AN OPTIMISATION MODEL AND ALGORITHMS FOR SOLVING THE MULTIMODAL NETWORK DESIGN PROBLEM IN REGIONAL CONTEXTS

GALLO, Mariano; Department of Engineering, University of Sannio, Benevento, Italy, gallo@unisannio.it

D'ACIERNO, Luca; Department of Transportation Engineering, 'Federico II' University of Naples, Italy, luca.dacierno@unina.it

MONTELLA, Bruno; Department of Transportation Engineering, 'Federico II' University of Naples, Italy, bruno.montella@unina.it

ABSTRACT

In this paper we focus on the multimodal network design problem that consists in designing jointly road and transit systems, assuming elastic demand at least at the mode choice level. We refer to regional contexts where a planner may have financial resources to be invested for improving the mobility of a wide area and have to decide how these resources should be allocated between transit and road systems. We propose an optimisation model for solving the problem, whereby we introduce an objective function that takes into account different objectives of the problem (reduction in user costs, reduction in external costs, etc.) and all constraints to be considered (budget constraints, capacity constraints, assignment constraints, etc.). We then propose a meta-heuristic solution algorithm for solving the problem and test it on a trial and a real-scale network.

Keywords: Multimodal network design, Elastic demand, Transportation, Scatter search

1. INTRODUCTION

The sound investment of public money in transportation systems is a major objective in transport policy since local administrations or central governments generally have limited financial resources at their disposal. Such limited resources need to be optimised to improve the global mobility of the area, taking account not only of user requirements but also of benefits for the whole society.
With regard to ground transportation, road networks and transit services are the main investment sectors and the politician has to choose where and how to invest public money.

In this paper we formulate a Multimodal Network Design Problem (MNDP) where the objective is to optimise investments in road networks and transit services, allowing for both user and society benefits. This problem may be seen as a Supply Design Problem (SDP) (see also Cascetta, 2009) under the assumption of elastic demand (at least at the mode choice level).

The (monomodal) Network Design Problem (NDP) has been widely studied. Both the Transit Network Design Problem (TNDP) and the Road Network Design Problem (RNDP) have attracted considerable attention in the literature.

Some recent literature reviews on TNDPs can be found in Guihaire and Hao (2008), Desaulniers and Hickman (2007); a less recent review was reported in Chua (1984). In this context, recent contributions on transit elastic demand have been made by Lee and Vuchic (2005), Cipriani et al. (2006), Fan and Machemehl (2006) and Gallo et al. (2009). As regards the Road Network Design Problem, more recent models and methods proposed for road networks include Herrmann et al. (1996), Solanki et al. (1998), Cho and Lo (1999), Cruz et. al. (1999), Meng et al. (2001), Meng and Yang (2002), Drezner and Wesolowsky (2003), Chiou (2005), Gao et al. (2005), Poorzahedy and Abulghasemi (2005), Cantarella et al. (2006), Cantarella and Vitetta (2006), Ukkusuri et al. (2007), Poorzahedy and Rouhani (2007) and Gallo et al. (2010).

This paper is structured as follows: Section 2 focuses on the model formulation; the solution algorithm is proposed in Section 3; numerical results on a real-scale network are reported in Section 4; Section 5 draws the main conclusions.

2. OPTIMISATION MODEL

A Multimodal Network Design Problem (MNDP) can be generally formulated by the following constrained optimisation model:

$$[\boldsymbol{x}\text{^}, \boldsymbol{y}\text{^}] = \text{Arg}_{\,\boldsymbol{x},\boldsymbol{y}} \min w(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_m{}^*) \tag{1}$$

subject to:

$$\boldsymbol{x} \in X \tag{2}$$

$$\boldsymbol{y} \in Y \tag{3}$$

$$\boldsymbol{f}_m{}^* = \Lambda(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_m{}^*, \boldsymbol{d}_m(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_m{}^*)) \tag{4}$$

where:

$\boldsymbol{x}$ is the vector of road decision variables;
$\boldsymbol{y}$ is the vector of transit decision variables;
$\boldsymbol{x}$^ is the optimal solution for $\boldsymbol{x}$;
$\boldsymbol{y}$^ is the optimal solution for $\boldsymbol{y}$;
$w(\cdot)$ is the objective function;
$\boldsymbol{f}_m{}^*$ is the multimodal equilibrium flow vector;
$\Lambda(.)$ represents the multimodal assignment function;
$\boldsymbol{d}_m$ is the multimodal demand vector;
X represents the feasible set for $\boldsymbol{x}$;

Y represents the feasible set for $\boldsymbol{y}$.

Eqns (2) and (3) summarise all constraints on decision variables. Eqn (4) represents the demand-supply consistency constraint that is in this case a multimodal assignment constraint; this constraint links the descriptive variables, $\boldsymbol{f}_m$*, to the decisional ones, $\boldsymbol{x}$ and $\boldsymbol{y}$, simulating user behaviour jointly as regards mode choice and path choice on the multimodal network.

The multimodal demand vector, $\boldsymbol{d}_m(.)$, arranges the transportation demand vectors for each transportation system; it depends on the decisional variables, $\boldsymbol{x}$ and $\boldsymbol{y}$, and on the multimodal equilibrium flows, $\boldsymbol{f}_m$*. Vector $\boldsymbol{f}_m$* can be estimated by adopting the multimodal equilibrium assignment model proposed by D'Acierno et al. (2002).

Given a transportation supply layout (i.e. given vectors $\boldsymbol{x}$ and $\boldsymbol{y}$), under some assumptions on cost functions and demand models, it may be proved that the multimodal equilibrium flow vector $\boldsymbol{f}_m$* exists and is unique (Cantarella, 1997; D'Acierno et al., 2002; Cascetta, 2009). Therefore, eqn (4) can be considered an application: to each supply configuration, identified by vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, corresponds one and only one multimodal equilibrium link flow vector $\boldsymbol{f}_m$*.

The MNDP consists in searching, among all feasible supply configurations ($\boldsymbol{x}$, $\boldsymbol{y}$), for the one ($\boldsymbol{x}$^, $\boldsymbol{y}$^) which corresponds to the optimal value of the objective function, $w(\cdot)$.

In this paper we focus on the problem of optimising regional investments on transportation systems. The aim is to design jointly the road infrastructures to improve (on a known road network) and the frequencies for a regional metro system (on a known line network), optimising total costs and taking demand elasticity into account.

