

A ROBUST ALGORITHM TO SOLVE THE SIGNAL SETTING PROBLEM CONSIDERING DIFFERENT TRAFFIC ASSIGNMENT APPROACHES

LUDOVICA ADACHER ^{a,*}, ANDREA GEMMA ^a

^aDepartment of Engineering
Roma Tre University, via della Vasca Navale 79, 00146 Rome, Italy
e-mail: {adacher, andrea.gemma}@dia.uniroma3.it

In this paper we extend a stochastic discrete optimization algorithm so as to tackle the signal setting problem. Signalized junctions represent critical points of an urban transportation network, and the efficiency of their traffic signal setting influences the overall network performance. Since road congestion usually takes place at or close to junction areas, an improvement in signal settings contributes to improving travel times, drivers' comfort, fuel consumption efficiency, pollution and safety. In a traffic network, the signal control strategy affects the travel time on the roads and influences drivers' route choice behavior. The paper presents an algorithm for signal setting optimization of signalized junctions in a congested road network. The objective function used in this work is a weighted sum of delays caused by the signalized intersections. We propose an iterative procedure to solve the problem by alternately updating signal settings based on fixed flows and traffic assignment based on fixed signal settings. To show the robustness of our method, we consider two different assignment methods: one based on user equilibrium assignment, well established in the literature as well as in practice, and the other based on a platoon simulation model with vehicular flow propagation and spill-back. Our optimization algorithm is also compared with others well known in the literature for this problem. The surrogate method (SM), particle swarm optimization (PSO) and the genetic algorithm (GA) are compared for a combined problem of global optimization of signal settings and traffic assignment (GOSSTA). Numerical experiments on a real test network are reported.

Keywords: genetic algorithms, surrogate method, traffic signal synchronization, traffic assignment, simulation.

1. Introduction

Transportation is involved with the movement of people and material from one place to another, from an origin to a destination. The performance of a traffic network can be influenced through several types of actions or decision variables. Some of these pertain to changing the load pattern on the network, through demand management actions, including attempts to route vehicles optimally through the network (Góngora and Rosenblueth, 2015; Klaučo *et al.*, 2016); others are concerned with how the traffic flow is controlled through the network components, such as junction utilization through signal control. Although the potential of explicitly combining both types of actions, especially joint signal control and route assignment, long has been suggested, most of the research and virtually all the approaches used in practice have followed one of two schemes: (a) considering signal control to be fixed and using traffic assignment

as decision variables (traffic assignment models), and (b) considering traffic assignment (loading pattern) to be fixed and using signal control as decision variables (signal optimization models). Nonetheless, several researchers have recognized the interaction between signal control and traffic assignment.

Usual traffic signal optimization methods seek either to maximize the green bandwidth or to minimize a general objective function that typically includes delays, number of stops, fuel consumption and some external costs like pollutant emissions. A major objective of traffic signal synchronization at an intersection is to clear maximum traffic throughput in a given length and time with the least number of accidents, at maximum safe speed and with minimum delay (see Goliya and Jain, 2012; Adacher, 2012). It has been widely accepted that improving the traffic flow is one of the strategies to reduce vehicle emissions and fuel consumption. In urban areas, frequent stop-and-go driving and excessive

*Corresponding author

speed variations contribute to higher fuel consumption and emissions (see Adacher and Cipriani, 2010; Xiaojian *et al.*, 2015; Stevanovic *et al.*, 2009; Khaki and Pour, 2014).

In order to improve these strategies, several authors combined in a different way the two synchronization approaches, which are the minimum delay and the maximal bandwidth. Cohen (1983) used the maximal bandwidth as the initial solution of the former problem; Cohen and Liu (1986) constrained the solution of the former problem to fulfill maximum bandwidth; Hadi and Wallace (1993) used the bandwidth as the objective function; Malakapalli (1993) added a simple delay model to the maximal bandwidth algorithm; Adacher and Cipriani (2010) expressed the delay at nodes as a closed form function of the maximal bandwidth solution. It is important to notice that the minimum delay is related to physical variables that are to be minimized; anyway, it is a non-convex problem and the existing solution methods do not guarantee to achieve the optimal solution. The maximal bandwidth method maximizes an opportunity of progression for drivers and does not necessarily reduce delays; nevertheless, it is a quasi-concave problem and efficient solving algorithms exist to find an optimal solution.

Traffic assignment on its own is also a well-studied problem. There are strict mathematical approaches using network flow theory. This branch of research started with the seminal work of Ford and Fulkerson in 1956, and recent results primarily focus on various aspects of traffic, e.g., a new concept of fairness in networks with congestion and flows over time with load dependent transit times (Köhler *et al.*, 2008). More application-oriented concepts include dynamic traffic assignment techniques (Szeto and Lo, 2006), and simulation based solutions in (Nagel and Flötteröd, 2012).

In contrast, the combined problem of traffic signal optimization and traffic assignment is less well-developed compared with split problems. Allsop and Charlesworth (1977) recognized the feedback between an optimized coordination and traffic assignment. They proposed an iterative approach where signal timings are optimized with TRANSYT. Afterwards, equilibrium traffic assignment is computed. These steps are repeated until no change in the coordination occurs. Using iterative heuristic methods, one may only hope to derive a local optimum.

Since the first platoon dispersion model introduced by Robertson (1969), progressively more complex models have been developed. Park *et al.* (1999) introduced a genetic algorithm-based traffic signal optimization program for over-saturated intersections consisting of two modules: a genetic algorithm optimizer and a mesoscopic simulator. Sun *et al.* (2006) proposed a bi-level programming formulation and a heuristic solution

approach for dynamic traffic signal optimization in networks with time-dependent demand and stochastic route choice. Chang and Sun (2004) proposed a dynamic method to control an oversaturated traffic signal network by utilizing a bang-bang-like model for oversaturated intersections and TRANSYT for unsaturated intersections. Chiou (2005) presented a bilevel formulation based on this approach. Smith (2006) also suggested a bi-level optimization, where the equilibrium property of the flow is preserved during consecutive iterations. Recent results were also obtained using genetic programming (Ceylan and Bell, 2004; Teklu *et al.*, 2007). Van den Berg *et al.* (2008) proposed a hybrid approach using methods from both optimal control and mixed integer linear programming. However, Smith (2015) summarizes the situation as follows: "At the moment, in practice, traffic signal timings are designed or optimized without systematically seeking to influence route choices beneficially." Substantially, there are two approaches in modelling signal settings and link flows in this combined problem. In both methods, it is recognized that link flows are affected by changes in signal settings and thus the flows and the settings are treated independently; instead, the former are regarded as functions of signal settings. First, a mutually consistent approach is adopted in which signal settings and link flows are calculated alternately and obtained respectively by signal setting optimization for given link flows and by user equilibrium traffic assignment for given signal settings. Second, a bi-level programming technique is adopted in which we may regard signal setting optimization as an upper level problem while the problem for user equilibrium traffic assignment by which link flows are obtained can be regarded as a lower level problem.

