conversion parameter for link  $k \rightarrow m$

$Q_m$  : total exogenous flow at node  $m$

$F_m^{Max}$  : exogenous maximum feasible transshipment capacity of hub  $m$

$cF_m$  : fixed hub cost at node  $m$

$ca_m$  : capacity unit cost of hub  $m$

$CO_m$  : time-cost conversion parameter for hub  $m$

$v$ : parameter in Davidson's LPF

$t_{km}^0$  : exogenous free-flow time on link  $k \rightarrow m$

$d_{km}^{rsj}$  : incidence parameter,  $d_{km}^{rsj} = 1$ , if link  $k \rightarrow m$  is on path  $(r,s,j)$ ;  $= 0$ , otherwise

$b_0, b_1$  : exponents in the link and hub capacity cost functions

## Functions

$CA_{km}(F_{km}) = ca_{km}(F_{km})^{b_0}$  : capacity cost function for link  $k \rightarrow m$

$CA_m(F_m) = ca_m(F_m)^{b_1}$  : hub capacity cost function for node  $m$

$CO_{km}(X_{km}, F_{km})$  : operating cost function for link  $k \rightarrow m$

$CO_m(Z_m, F_m)$  : hub operating cost function for node  $m$

$t_{km} = t_{km}^0 [ (1 + vX_{km}) / (F_{km} - X_{km}) ]$  : travel time for link  $k \rightarrow m$  (Davidson's LPF)