2.1. Problem definition

Since our focus is on optimising limited resources to improve mobility at a regional level, we consider that a maximum amount of resources is available and can be used to improve the road network and/or the transit systems in an extra-urban context. We assume that the road network is known and that the improvements consist in enhancing the performance of some existing roads in terms of capacity and free-flow speed; in this paper we do not consider investments for building new roads.

We make the further assumption that the topological configuration of the transit network is known, in terms of lines and that network improvements consist in increasing the frequencies of major transit lines, as well as rail lines and main bus lines. It is evident that in this case we need to consider demand as elastic, at least at the mode choice level, and we have to simulate car and transit systems jointly in a multimodal context.

2.2. Decision variables

The decision variables represent the improving interventions on road infrastructures and on transit services. We introduce two vectors of decision variables:

- $\boldsymbol{x}$ is the vector of road decision variables, x_i; this vector is composed by as many elements as the road infrastructures to be improved. These variables are binary (0/1) and each variable, x_i, assumes the value 1 if the improving intervention is adopted on the infrastructure, and 0 otherwise;

– ***y*** is the vector of transit decision variables, y_j; this vector is composed by as many elements as the transit lines to be improved. These variables are discrete (1, 2, …, y_j^{max}) and each variable, y_j, assumes the value of the hourly frequency adopted on the line.

In this paper we refer to a single hour of operation (e.g. peak hour). Extension to several operation periods, with different demand and supply features, will be the subject of further research.

2.3. Constraints

In this problem several constraints can be identified; some constraints refer only to road decision variables, some to transit decision variables and others to the whole problem.

2.3.1. Constraints on road decision variables

The constraints on road decision variables are the following:

$$x_i = 0/1 \qquad \forall\ i \in I$$

where x_i represents the decision variable for road i and I is the set of roads on which it is possible to intervene. This constraint introduces the binary nature of the variables. No other constraints that regard only the road decision variables are identified.

2.3.2. Constraints on transit decision variables

The constraints on transit decision variables are the following:

$$y_j = \text{integer} \qquad \forall\ j \in J$$

where y_j represents the decision variable for transit line j and J is the set of transit lines on which it is possible to intervene. This constraint introduces the integer nature of the variables. Since these variables represent the hourly frequencies of lines, they should be assumed continuous, but for algorithmic reasons we prefer to consider them as discrete.

$$y_j^{now} \le y_j \le y_j^{max} \qquad \forall\ j \in J$$

where y_j^{now} is the frequency of line j in the current configuration and y_j^{max} is the maximum value of the frequency allowed on line j.

$$\Sigma_{j\in JR}\ \text{Ceiling}(y_j \cdot rt_j) \le NR_{max}$$

where rt_j is the route time of rail line *j*, NR_{max} is the maximum available number of trains and *JR* is the subset of *J* that includes only the rail lines. This constraint is not considered if the possible transit investments also include purchasing new trains.

$$\Sigma_{j\in JB} \text{Ceiling}(y_j \cdot rt_j) \leq NB_{max}$$

where rt_j is the route time of bus line *j*, NB_{max} is the maximum available number of buses and *JB* is the subset of *J* that includes only the bus lines. The function Ceiling(*x*) gives the smallest integer ≥ *x*: the formula expresses the need to round up to the next integer the number of buses or trains for operating the line. This constraint is not considered if the possible transit investments also include purchasing new buses.

$$\Sigma_{j\in JR}\, y_j \cdot L_j \leq TKR_{max}$$

where L_j is the length of line *j* and TKR_{max} is the maximum number of train-km that can be operated. This constraint is considered only if this limit on train-km exists.

$$\Sigma_{j\in JB}\, y_j \cdot L_j \leq TKB_{max}$$

where L_j is the length of line *j* and TKB_{max} is the maximum number of bus-km that can be operated. This constraint is considered only if this limit on bus-km exists.

$$\max_{s,j} f_{m\, s,j} \leq y_j \cdot TCap_j \qquad \forall\, j \in J$$

where $f_{m\, s,j}$ is the flow on section *s* of line *j* and $TCap_j$ is the capacity of a vehicle that is operated on line *j*. These constraints should be almost always inactive since the minimum value of frequency is the current one.

2.3.3. Constraint on the whole problem

The constraint on the whole problem is the following:

$$\Sigma_i\, x_i\, cr_i + \Sigma_j\, (y_j - y_j^{now})\, L_j\, ckm_j \leq B$$

where cr_i represents the cost of the improving intervention on road *i* and ckm_j is the cost per veh-km of line *j*. Obviously, since the cost of transit systems refers to an hour, also the term cr_i has to refer to an hour, as a function of the useful life of the facility and of maintenance costs. *B* represents, similarly, the total available budget per hour.

2.3.4. Demand-supply consistency constraint

The demand-supply consistency constraint, eqn (4), links descriptive variables (flows on car and transit systems) to decision ones. In this problem, where we assume that the demand is variable only at the mode choice level and where only two modes are available (for extra-urban trips the pedestrian mode is unavailable), the constraint can be formulated as follows:

$$[\boldsymbol{f}_t^*, \boldsymbol{f}_c^*] = \Lambda(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_t^*, \boldsymbol{f}_c^*, \boldsymbol{d}_m(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_t^*, \boldsymbol{f}_c^*))$$

where, in addition to the terms already defined:
$\boldsymbol{f}_t^*$ is the equilibrium flow vector of the transit system;
$\boldsymbol{f}_c^*$ is the equilibrium flow vector of the private car system.

A detailed description of the multimodal assignment model and the algorithms for solving it are reported in D'Acierno *et al.* (2002).

2.4. Objective function

The choice of the objective function in a multimodal network design problem is a critical point mainly if, as in the defined problem, significant economic resources have to be allocated. Indeed, several aspects of transport policy have to be considered: user costs, external costs, resources, etc.
The costs that we considered in the objective function are the following:

- private car user costs, C_c;
- transit user costs, C_t;
- resources invested, R;
- external costs, EC.

Private car user costs are the total costs incurred by car users on the road network. Such costs depend on road performance, which also depends on (multimodal equilibrium) traffic flows:

$$C_c = \Sigma_l\, c_l(\boldsymbol{x}, fc_l^*(\boldsymbol{x}, \boldsymbol{y})) \cdot fc_l^*(\boldsymbol{x}, \boldsymbol{y})$$

where:
fc_l^* is the equilibrium flow on road link l, element of the equilibrium vector $\boldsymbol{f}_c^*$;
$c_l(.)$ is the cost function on road link l.