In this paper, we adopt an iterative procedure to solve the problem by alternately updating signal settings based on fixed flows and traffic assignment based on fixed signal settings (see Fig.1). Our iterative procedure optimizes the signal setting with the surrogate method and calculates delay by a traffic assignment model. It is stopped if for a fixed number of times the objective function does not improve. Our objective function is the minimum total delay. The total delay is a non-convex function, and the optimal signal setting feasible set comprehends non-linear constraints and integer variables. For this type of optimization problems, a closed-form objective function may not exist, and therefore it has to be estimated via simulation. The simulation gives the value of the objective function and then it needs to integrate an algorithm to find the best solution.

We apply the surrogate method to signal setting optimization, and two different types of traffic assignment are considered. The aim of this paper is to validate the capacity of convergence of the surrogate method compared with well-known optimization algorithms for

this problem. The comparison highlights the robustness of the SM in relation to the other algorithms.

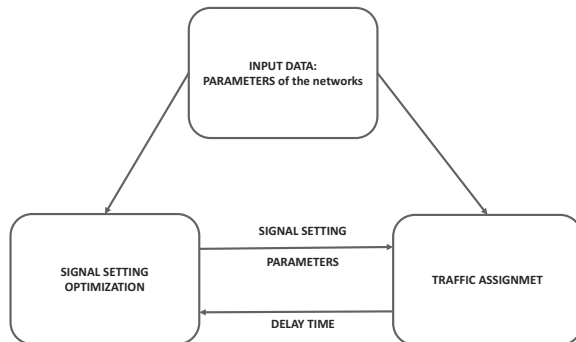


Fig. 1. Scheme of our approach.

In Section 2 we present the combined problem of global optimization of signal settings and traffic assignment (GOSSTA). We propose an optimization method for the signal setting considering two different types of assignment: one based on the user equilibrium and the other on a platoon model. This simulation model based on TRANSYT is employed to evaluate the delay in the network.

In Section 3 we present our optimization method. In Section 4 we consider the GOSSTA problem with the user equilibrium, and a comparison with the most utilized methods is reported. In Section 5 we discuss the GOSSTA problem with a simulation model. The simulation based on the platoon model produces the total delay on a road artery. A comparison with the most utilized methods is reported. In Section 6 some conclusions are drawn.

2. Combined problem of global optimization of signal settings and traffic assignment (GOSSTA)

Extensive research has been conducted on the combined problem of signal setting and network assignment. According to Cascetta *et al.* (1998), two sub-problems in the combined signal setting and traffic assignment can be identified: global optimization of signal settings (GOSSp) and local optimization of signal settings (LOSSp). The former minimizes the objective function describing the global network performance and the latter minimizes a local objective function. The signal setting can be applied to a single junction or to a network level. In the case of single junctions the interactions between successive junctions do not need to be considered. Coordinating traffic signals means coupling signals via a parameter called the offset. This quantity specifies how green phases

of different signals are shifted (or offset) to each other. Most prominent coordination objectives are the so-called green waves, where vehicles travel without being impeded by a signal showing red. Nevertheless, when considering networks of signals instead of arterial of signals, it is often impossible to adjust green waves for the whole network. Instead, the green wave goal has to be replaced by a more practical term such as minimum possible delay. Hereby, item delay refers to waiting times of vehicles facing red at the signals. The application of the signal setting can have a direct effect on the reduction of the vehicles waiting times at junctions (i.e., direct externals) and of other performance indicators of the junctions, such as queue lengths, fuel consumption and air pollution emissions (i.e., indirect externals), etc. Finally, the signal setting can be adopted in order to manage specific priority cases in the presence of weak users (e.g., pedestrians) or transit flows. In traffic engineering one considers traffic, i.e., vehicles, moving through a single and isolated junction, an arterial, which is a (possibly bi-directionally traversal) series of junctions, or through a whole network, i.e., an arbitrary set of junctions.

A traffic signal is described by its timing variables; first of all, by the time duration of its lights (red, yellow/amber, green). The sum of these times expresses the signal cycle, i.e., the time period before the same light turns on again. Having the cycle time, light duration is often expressed as its split, i.e., the ratio of cycle time. Considering more signals, we need to introduce the offset, i.e., the time period between a common reference instant and the cycle start. Depending on the goal, the problem formulation and thus the solving algorithm vary subsequently. Traffic lights improve drivers' safety, but they unavoidably introduce delays in travel times, too. Thus, minimizing the total delay, as a sum of all vehicles delays, may be an intuitive objective. Total delay minimization is a non-convex problem, and therefore the global optimum cannot be found analytically.

We aim to minimize the total delay. To do this, it is useful to remark that the users' total travel time can be expressed as a sum of two terms: the free flow travel time on the non-signalized network, which is a constant term, and an additional delay due to the interaction between flows and traffic signals, i.e., congestion, stops and queues. It is immediate that minimizing the total delay and minimizing the total travel time are equivalent. We assumed the following simplifications. First, only two lights signals are considered: an effective green, indicating the time period vehicle can actually cross the intersection, and a residual red, when vehicle intersection crossing is prohibited. Synchronization requires the cycle time to be the same for all signals or multiplies or fractions with each other, so that we consider the same cycle for all signals. These simplifications are very common in traffic works, so that the problem variables are the common

cycle time and the green ratio as well as offset for every synchronized intersection.

3. Proposed algorithm: The surrogate method

The surrogate problem methodology was initially developed for resource allocation problems of the form

$$\min_{x \in A_d} J_d(x), \tag{1}$$

where x is an N -dimensional decision vector with x_i denoting the number of resources that User i is assigned subject to a capacity constraint (Gokbayrak and Cassandras, 2002):

$$A_d = \left\{ x : x = [x_1, \dots, x_N]', \sum_{i=1}^N x_i = K, x_i \in \mathbb{Z}^+ \right\}, \tag{2}$$

and $J_d(x)$ is the cost incurred when the state is x . The integer capacity constraint is relaxed and a resulting surrogate problem is given by

$$\min_{\rho \in A_c} J_c(\rho) \tag{3a}$$

$$A_c = \left\{ \rho : \rho = [\rho_1, \dots, \rho_N], \sum_{i=1}^N \rho_i = K, \rho_i \in \mathbb{R}^+ \right\}. \tag{3b}$$

The basic idea of this method is to solve the continuous optimization problem above with standard stochastic approximation methods and establish the fact that, when (and if) a solution ρ^* is obtained, it can be mapped into some point $x = f(\rho^*) \in A_d$, which is in fact the solution to (1). The sequence of steps is reported as Algorithm 1.

