CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

Profit-Oriented Fixed-Charge Network Design with Elastic Demand

Carlos Armando Zetina
Ivan Contreras
Jean-François Cordeau

December 2018

CIRRELT-2018-48

Bureaux de Montréal :
Université de Montréal
Pavillon André-Aisenstadt
C.P. 6128, succursale Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Profit-Oriented Fixed-Charge Network Design with Elastic Demand

Carlos Armando Zetina[1,2,*], Ivan Contreras[1,2], Jean-François Cordeau[1,3]

1. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)
2. Department of Mechanical and Industrial Engineering, Concordia University, 1515 St-Catherine Street W., EV4 139, Montréal, Canada H3G 1M8
3. Department of Logistics and Operations Management, HEC Montréal, 3000 Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

Abstract. This paper extends classic fixed charge multicommodity network design by explicitly considering demand elasticity with respect to routing cost in a profit maximization context with service commitments. Demand quantity is determined by a spatial interaction model that accounts for routing costs thus capturing the trade-off between infrastructure investment, efficient routing, and increased revenue. A numerical example is presented to demonstrate the added value of incorporating demand elasticity in profit-oriented network design problems. An arc-based and a path-based formulation, both with the flexibility of incorporating O/D pair selection by means of network and data transformations, are presented. The arc-based formulation is solved using state-of-the-art global optimization software while the path-based formulation serves as the basis for a hybrid matheuristic that combines a slope scaling metaheuristic and column generation. Computational experience shows the hybrid matheuristic to be superior in terms of solution quality and computation time.

Keywords. Gravity model, networks, elastic demand, global optimization.

Acknowledgements. This research was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) (Grants 418609-2012 and 04959-2014). This support is gratefully acknowledged.

---

* Corresponding author: c_zetina@encs.concordia.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2018

1. Introduction

Fixed charge multicommodity network design (FMND) is a fundamental optimization problem arising in industries such as transportation and communications to capture the trade-off between strategic investment and operational efficiency. The problem consists of selecting a subset of potential arcs to be installed and to route the demand of commodities from several origin/destination (O/D) pairs using only the installed arcs. A fixed cost is incurred upon installing an arc and a unit transportation cost is paid for each unit of commodity transported through each arc. It generalizes a large class of well-known combinatorial problems such as the *shortest path problem*, the *traveling salesman problem*, the *uncapacitated lot-sizing problem*, and the *Steiner network design problem* (Holmberg and Hellstrand 1998, Zetina et al. 2018), and models a variety of problems in communications and transportation (Magnanti and Wong 1984, Minoux 1989).

In this paper, we extend the FMND by incorporating demand elasticity with respect to routing cost in a profit maximization context where a predetermined amount of revenue is received for each unit of demand routed. The problem considers the same decisions as the classic model: selecting a subset of arcs to install and routing the demand of O/D pairs using them. In addition, the decision maker selects which O/D pairs to serve subject to a service commitment constraint. Since demand quantity is a function that depends on route distance which in turn is determined by the solution of the model, the proposed framework endogenously captures the feedback loop between the network design decisions and demand quantities, leading to equilibrium-like conditions at its optimal solutions.

In the network optimization literature, demand quantities are exogenously estimated via historical information and given as fixed parameters to the problem. This makes the resulting optimization model highly dependent on the quality of the initial demand estimate. To circumvent this, researchers have developed models for FMND that account for demand uncertainty via robust optimization as in Lee et al. (2013) and Keyvanshokooh et al. (2016) or stochastic programming as in Rahmaniani et al. (2018). The former assumes demand realizations are within a predefined set and finds the best solution considering the worst possible demand realization occurs (risk-averse). The latter assumes that demand follows a given probability distribution and seeks the solution with the best expected value (risk-neutral).

Both approaches assume demand behaves statically in the sense that it does not depend on the network's configuration, i.e. demand is inelastic. This does not hold in many of the applications of FMND. An example is the construction of the U.S.A's interstate highway system which led to a significant increase in transportation of commodities among cities in particular within the "Sunbelt states" (*https://www.fhwa.dot.gov*). Another example is the shift in travel patterns when shorter flight routes are offered between cities (Boonekamp et al. 2018). The myopic perspective of ignoring demand elasticity compromises the applicability of FMND at the strategic level where decisions have long-lasting repercussions, a context in which its use is ubiquitous.

In the literature, there are three areas of network design in which elastic demand has been considered: transit network design, service network design, and network pricing.

Each models demand elasticity according to the level of detail needed for the problem. In transit network design, elastic demand is incorporated by means of traffic assignment problems (TAP), where demand information is obtained at the link level. Service network design and network pricing problems on the other hand use a distance decay function to estimate demand between O/D pairs.

Traffic assignment problems on congested networks were the earliest to incorporate elastic demand into network problems. The purpose of these models is to calculate the use of each link (road) on a network considering: 1) users are selfish and will use paths that minimize their travel time; 2) travel time over a link is inversely proportional to the number of users on it; and 3) the number of users going between two points in the network is a function of the travel time. Beckmann et al. (1956) presented a non-linear formulation whose optimal solution also solves the TAP, satisfying what later became termed as a "user-equilibrium". Solution algorithms that converge to this user-equilibrium for the fixed demand case were first proposed by Dafermos (1968) and LeBlanc et al. (1975), while Florian and Nguyen (1974) and Evans (1976) devised the first efficient algorithms for the elastic demand case. Numerous extensions have been proposed for both the static TAP (Matsoukis and Michalopoulos 1986) and the dynamic TAP (Peeta and Ziliaskopoulos 2001).

As mentioned before, the TAP is a subproblem of transit network design problems in which higher level decisions such as added road capacity, vehicle outgoing frequency, or vehicle sizes must be determined (Newell 1979). These problems are posed as bilevel programs in which the upper level seeks to maximize social benefit and the lower level corresponds to a TAP. Due to the difficulty of solving the TAP, in particular when considering elastic demand, and the added challenge of bilevel programming, most solution methods for these problems have been heuristic in nature (Cipriani et al. 2012).

Demand elasticity has also been incorporated into location models by means of distance decay functions, spatial interaction models, or user-equilibrum constraints to characterize demand loss due to travel time/cost, congestion and decline in utility. Two families of location problems have considered demand elasticity. The first is Competitive Facility Location in which the decision maker seeks to maximize the market share captured or minimize lost demand by strategically locating facilities for customers whose willingness to patronize it is sensitive to travel and waiting time. Solution algorithms and extensions of these can be found in Marianov et al. (2005), Berman et al. (2006), Aboolian et al. (2007) and Marianov et al. (2008). Other works with a specific application to healthcare service network design are Zhang et al. (2009) and Zhang et al. (2010) in which the former uses queueing theory to determine expected demand and the latter models demand elasticity via user-equilibrium conditions to account for willingness to participate in preventive healthcare.

The second family of location problems to consider demand elasticity is the profit maximizing service network design problem presented in Aboolian et al. (2012). The model seeks to determine the optimal facility locations and their corresponding service levels so as to maximize the profit obtained as the difference between the revenue received from the captured demand and the investment in infrastructure. The model accounts for sensitivity to both travel and waiting time where the latter is incorporated as constraints derived from well-known queueing theory results. It is solved to optimality by a successive improvement algorithm that removes non-optimal feasible solutions at each iteration. However, the solution time shows to be sensitive to the allowed minimum number of workers per facility.

The most recent area of network design to incorporate elastic demand is that of network pricing introduced by Labbé et al. (1998). This problem seeks to maximize the revenue raised from tolls placed on a network that must transport multiple commodities. Kuiteing et al. (2017) extended the original version to include elastic demand by means of a linearly decreasing function. The resulting problem is posed as a bilevel program and reformulated into a mixed integer quadratic program which is solved by a general-purpose solver. Kuiteing et al. (2018) later extended the model to include non-linear demand decay functions and proposed an exact method based on piecewise linear approximations of the demand function that asymptotically converges to the optimal solution.

These studies demonstrate the importance and impact of accounting for elastic demand in strategic network design problems both in "directed choice" models where a central decision maker establishes the O/D routes and in "user-choice" models where routing is determined by user-equilibria. Recently, Daganzo (2012) presented conditions under which demand estimation and system design can be done separately in public infrastructure network design through user-choice models. These problems seek to maximize social benefit by deciding on the system design, including its layout and control, and the prices to be charged for the service. We note that this result does not apply to the problem presented in this paper as it is a directed choice model placed in a more abstract context to allow its use in applications beyond public infrastructure.

Among the classes of problems reviewed, the most closely related to the proposed framework is that of Aboolian et al. (2012). It also captures the trade-off between additional investment and increased revenue. However, the presented model differs in that it assumes a maximum threshold of possible demand, and that elasticity is modeled by both a distance and congestion decay function. In addition, the inherent difference between location and network design problems makes the corresponding approaches significantly different. In location problems, locating a facility directly impacts the travel costs of nearby patrons. On the other hand, the effect on routing costs of installing an arc in a network is dependent on the other arcs that determine the shortest paths of commodities.

With respect to modeling demand elasticity, the methods used in the reviewed literature do not fit well with the assumptions and level of detail of FMND. Compared to the TAP, we do not require traffic levels per link nor does the model assume having the parameters necessary to formulate the equilibrium model. On the other hand, distance decay functions are useful when modeling lost demand whereas the goal of the proposed framework is to also capture the possibility of increasing demand quantity based on network decisions.

In this paper we propose the use of a gravity model to incorporate demand elasticity to routing cost into a profit-oriented variant of FMND. One of the key advantages of the use of the gravity model is that its simplest version allows for the incorporation of demand elasticity by using an O/D demand matrix as in the classic FMND. On the other hand, more sophisticated gravity models that consider other determining features for demand prediction can be easily incorporated by doing the necessary calibration. This allows for a wide spectrum of possible gravity models to predict demand being incorporated into FMND. The gravity model also generalizes distance decay functions as it does not assume an estimated maximum possible demand. Instead, the maximum possible demand is implicitly determined by the underlying network's shortest distances. In addition, we allow the decision maker to choose which O/D pairs will be served subject to a service commitment constraint that enforces a minimum number of them to be routed. We present two non-linear mixed integer programming formulations obtained by incorporating a general form of the gravity model and demonstrate the added value of incorporating demand elasticity by comparing solutions obtained from the proposed model and from its inelastic version. Both formulations are able to model O/D pair selection by means of simple network transformations. Finally, we present solution algorithms and a computational comparison of their performance with respect to solution quality and computation time.