In this equation the dependence of equilibrium traffic flows on decision variables, $\boldsymbol{x}$ and $\boldsymbol{y}$, is explicitly indicated.
Transit user costs are the total costs incurred by transit users on the transit network (both bus and rail services). We assume that these costs are not dependent on transit user flows (uncongested network) but only on transit decision variables:

$$C_t = \Sigma_m\, c_m(\boldsymbol{y}) \cdot ft_m^*(\boldsymbol{x}, \boldsymbol{y})$$

where:
ft_m^* is the flow on transit link m, element of the equilibrium vector $\boldsymbol{f}_t^*$;
$c_m(.)$ is the cost on road link l.

The resources invested in the project can be expressed as (see budget constraint):

$$R = \Sigma_i\, x_i\, cr_i + \Sigma_j\, (y_j - y_j^{now})\, L_j\, ckm_j$$

Finally, the external costs can be calculated as:

$$EC = Rec_{km} \cdot \Sigma_l\, fc_l^*(\boldsymbol{x}, \boldsymbol{y}) \cdot L_l + TRec_{km} \cdot (\Sigma_{j\in JR}\, y_j \cdot L_j) + TBec_{km} \cdot (\Sigma_{j\in JB}\, y_j \cdot L_j)$$

where, in addition to the terms already defined:

Rec_{km} is the average external cost produced by a car travelling 1 km (€/km);
L_l is the length of road link l;
L_j is the length of transit link l;
$TRec_{km}$ is the average external cost produced by a train travelling 1 km (€/km);
$TBec_{km}$ is the average external cost of a bus travelling 1 km (€/km).

All these terms of the objective function may be weighted to allow for the relative importance given to each of them. Hence the objective function can be summarised as follows:

$$w(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_t^*, \boldsymbol{f}_c^*) = \beta_1 \cdot C_c(\boldsymbol{x}, fc_l^*(\boldsymbol{x}, \boldsymbol{y})) + \beta_2 \cdot C_t(\boldsymbol{y}, ft_m^*(\boldsymbol{x}, \boldsymbol{y})) + \beta_3 \cdot R(\boldsymbol{x}, \boldsymbol{y}) + \\ + \beta_4 \cdot EC(\boldsymbol{x}, \boldsymbol{y}, fc_l^*(\boldsymbol{x}, \boldsymbol{y}))$$

In this objective function transit ticket revenues are not considered since they represent a cost for transit users but a benefit for society, annulling each other.

2.5. Whole model formulation

The model formulated is a constrained integer optimisation model that can be expressed as:

$$[\boldsymbol{x}^\wedge, \boldsymbol{y}^\wedge] = \text{Arg} \min_{\boldsymbol{x},\boldsymbol{y}} \; (\beta_1 \cdot C_c(\boldsymbol{x}, fc_l^*(\boldsymbol{x}, \boldsymbol{y})) + \beta_2 \cdot C_t(\boldsymbol{y}, ft_m^*(\boldsymbol{x}, \boldsymbol{y})) + \beta_3 \cdot R(\boldsymbol{x}, \boldsymbol{y}) + \\ + \beta_4 \cdot EC(\boldsymbol{x}, \boldsymbol{y}, fc_l^*(\boldsymbol{x}, \boldsymbol{y})))$$

subject to:

$$x_i = 0/1 \qquad \forall\, i \in I$$
$$y_j = \text{integer} \qquad \forall\, j \in J$$
$$y_j^{now} \le y_j \le y_j^{max} \qquad \forall\, j \in J$$
$$\Sigma_{j\in JR}\, \text{Ceiling}(y_j \cdot rt_j) \le NR_{max}$$
$$\Sigma_{j\in JB}\, \text{Ceiling}(y_j \cdot rt_j) \le NB_{max}$$
$$\Sigma_{j\in JR}\, y_j \cdot L_j \le TKR_{max}$$
$$\Sigma_{j\in JB}\, y_j \cdot L_j \le TKB_{max}$$
$$\max_{s,j} f_{m\, s,j} \le y_j \cdot TCap_j \qquad \forall\, s, j$$
$$\Sigma_i\, x_i\, cr_i + \Sigma_j\, (y_j - y_j^{now})\, L_j\, ckm_j \le B$$
$$[\boldsymbol{f}_t^*, \boldsymbol{f}_c^*] = \Lambda(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_t^*, \boldsymbol{f}_c^*, \boldsymbol{d}_m(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{f}_t^*, \boldsymbol{f}_c^*))$$

Some features of this model are:

- some decision variables are binary and others discrete;
- the objective function is neither linear nor convex (except in particular cases);
- elastic demand assignment must be performed to evaluate each solution;
- some constraints are non-linear;
- the assignment constraint is not expressible in a closed form;
- the problem is NP-Hard.

These features lead to the need to define efficient heuristic algorithms that should minimise the number of solutions to evaluate.

3. SOLUTION ALGORITHM

The algorithm proposed for solving the Multimodal Network Design Problem (MNDP) is based on the meta-heuristic technique called *Scatter Search* (see Laguna, 2002, and Glover et al., 2003). This allows discrete optimisation models to be solved, overcoming the boundaries of local optimisation. The method applies the optimal solution search to regions that are not explored by discrete local search algorithms (e.g. neighbourhood search techniques). A scatter search method for solving a general network design problem for undirected networks is proposed by Alvarez et al. (2005) while an application to Urban Network Design Problem was recently proposed by Gallo et al. (2010).
Scatter Search is described in the following subsections, where the method for the MNDP is also specified. Section 3.1 introduces preliminary definitions, section 3.2 briefly describes the Neighbourhood Search method, which is an important subroutine of the proposed Scatter Search and, finally, section 3.3 illustrates the general framework and the steps of Scatter Search.