Note, however, that the sequence $\{\rho_k\}$, $k = 1, 2, \dots$ generated by an iterative scheme for solving (3a) consists of real-valued allocations which are unfeasible, since the actual system involves only discrete resources. Thus, a key feature of our algorithm is that at every step k of the iteration scheme involved in solving (3a) the discrete state is updated through $x_k = f_k(\rho_k)$ as ρ_k is updated. This has the following two advantages:

- the cost of the original system is continuously adjusted (in contrast to an adjustment that would only be possible at the end of the surrogate optimization process);
- it allows us to make use of information typically employed to obtain cost sensitivities from the *actual* operating system at every step of the process.

We can therefore see that this scheme is intended to combine the advantages of a stochastic approximation type of algorithm with the ability to obtain sensitivity estimates with respect to discrete decision variables.

Two sequential operations are then performed at the k -th step:

1. The continuous state ρ_k is updated through

$$\rho_{k+1} = \pi_{k+1}[\rho_k - \eta_k \nabla L_c(\rho_k)], \tag{4}$$

where $\pi_{k+1} : \mathbb{R}^N \rightarrow A_c$ is a *projection function* and η_k is a *step size* parameter.

2. The newly determined state of the surrogate system, ρ_{k+1} , is transformed into an actual feasible discrete state of the original system through

$$x_{k+1} = f_{k+1}(\rho_{k+1}), \tag{5}$$

where $f_{k+1} : A_c \rightarrow A_d$ is a *projection mapping* of feasible continuous states to feasible discrete states.

One can recognize in Step 1 the form of a stochastic approximation algorithm that generates a sequence $\{\rho_k\}$ aimed to solve the surrogate problem. However, there is an additional operation, Step 2, for generating a sequence $\{x_k\}$ which we would like to converge to x^* . It is important to note that $\{x_k\}$ corresponds to feasible realizable states based on which one can evaluate estimates $\nabla L_c(\rho_k)$, calculated on an actual system x_k (not the surrogate state ρ_k , see Step 3). We can therefore see that this scheme is intended to combine the advantages of the stochastic approximation type of algorithm with the ability to obtain sensitivity estimates with respect to discrete decision variables.

4. Traffic signal setting adopting the user equilibrium

In our previous study (Adacher *et al.*, 2015), we considered the user equilibrium for the traffic assignment problem and the green ratio for the signal traffic problem. The light duration of traffic lights is often expressed as the ratio of the cycle time. We considered only two lights (i.e., green and red); obviously, the green ratio plus the red ratio give the 100% of cycle. The GOSSTA problem can be formulated as finding a vector of green split ratios g^* that minimizes the objective function J which expresses the global performance of the network under descriptive user equilibrium conditions and subject to admissibility constraints.

For simplicity, we assume here the objective function J as the total travel time on the network,

$$J(g, q) = \sum_l (c_a, g_a, q_a), \tag{6}$$

where

Algorithm 1. Steps of the surrogate method.

Step 0. Initialize $\rho_0 = x_0$ and perturb ρ_0 to have all components non-integer.

For any iteration $k = 0, 1, \dots$ repeat the following steps.

Step 1. Determine the selection set $S(\rho_k)$ using these steps:

Initialize

$$I = \{1, \dots, N\},$$

$$v = \rho - \lfloor \rho \rfloor.$$

Repeat the following steps until $I \neq \emptyset$:

$$i = \arg \min_{j \in I} (v_j),$$

$$y_i = v_i,$$

$$W_i = \sum_{j \in I} e_j,$$

$$v = v - y_i W_i,$$

$$I = I \setminus \{i\},$$

$$S(\rho_k) = \{W_i - \lfloor \rho \rfloor, i = 0, \dots, N\}.$$

Step 2. Select a transformation function f_k such that $x_k = f_k(\rho_k) = \arg \min_{r \in S(\rho_k)} \|x - \rho_k\|$.

Step 3. Estimate the gradient

$$\nabla L_c(\rho_k) = [\nabla_1 L_c(\rho_k), \dots, \nabla_N L_c(\rho_k)]^T,$$

using

$$\nabla_j L_c(\rho_k) = L_d(x^j) - L_d(x^k),$$

where k satisfies $x^j - x^k = e_j$ (the j -th versor).

Step 4. Update state: $\rho_{k+1} = \pi_{k+1}[\rho_k - \eta_k \nabla L_c(\rho_k)]$.

Step 3. If some stopping condition is not satisfied, repeat Steps 1–5 for $k + 1$. Otherwise, set ρ^* .

- g = green split vector,
- q = link flow vector,
- c_a = travel time on link a (min),
- q_a = flow on link a (veh/h),
- g_a = green time ratio for link a (integer).

The green time ratio exists only for links connected with the signalized intersection (i.e., intersection with the traffic light)

Thus, the GOSSTA problem is written as follows:

$$g^* = \arg \min_g \sum_a c_a (q_a^*(g), g_a) q_a^*(g) \quad (7)$$

subject to the constraints

$$g_{a_{\min}} \leq g_a \leq g_{a_{\max}}, \quad q_a \geq 0, \quad g_a \geq 0, \quad (8)$$

where

- q_a^* = user equilibrium flow on link a (veh/h),
- $g_{a_{\min}}$ = minimum green time ratio for signalized link a fulfilling the capacity constraint,
- $g_{a_{\max}}$ = maximum green time ratio for signalized link a fulfilling the capacity constraint.

Note that $g_{a_{\min}}$ and $g_{a_{\max}}$ guarantee a minimum of green and red for a traffic light.

The objective function is the overall *mean travel time* (or delay) of cars, denoted by $J(g_1, \dots, g_N)$. The (integer) green split ratio vector (g_1, \dots, g_N) is iteratively adjusted for N signalized links based on data directly observed and aiming at minimizing the global performance of the network. The comparison is made on a large size network relative to a North-East area of the city of Rome called Heaven (a healthier environment through the abatement of vehicle emission and noise). It is composed of 51 centroids, 300 nodes, 870 links and 70 signalized junction (isolated), 20 under control. The delay function utilized in signalized approaches is a linear adaptation of the Webster delay formula, while link performance functions are usual BPR volume-delay curves. All junctions are supposed to have a fixed cycle length (100 s), 2 phases, neither lost time nor shared lanes and permitted movements. A static deterministic user equilibrium has been considered for traffic assignment, solved using the usual Frank–Wolfe algorithm.