The rest of the paper is organized as follows. In Section 2, we provide some preliminaries on the use of gravity models, present the notation, formally describe the problem, and give a numerical example that demonstrates the value of incorporating demand elasticity. Section 3 presents an arc-based and a path-based formulation and the transformations necessary to incorporate O/D pair selection and service commitment constraints. Section 4 details the components of the hybrid matheuristic used to solve the path-based formulation while Section 5 compares its performance with that of solving the arc-based formulation with a state-of-the-art global optimization solver. Finally, Section 6 provides conclusions and future lines of research.

2. Problem definition

As the proposed framework is based on the FMND, we adopt the same notation. The problem is defined on a directed graph $G = (N, A)$ with node set N and arc set A. We assume the existence of a set of O/D pairs, which we denote by K, between which demand must be routed on a single path. Each O/D pair will also be referred to as a commodity. Each arc has a corresponding fixed installation cost $f_{ij} \geq 0$ and a unit transportation cost $c_{ij}^k > 0$ for each commodity $k \in K$. A revenue of $\alpha_k \geq 0$ is received for every unit of commodity $k \in K$ that is routed.

There is an added value in allowing the decision maker to select which commodities to route in a profit-oriented problem. A commodity may be left unserved if the resulting installation and operational costs do not compensate the obtained revenue in the overall network design. This additional decision level is incorporated in the problem along with a service commitment constraint that enforces that at least Γ O/D pairs are routed, where $\Gamma \in \mathbb{Z}^+$. Considering all previously mentioned characteristics, the proposed problem consists of finding the network configuration that maximizes the total profit obtained from routing at least Γ of a set of given O/D pairs. The flexibility of the resulting problem allows it to be used in both regulated industries where service commitments are imposed and unregulated industries where the service provider has complete freedom to selfishly pursue profit.

Given that the total revenue depends on the demand quantities of the commodities routed, the role of demand elasticity to routing cost directly impacts our objective. The proposed problem implicitly seeks an equilibrium between spending on network infrastructure to increase demand quantities of commodities while ensuring maximum profitability in the overall endeavor. We next introduce the preliminaries of gravity models, the tool used to model demand elasticity to routing cost.

2.1. Gravity models

Based on Newton's law of universal gravitation (Newton 1687), gravity models have spread to other fields in the social sciences (Haynes and Fotheringham 1984). Since the late 19th century, ideas based on gravity models have been used to explain principles of social science (Carey 1858), to define laws of migration (Ravenstein 1885), to explain retail gravitation (Reilly 1929), to model consumer behavior (Huff 1964), and to analyze traffic patterns (Wilson 1967). In the optimization community, Huff-like models have been used in competitive facility locations (Eiselt et al. 1993, Aboolian et al. 2007, Fernández and Hendrix 2013) with both market share capture and profit maximization objectives.

The original idea behind gravity models was to measure the interaction between two locations as being directly proportional to their size and inversely proportional to the distance between them. The model's simplified form led to skepticism about its prediction capabilities (Jensen-Butler 1972). As a response, theoretical refinements were made to improve its reliability as a prediction tool. Wilson (1971) presented gravity models that considered information restrictions and provided a corresponding taxonomy for the resulting families of gravity models: *unconstrained*, *production constrained*, *attraction constrained*, and *production-attraction constrained*. Later, Senior (1979) provided a means of extracting information at a disaggregated level based on entropy maximization. Throughout the years, other refinements have been proposed that allow for the inclusion of industry-specific features into the gravity model (Fotheringham and O'Kelly 1989). Today, refined and meticulously calibrated gravity models such as those presented in Hodgson (1990), Grosche et al. (2007), Lesage and Polasek (2008), Zhong et al. (2018) and Boonekamp et al. (2018) provide reliable estimates of interactions between locations.

In general, gravity models assume the interactions W_{ij} from city i to j can be estimated as $W_{ij} = f(\bar{U}_i, \bar{V}_j, \bar{D}_{ij})$ where $\bar{U}_i$ is a vector of origin features, $\bar{V}_j$ is a vector of destination features, $\bar{D}_{ij}$ represents a set of separation attributes, one of which is the travel cost, and $f(.)$ is a real-valued function (Fotheringham and O'Kelly 1989). To ease notation, we assume the following simplified version of the gravity model:

$$W_{ij} = \frac{P_i P_j}{(d_{ij})^r}, \tag{1}$$

where P_i is a weight attributed to the population size at location i, d_{ij} represents the distance between them, and $r \geq 0$ is an exponent that models the sensitivity to distance.

As seen in Black (1973), depending on the context in which the gravity model is used, different values of r have been shown to better approximate demand patterns. In practice, its value often varies between 0.5 and 2.0 (Fotheringham and O'Kelly 1989).

The formulations and solution algorithms presented in this paper are also compatible with more sophisticated gravity models. Note that the components of $\bar{U}_i$ and $\bar{V}_j$ are not affected by the decisions taken in network design problems. In fact, the only feature affected by network design decisions is the distance between the locations. Therefore, the addition of features would appear in the proposed optimization model as constants. On the other hand the variability of the exponent r is accounted for and, as will be seen in Section 4, the presented solution algorithm easily adapts to any value of r.

2.2. Profit-oriented network design with elastic demand

We next present the objective function of the problem resulting from incorporating (1) into FMND where distances considered are based on routing costs. Let the variable $x_{ij}^k \in \{0,1\}$ represent whether arc $(i,j) \in A$ is used in the route of commodity $k \in K$ while y_{ij} represents the activation of arc $(i,j) \in A$. To model revenue, we convert the demand quantity into a monetary value by multiplying it by the per unit revenue $\alpha_k \geq 0$ received for each unit of commodity $k \in K$ routed. We assume that the routing costs used in the gravity model are directly proportional to the transportation costs incurred c_{ij}^k, i.e. $d_{ij} = \tau c_{ij}^k$, where $\tau > 0$ is a transportation cost scaling factor. Substituting the gravity model form (1) for the fixed demand parameter W_k of the FMND, we obtain the following objective function for the profit-oriented network design problem with elastic demand (POFMND-E):

$$(OF) \quad \sum_{k\in K} \frac{\alpha_k P_{o(k)} P_{d(k)}}{(\tau \sum\limits_{(i,j)\in A} c_{ij}^k x_{ij}^k)^r} - \sum_{k\in K} \frac{P_{o(k)} P_{d(k)}}{(\tau \sum\limits_{(i,j)\in A} c_{ij}^k x_{ij}^k)^r} \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k - \sum_{(i,j)\in A} f_{ij} y_{ij} \tag{2}$$

$$= \sum_{k\in K} P_{o(k)} P_{d(k)} \frac{[\alpha_k - \sum\limits_{(i,j)\in A} c_{ij}^k x_{ij}^k]}{(\tau \sum\limits_{(i,j)\in A} c_{ij}^k x_{ij}^k)^r} - \sum_{(i,j)\in A} f_{ij} y_{ij} \tag{3}$$

$$= \sum_{k\in K} PP_k(x) - \sum_{(i,j)\in A} f_{ij} y_{ij}. \tag{4}$$

The first, second and third terms of (2) correspond to the total revenue, transportation cost and investment cost, respectively. Note that unlike classic multicommodity network design problems, in OF the demand quantities between O/D pairs will depend on the cost of the routes used. This models the effect the decision maker's choice of routes has on demand quantities, i.e. demand elasticity to routing cost.

Simplifying, we obtain (3) from which we can more clearly see the profit maximization and demand elasticity characteristics of the non-linear objective function. Note that each addend of the first term, rewritten as $PP_k(x)$ in (4), lends itself to the interpretation of "partial profit" obtained from serving commodity $k \in K$. For each O/D pair, it calculates the difference between the per unit revenue and transportation cost, multiplied by the demand quantity obtained according to the gravity model.

Since no assumptions, except non-negativity, are imposed on the value of the per unit revenue α_k of a commodity, the model is capable of handling cases in which for a given commodity k, $PP_k(x) < 0$. In other words a loss is incurred. One such case is when the per unit revenue of a commodity is strictly less than the cheapest route over the complete underlying network, p_s^k.

An important characteristic of $PP_k(x)$ is that it is the composition of two functions $d_k(x) : \mathbb{R}^n \to \mathbb{R}$ and $PP_k(d_k) : \mathbb{R} \to \mathbb{R}$, where $d_k(x) = \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k$ and $PP_k(d_k) = P_{o(k)} P_{d(k)} \frac{[\alpha_k - d_k]}{(\tau d_k)^r}$. The following proposition uses this to provide insights on the shape of the function $PP_k(x)$.

Proposition 1. *For $r > 1$,*

a) $d_k(x) = \frac{r\alpha_k}{r-1}$ is a minimum of $PP_k(d_k(x))$,

b) $PP_k(x)$ is convex for $0 < d_k(x) \leq \frac{(r+1)\alpha_k}{(r-1)}$,

c) $PP_k(x)$ is concave for $d_k(x) \geq \frac{(r+1)\alpha_k}{(r-1)}$.

Proof Given that $PP_k(d_k) : \mathbb{R} \to \mathbb{R}$, i.e. the partial profit of commodity k as a function of distance $d_k > 0$, is a twice differentiable univariate function, we can calculate its first and second order derivatives. We denote $f'(x)$ and $f''(x)$ as the first and second order derivatives of $f(x)$, respectively. Using univariate calculus we obtain the following differential information for $PP_k(d_k)$:

$$PP_k'(d_k) = P_{o(k)} P_{d(k)} \frac{(d_k(r-1) - r\alpha_k)}{\tau^r d^{r+1}}, \tag{5}$$

$$PP_k''(d_k) = P_{o(k)} P_{d(k)} r \frac{((r+1)\alpha_k - (r-1)d_k)}{\tau^r d^{r+2}}. \tag{6}$$

From (5) and (6) we observe that for any $r > 1$:

- $PP_k''(d_k) \geq 0$ when $d_k \leq \frac{(r+1)\alpha_k}{(r-1)} \implies PP_k(d_k)$ is convex for $d_k \in (0, \frac{(r+1)\alpha_k}{(r-1)}]$.
- $PP_k''(d_k) < 0$ when $d_k > \frac{(r+1)\alpha_k}{(r-1)} \implies PP_k(d_k)$ is concave for $d_k \in (\frac{(r+1)\alpha_k}{(r-1)}, \infty)$.
- $PP_k(d_k)$ obtains a minimum of $\frac{-P_{o(k)} P_{d(k)} \alpha_k (r-1)^{(r-1)}}{(r\tau\alpha_k)^r}$ at $d_k = \frac{r\alpha_k}{r-1}$ since $PP_k'(\frac{r\alpha_k}{r-1}) = 0$ and $PP_k''(\frac{r\alpha_k}{r-1}) > 0$.