3.1. Preliminary definitions

We indicate as $\boldsymbol{s} \in S$ a solution of a discrete optimisation problem, such as the MNDP, where S is the set of solutions. To each solution $\boldsymbol{s}$ a set of solutions $N(\boldsymbol{s}) \subset S$ is associated, called neighbourhood of $\boldsymbol{s}$. Solution $\boldsymbol{s}$ is called the centre of the neighbourhood $N(\boldsymbol{s})$. Each solution $\boldsymbol{s}' \in N(\boldsymbol{s})$, called neighbour, is obtained from solution $\boldsymbol{s}$ by an elementary operation called a move; a move changes only one value of a variable of solution $\boldsymbol{s}$, generating the next solution $\boldsymbol{s}'$. Usually it is assumed that the neighbourhoods are symmetrical, that is: if $\boldsymbol{s}' \in N(\boldsymbol{s})$ then $\boldsymbol{s} \in N(\boldsymbol{s}')$.
A solution $\boldsymbol{s}_{loc}\boldsymbol{^\wedge} \in S$ is a local optimum if the objective function value $w(\boldsymbol{s}_{loc}\boldsymbol{^\wedge})$ is less [greater] than, or equal to, in a minimisation [maximisation] problem, objective function values corresponding to all solutions belonging to its neighbourhood:

$$w(\boldsymbol{s}_{loc}\boldsymbol{^\wedge}) \leq w(\boldsymbol{s}') \quad \forall\ \boldsymbol{s}' \in N(\boldsymbol{s})$$
$$[w(\boldsymbol{s}_{loc}\boldsymbol{^\wedge}) \geq w(\boldsymbol{s}') \quad \forall\ \boldsymbol{s}' \in N(\boldsymbol{s})]$$

The distance of solution $\boldsymbol{s}''$ from solution $\boldsymbol{s}'$ is the minimum number of moves needed to transform solution $\boldsymbol{s}''$ into solution $\boldsymbol{s}'$; the distance is indicated with $D(\boldsymbol{s}''-\boldsymbol{s}')$. For symmetrical neighbourhoods it will be:

$$D(\boldsymbol{s}''-\boldsymbol{s}') = D(\mathbf{s}'-\mathbf{s}'') \qquad \forall\ \boldsymbol{s}', \boldsymbol{s}'' \in S$$

Any solution belonging to a neighbourhood has a distance equal to 1 from the centre:

$$D(\boldsymbol{s}'-\boldsymbol{s}) = 1 \qquad \forall\ \boldsymbol{s}' \in N(\boldsymbol{s})$$

For the MNDP a solution $\boldsymbol{s}$ is equal to $[\boldsymbol{x}, \boldsymbol{y}]$ where the components of $\boldsymbol{x}$ and $\boldsymbol{y}$ are defined previously. Conventionally, we assume as positive a move that converts the value of a variable x_i from 0 to 1 or that increases the value of a variable y_j; the opposite moves are assumed negative.

3.2. Neighbourhood Search

If $\boldsymbol{s}^k$ is a solution, the Neighbourhood Search generates the following solution $\boldsymbol{s}^{k+1}$ such that:

$$\boldsymbol{s}^{k+1} \in N(\boldsymbol{s}^k)$$

and $\boldsymbol{s}^{k+1}$ respects a specified rule. One of the most commonly adopted rules for generating the following solution is the *steepest descent method*; it examines all neighbours, calculating their objective function values, and chooses the following solution as the one with the best value:

$$w(\boldsymbol{s}^{k+1}) = \text{Min}\ \{w(\boldsymbol{s}); \qquad \forall\ \boldsymbol{s} \in N(\boldsymbol{s}^k)\}$$

The procedure then generates at each iteration a solution better than the previous one, choosing, among all solutions belonging to the neighbourhood, the one with the best objective function value. The procedure ends when solution $\boldsymbol{s}^k$ is a local optimum, that is when:

$$w(\boldsymbol{s}^k) \leq w(\boldsymbol{s}) \qquad \forall\ \boldsymbol{s} \in N(\boldsymbol{s}^k)$$

This method is not suitable for our problem if the network has real dimensions. Indeed, in this case the variables are very numerous and the neighbourhoods are very wide; evaluating at each step the objective function for all neighbours is not compatible with acceptable computation times, since each objective function evaluation requires a multimodal assignment (4).

In this paper, in order to reduce the computation times, we propose to use a random method for generating the following solution, that we call the *random descent method*. This method randomly extracts a solution from the neighbourhood and evaluates its objective function; if the new solution is better than the current one, it becomes the current solution; otherwise,

another neighbourhood solution is randomly extracted and so on, until a local optimum is found. If no neighbourhood solutions improve the objective function, the last solution is a local optimum. This method has been already applied in Gallo et al. (2010) for the Urban Network Design Problem with significant benefits in terms of computing times.
For solving our problem, a first random extraction is performed extracting a road variable x_i and changing the value of this variable, from 0 to 1 or from 1 to 0; if the new solution is better than the previous one, the next extraction will regard a transit variable y_j. Otherwise another random extraction of a road variable will be performed until an improved solution has been found. Analogously, when a transit variable is extracted, it will be changed with a positive or a negative move (randomly decided); if the new solution is better than the previous one, the next extraction will regard a road variable x_i. Otherwise another random extraction of a transit variable will be performed until an improved solution is found. The algorithms will stop when no improvements are possible: each variable modification will worsen the objective function. Hence the last solution is a local optimum.
In this way, for every two improvements of the solution one will regard the road system and the other the transit system. Only if no improvements are found for one of the two systems will the subsequent move regard only the same system. This alternate method avoids one system being examined more than the other if the number of system variables is not equilibrated: sometimes, variables x_i can be significantly more numerous than variables y_j if transit lines are few while there are many improvable road links.

3.3. Scatter Search

Scatter Search is a *metaheuristic* technique for solving complex combinatorial optimisation problems. It can be adapted in several ways to several kinds of optimisation problems by suitably defining the criteria used in the *phases* of the solution procedure. A phase of Scatter Search is a mathematical or algorithmic subroutine that operates on a solution subset generating another solution subset. Each phase (or also the sequence of phases) can be defined in different ways depending on the specific problem. Below, the phases of the scatter search are examined and adapted to the MNDP.

Phase 1 – Starting set generation

In this phase a set of solutions is generated which should have a high level of *diversity* so as to cover different parts of the solution set. The subroutine that allows us to obtain the starting set is also called the *Diversification Generation Method*, which can generally differ for each specific problem. This routine is applied in our problem as follows:

- we define a mother solution as the initial configuration of the multimodal transportation system; in this solution all road variables x_i are equal to zero and all transit variables y_j are equal to y_j^{now};
- from this solution, other solutions, at fixed a priori distances from the mother solution, are randomly generated; they are called base solutions;
- infeasible solutions are eliminated and substituted with other solutions (randomly generated) at the same distance from the mother one;

- the mother solution and the base solutions constitute the starting set.