The objective function for this real network shows the existence of many local minima, but with small oscillations of the objective function. Consequently, the convergence capacity of the algorithms is important to find the best solution.

We compared our method (SM) with the most promising algorithms known for this problem: particle swarm optimization (PSO) and the genetic algorithm (GA). The results are summarized in Fig. 2; the SM is more effective than PSO: specifically, irrespective of the initial point, the SM always allows reaching better solutions, narrowed within a small convergence area (around 42000 veh×min). Such results point out the capabilities of the SM to escape from local minima. In contrast, this occurrence does not always happen with PSO, which may be trapped in a poor solution when the initial condition is represented by a bad point (very far from a good solution because of discontinuities and deep local valleys). The comparison of all the presented algorithms is reported in Fig. 2. From a numerical point of view, when the starting point is near 60000, the improvement of the SM with respect to PSO amounts

to about 6%. PSO gives worse results, but the GA behavior comparable with that of the SM, which gives best performance. Both the algorithms, SM and GA, are independent from the different initial points and present similar behaviors. The SM yields better solutions, by around 2%, but the GA needs fewer iterations to reach the same solutions. It also implies reduction in computational time. In particular, the SM improvement is around 2% but the reduction in the calculation time settles at 20%. The results put in evidence that this number of iterations is necessary for the SM to reach the best results, but certainly fewer iterations are needed to reach values different from the optimum, but close to it.

We want to analyze the behavior of these algorithms when the objective function has more deep local valleys. For this reason, we consider a more complex problem.

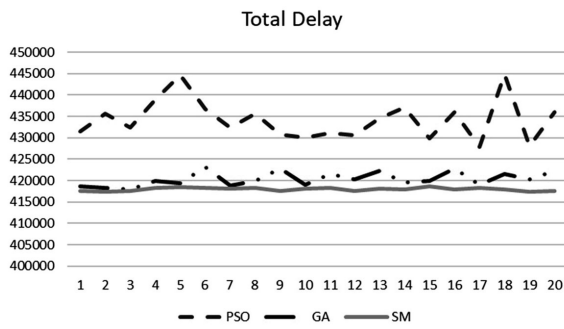


Fig. 2. Comparison of the algorithms for 20 trials.

5. Traffic signal setting adopting the platoon model

In urban networks most delays are spent on the main arterial roads, as there are larger flows and thus higher congestion. Therefore, our objective is fastening flows along these corridors.

The model used in this paper calculates the delay by vehicles for particular traffic light synchronization along a main road. This model is based on the platoon representation and simulates the congested road with propagation of the queue between nodes (spill-back phenomena). In this model we also take into account the delay of the employed approaches. We also utilize the platoon model to deal with even non-stationary traffic demand and nonsynchronized signal settings. The model is rather similar to the well-established TRANSYT solving procedure, which introduces some additional flexibility. This platoon model is studied by Fusco *et al.* (2013), who assess the consistency of macroscopic models with dynamic traffic assignment models based on the so-called mesoscopic approach that reproduces the longitudinal interaction between vehicles at an aggregate level and simulates individual behavior

at nodes, including route choice. The authors introduce and compare different approaches to model the traffic flow: the platoon progression model on a synchronized artery, a generalization of the cell transmission model, a micro-simulation queuing model, and two dynamic traffic assignment models. In numerical applications to a real-size artery, they observed good correspondence between the platoon progression model and dynamic traffic simulation models. For these reasons, we consider the platoon simulation model for the assignment problem.

The node delay model computes delays at every approach of the artery by checking, for each arriving platoon, which condition occurs among A, B, C; see Fig. 3. It presents an example of delay, according to the classification of platoons by Colombaroni *et al.* (2009).

Delay computation requires an iterative procedure that classifies the different platoons progressively. This procedure involves few iterations, because the platoons can both catch up each other along the links and recombine themselves at nodes, when more platoons arrive during the red phase. For more details, see the work of (Adacher and Gemma, 2016).

The objective function is defined as a linear combination of the total delay in each direction of the artery and it is calculated via simulation,

$$\begin{aligned}
 J = & (1 - w_t) \left(w_a \sum_{i=1}^n \omega_i D_i^{(a)} + (1 - w_a) \sum_{i=1}^n \omega_i D_i^{(-a)} \right) \\
 & + w_t \sum_{i=1}^n \omega_i D_{i_h}^{(t)} = J(D),
 \end{aligned}
 \tag{9}$$

where

- $D_i^{(a)}$ = the total delay at node i in the main direction of artery a ,
- $D_i^{(-a)}$ = the total delay at node i in the opposite direction of artery a ,
- $D_{i_h}^{(t)}$ = the total delay at node i of queue h in lateral approach t ,
- w_t = the weight of the delay in lateral approaches,
- w_a = the weight of the delay in the main direction of artery a ,
- ω_i = the weight of node i .

Analyzing a road artery, the problem of minimum travel time or minimum delay for traffic signal synchronization (MDP) can be expressed as follows:

$$\min J(D) = \min f(C, g, \theta, L, s, X, Q), \tag{10}$$

subject to

$$0 \leq \theta_i < C_i,$$

$$C_{\min} \leq C_i \leq C_{\max},$$

$$\max\{\gamma_{i,a}C_i\} \leq g_i \leq C_i - L_i - \max\{\gamma_{i,t}C_i\},$$

where

- $C_i[s]$ = the traffic light cycle for intersection i , (the traffic light cycle is defined as any complete sequence of the switch on (and off) traffic lights at the end of which the same configuration of the lights is returned as the one at the beginning of the sequence),
- $\theta_i[s]$ = the offsets for intersection i ,
- $g_i[ad]$ = the green time ratio of node i ,
- $L_i[s]$ = the time loss at node i (the time during which the intersection is not fully used).

The time lost is mainly due to three contributions:

- transient states of vehicles in the queue at the beginning of the green phase,
- transient states of exiting vehicles at the end of the green phase and during the yellow phase,
- the time between the end of yellow and the beginning of green of the next phase (the lost times at the beginning and at the end of green are used to determinate the duration of effective green),
- $\gamma_{i,a}, \gamma_{i,t}$ [dimensionless] = saturation degree of approach a along the artery and of traversal approach t of i node (the saturation degree is the ratio between traffic and saturation flows; this quantity is an indicator of the congestion level),
- s [veh/s] = saturation flow vector for each arc (the saturation flow is the maximum number of vehicles that can cross a stop intersection line per unit time, in the presence of continuous queue; the saturation flow depends on the geometric characteristics of the intersection, on flow composition and on the control traffic lights),
- X = urban artery geometry,
- Q [veh/s] = demand level or vehicular flow (it defines the vehicles passing through a section within unit time).