Given that $d_k(x)$ is an affine transformation and $PP_k(x)$ is the composition of $PP_k(d_k)$ and $d_k(x)$, we conclude that $PP_k(d_k)$ is convex when $0 < d_k(x) \leq \frac{(r+1)\alpha_k}{(r-1)}$ and concave when $d_k(x) \geq \frac{(r+1)\alpha_k}{(r-1)}$ (Boyd and Vandenberghe 2004, section 3.2.2) and is therefore a non-convex function throughout the domain $x \in \mathbb{R}^{|A|}$. ∎

The case of $0 \leq r \leq 1$ merits special attention because the function $PP_k(x)$ possesses properties that can be exploited from a mathematical programming perspective. Although the scope of this study is to present a general solution methodology adaptable to any $r \geq 0$, we present these special characteristics for completeness.

Substituting $0 \leq r \leq 1$ into (6) we note that $PP''_k(d_k) \geq 0$ for all $d_k \in \mathbb{R}$. This implies that $PP_k(d_k)$ is a convex function throughout the entire domain of $x \in \mathbb{R}$. In fact, as $r \mapsto 0$, $PP_k(d_k)$ tends to become more linear, achieving linearity at $r = 0$. In addition, not only is $PP_k(d_k)$ convex throughout its domain, it also does not attain its minimum. This can be seen by substituting $0 \leq r \leq 1$ into the expression $(5) = 0$ whose solutions give the stationary points of $PP_k(d_k)$. We note that a key assumption of these results is that $c^k_{ij} > 0$ for all $k \in K$ and $(i, j) \in A$, since routes with a cost of zero are undefined in the presented definition of partial profit. The advantage of convexity lies in the fact that there exist efficient ways of dealing with OF by using subgradients to under-approximate each $PP_k(x)$.

Figure 1 shows the shape of $PP_k(d_k)$ with respect to per unit routing cost d_k for three parameter values of r. For a given $k \in K$, note that the partial profit $PP_k(d_k)$ may take negative values, i.e. incur a loss, when the per unit routing cost d_k is greater than α_k. After attaining its minimum, which as seen from our previous analysis will always be a maximum loss, if the demand sensitivity parameter $r > 1$ then an increase in per unit routing cost decreases the loss. Therefore in this particular case, after a certain point, there is an incentive for the decision maker to route commodities on a longer path so as to dissuade demand of a particular commodity whose efficient routing does not globally compensate the revenue obtained.

On the other hand, if $r \in [0, 1]$ then the partial profit $PP_k(d_k)$ is a convex function that keeps decreasing as the routing cost increases therefore there would be no incentive to route commodities through a longer route. This flexibility in modeling demand behavior and its effect on the partial profit is one of the most important reasons for using the gravity model. It generalizes schemes such as demand decay functions and is able to capture phenomena such as the effect of hyper-sensitivity to routing costs ($r > 1$) on partial profit that may be missed from more conventional demand functions. Finally, we point out that in the context of our problem, the values $d_k(x)$ are bounded above and below by the most expensive (p^k_l) and the cheapest (p^k_s) possible routes, respectively, in the complete underlying graph. Therefore, depending on their values, the corresponding problem may consider only the convex part of PP_k when $r > 1$. However, in general, the problem to be solve is a non-linear optimization problem for all $r \geq 0$.

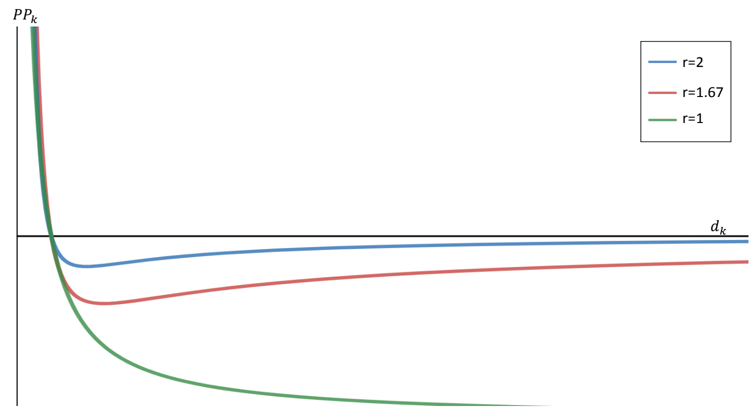


Figure 1: Shape of $\mathrm{PP}_k(\mathrm{d}_k)$

These nuances capture other levels of trade-off inherent to profit maximization where taking losses or offering worse service levels for some commodities compensates the overall profitability of an enterprise's operation. The proposed framework can also be used as a means of finding how to create commodity groups with better operation synergy, i.e. considering them together leads to decisions that do not dissuade demand by offering poorer service.

Comparing the shape of the partial profit for varying values of r, we note that the minimum value of partial profit is larger for greater values of r. This comes from the fact that larger values of r represent greater demand sensitivity to routing cost leading to fewer units of the commodity being routed at the minimum value of partial profit and therefore a smaller loss.

3. Formulations

We next propose two non-linear mixed integer programming formulations for the *profit-oriented network design problem with elastic demand* (POFMND-E) in which all commodities must be served. These formulations can be easily adapted to incorporate O/D pair selection decisions and service commitments. Both formulations fall in the domain of global optimization problems. Given the limited resources to efficiently solve these types of problems, we use one of the formulations as a base to develop a hybrid matheuristic that is applicable to any $r \geq 0$. It exploits known results of similar network design problems and converges in a reasonable amount of computation time. The other formulation serves as a benchmark for solving the POFMND-E with O/D pair selection and service commitments using a state-of-the-art general purpose global optimization software.

3.1. Arc-based formulation

Our arc-based formulation is based on the well-known strong formulation for the uncapacitated FMND with the difference that the routing variables x_{ij}^k are binary instead of continuous. In the case of POFMND-E, when the per unit revenue α_k of a commodity k is less than the cheapest possible route p_s^k, the model then seeks to make the loss resulting from $PP_k(x)$ as small as possible by increasing the length of the route as much as possible. Since flow conservation constraints alone do not prohibit circuits, formulations based only on the constraints of the classic uncapacitated FMND formulation lead to solutions having routes with sub-circuits within the path and isolated from it.

Given the inability of most global optimization software to allow for cut callbacks, we implement the subtour elimination constraints (SEC) via a modified version of the well-known, but weak, Miller-Tucker-Zemlin (MTZ) subtour elimination constraints (Miller et al. 1960). These require an additional set of variables u_i^k for each $i \in N, k \in K$ and that x_{ij}^k be binary leading to the following arc-based formulation. The variables u represent the order in which all nodes except the depot are visited.

$$(P_1) \quad \text{maximize} \sum_{k \in K} P_{o(k)} P_{d(k)} \frac{[\alpha_k - \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k]}{(\tau \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k)^r} - \sum_{(i,j)\in A} f_{ij} y_{ij} \tag{7}$$

$$\text{subject to} \quad \sum_{j \in N:(j,i)\in A} x_{ji}^k - \sum_{j \in N:(i,j)\in A} x_{ij}^k = \begin{cases} -1 & \text{if } i = o_k \\ 1 & \text{if } i = d_k \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in N, \forall k \in K \tag{8}$$

$$x_{ij}^k \le y_{ij} \quad \forall (i,j) \in A, k \in K \tag{9}$$

$$u_{o_k}^k = 1 \quad \forall k \in K \tag{10}$$

$$u_i^k - u_j^k + 1 \le (N-1)(1 - x_{ij}^k) \quad \forall (i,j) \in A, k \in K \tag{11}$$

$$2 \le u_i^k \le N \qquad \forall i \in N \backslash \{o_k\}, k \in K \tag{12}$$

$$x_{ij}^k \in \{0,1\} \qquad \forall (i,j) \in A, k \in K \tag{13}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in A. \tag{14}$$

The objective function (7) seeks to maximize profit, while constraints (8) and (9) are the flow conservation constraints and binding constraints from the classic uncapacitated FMND, respectively. Finally, constraints (10) and (11) are the MTZ subtour elimination constraints, while (12)-(14) are the variable definitions.

As a required input of the algorithms used in global optimization software, we provide the lower and upper bounds for the values of each $PP_k(x)$ as detailed in Section 2.2. This model is solved with the general-purpose global optimization solver Baron 18.8.23 (Tawarmalani and Sahinidis 2005, Sahinidis 2017, Kılınç and Sahinidis 2018) accessed through the AMPL modeling language.

3.2. Path-based formulation and pricing problem

We next present a path-based formulation of the problem. Let Θ_k^μ denote a binary variable whose value is equal to 1 if path μ is used for commodity k, and define the parameter $v_k^\mu(i,j)$=1 if arc (i,j) belongs to path μ for commodity k, 0 otherwise. Finally, let Ω_k denote the set of simple paths from $o(k)$ to $d(k)$ and Ω represent the union of these over $k \in K$. With this notation we have the following path-based formulation for the POFMND-E.

$$(P_2) \qquad \text{maximize} \sum_{k \in K} \sum_{\mu \in \Omega_k} [P_{o(k)} P_{d(k)} \frac{[\alpha_k - \sum\limits_{(i,j) \in A} c_{ij}^k v_k^\mu(i,j)]}{(\tau \sum\limits_{(i,j) \in A} c_{ij}^k v_k^\mu(i,j))^r}]\Theta_k^\mu - \sum_{(i,j) \in A} f_{ij} y_{ij} \tag{15}$$

$$\text{subject to } [\lambda] \quad \sum_{\mu \in \Omega_k} v_k^\mu(i,j)\Theta_k^\mu \leq y_{ij} \qquad \forall (i,j) \in A, \forall k \in K \tag{16}$$

$$[\mu] \quad \sum_{\mu \in \Omega_k} \Theta_k^\mu = 1 \qquad \forall k \in K \tag{17}$$

$$\Theta_k^\mu \in \{0,1\} \qquad \forall (i,j) \in A. \tag{18}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in A. \tag{19}$$

The objective function (15) calculates the total profit obtained from the selected network configuration. Constraint set (16) ensures that all design variables of the arcs used in the routing take value one while constraints (17) ensure that each commodity is routed through one path. Finally, $\lambda_{ij}^k \geq 0$, $\forall (i,j) \in A$, $\forall k \in K$ and $\mu_k \in \mathbb{R}$, $\forall k \in K$ are the dual variables corresponding to (16) and (17), respectively. We point out that, unlike P_1, the presented path-based formulation is a binary linear formulation.