For generating base solutions, we propose to adopt the following procedure. Let n_x be the number of road variables x_i, n_y the number of transit variables y_j, N_D the number of solution subsets that we want to generate at several distances from the mother solution and N_S the number of solutions in each subset. Obviously, the number of generated base solutions is equal to $N_D \cdot N_S$. The decision about adopting N_D and N_S may influence the quality of the final solution and computing times; we assume in this paper N_D equal to the integer part of $(n_x + n_y)/10$ and that N_S equal to the integer part of $(n_x + n_y)/25$, with a minimum equal to 2. For $(n_x + n_y) = 50$ the method will generate 10 base solutions, for $(n_x + n_y) = 100$ the method will generate 40 base solution while for $(n_x + n_y) = 500$ the method will generate 1000 base solutions.
The N_D subsets are generated by randomly extracting a number z between 1 to 10 and changing in the mother solution the values of $z \cdot n$ variables (with $n = 1, 2, \ldots, 10$) N_S times so as to generate on the whole $N_D \cdot N_S$ base solutions; the variables to change and the corresponding moves are randomly extracted.

Phase 2 – Improvement in current solutions

In this phase, from any current solution an improved solution is generated by an algorithmic subroutine that is also called the *Improvement Method*. Several improvement methods can be adopted; in this paper we adopt the proposed *random descent method* introduced in subsection 3.2. The improved solutions are local optima.

Phase 3 – Reference set generation or updating

A *reference set* is generated by selecting all improved solutions (local optima) generated in the previous phase or, if they are too numerous, only part of them; in this second case, the selection should take account of objective function values (*good solutions*) and diversity (*scattered solutions*). Especially for real networks, it is preferable to limit the dimension of the reference set by fixing the maximum number of its solutions. In this case a maximum number of good solutions and scattered solutions may be fixed; the reference set will consist of good solutions with better values of the objective function, and scattered solutions with maximum distances from the best solution, until the maximum number of solutions is reached.
Operating in this way, the reference set will comprise good solutions, as regards the objective function value, and scattered solutions that allow the search to be extended to regions that cannot otherwise be explored. The subroutine that generates, at the first iteration, or updates, in subsequent iterations, the reference set is called the *Reference Set Update Method*.

Phase 4 – Solution subset generation

In this phase some solution subsets are generated, consisting of two or more solutions belonging to the reference set, which will be *combined* in the subsequent phase to generate other solutions. If $n_{ref\ set}$ indicates the number of solutions in the reference set, the following subsets can be generated:

(i) all subsets containing 2 solutions;
(ii) subsets containing 3 solutions obtained by adding to 2-solution subsets the best solution among those not contained in them;
(iii) the subset containing 4 solutions obtained by adding to 3-solution subsets the best solution among those not contained in them;
(iv) and so on until a subset containing $n_{ref\ set}$ solutions is generated.

As for reference set generation, if the reference set solutions are numerous, the maximum number of subsets can be fixed a priori and can be randomly generated.

Phase 5 – Solution combination

In this phase, the solutions of each subset are combined. The *Solution Combination Method* may differ depending on the kind of problem, and usually leads to generate one solution from each subset. The method generally associates a *score* to each value that can be assumed by a variable; this score has to take account of objective function values of solutions of the subset and of the times that the specific value is assumed by the variable. The combined solution obtained from the subset will be that in which every variable assumes the value with the best score.
In our problem the new solution is generated as follows:

(a) a *solution score* is associated to each solution of the subset as follows:
the ratio between the objective function value corresponding to the solution and the sum of objective function values of all subset solutions is calculated (objective function ratio); since we are dealing with a minimisation problem the solution score is calculated as 1 less the objective function ratio;
(b) a *variable value score* is associated to each value that can be assumed by a variable x_i or y_j as the sum of the relative values of solutions in which that variable assumes that specific value;
(c) the combined solution is generated such that any variable, x_i or y_j, assumes the best variable value score.

The solutions obtained in phase 5 are improved (phase 2), generating a new reference set. The procedure ends when the reference sets in two successive iterations are equal or when a fixed a priori number of iterations is reached. All solutions belonging to the last reference set are local optima. Among them, the one with the best objective function value can be chosen.

4. Numerical results

In this section we analyse the proposed algorithm in three cases:

- a trial network in order to compare the exhaustive approach with the proposed algorithm;
- a real-scale network in order to compare two kinds of planning strategies: optimisation of rail frequencies and joint optimisation of both rail and bus frequencies;
- a real-scale network in order to test algorithm performances in the case of the following planning strategy: joint optimisation of rail frequencies, bus frequencies and improvements in road features.

In the first case, the proposed model and algorithm were tested on a trial network (shown in Fig. 1), whose features are reported in Table I in order to provide a comparison between an exhaustive approach and the proposed algorithm. Indeed, in the case of five frequencies to be optimised (2 rail lines and 3 bus lines), whose feasible values are indicated in Table II, and five roads that can be improved (shown in Fig. 2), whose features are indicated in Table III), the number of feasible solutions is equal to $10^5 \cdot 2^5$ = 3,200,000.
Numerical results in terms of algorithm approaches are shown in Tables IV and V. In particular, the application of the exhaustive approach required about 9.91 h of calculation time using a PC Intel Core 2 Quad 2.40 GHz, while the meta-heuristic approach provided the same results by analysing only 65 solutions in about one second.