In this work the decision variables that have been used to minimize the delay are three vectors: *cycle* C , *green time ratio* g and *offsets* θ .

The travel time of a road section is closely linked to the geometry of the road itself and to the configuration of traffic light plans.

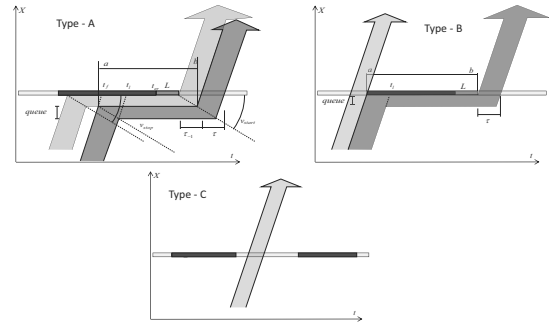


Fig. 3. Example of delay.

The saturation flow and geometry are usually studied and designed in earlier phases and they often cannot be changed *a posteriori*. In this work they are fixed. Compared with traditional delay formulations, this model includes both upstream and downstream node delay. The model used in this paper calculates the delay given a particular traffic light synchronization along a main road. In this model we take into account the delay of vehicles on the artery and also those coming from the subsequent approaches of the streets; see Fig. 4.

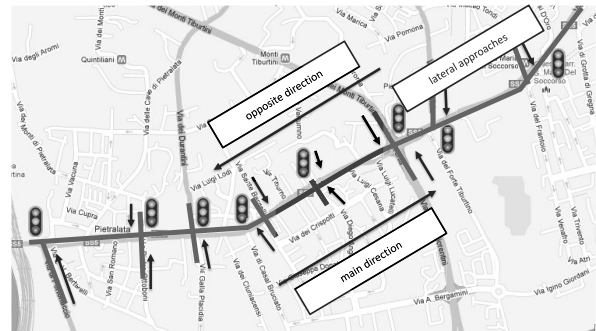


Fig. 4. Example of synchronized arteries while taking into account lateral approaches.

5.1. Analysis of the objective function. The total delay is a non-convex function, and an optimal feasible set of signal settings comprehends non-linear constraints and sometimes integer variables (phase optimization). Finding an optimal solution for not very small and simple networks may take long time, wherever possible.

On urban networks most delays are spent on the main arterial roads, as there are larger flows and thus higher congestion. So, our objective is fastening flows along these corridors.

The minimum delay problem can be formulated as finding out the cycle C^* (we consider the same cycle for

all junctions), the vector of green split ratios g^* and the vector of offsets θ^* that minimize an objective function $J(D)$, which expresses the total delay time of the network. All the components of these vectors are integer.

The shape of the objective function with respect to the cycle is quasi-convex and the related green vector is substantially quasi-convex. The real problem is the offset—the shapes presents many deep valleys. In Fig. 5 the shapes related to green time variation and offset variation are depicted. To visualize the characteristics of the solution space, it is not necessary to plot the average system time for all possible combinations of cycle/green time ratio/offset. To show the characteristics of the solution space, the cycle, the vector of the offset and all green time ratios are kept constant, except for one component. This is made for different values of green/offset/cycle that are kept constant.

5.2. Surrogate method considering the platoon model. The GOSSTA problem consists of two problems: optimization of signal settings (fixing cycle, green splits and offset) and calculation of the delay by simulation (platoon model), see Fig. 6. The longer the decision vector, the more time the surrogate method needs to find an optimal solution. Owing to the quasi convexity of the objective function with respect to the cycle, we fix the cycle using the binary search. We tackle the GOSSTA problem using an *on-line* optimization algorithm, i.e., for a fixed value of the cycle, we iteratively adjust the (integer) green split ratios g_1, \dots, g_n and the offsets $\theta_1, \dots, \theta_n$ for n intersection links based on data directly observed and aiming at minimizing the global performance of the network and the overall mean delay (or delay) of cars, denoted by $J(D) = J(C, g_1, \dots, g_n, \theta_1, \dots, \theta_n)$. The cycle is analyzed by a binary search; for a given fixed value of the cycle C_i , g and θ are evaluated based on the surrogate method. We denote by x the state variable of the surrogate method. The surrogate method is applied twice in sequence, first $x = \theta$, and when an optimal value of the offset is defined, then the state becomes $x = g$.

For a fixed C_i , the problem can be formulated as follows:

$$\min_{(g, \theta) \in A_d} J_d(C_i, g, \theta), \tag{11}$$

subject to the capacity constraint

$$A_d = \{g := [g_1, \dots, g_n]^T, g_{\min} \leq g_i \leq g_{\max}, g_i \in \mathbb{Z}^+; \theta := [\theta_1, \dots, \theta_n]^T, 0 \leq \theta_i \leq 1, \theta_i \in \mathbb{Z}^+\}, \tag{12}$$

where

- C_i = is the cycle fixed by the binary search (it is the same for all intersection links i),

- g = is an n -dimensional decision vector with $g_i \in \mathbb{Z}^+$ denoting the green time ratio for intersection link i ,
- θ = is an n -dimensional decision vector with $\theta_i \in \mathbb{Z}^+$ denoting the offset for intersection link i ,

and $J_{(C, g, \theta)} = J_{(D)}$, which is the total travel time on the network when the variables (green split vector, offsets and cycle) are fixed. For more details, see the works of Adacher (2012) as well as Adacher and Gemma (2016).

We tested four major roads of Rome. Their characteristics are reported Table in 1.

Table 1. Characteristics of instances. Capacity intersection is given in [veh/h]. The flow inside the main road is given in [veh/h]. The cross flow represents vehicles wanting to cross the artery [veh/s].

Signals	Type	Capacity	Flow	Cross flow	Way
4	ring road	≈ 3000	≈ 2500	> 1000	two
4+ 2 pedes.	radial road	≈ 3500	≈ 2000	< 1000	two
5	ring road	≈ 3500	≈ 2000	> 1000	one
8	radial road	≈ 5000	> 3000	< 1000	two

For example, Via Tiburtina is a 3 km long urban artery in Rome, containing 8 signalized intersections. During rush hours it is usually heavily congested, with an average speed of about 8 km/h in the direction of the town center and about 16 km/h in the opposite direction.