Given the exponential number of potential paths that can be used to route each commodity, traditionally these are added on the fly to P_2 by means of a column generation algorithm which adds columns based on the solution of a pricing problem that determines whether new columns are needed to calculate the linear relaxation of P_2. In this case, the pricing problems of POFMND-E decompose to one for each commodity $k \in K$. These have the following form for each $k \in K$:

$$(Pr_2^k) \quad \text{maximize} \quad [P_{o(k)} P_{d(k)} \frac{[\alpha_k - \sum\limits_{(i,j) \in A} c_{ij}^k v_k^\mu(i,j)]}{(\tau \sum\limits_{(i,j) \in A} c_{ij}^k v_k^\mu(i,j))^r}] - \sum_{(i,j) \in A} \lambda_{ij} v_k^\mu(i,j) \tag{20}$$

$$\text{subject to} \quad \sum_{j \in N:(j,i) \in A} v_k^\mu(j,i) - \sum_{j \in N:(i,j) \in A} v_k^\mu(i,j) = \begin{cases} -1 & \text{if } i = o_k \\ 1 & \text{if } i = d_k \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in N \tag{21}$$

$$u_{o_k}^k = 1 \quad \forall k \in K \tag{22}$$

$$u_i^k - u_j^k + 1 \leq (N-1)(1 - v_k^\mu(i,j)) \quad \forall (i,j) \in A \tag{23}$$

$$2 \leq u_i^k \leq N \qquad \forall i \in N \backslash \{o_k\} \tag{24}$$

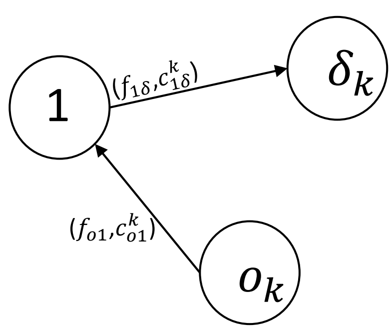
$$v_k^\mu(i,j) \in \{0,1\} \qquad \forall (i,j) \in A. \tag{25}$$

The objective function (20) seeks to maximize the reduced cost of the column to be added while constraints (21) are the classic flow conservation constraints. Constraints (22) and (23) are the MTZ subtour elimination constraints. Note that Pr_2^k is similar to P_1 in that it is a non-linear formulation, with binary variables and MTZ subtour elimination constraints. However, Pr_2^k is significantly smaller since it is defined for each $k \in K$ and the fixed costs are replaced by the dual variables of (16). It is a modified shortest path problem with a non-linear objective function. The difficult non-convexity present in P_1 is transferred to the pricing problem Pr_2^k. Our efforts will therefore be placed in finding alternative ways of adding new columns to P_2 by heuristically solving an easier problem based on Pr_2^k.

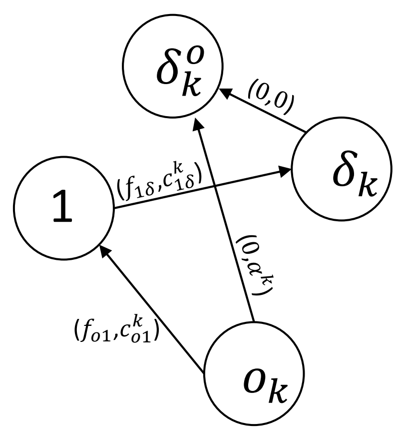
3.3. Incorporating O/D pair selection and service commitments

Formulations P_1 and P_2 do not account for O/D pair selection and service commitment. Both formulations allow for the convenient incorporation of both considerations by either performing simple transformations on the network or defining artificial variables, and adding a knapsack-type constraint, respectively. For P_1 to allow O/D pair selection, the only modification required is done to the underlying graph used in the formulation. The following network transformation incorporates O/D pair selection into P_1.

Let $\Delta = \{i \in N | \exists (o,d) \in K \text{ such that } i = d\}$ denote the set of nodes that are destinations of some commodity $k \in K$. For each node $\delta \in \Delta$ create an artificial node δ^o and an arc $(\delta, \delta^o) \in A$ with $f_{\delta\delta^o} = 0$ and $c^k_{\delta\delta^o} = 0$ for all $k \in K$. In addition, redefine all commodities (o, δ) as $(o, \delta^o) \in K$ and add another artificial arc $(o, \delta^o) \in A$ with $f_{\delta\delta^o} = 0$ and $c^k_{\delta\delta^o} = \alpha_k$ for all $k \in K$. Figure 2 illustrates the proposed transformation.



(a) Original network



(b) Transformed network

Figure 2: Network transformation to allow O/D pair selection

Note that by carrying out this transformation and using the resulting network in P_1, its solution accounts for the decision maker selecting which commodities to route. If a commodity $\bar{k} \in K$ used the arc $(o, \delta^o) \in A$, it obtains no revenue and incurs no fixed cost. In other words, the model has selected not to route commodity $\bar{k} \in K$. On the other hand, if a commodity uses the arc $(\delta, \delta^o) \in A$, its routing, fixed cost, and corresponding revenue remain the same as if it was routed from (o, δ) in the original network.

Finally, to impose a service commitment constraint in P_1 stating that at least Γ commodities are to be routed, the following knapsack type constraint on the arc variables of the transformed network should be added:

$$\sum_{(o,\delta^o)\in K} x^{o\delta^o}_{o\delta^o} \leq |K| - \Gamma. \tag{26}$$

Incorporating O/D pair selection in P_2 requires significantly less work. The addition of an empty route $\gamma_{\bar{k}}$, with $PP_{\bar{k}(x)} = 0$ for each commodity $\bar{k} \in K$ into the pool of routes Ω_k is enough to account for O/D pair selection. The associated binary variables for these empty routes will be denoted as $\Theta^{\mu^0}_{\bar{k}}$. If in a solution, the associated $\Theta^{\mu^0}_{\bar{k}}$ is equal to 1, then commodity $\bar{k}$ is not routed.

Imposing a service commitment in P_2 where at least Γ commodities are to be routed is also done by incorporating a knapsack-type constraint. In this case, the inequality is defined over the artificial variables $\Theta^{\mu^0}_{\bar{k}}$ of the empty routes and has the following form:

$$\sum_{k\in K} \Theta^{\mu^0}_{k} \leq |K| - \Gamma. \tag{27}$$

The network transformations or empty routes can be added as a preprocessing step leaving the incorporation of either inequality (26) or (27) as the only modification of P_1 and P_2, respectively, to incorporate O/D pair selection and service commitment. These transformations allow us to formulate the POFMND-E with commodity selection and a service commitment of Γ, denoted as POFMND-E(Γ). $P_1(\Gamma)$ is defined as P_1 with the addition of constraint (26) while $P_2(\Gamma)$ is defined as P_2 with the addition of constraint (27).

In the interest of brevity, particular attention will be placed on the two extreme cases $\Gamma = |K|$ and $\Gamma = 0$ referred to as variants I and II, respectively. Note that for both cases, one can omit the service commitment constraints (26) or (27). In variant I, it suffices to not carry out the network transformation or not define the empty routes in P_1 or P_2, respectively. In variant II, the service commitment constraints become redundant as the right hand side value $|K|$ is the maximum possible value that can be taken.

We next present a numerical example to show the value of incorporating demand elasticity in profit-oriented network design. Although the example presents a comparison for variant I, a similar conclusion can be derived for the general POFMND-E(Γ).

3.4. The value of considering demand elasticity

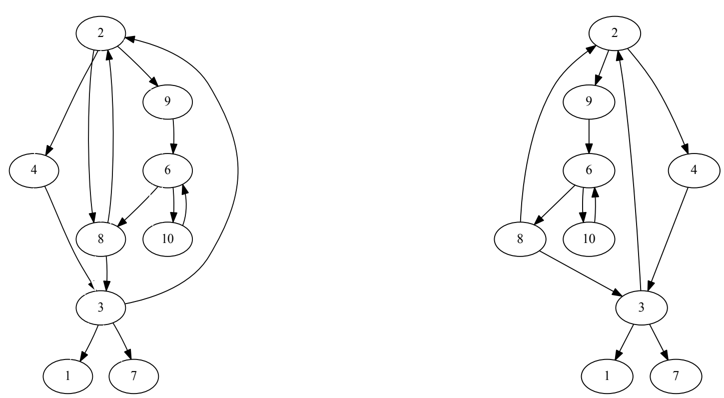
To demonstrate the value of considering demand elasticity within POFMND, we solve variant I for a small example of a network with 10 nodes numbered from one to ten, 35 arcs and ten commodities. Table 1 details the parameters of the ten commodities: origin, destination, population at origin, population at destination, per unit revenue and cheapest route over the complete network, respectively.

Table 1: Commodity parameters of example

o_k	d_k	P_o	P_d	α_k	p_s^k
2	8	524	1,076	93	93
3	1	228	744	141	76
4	2	792	524	210	122
4	7	792	160	189	106
6	1	640	744	245	141
6	10	640	448	53	41
8	1	1,076	744	220	117
8	6	1,076	640	157	93
9	8	292	1,076	215	123
10	4	448	792	304	164

As mentioned in Section 2, POFMND-E obtains solutions with equilibrium-like characteristics between demand quantity and routing cost. Therefore, to formulate an equivalent profit-oriented FMND with inelastic demand (POFMND-I), we must assume a fixed demand quantity for each commodity. A demand quantity value that exploits all information in the instance data is to assign the demand quantity between nodes i and j as $W_{ij} = \frac{P_i P_j}{(p_s^k)^r}$. This is an optimistic value of demand quantity as it assumes all commodities are routed along the path with least travel cost over the complete underlying network. Note that by fixing demand quantity, the model becomes a binary linear program which can then be solved by any of a plethora of available integer linear programming tools.

Figure 3a is the optimal solution POFMND-I while Figure 3b is the optimal solution of POFMND-E.



(a) POFMND-I (b) POFMND-E

Figure 3: Optimal solutions POFMND where all commodities are served

We note that at a design level, the optimal solution POFMND-I has one additional arc from node $2 \rightarrow 8$ as a result of routing commodity (2,8) directly from origin to destination. POFMND-E instead routes this commodity along nodes $2 \rightarrow 9 \rightarrow 6 \rightarrow 8$ with a routing cost of 136. POFMND-I routes all commodities except (6,1) along their cheapest possible path, p_s^k, therefore their corresponding demand quantity values coincided with that obtained from the presented gravity model. To ensure the solutions of both models are comparable, we substitute the solution obtained from POFMND-I into the objective function of POFMND-E. The objective function value obtained from POFMND-I is 9.55% worse than that of POFMND-E. This comes as a result of POFMND-I not having the flexibility to identify the overall benefit of routing commodity (2,8) along a more costly, unprofitable route with less demand. We next present a hybrid matheuristic algorithm for solving P_2.