Table I – Trial network features

<table>
<tr><td>8
Centroids</td><td>9
Road nodes</td><td>4
Rail nodes</td><td>2
Rail lines</td><td>3
Bus lines</td><td rowspan="2">4
(Transfer)
pedestrian
links</td></tr>
<tr><td>8
Connectors</td><td>13
Road sections</td><td>3
Rail sections</td><td>4
Rail stations</td><td>7
Bus stops</td></tr>
</table>

Table II – Feasible frequencies

Feasible frequency [veh./h]	1	2	3	4	5	6	8	10	12	15
Vehicle headway [min.]	60	30	20	15	12	10	7.5	6	5	4

Table III – Feasible road improvements

Road number	Free-flow speed [km/h]		Capacity [veh/h]		Improvement costs [€/km · h]
	Current	Improved	Current	Improved	
1	60	90	2,000	3,000	500
2	60	90	2,000	3,000	500
3	60	90	2,000	3,000	500
4	60	75	2,000	2,500	375
5	60	75	2,000	2,500	375

Table IV – Comparison between algorithm approaches in the case of the trial network

Algorithm approach	Examined solutions [#]	Calculation times [s]	Objective function values [€/h]	Mass transit travel demand [%]
Exhaustive	3,200,000	35,668	175,953	21.59%
Neighbourhood Search	65	1	175,953	21.59%

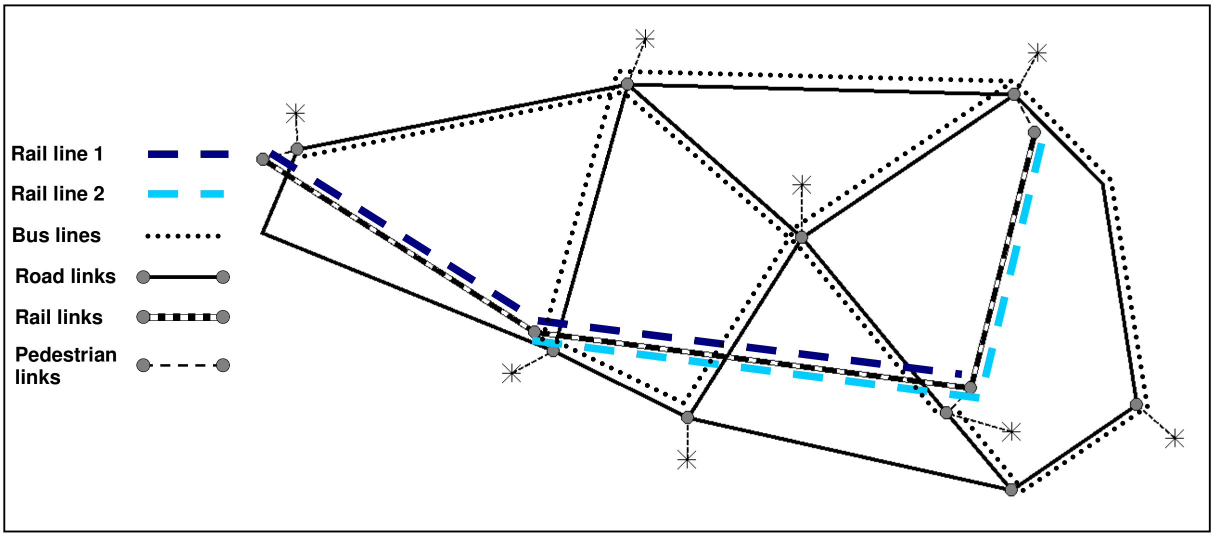

Figure 1 – Transit network

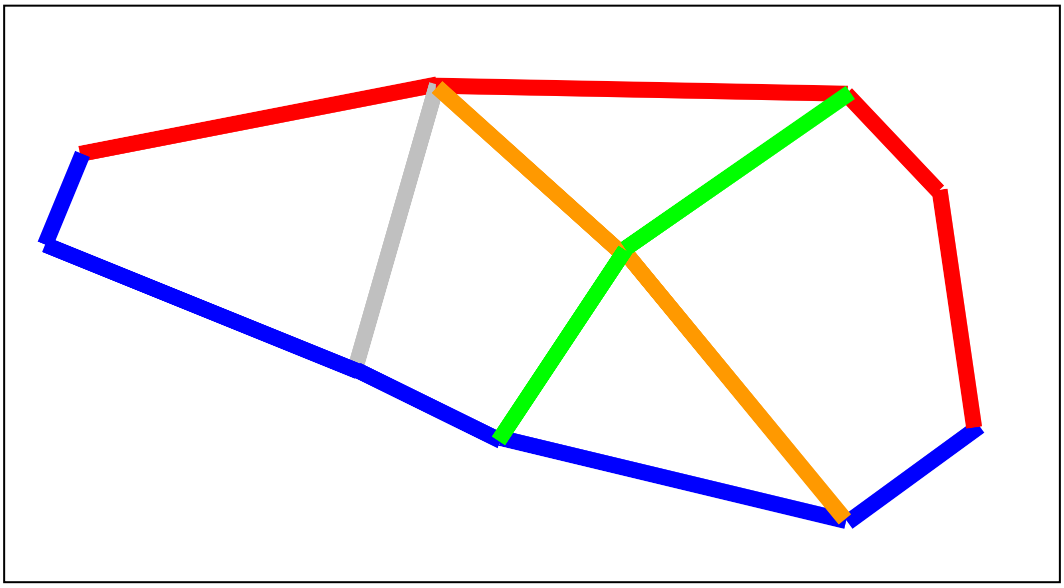

Figure 2 – Roads that can be improved

Table V – Optimal solutions

	Mass-transit frequencies					Road improvements				
Algorithm approach	Y1	Y2	Y3	Y4	Y5	X1	X2	X3	X4	X5
Exhaustive	1	1	15	15	12	0	1	1	0	0
Neighbourhood Search	1	1	15	15	12	0	1	1	0	0

The main result of the test on the trial network was that the proposed meta-heuristic approach allows a significant reduction to be obtained in calculation times whilst providing the same results. Although the goodness of results cannot be stated a priori in the case of the meta-heuristic approach, the great reduction in calculation times suggests the use of this class of solution algorithms for analysing real-scale networks in reasonable times. Therefore the other two analyses were performed by using only the meta-heuristic approach.
The second test was performed in the case of a real-scale network (shown in Fig. 3, whose features are reported in Table VI) in order to ascertain the applicability of the proposed model and algorithm. In this case, we compared results obtained by optimising only rail frequencies with optimising both rail and bus frequencies. The results are shown in Tables VII and VIII.