In Fig. 7 two of the four test networks are depicted. Via Togliatti is an example of the test network utilized in this study; it has 4 signalized intersections and 2 pedestrian intersections, and all the direction flows are particularly depicted.

This model is characterized by many parameters, and so careful calibration of the arterial model is required to fit observed traffic counts. Considering the rush hour we gave the priority at the artery to the lateral approach ($w_a = 0,6$ and $w_t = 0,4$). Three demand scenarios were considered to verify the robustness of the synchronization solution with respect to possible demand fluctuations. Starting from the actual average demand, two other scenarios, high and low, were obtained by increasing and reducing the average demand level to +15% and -15%, respectively. The simulation results highlight the fact that the optimizing procedure improves the average unitary delay at the nodes from 50% to 15%. All the algorithms and the models were developed with Matlab 2009 and the tests were made on an Intel i7-3770 3.4 GHz computer with 16 GB of RAM.

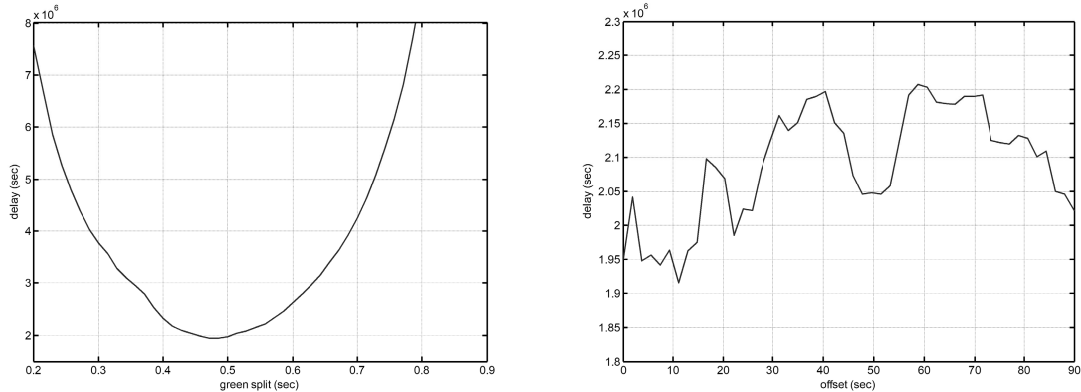


Fig. 5. Shape of the objective function considering the duration of green light and the offset.

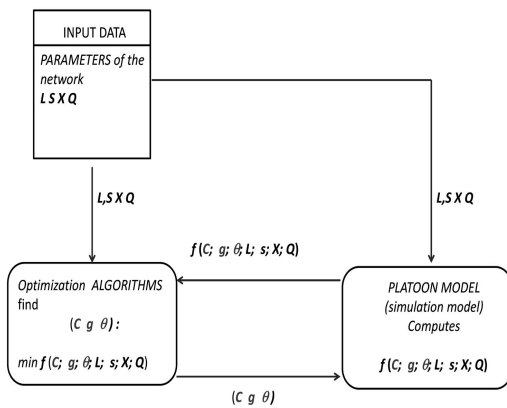


Fig. 6. Flow chart of our approach to solve total delay problem.

On the basis of the analysis and simulation, the following settings were adopted for the PSO algorithm: number of iterations—300, number of agents—25, speed parameter—0.8, best local parameter—1.9, best global parameter—1.5, random parameter—1.0, inertia parameter—1.0. The following settings were adopted for the GA: number of iterations—300, population size—440, crossover rate—0.8, mutation rate—0.1, elite rate—0.1. Two different stopping conditions are considered, the algorithm stops if the number of iterations is reached or if an improvement in the solution is no more than 0.1.

Considering the performance obtained for the first case (i.e., user equilibrium), we compare the SM and the GA, the two most promising algorithms.

In Fig. 8 the average behavior of the two algorithms (SM and GA) is depicted related to the four real test

networks. In this figure there we report the objective function value (total average delay) related to the number of iterations of the algorithm. For each iteration the SM needs $2 \times (N + 1)$ fitness evaluations (N is the number of signalized intersections). The GA for each iteration needs 25 fitness evaluations. The calculation of the fitness is the heaviest part for each algorithm. The platoon model takes from 0.5 to 1 second to simulate one hour of traffic and to calculate our fitness. It is also important to notice that the quality of the initial population affects significantly the GA and PSO convergence. For this reason 70 different trials are tested for each test networks. For some start points the reduction of total delay reaches 27%.

The SM gives a reduction in the total average delay around 15% compared with the GA and 30% compared with the actual average demand.

In Fig. 9 we depict the average behavior of the different algorithms related to the four real test networks, and the average delay related to the number of fitness evaluations. In fact, the presented algorithms work in different ways but all are based on the calculated fitness. Based on these tests, PSO and SM exhibit more or less the same behavior and give better results. They offer a reduction in the total delay around 15% compared with the GA and 30% compared with the actual average demand. In terms of efficacy, it is possible notice that the SM when compared with always yields better performance. In fact, around 600 steps are sufficient for the convergence of the SM when compared with 1220 for PSO. The SM has the capacity to find optimal solutions independently of the initial points.

6. Conclusions

In this paper, we have compared some known algorithms to the generalized ‘surrogate problem’ methodology that

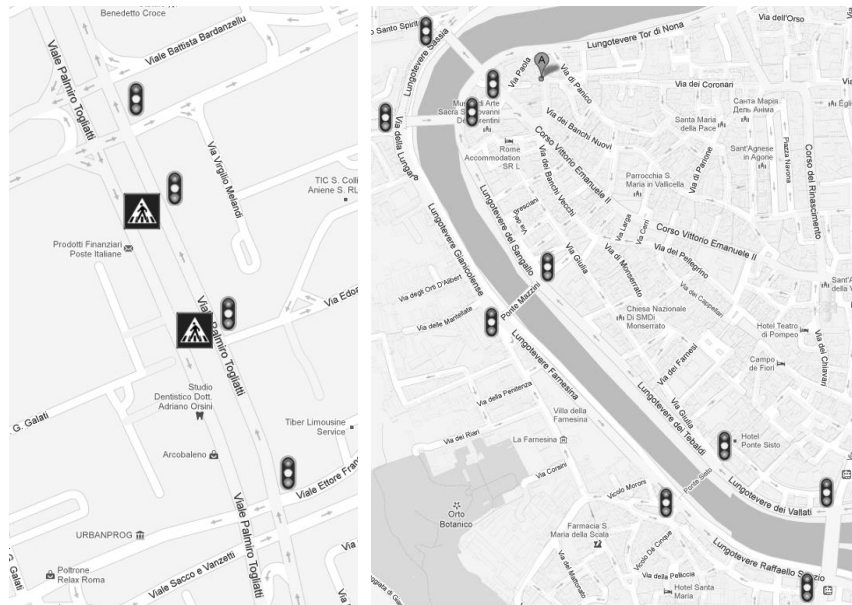


Fig. 7. Example of different instances.

is based on an on-line control scheme which transforms the problem into a ‘surrogate’ continuous optimization one and proceeds to solve the latter using standard gradient-based approaches while simultaneously updating both actual and surrogate system states.