4. Solving the path-based formulation

To exploit existing algorithms for classic network design problems, we replace the use of Pr_2^k with the classic uncapacitated FMND formulation to generate new paths. By doing this, we are no longer able to obtain dual bounds to assess the solution quality. However, given that our purpose is to develop a fast heuristic adaptable to any value of $r \geq 0$, the loss of dual bounds does not hinder our purpose. On the other hand, the classic uncapacitated FMND is itself solved by a metaheuristic since its main purpose is to generate a varied set of paths. As we will see later on, the solutions obtained from our proposed heuristic are optimal for most cases for which the general purpose global optimization solver was able to prove optimality.

4.1. A hybrid matheuristic for the path-based formulation

We begin by presenting the logic behind the tools used for the matheuristic part of the proposed solution algorithm. For each $k \in K$ let $W_k = \frac{P_{o(k)}P_{d(k)}}{(\tau \sum_{(i,j)\in A} c_{ij}^k v_k^{\mu}(i,j))^r}$ then:

$$\text{maximize} \sum_{k\in K} P_{o(k)}P_{d(k)} \frac{[\alpha_k - \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k]}{(\tau \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k)^r} - \sum_{(i,j)\in A} f_{ij} y_{ij} \tag{28}$$

$$\iff \text{maximize} \sum_{k\in K} W_k[\alpha_k - \sum_{(i,j)\in A} c_{ij}^k x_{ij}^k] - \sum_{(i,j)\in A} f_{ij} y_{ij} \tag{29}$$

$$\iff \text{minimize} \sum_{(i,j)\in A} f_{ij} y_{ij} + \sum_{k\in K} \sum_{(i,j)\in A} W_k c_{ij}^k x_{ij}^k \tag{30}$$

where the equivalence between (29) and (30) comes from the fact that $W_k\alpha_k$ is now a constant. Expression (30) is in fact the objective function of the uncapacitated FMND and methods for this problem can thus be applied to obtain new routes, update W_k based on the best routes found, and then repeat.

Given that the uncapacitated FMND instance obtained assumes a particular value for W_k, it is not in our interest to solve it exactly. A more suitable strategy is to use a metaheuristic that generates diverse routes for each commodity. In this spirit, we use a slope-scaling metaheuristic (SS) based on that presented in Crainic et al. (2004) for the capacitated FMND. This heuristic will be detailed later on in this section.

We next outline the algorithm which embeds the slope-scaling metaheuristic within two loops, the first updates the assumed W_k while the second solves a relaxation of the path-based formulation and updates the transportation costs of the network used for the slope-scaling.

We begin by initializing W_k based on the maximum possible demand, ∇_k, calculated by substituting the cheapest route p_s^k into (1). If the per unit revenue α_k of commodity k is less than the cheapest route p_s^k, we initialize W_k as $0.2\nabla_k$. Otherwise we initialize it as $W_k = \nabla_k$.

The slope-scaling metaheuristic is then called and all routes generated during the process are stored in Ω with the proper evaluation of $PP_k(x)$. W_k is then updated based on the routes of the best solution found, with respect to the uncapacitated FMND, and the process is then repeated. This loop is terminated when the values of W_k no longer change. We then add all generated paths to the formulation $\mathrm{P}_2(\Gamma)$ and solve its linear relaxation. The values of λ in $\mathrm{P}_2(\Gamma)$ are then used to update the transportation costs of each commodity $k \in K$ on each arc $(i,j) \in A$ as $c_{ij}^k = c_{ij}^k + \lambda_{ij}^k$ and the entire process is repeated.

The purpose of this algorithm is to keep generating new routes based on modified demand quantities and transportation costs. Greater diversification leads to a richer pool Ω of routes in the master problem $\mathrm{P}_2(\Gamma)$. We denote this problem as $\mathrm{P}_2(\Gamma, \Omega(t))$. The final loop is terminated when no new paths have been generated. After this step, we then proceed to solve the restricted master problem $\mathrm{P}_2(\Gamma, \Omega(t))$ as an integer program to obtain a heuristic solution. Algorithm 1 summarizes the proposed hybrid matheuristic.

Algorithm 1 Hybrid matheuristic for POFMND-E

```
Initialization: t = 0; Ω(t) = ∅; W_k(0) = ∇_k if α_k > p_s^k, W_k(0) = 0.2∇_k otherwise;
do
    do
        Execute SS; add all generated routes μ_k to Ω_k(t); best solution=(ȳ, x̄).
        Update W_k(t+1) = (P_o(k) P_d(k)) / (τ Σ_{(i,j)∈A} c_ij^k x̄_ij^k)^r
        t=t+1
    while (∃k such that |W_k(t+1) − W_k(t)| > 0)
    Solve LP of restricted master problem P_2(Γ, Ω(t)); obtain dual variables λ.
    Update c_ij^k = c_ij^k + λ_ij^k
while (|Ω(t)| > |Ω(t−1)|)
Solve restricted master problem P_2(Γ, Ω(t)) with integrality constraints.
```

We point out that in general the metaheuristic SS can be substituted by any metaheuristic that generates a rich variety of paths. Other metaheuristics such as local and neighborhood searches can also prove effective if embedded within this hybrid matheuristic. The key to the complete procedure is to keep information at the path level therefore allowing P_2 the freedom to choose paths that were previously not considered together during the metaheuristic phase. A similar logic is used in Zetina et al. (2018).

4.2. A Slope Scaling Metaheuristic

Slope scaling was first presented in Yaged (1971) and later in Kim and Pardalos (1999, 2000) as a heuristic to solve network optimization problems. Crainic et al. (2004) improved it by adding Lagrangean perturbation and long term memory to help in diversifying and intensifying the search. The method is based on the idea that there exists a linear program of the form

$$\text{(SS)} \quad \min \sum_{k \in K} \sum_{(i,j) \in A} \hat{c}_{ij}^k x_{ij}^k \tag{31}$$

$$\text{s.t.} \quad \sum_{j \in N} x_{ji}^k - \sum_{j \in N} x_{ij}^k = \begin{cases} -W^k & \text{if } i = o_k \\ W^k & \text{if } i = d_k \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in N, \forall k \in K \tag{32}$$

$$0 \le x_{ij}^k \le W^k \qquad \forall (i,j) \in A, k \in K, \tag{33}$$

that obtains the same optimal solution as the uncapacitated FMND. The algorithm attempts to estimate the $\hat{c}_{ij}^k$ for which this equivalence holds by defining it as $\hat{c}_{ij}^k = c_{ij}^k + \rho_{ij}^k$, where ρ_{ij}^k is a slope scaling factor that estimates the contribution of the fixed costs. An initial ρ_0 is chosen to begin the algorithm. At each iteration t, SS$(\rho(t))$ is solved and its solution is used to obtain ρ_{t+1}.

For our implementation, SS is split into $|K|$ shortest path problems with arc lengths of $\hat{c}_{ij}^k$. We use a multi-start method with different initial values of ρ based on the fixed cost and demand quantity. Let $\bar{W}^k = \sum_{k \in K} W^k/|K|$, i.e. the average demand quantity of the commodities. The initial values for ρ are

- $f_{ij}/\sum_{k \in K} W^k$: the fixed cost divided by the total demand quantity;
- $f_{ij}/\max_{k \in K} W^k$: the fixed cost divided by the largest demand quantity;
- $f_{ij}/\bar{W}^k$: the fixed cost divided by the average demand quantity;
- $f_{ij}/[(\max_{k \in K} W^k - \bar{W}^k)/2]$: the fixed cost divided by the mid-point between the average and maximum demand quantities;
- $f_{ij}/[(\sum_{k \in K} W^k - \bar{W}^k)/2]$: the fixed cost divided by the mid-point between the average and total demand quantities;
- $f_{ij}/[(\sum_{k \in K} W^k - \max_{k \in K} W^k)/2]$: the fixed cost divided by the mid-point between the total and maximum demand quantities.

Upon obtaining the optimal solution $\tilde{x}$ of SS$(\rho(i))$, the slope scaling factor is updated as

$$\rho_{ij}^k(i+1) = \begin{cases} \frac{f_{ij}}{\sum_{k \in K} \tilde{x}_{ij}^k} & \text{if } \sum_{k \in K} \tilde{x}_{ij}^k > 0 \\ \rho_{ij}^k(i) & \text{otherwise.} \end{cases} \tag{34}$$

This process is continued until a given number of iterations (T_{SS}) have been performed. Note that upon solving SS, a feasible solution can be constructed for FMND by fixing to 1 the arcs through which some flow has been sent and solving $|K|$ shortest path problems over this subgraph. To improve the quality of the solution, we then remove any arcs of the subgraph that have not been used in the shortest path of at least one of the commodities.

When two successive iterations obtain the same solution $\tilde{x}$, then the procedure will not produce any new distinct solutions. This may occur before having performed the T_{SS} iterations. To aid in diversifying the metaheuristic's search, we implement a perturbation tool similar to that of Crainic et al. (2004). When two successive iterations of solving SS obtain the same optimal value or a determined number of iterations T_{pert} without an improved solution have passed, we then solve a shortest path problem for each $k \in K$ and use the corresponding dual variables (λ_k, μ_k) of the classic shortest path formulation to update the slope scaling factor as $\rho_{ij}^k = -\lambda_j^k + \lambda_i^k + \mu_{ij}^k$ and continue iterating until a maximum number T_{SS} of SS models have been solved.

Finally, as in Crainic et al. (2004), we implement a long term memory mechanism in which we keep statistics throughout the history of the search. These statistics are used to update ρ and restart the process. Based on whether the current round of the algorithm produced an improved best solution, we choose to update ρ in such a way to intensify or diversify the search. The statistics kept for each $(i, j) \in A$ and $k \in K$ up to iteration T are the number of iterations for which $\tilde{x}_{ij}^k > 0$ $(n_{ij}^k(T))$, average number of commodities routed through each arc $(\bar{x}_{ij}(T))$ and maximum number of commodities routed through each arc $(\hat{x}_{ij}^k(T))$.

After performing T_{SS} iterations with the corresponding dual perturbations along the way, we calculate for each $(i, j) \in A$, $v_{ij} = \bar{x}_{ij}(T_{SS})/\hat{x}_{ij}(T_{SS})$ or $v_{ij} = 0$ if $\bar{x}_{ij}(T_{SS}) = 0$. Here, v_{ij} measures the variability of the number of commodities sent through arc (i, j) throughout the last T_{SS} iterations. Hence, $v_{ij} \approx 1$ means the number of commodities sent through this arc has been stable throughout the process, while $v_{ij} \approx 0$ shows high variability or no commodities sent at all. During intensification, variables with stable behavior are favored while the opposite is done when a diversification step is taken.