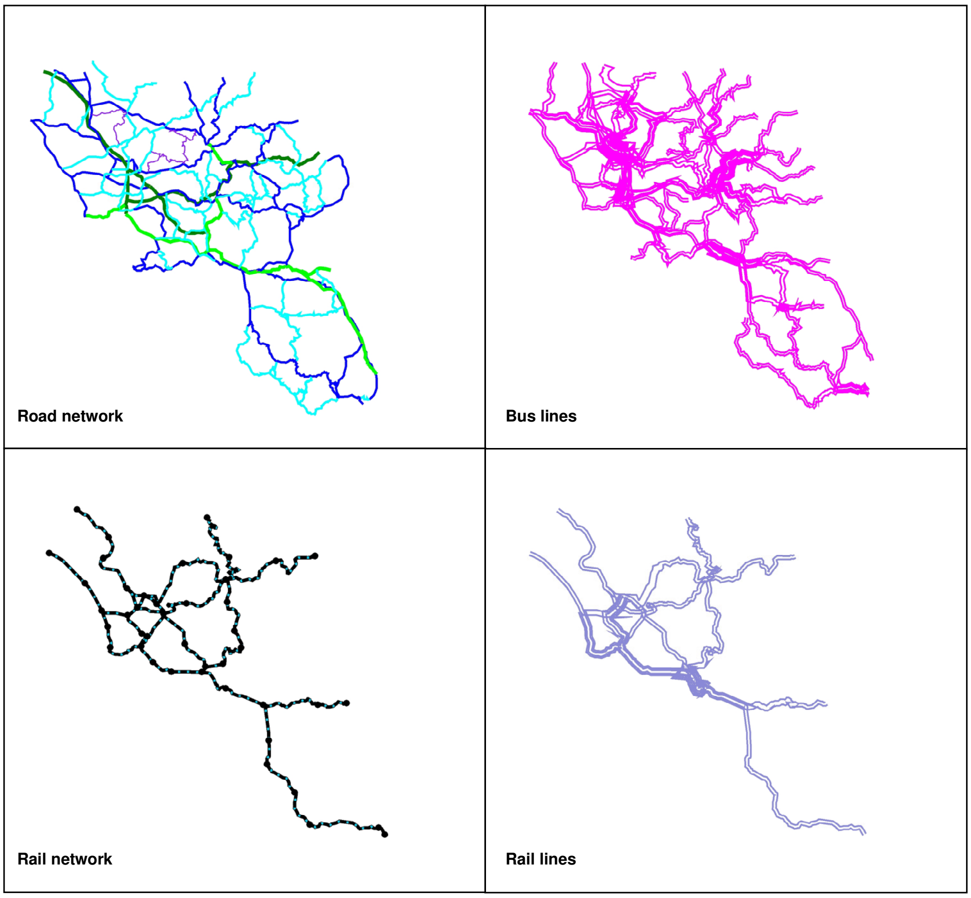

Figure 3 – Real-scale network

Table VI – Real-scale features

91 Centroids	262 Road nodes	43 Rail nodes	14 Rail lines	47 Bus lines	104 (Transfer) pedestrian links
161 Connectors	382 Road sections	49 Rail sections	43 Rail stations	120 Bus stops	

Table VII – Solution comparisons (Part 1)

Optimised variables	Bus-km/h	Train-km/h	Required bus number [#]	Required train number [#]	Mass-transit operational costs [€/h]
Only rail frequencies	21,215	10,381	611	276	205,879 €/h
Rail and bus frequencies	50,802	5,840	1,333	153	225,046 €/h

Table VIII – Solution comparisons (Part 2)

Optimised variables	Mass transit travel demand [%]	Objective function value [€/h]	Variable number [#]	Examined solutions [#]	Calculation times [min.]
Only rail frequencies	43.76%	13,693,938	14	158	349
Rail and bus frequencies	46.72%	12,878,226	61	1,290	2,969

In the first case, the number of design variables is equal to 14 and the number of feasible solutions is equal to 10^{14}, while in the second case the number of design variables is 61 and there are 10^{61} related feasible solutions. In the first case, the algorithm analysed 158 solutions in about 5.82 hours while in the second case the analysis of 1,290 solutions required 49.48 hours. It is worth noting that, since in the first case the algorithm analysed one solution every 2.21 minutes, application of an exhaustive approach would require about 420 million years as calculation times.
Comparison between the two planning strategies yielded the following results:

- the design of both rail and bus frequencies allows the system to be better optimised since the system is less constrained: the objective function value (which represents the sum of all system costs) is reduced by 5.96% and the mass transit travel demand is increased by 6.77%;
- the second strategy requires an increase in bus services in terms of bus-km/h (about +139%) and number of buses used (about +118%), while it requires a decrease in rail services in terms of train-km/h (about -44%) and number of trains used (about -45%). The difference in service costs (2.81 €/bus-km vs. 14.09 €/train-km) means that in the

second case we obtain an increase in public transport operating costs of 19,167 €/h which corresponds to a percentage increase equal to +9.31%;

- the increase in public transport costs allows travel times to be reduced such that a percentage of users, who in the first case were travelling on the road system, decide to change their modal choice and travel by public transport. It is worth noting that, by switching, these users lead to reduced road congestion, better performances being yielded for both road and bus systems. This means that, although in the second case an increase is required in public transport costs, total system costs decrease (as shown by the objective function reduction).

The last test was performed on the previous real-scale network and implemented by jointly designing rail frequencies, bus frequencies and road system improvements. In particular, there were 61 public transport frequencies to be optimised (according to values indicated in Table II), and 20 roads to be improved. Road improvements regard main regional roads. For all roads, improvement yields an increase in free-flow speed from 70 km/h to 110 km/h and an increase in capacity from 4,000 to 4,500 veh/h; improvement costs are assumed equal to 70 €/km/h. In this case the feasible solutions were $10^{61} \cdot 2^{20} = 1.05 \cdot 10^{67}$. However, application of the proposed algorithm required the analysis of 2,293 solutions in about 86.21 hours.

Our results, summarised in Tables IX and X, show that the design of all variables (i.e. rail and bus frequencies jointly with road improvements) provides an increase in bus services of about 56% in terms of bus-km/h and 42% in terms of number of buses required, and a decrease in train services of about 40% in terms of train-km/h and 36% in terms of number of trains required. Globally, there is a 5.69% decrease in public transport operating costs. The results suggest improvements be made on six roads, corresponding to a road improvement cost of 48,155 €/h.

Moreover, the design of all considered variables yielded a 4.33% increase in public transport travel demand and a 5.57% decrease in total system costs (i.e. objective function).