The major objective of this paper was to compare the computational effectiveness and efficiency of the GA, SM and PSO using the testing approach. A motivation was to validate efficiency and efficacy of the SM with respect to the GA, the most utilized algorithm for this type of problem. Additionally, we aimed to analyze the performance of PSO, an emerging algorithm for the signal setting problem. Numerical experiments were conducted on real test networks for different dimensions and characteristics. We also analyzed a more realistic approach considering the most important variables for the signal setting and a dynamic model for the assignment. In this case, the analysis was made only considering the green time ratio and a user equilibrium, where the surrogate algorithm gives satisfying results in terms of efficiency. We generalized the problem by considering also the cycle and the offset with an assignment by the platooning model.

The surrogate method seems to be the most robust algorithm. It is independent of the objective function trend. Despite the discontinuous and noisy character of the real case problem, the SM seems to be more effective than GA taking advantage of its capabilities to escape from local minima. Differently, this feature was not detected by the GA, which may be trapped in a poor

solution when starting the procedure from a bad point placed in a region very far from a good solution and characterized by discontinuities and deep local minima. The SM is capable to escape from local minima and converge to the best solution. Also the PSO algorithm seems to be effective for this kind of problem, but less efficient than the SM.

References

- Adacher, L. (2012). A global optimization approach to solve the traffic signal synchronization problem, *Procedia-Social and Behavioral Sciences* **54**: 1270–1277.
- Adacher, L. and Cipriani, E. (2010). A surrogate approach for the global optimization of signal settings and traffic assignment problem, *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Funchal, Portugal, pp. 60–65.
- Adacher, L., Cipriani, E. and Gemma, A. (2015). The global optimization of signal settings and traffic assignment combined problem: A comparison between algorithms, *Advances in Transportation Studies* **36**: 35–48.
- Adacher, L. and Gemma, A. (2016). An optimizing algorithm to minimize the delay signal setting problem, *3rd International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, Crete, Greece, pp. 197–202.
- Allsop, R. and Charlesworth, J. (1977). Traffic in a signal-controlled road network: An example of different signal timings including different routing, *Traffic Engineering & Control* **18**(5): 262–265.
- Cascetta, E., Gallo, M. and Montella, B. (1998). Optimal signal setting on traffic networks with stochastic equilibrium

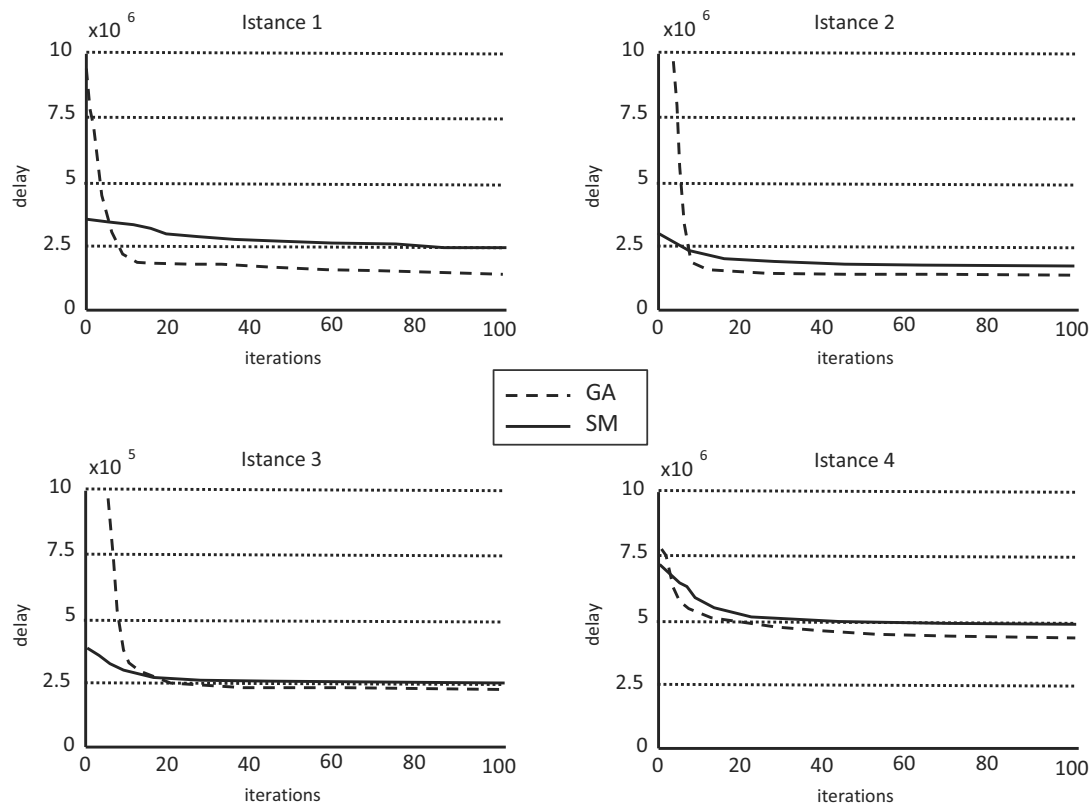


Fig. 8. Comparison of the SM and the GA.

assignment, *Proceedings of TRISTAN III, San Juan, Puerto Rico*, pp. 17–23.