An intensification update to ρ is done if in the last cycle an improved best solution was obtained. Otherwise, a diversification step is taken. A limit of div_{max} and int_{max} diversification and intensification updates, respectively, are applied throughout the algorithm. The updates for each scheme are presented below where $\bar{n}$ and S_n are the average and standard deviation of n_{ij}^k, respectively.

- Normalize $\rho_{ij}^k := \rho_{ij}^k - \min\limits_{(i,j) \in A, k \in K} \rho_{ij}^k$ so $\rho_{ij}^k \geq 0 \; \forall (i, j) \in A, k \in K$.
- To apply the intensification scheme, $\forall (i, j) \in A, k \in K$

- If $n_{ij}^k \geq \bar{n} + S_n$ then $\rho_{ij}^k := \rho_{ij}^k(1 - v_{ij})$
- If $n_{ij}^k \leq \bar{n}$ then $\rho_{ij}^k := \rho_{ij}^k(2 - v_{ij})$
- Else $\rho_{ij}^k := \rho_{ij}^k$.

• To apply the diversification scheme, $\forall(i,j) \in A, k \in K$

- If $n_{ij}^k \geq \bar{n} + S_n$ then $\rho_{ij}^k := \rho_{ij}^k(1 + v_{ij}^k)$
- If $n_{ij}^k \leq \bar{n}$ then $\rho_{ij}^k := \rho_{ij}^k(v_{ij}^k)$
- Else $\rho_{ij}^k := \rho_{ij}^k$.

This metaheuristic is called several times within our hybrid matheuristic to generate new paths for which we evaluate the true value of $PP_k(x)$. For this reason, it is stopped prematurely if two consecutive dual perturbations do not lead to any new paths. Our heuristic is run with the parameter values $T_{SS} = 30, T_{DP} = 5, div_{max} = 10, int_{max} = 2$. Figure 4 graphically summarizes the slope scaling with dual perturbation and long term memory algorithm sequentially applied to each of the six initial ρ values.

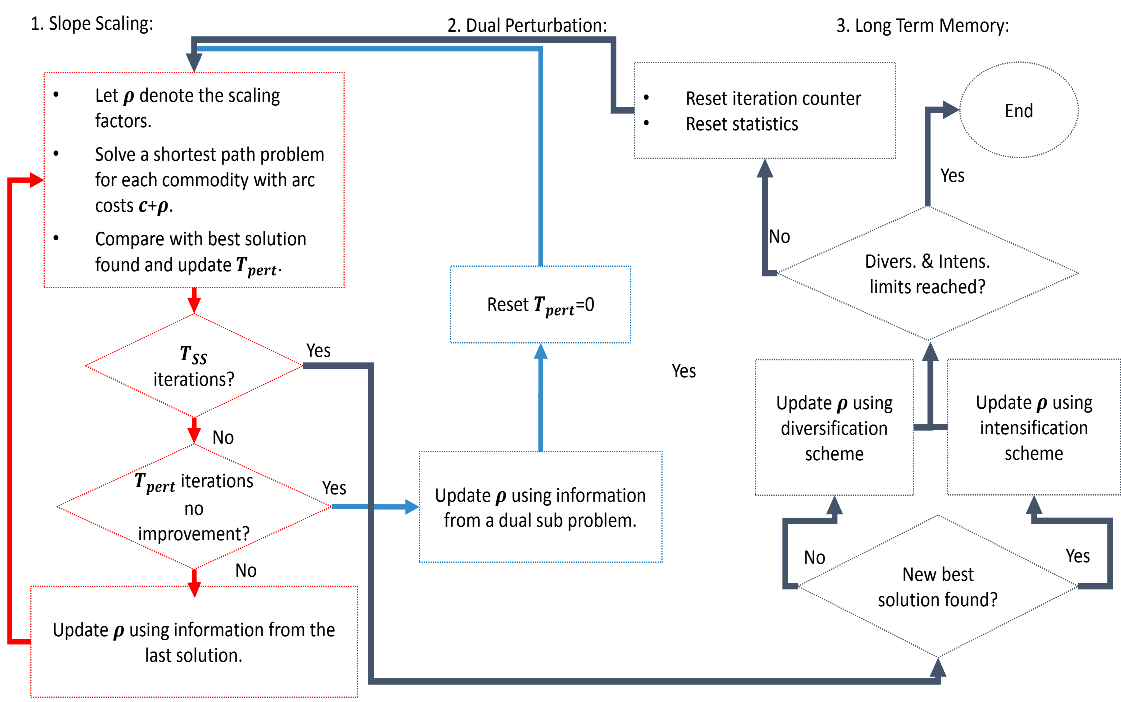


Figure 4: Slope scaling heuristic for MUFND

5. Computational Experiments

We test the computational efficiency of the proposed formulations and solution algorithms using the well-known "Canad" multicommodity capacitated network design testbed (Crainic et al. 2001). The population weights of each node P_i are calculated as the sum of all inbound and outbound demand in each node multiplied by four. To adjust for nodes $i \in N$ with no inflow or outflow, we assign them population values of P_{min} which denotes the minimum non-zero population calculated. Per unit revenue of each commodity is $\alpha_k = p_s^k + \sigma p_s^k$ where σ is a random number between 0 and 1 and p_s^k is the cheapest possible route, i.e. the shortest path over the complete network. Finally, to adjust for the population factors P_i for $i \in N$, the value of the fixed costs of an arc (i, j) is adjusted based on the original value of the Canad instance $\bar{f}_{ij}$ as follows: $f_{ij} = \frac{\bar{f}_{ij} P_i P_j}{\nu_{ij}}$ where $\nu_{ij} = P_{min} + .75[\max\{P_i, P_j\} - P_{min}]$. In total our testbed is comprised of 85 instances ranging from small to medium scale. The computational experiments will first focus on the two extreme variants of POFMND-E(Γ), where variant I refers to $\Gamma = |K|$ and variant II refers to $\Gamma = 0$. As previously mentioned, in these extreme cases one can omit the service commitment constraint.

The arc-based formulation P_1 is solved using the branch-and-reduce algorithm implemented in the general purpose global optimization software Baron 18.8.23 (Tawarmalani and Sahinidis 2005, Sahinidis 2017, Kılınç and Sahinidis 2018) through its AMPL interface. The hybrid matheuristic is coded in C using CPLEX 12.7.0 to solve the restricted master problem and the shortest path subproblems within the slope-scaling metaheuristic. For a fair comparison, all use of CPLEX was limited to one thread and the traditional MIP search strategy. All experiments were executed on an Intel Xeon E5 2687W V3 processor at 3.10 GHz under Linux environment with a time limit of two hours. Finally, we fix the parameters of POFMND-E to $\tau = 1$ and $r = 1.7$. Tables 2 and 3 compare the performance of P_1 solved using Baron and the proposed hybrid matheuristic for variant I and II, respectively.

The first three columns of Table 2 contain the instance class (Class), dimensions (N,A,K) representing the number of nodes, arcs and commodities respectively, and the number of instances in each group (Nb.). The next two columns contain the number of instances for which a feasible solution was found by the hybrid matheuristic and its average time to completion in seconds. The following three columns correspond to the number of instances for which solving P_1 with Baron found a feasible and optimal solution, respectively, while the last column details the average time taken.

Table 2: Performance comparison- Variant I

			Hybrid Matheuristic		Baron		
Class	**(N,A,K)**	**Nb.**	**Feasible Sols**	**Seconds**	**Feasible Sols**	**Optimal Sols**	**Seconds**
I	20,230,40	3	3	0.47	3	0	7,200.00
	20,230,200	4	4	1.17	0	0	7,200.00
	20,300,40	4	4	0.50	4	0	7,200.00
	20,300,200	4	4	1.62	0	0	7,200.00
	30,520,100	4	4	2.70	0	0	7,200.00
	30,520,400	4	4	7.64	0	0	7,200.00
	30,700,100	4	4	3.09	0	0	7,200.00
	30,700,400	4	4	9.73	0	0	7,200.00
	Subtotal	**31**	**31**	**3.45**	**7**	**0**	**7,200.00**
III-A	10,35,10	3	3	0.04	3	3	99.67
	10,35,25	3	3	0.08	3	1	4,807.00
	10,35,50	3	3	0.12	3	0	7,200.00
	10,60,10	3	3	0.03	3	3	13.00
	10,60,25	3	3	0.14	3	1	4,862.00
	10,60,50	3	3	0.23	3	0	7,200.00
	10,85,10	3	3	0.04	3	1	4,808.00
	10,85,25	3	3	0.14	3	0	7,200.00
	10,85,50	3	3	0.29	3	0	7,200.00
	Subtotal	**27**	**27**	**0.12**	**27**	**9**	**4,821.81**
III-B	20,120,40	3	3	0.78	3	0	7,200.00
	20,120,100	3	3	2.36	3	0	7,200.00
	20,120,200	3	3	4.04	2	0	7,200.00
	20,220,40	3	3	1.94	3	0	7,200.00
	20,220,100	3	3	2.42	3	0	7,200.00
	20,220,200	3	3	5.22	0	0	7,200.00
	20,320,40	3	3	1.66	3	0	7,200.00
	20,320,100	3	3	3.25	3	0	7,200.00
	20,320,200	3	3	5.24	0	0	7,200.00
	Subtotal	**27**	**27**	**2.99**	**20**	**0**	**7,200.00**
	Total	**85**	**85**	**2.25**	**54**	**9**	**6,551.46**

As seen from Table 2 not only is the hybrid matheuristic able to find a feasible solution for all instances compared to only 54/85 for solving P_1 with Baron, it also requires significantly less time. In addition, for nine of the small instances of Class III-A, Baron's branch-and-reduce method applied to P_1 is able to find and prove an optimal solution. These solutions coincide with that obtained by our hybrid matheuristic for eight of the nine instances while for the remaining instance, the solution found by the hybrid matheuristic is 0.29% away from the optimal. Similar behavior can be observed when solving variant II of POFMND-E that allows for the decision maker to freely select a subset of commodities to route. Table 3 presents a summary of the performance of both algorithms applied to variant II with the same column definitions as Table 2.