Table IX – Solution comparisons (Part 1)

Optimised variables	Bus-km/h	Train-km/h	Required bus number [#]	Required train number [#]	Mass-transit operational costs [€/h]
Initial solution	31,823	11,600	924	292	252,869
Final solution	49,761	7,001	1,308	186	238,470

Table X – Solution comparisons (Part 2)

Optimised variables	Mass transit travel demand [%]	Number of improved roads [#]	Road improvement costs [€/h]	Objective function value [€/h]
Initial solution	44.15%			13,617,713
Final solution	46.06%	6	48,155	12,859,205

Finally, by comparing the results of the strategy of design both road and transit systems (results are reported in Tables IX and X) with the design of only public transport (results reported in Tables VII and VIII), we obtain that a multimodal strategy in the transportation system design provides:

- a reduction in bus services of 2.05% in terms of bus-km/h and 1.88% in terms of required buses;
- an increase in rail services of 19.88% in terms of train-km/h and 21.57% in terms of required trains;
- a 5.97% increase in public transport operational costs which (with the costs for improving roads) yields a reduction in user travel times such that total system costs (i.e. the objective function values) decrease by 19,021 €/h.

5. Conclusions and research prospects

This paper proposed a model and an algorithm for solving the multimodal network design problem in a regional context where a planner has to evaluate the optimal allocation of financial resources between transit and road systems. Numerical results showed that also in simple cases an exhaustive approach can be considered prohibitive due to the huge number of alternative solutions to be examined and the calculation times involved. Moreover, in the case of real-scale networks, solutions may be obtained in reasonable times with the use of meta-heuristic algorithms.

With a view to future research, we suggest application of the proposed methodology to other real-scale networks in order to verify if it is possible to generalise considerations proposed in the paper. Further analysis could usefully test the proposed methods by varying the number of starting set solutions in order to search for the best trade-off between computation times and goodness of solution and compare the proposed method with other meta-heuristic approaches found elsewhere.

References

Alvarez, A. M., Gonzalez-Velarde, J. L. and K. De-Alba (2005). Scatter Search for network design problem. Ann. Oper. Res., 138, 159-178.

Cantarella, G. E. and A. Vitetta (2006). The multi-criteria road network design problem in an urban area. Transportation, 33, 567-588.

Cantarella, G. E., Pavone, G. and A. Vitetta (2006). Heuristics for urban road network design: lane layout and signal settings. Eur. J. Oper. Res., 175, 1682-1695.

Cantarella, G.E. (1997). A general fixed-point approach to multimodal multi-user equilibrium assignment with elastic demand. Transport. Sci., 31, 107-128.

Cascetta, E. (2009). Transportation systems analysis: models and applications. Springer, New York (NY), USA.

Chiou, S. W. (2005). Bilevel programming for the continuous transport network design problem. Transport. Res. B, 39, 361-383.

Cho, H. J. and S. C. Lo (1999). Solving bilevel network design problem using linear reaction function without nondegeneracy assumption. Transport. Res. Rec., 1667, 96-106.

Chua, T.A. (1984). The planning of urban bus routes and frequencies: a survey. Transportation, 12, 147–172.

Cipriani, E., Petrelli, M. and G. Fusco (2006). A multimodal transit network design procedure for urban areas. Adv. Transp. Stud., 10, 5-20.

Cruz, F. R. B., MacGregor Smith, J. and G. R. Mateus (1999). Algorithms for a multi-level network optimization problem. Eur. J. Oper. Res., 118, 164-180.

D'Acierno, L., Montella, B. and M. Gallo (2002). Multimodal assignment to congested networks: fixed-point models and algorithms. Proceedings of 'European Transport Conference 2002', Cambridge, United Kingdom.

Desaulniers, G. and M. Hickman (2007). Public transit. In: Handbooks in Operation Research and Management Science (G. Laporte, C. Bamhart, eds.), vol. 14, pp. 69-120. Elsevier, Amsterdam, The Netherlands.

Drezner, Z. and G. O. Wesolowsky (2003). Network design: selection and design of links and facility location. Transport. Res. A, 37, 241-256.

Fan, W. and R. Machemehl (2006). Optimal transit route network design problem with variable transit demand: genetic algorithm approach. J. Transp. Eng., 132, 40-51.

Gallo, M., D'Acierno, L. and B. Montella (2010). A meta-heuristic approach for solving the Urban Network Design Problem. Eur. J. Oper. Res., 201, 144-157.

Gallo, M., Montella, B. and L. D'Acierno (2009). The transit network design problem with elastic demand and internalisation of external costs. Proceedings of EURO XXIII – 23rd European Conference on Operational Research 'OR creating competitive advantage', Bonn, Germany, July 2009.

Gao, Z., Wu, J. and H. Sun (2005). Solution algorithm for the bi-level discrete network design problem. Transport. Res. B, 39, 479-495.

Glover, F., Laguna, M. and R. Martì (2003). Scatter search. In: Advances in evolutionary computation: theory and applications (A. Ghosh, S. Tsutsui, eds.), pp. 519-537. Springer-Verlag, New York (NY), USA.

Guihaire, V. and J.-K. Hao (2008). Transit network design and scheduling: A global review. Transport. Res. A, 42, 1251-1273.

Herrmann, J. W., Ioannou, G., Minis, I. and J. M. Proth (1996). A dual ascent approach to the fixed-charge capacitated network design problem. Eur. J. Oper. Res., 95, 476-490.

Laguna, M. (2002). Scatter search. In: Handbook of applied optimization (P. M. Pardalos, M. G. C. Resende, eds.), pp. 183-193. University Press, Oxford, United Kingdom.

Lee, Y.J. and V. R. Vuchic (2005). Transit network design with variable demand. J. Transp. Eng., 131, 1-10.

Meng, Q. and H. Yang (2002). Benefit distribution and equity in road network design. Transport. Res. B, 36, 19-35.

Meng, Q., Yang, H. and M. G. H. Bell (2001). An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. Transport. Res. B, 35, 83-105.

Poorzahedy, H. and F. Abulghasemi (2005). Application of Ant System to network design problem. Transportation, 32, 251-273.

Poorzahedy, H. and O. M. Rouhani (2007). Hybrid meta-heuristic algorithms for solving network design problem. Eur. J. Oper. Res., 182, 578-596.

Solanki, S. R., Gorti, J. K. and F. Southworth (1998). The highway network design problem. Transport. Res. B, 32, 127-140.

Ukkusuri, S. V., Mathew, T. V. and S. Travis Waller (2007) Robust transportation network design under demand uncertainty. Comput. Aided Civil Inf., 22, 6-18.