- Ceylan, H. and Bell, M. (2004). Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing, *Transportation Research B: Methodological* **38**(4): 329–342.
- Chang, T. and Sun, G. (2004). Modeling and optimization of an oversaturated signalized network, *Transportation Research B: Methodological* **38**(8): 687–707.
- Chiou, S. (2005). Joint optimization for area traffic control and network flow, *Computers & Operations Research* **32**(11): 2821–2841.
- Cohen, S. (1983). Concurrent use of MAXBAND and TRANSYT signal timing programs for arterial signal optimization, *Transportation Research Record* (906): 81–84.
- Cohen, S. and Liu, C. (1986). The bandwidth-constrained TRANSYT signal-optimization program (discussion and closure), *Transportation Research Record* (1057): 1–7.
- Colombaroni, C., Fusco, G., Gemma, A., Demiralp, M., Baykara, N. and Mastorakis, N. (2009). Optimization of traffic signals on urban arteries through a platoon-based simulation model, *WSEAS International Conference, Istanbul, Turkey*, pp. 450–455.
- Fusco, G., Bielli, M., Cipriami, E., Gori, S. and Nigro, M. (2013). Signal settings synchronization and dynamic traffic modelling, *European Transport* **53**, Paper no. 7.
- Gokbayrak, K. and Cassandras, C.G. (2002). Generalized surrogate problem methodology for online stochastic discrete optimization, *Journal of Optimization Theory and Applications* **114**(1): 97–132.
- Goliya, H. and Jain, N. (2012). Synchronization of traffic signals: "A case study—eastern ring road, Indore", *International Journal of Advanced Technology in Civil Engineering* **1**(2): 1–7.
- Góngora, P.A. and Rosenblueth, D.A. (2015). A symbolic shortest path algorithm for computing subgame-perfect Nash equilibria, *International Journal of Applied Mathematics and Computer Science* **25**(3): 577–596, DOI: 10.1515/amcs-2015-0043.
- Hadi, M. and Wallace, C. (1993). Algorithm to optimize signal phasing and transportation research record, *Transportation Research Record* (1421): 104–112.
- Khaki, A. and Pour, P. (2014). The impacts of traffic signal timings optimization on reducing vehicle emissions and

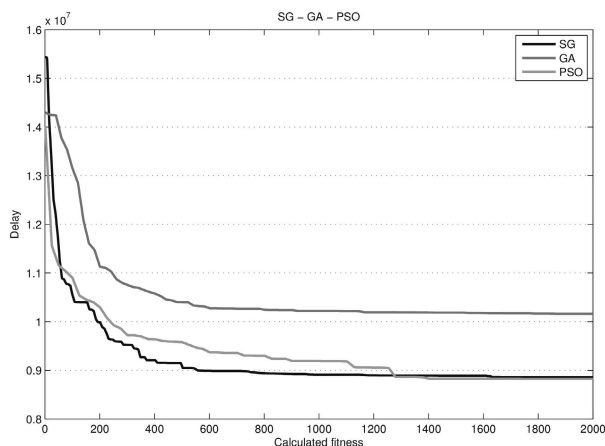


Fig. 9. Comparison between the different algorithms presented.

fuel consumption by aimsun and synchro software's (Case study: Tehran intersections), *International Journal of Civil and Structural Engineering* **5**(2): 144.

- Klaučo, M., Blažek, S. and Kvasnica, M. (2016). An optimal path planning problem for heterogeneous multi-vehicle systems, *International Journal of Applied Mathematics and Computer Science* **26**(2): 297–308, DOI: 10.1515/amcs-2016-0021.
- Köhler, E., Möhring, R., Nökel, K. and Wunsch, G. (2008). Optimization of signalized traffic networks, in H.J. Krebs and W. Jäger (Eds.), *Mathematics—Key Technology for the Future*, Springer, Berlin/Heidelberg, pp. 179–188.
- Malakapalli, M.M. (1993). Enhancements to the PASSER II-90 delay estimation procedures, *Transportation Research Record* (1421): 94–103.
- Nagel, K. and Flötteröd, G. (2012). Agent-based traffic assignment: Going from trips to behavioural travelers, *12th International Conference on Travel Behaviour Research, Jaipur, Rajasthan, India*, pp. 261–294.
- Park, B., Messer, C. and Urbanik, T. (1999). Traffic signal optimization program for oversaturated conditions: Genetic algorithm approach, *Transportation Research Record* (1683): 133–142.
- Robertson, D. (1969). *Transyt: A traffic network study tool* ministry of transport, *RRL Report LR253*, Crowthorne, Berkshire.
- Smith, M. (2006). Bilevel optimisation of prices and signals in transportation models, in S. Lawphongpanich *et al.* (Eds.), *Mathematical and Computational Models for Congestion Charging*, Springer, Boston, MA, pp. 159–199.
- Smith, M. (2015). Traffic signal control and route choice: A new assignment and control model which designs signal timings, *Transportation Research C: Emerging Technologies* **58**: 451–473.
- Stevanovic, A., Stevanovic, J., Zhang, K. and Batterman, S. (2009). Optimizing traffic control to reduce fuel consumption and vehicular emissions: Integrated approach with VISSIM, CMEM, and VISGAOST, *Transportation Research Record: Journal of the Transportation Research Board* (2128): 105–113.
- Sun, D., Benekohal, R.F. and Waller, S. (2006). Bi-level programming formulation and heuristic solution approach for dynamic traffic signal optimization, *Computer-Aided Civil and Infrastructure Engineering* **21**(5): 321–333.
- Szeto, W. and Lo, H. (2006). Dynamic traffic assignment: Properties and extensions, *Transportmetrica* **2**(1): 31–52.
- Teklu, F., Sumalee, A. and Watling, D. (2007). A genetic algorithm approach for optimizing traffic control signals considering routing, *Computer-Aided Civil and Infrastructure Engineering* **22**(1): 31–43.
- Van den Berg, M., De Schutter, B., Hellendoorn, J. and Hegyi, A. (2008). Influencing route choice in traffic networks: A model predictive control approach based on mixed-integer linear programming, *IEEE International Conference on Control Applications, CCA 2008, San Antonio, TX, USA*, pp. 299–304.
- Xiaojian, H., Jian, L., Wang, W. and Zhirui, Y. (2015). Research article traffic signal synchronization in the saturated high-density grid road, *Computational Intelligence and Neuroscience* **2015**, Article ID: 532960.



Ludovica Adacher received a *Laurea* degree in 1997 in computer science and a PhD in system engineering in 2002 from the University of Rome “La Sapienza”. She joined Roma Tre University as an assistant professor in 2003. Her research interests are in the field of decentralized control, optimization and simulation of logistic and transportation problems.



Andrea Gemma holds a PhD degree in computer science from the Department of Engineering, Roma Tre University, Rome, Italy. Currently, he is working on a research grant in civil engineering, and cooperates with various private and public companies as a computer science engineer. His know-how focuses mainly on transport system planning, intelligent transportation systems, and geographic information systems.

Received: 6 October 2016

Revised: 14 March 2017

Re-revised: 4 May 2017

Accepted: 4 May 2017