Table 3: Performance comparison- Variant II

		Hybrid Matheuristic		Baron		
Class	**Nb.**	**Feasible Sols**	**Seconds**	**Feasible Sols**	**Optimal Sols**	**Seconds**
Class I	31	31	3.29	11	0	6,769.90
Class III-A	27	27	0.13	27	14	3,511.56
Class III-B	27	27	2.81	22	0	6,654.04
Total	**85**	**85**	**2.14**	**60**	**14**	**5,698.09**

We next analyze the quality of the solutions found. Based on the results seen in Tables 2 and 3, we note that solving P_1 with Baron requires a significant amount of computation time. To assess whether this effort is compensated by better solution quality, we analyze the instances for which both algorithms obtained a feasible solution within the time limit. The study is limited to instances for which optimality of the solution was not proven for the classes I and III-B, i.e. the execution of Baron reached the time limit before proving optimality. Table 4 details the objective function value, in millions, of the feasible solutions found in Table 2 by the hybrid matheuristic and Baron, respectively. The last column contains the relative difference between them calculated as $\frac{100*|HM-Bar|}{|HM|}$ where HM refers to the value of the solution from the hybrid matheuristic and Bar refers to the value of the solution from Baron.

Table 4: Comparison of solution quality-Variant I

		Hybrid Matheuristic	Baron	
Class	**instance**	**Obj (Millions)**	**Obj Baron (Millions)**	**Rel. Difference**
I	c33.dat	229.95	9.60	95.83
	c35.dat	106.99	26.14	75.57
	c36.dat	132.14	12.24	90.74
	c41.dat	225.16	(3.92)	101.74
	c42.dat	245.59	(23.44)	109.54
	c43.dat	213.82	(42.58)	119.92
	c44.dat	173.11	(62.10)	135.88
III-B	r10.1.dow	1.20	0.60	50.26
	r10.2.dow	0.86	0.12	86.07
	r10.3.dow	0.52	(0.24)	145.67
	r11.1.dow	10.65	(1.43)	113.41
	r11.2.dow	9.18	(1.87)	120.38
	r11.3.dow	7.61	(2.38)	131.21
	r12.2.dow	89.94	(15.64)	117.39
	r12.3.dow	84.47	(17.19)	120.35
	r13.1.dow	1.56	(0.21)	113.50
	r13.2.dow	1.18	(0.41)	134.88
	r13.3.dow	0.76	(0.60)	177.92
	r14.1.dow	15.66	(5.06)	132.33
	r14.2.dow	14.01	(5.34)	138.13
	r14.3.dow	12.04	(5.32)	144.17
	r16.1.dow	1.50	(0.48)	132.16
	r16.2.dow	1.09	(0.53)	148.39
	r16.3.dow	0.67	(0.61)	191.48
	r17.1.dow	19.32	(6.22)	132.19
	r17.2.dow	17.38	(5.41)	131.14
	r17.3.dow	15.18	(5.97)	139.32
Total		**60.43**	**(5.86)**	**109.70**

Table 4 shows that the solutions found by the proposed hybrid matheuristic are of significantly better quality. It finds profitable solutions for all instances in an average of 2.25 seconds as shown in Table 2. The results show that not only is it difficult for Baron to find feasible solutions with formulation P_1 but also that the solutions found are of poor quality. This comes as a result of the solver not exploiting the network structure of the problem whereas the use of our slope scaling heuristic bypasses the difficulty presented by the non-convexities of Pr_2^k to generate good paths.

The proposed hybrid matheuristic also obtains superior solutions than solving P_1 with Baron when solving variant II. Table 5 contains the instance class (Class), dimensions (N,A,K) representing the number of nodes, arcs and commodities respectively, and the number of instances considered in each group (Nb.). The following two columns represent the average objective function value of the best found solution for each instance group while the last column represents the average relative difference between them calculated as in Table 4.

Table 5: Comparison of solution quality-Variant II

Class	(N,A,K)	Nb.	Hybrid Matheuristic Obj (Millions)	Baron Obj (Millions)	Rel. Difference
I	20,230,40	2	119.57	51.15	58.96
	20,230,200	1	576.78	576.83	0.01
	20,300,40	2	229.72	58.79	74.25
	30,520,100	2	26.25	13.91	46.18
	Subtotal	**7**	**189.69**	**117.79**	**51.25**
III-B	20,120,40	2	0.82	0.77	6.00
	20,120,100	2	8.49	8.27	2.46
	20,120,200	1	84.69	5.57	93.42
	20,220,40	3	1.22	0.46	61.78
	20,220,100	2	14.88	7.90	49.47
	20,320,40	2	0.93	0.00	99.50
	20,320,200	3	159.54	52.71	67.04
	Subtotal	**15**	**41.15**	**13.27**	**52.98**
	Total	**35**	**88.41**	**46.52**	**52.43**

As opposed to the previous analysis, the solutions found by Baron are all profitable. This comes from the added flexibility of being able to freely select which commodities to route. In addition, the difference between the best solutions found by both algorithms is less significant than for variant I. Unlike the behavior seen when solving variant I, Baron was able to find a better solution than the proposed hybrid matheuristic for an instance in the 20,230,200 group of class I. This shows the increased difficulty of solving variant I with general purpose global optimization solvers. Being obliged to route all commodities forces the decision maker to consider influencing demand quantity of commodities with lower margins.

We next compare the results and solution process of the hybrid matheuristic for both variants. Table 6 details for each instance group, the number of instances, average objective function value of the best solution found in millions (Obj Millions), the average CPU time (Seconds), and average number of master problem iterations (CG Iters) of variant I and II. In addition, the last columns present the average % of commodities not routed in the solution obtained from variant II (% Unserved) and the average relative increase in profit between the solutions obtained from variants I and II.

Table 6: Variant I vs II

			Variant I			Variant II				
Class	(N,A,K)	Nb.	Obj (Millions)	Seconds	CG Iters	Obj (Millions)	Seconds	CG Iters	% Unserved	% Profit Inc.
I	20,230,40	3	156.36	0.47	3.00	156.37	0.35	3.00	4.19	0.00
	20,230,200	4	548.92	1.17	3.00	548.96	1.24	3.00	24.00	0.01
	20,300,40	4	214.42	0.50	3.25	214.43	0.47	3.00	7.50	0.01
	20,300,200	4	667.60	1.62	3.00	667.63	1.55	3.00	31.00	0.00
	30,520,100	4	23.21	2.70	3.00	23.22	2.72	3.25	16.25	0.04
	30,520,400	4	2,132.69	7.64	3.00	2,132.71	6.17	3.00	24.44	0.00
	30,700,100	4	29.70	3.09	3.00	29.70	2.63	3.00	16.75	0.02
	30,700,400	4	2,427.40	9.73	3.00	2,427.43	10.51	3.00	30.19	0.00
	Subtotal	**31**	**795.00**	**3.45**	**3.03**	**795.01**	**3.29**	**3.03**	**19.78**	**0.01**
III-A	10,35,10	3	0.05	0.04	3.33	0.06	0.02	3.00	26.67	45.65
	10,35,25	3	0.68	0.08	3.00	0.70	0.09	3.33	16.00	2.65
	10,35,50	3	4.53	0.12	3.00	4.55	0.14	3.33	3.33	0.40
	10,60,10	3	0.10	0.03	3.67	0.11	0.03	3.00	20.00	9.35
	10,60,25	3	0.75	0.14	4.00	0.77	0.17	3.00	10.67	3.07
	10,60,50	3	6.90	0.23	3.33	6.90	0.31	3.33	0.00	0.00
	10,85,10	3	0.06	0.04	3.67	0.07	0.03	3.33	26.67	31.71
	10,85,25	3	1.05	0.14	4.00	1.05	0.12	3.00	5.33	0.51
	10,85,50	3	6.19	0.29	3.00	6.23	0.25	3.00	8.67	0.71
	Subtotal	**27**	**2.26**	**0.12**	**3.44**	**2.27**	**0.13**	**3.15**	**13.04**	**10.45**
III-B	20,120,40	3	0.86	0.78	5.00	0.95	0.69	3.33	28.33	15.02
	20,120,100	3	9.15	2.36	3.33	9.22	2.39	3.00	9.33	0.83
	20,120,200	3	89.67	4.04	4.00	89.76	3.72	3.33	3.33	0.11
	20,220,40	3	1.17	1.94	5.00	1.22	1.98	4.33	25.00	6.68
	20,220,100	3	13.90	2.42	3.33	14.04	2.39	3.00	9.67	1.10
	20,220,200	3	131.42	5.22	3.33	131.53	4.71	3.67	4.50	0.09
	20,320,40	3	1.08	1.66	4.67	1.12	1.11	3.33	18.33	5.05
	20,320,100	3	17.29	3.25	3.33	17.45	3.78	3.00	9.00	0.99
	20,320,200	3	159.19	5.24	3.00	159.54	4.51	3.00	4.33	0.23
	Subtotal	**27**	**47.08**	**2.99**	**3.89**	**47.20**	**2.81**	**3.33**	**12.43**	**3.34**
	Total	**85**	**305.61**	**2.25**	**3.44**	**305.66**	**2.14**	**3.16**	**15.30**	**4.39**

One of the important characteristics of the proposed hybrid matheuristic is its ability to adapt to varying values of r and variants of the problem without it significantly modifying its performance. Table 6 confirms the latter to be true. Both require similar computation time and number of master problem iterations. Allowing the decision maker to freely choose which commodities to route (variant II), leads to an average of 4.39% increase in profit obtained by choosing not to route an average of approximately 15% of the commodities.

We point out that this marginal increase in profitability is highly dependent on the instance data. Since in the generated instances there are no commodities whose revenue is smaller than the shortest possible path, it is not surprising that on average there is not a significant difference between the profit of the best solutions for variants I and II. The same would not be said if the per unit revenue of some commodities was less than their corresponding cheapest route.

We next analyze one of the instances in the testbed with highest percentage of commodities being left unserved (50%) for variant II. The instance consists of ten nodes, 35 arcs and ten commodities. We solve the corresponding POFMND-E(Γ) for each service commitment level $\Gamma \in \{1, 2, .., 10\}$. Figure 5 plots the total profit (in thousands) of the optimal solution against the service commitment value Γ.

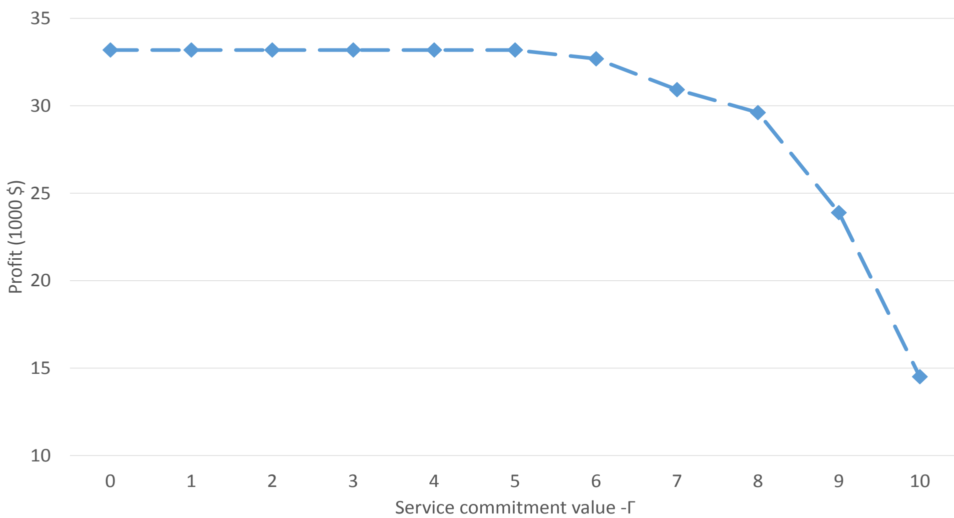


Figure 5: The effect of imposing service commitment constraints

Note that imposing a service commitment value of up to 5 commodities has no effect on the profit obtained. This coincides with what was seen from the result of variant II where 50% of the commodities were served despite no service commitment being imposed. As the service commitment value increases, the maximum profit obtained decreases. This decrease becomes more pronounced as the values of Γ are closer to $|K|$ with a decrease of 39.3% in profit when increasing Γ from $|K|-1$ to $|K|$. This result suggests that the decision maker should seek to arrange service commitment levels close to that obtained from variant II since these have a marginal effect on profit.

6. Conclusion

We have extended the classic fixed charge multicommodity network design problem by incorporating demand elasticity to travel cost in a profit-oriented problem by means of the gravity model. The proposed problem allows the decision maker to choose which O/D pairs will be served subject to a service commitment constraint. The resulting model captures additional levels of trade-off missing in classic fixed charge network design such as the effect of efficient routes on expected demand. We proposed two non-linear mixed integer programming formulations that model the profit-oriented fixed-charge multicommodity network design problem with elastic demand. We showed how both can incorporate O/D pair selection without modifying the formulations by carrying out simple network transformations or adding artificial variables. We also note that the inclusion of a service commitment constraint that requires a minimum number of O/D pairs be served can be done via an additional knapsack-type constraint. We proposed a flexible hybrid matheuristic capable of solving the problem for varying gravity model parameters and service commitment values. Computational results show this algorithm to be superior in terms of both solution time and quality when compared to the use of a general purpose global optimization software. The results also give managerial insights on the establishment of service commitments. The proposed framework allows decision makers to benefit from the added value of incorporating demand elasticity in the optimization of their strategic network design.

References

Aboolian R, Berman O, Krass D (2007) Competitive facility location and design problem. *European Journal of Operational Research* 182(1):40–62.

Aboolian R, Berman O, Krass D (2012) Profit maximizing distributed service system design with congestion and elastic demand. *Transportation Science* 46(2):247–261.

Beckmann M, McGuire C, Winsten C (1956) *Studies in the Economics of Transportation* (Yale University Press).

Berman O, Krass D, Wang J (2006) Locating service facilities to reduce lost demand. *IIE Transactions* 38(11):933–946.

Black WR (1973) An analysis of gravity model distance exponents. *Transportation* 2(3):299–312.

Boonekamp T, Zuidberg J, Burghouwt G (2018) Determinants of air travel demand: The role of low-cost carriers, ethnic links and aviation-dependent employment. *Transportation Research Part A: Policy and Practice* 112:18–28.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, New York, NY).

Carey HC (1858) *Principles of social science* (J.B. Lippincott & Co , Philadelphia ,PA).

Cipriani E, Gori S, Petrelli M (2012) Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies* 20(1):3–14.

Crainic TG, Frangioni A, Gendron B (2001) Bundle-based relaxation methods for multicommodity capacitated fixed charge network design. *Discrete Applied Mathematics* 112(1–3):73–99.

Crainic TG, Gendron B, Hernu G (2004) A slope scaling/Lagrangean perturbation heuristic with long-term memory for multicommodity capacitated fixed-charge network design. *Journal of Heuristics* 10(5):525–545.

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

Daganzo CF (2012) On the design of public infrastructure systems with elastic demand. *Transportation Research Part B: Methodological* 46(9):1288–1293.

Eiselt HA, Laporte G, Thisse JF (1993) Competitive location models: A framework and bibliography. *Transportation Science* 27(1):44–54.

Evans SP (1976) Derivation and analysis of some models for combining trip distribution and assignment. *Transportation Research* 10(1):37–57.

Fernández J, Hendrix EM (2013) Recent insights in Huff-like competitive facility location and design. *European Journal of Operational Research* 227(3):581–584.

Florian M, Nguyen S (1974) A method for computing network equilibrium with elastic demands. *Transportation Science* 8(4):321–332.

Fotheringham AS, O'Kelly ME (1989) *Spatial interaction models: formulations and applications*, volume 1 (Kluwer academic publishers, Dordrecht).

Grosche T, Rothlauf F, Heinzl A (2007) Gravity models for airline passenger volume estimation. *Journal of Air Transport Management* 13(4):175–183.

Haynes K, Fotheringham A (1984) *Gravity and Spatial Interaction Models*. Scientific Geography Series (SAGE Publications, Beverly Hills, CA).

Hodgson MJ (1990) A flow-capturing location-allocation model. *Geographical Analysis* 22(3):270–279.

Holmberg K, Hellstrand J (1998) Solving the uncapacitated network design problem by a Lagrangean heuristic and branch-and-bound. *Operations Research* 46(2):247–259.

Huff DL (1964) Defining and estimating a trading area. *Journal of Marketing* 28(3):34–38.

Jensen-Butler C (1972) Gravity models as planning tools: A review of theoretical and operational problems. *Geografiska Annaler. Series B, Human Geography* 54(1):68–78.

Keyvanshokooh E, Ryan SM, Kabir E (2016) Hybrid robust and stochastic optimization for closed-loop supply chain network design using accelerated Benders decomposition. *European Journal of Operational Research* 249(1):76–92.

Kim D, Pardalos PM (1999) A solution approach to the fixed charge network flow problem using a dynamic slope scaling procedure. *Operations Research Letters* 24(4):195–203.

Kim D, Pardalos PM (2000) Dynamic slope scaling and trust interval techniques for solving concave piecewise linear network flow problems. *Networks* 35(3):216–222.

Kılınç MR, Sahinidis NV (2018) Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON. *Optimization Methods and Software* 33(3):540–562.

Kuiteing AK, Marcotte P, Savard G (2017) Network pricing of congestion-free networks: The elastic and linear demand case. *Transportation Science* 51(3):791–806.

Kuiteing AK, Marcotte P, Savard G (2018) Pricing and revenue maximization over a multicommodity transportation network: the nonlinear demand case. *Computational Optimization and Applications* In Press.

Labbé M, Marcotte P, Savard G (1998) A bilevel model of taxation and its application to optimal highway pricing. *Management Science* 44(12-part-1):1608–1622.

LeBlanc LJ, Morlok EK, Pierskalla WP (1975) An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9(5):309–318.

Lee C, Lee K, Park S (2013) Benders decomposition approach for the robust network design problem with flow bifurcations. *Networks* 62(1):1–16.

Lesage JP, Polasek W (2008) Incorporating transportation network structure in spatial econometric models of commodity flows. *Spatial Economic Analysis* 3(2):225–245.

Magnanti TL, Wong RT (1984) Network design and transportation planning: Models and algorithms. *Transportation Science* 18(1):1–55.

Marianov V, Ríos M, Icaza MJ (2008) Facility location for market capture when users rank facilities by shorter travel and waiting times. *European Journal of Operational Research* 191(1):32–44.

Marianov V, Ríos, Miguel, Barros, Francisco Javier (2005) Allocating servers to facilities, when demand is elastic to travel and waiting times. *RAIRO-Oper. Res.* 39(3):143–162.

Matsoukis EC, Michalopoulos PC (1986) Road traffic assignment—a review. *Transportation Planning and Technology* 11(2):117–135.

Miller CE, Tucker AW, Zemlin RA (1960) Integer programming formulation of traveling salesman problems. *J. ACM* 7(4):326–329.

Minoux M (1989) Networks synthesis and optimum network design problems: Models, solution methods and applications. *Networks* 19(3):313–360.

Newell GF (1979) Some issues relating to the optimal design of bus routes. *Transportation Science* 13(1):20–35.

Newton I (1687) *Philosophiae Naturalis Principia Mathematica* (Londini :Jussu Societatis Regiae ac Typis Josephi Streater).

Peeta S, Ziliaskopoulos AK (2001) Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics* 1(3):233–265.

Rahmaniani R, Crainic TG, Gendreau M, Rei W (2018) Accelerating the Benders decomposition method: Application to stochastic network design problems. *SIAM Journal on Optimization* 28(1):875–903.

Ravenstein EG (1885) The laws of migration. *Journal of the Statistical Society of London* 48(2):167–235.

Reilly WJ (1929) *Methods for the Study of Retail Relationships* ([Austin, Tex.] The University).

Sahinidis NV (2017) *BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs,* User's Manual.

Senior ML (1979) From gravity modelling to entropy maximizing: a pedagogic guide. *Progress in Geography* 3(2):175–210.

Tawarmalani M, Sahinidis NV (2005) A polyhedral branch-and-cut approach to global optimization. *Mathematical Programming* 103:225–249.

Wilson AG (1967) A statistical theory of spatial distribution models. *Transportation Research* 1(3):253–269.

Wilson AG (1971) A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space* 3(1):1–32.

Yaged B (1971) Minimum cost routing for static network models. *Networks* 1(2):139–172.

Zetina CA, Contreras I, Cordeau JF (2018) Exact algorithms for the multicommodity uncapacitated fixed-charge network design problem. Submitted to Computers & Operations Research.

Zhang Y, Berman O, Marcotte P, Verter V (2010) A bilevel model for preventive healthcare facility network design with congestion. *IIE Transactions* 42(12):865–880.

Zhang Y, Berman O, Verter V (2009) Incorporating congestion in preventive healthcare facility network design. *European Journal of Operational Research* 198(3):922–935.

Zhong H, Wang J, Yip TL, Gu Y (2018) An innovative gravity-based approach to assess vulnerability of a hazmat road transportation network: A case study of Guangzhou, China. *Transportation Research Part D: Transport and Environment* 62:659–671.