This is a post-peer-review, pre-copyedit version of an article published in Transportation Research Part C The final authenticated version is available online at: http://dx.doi.org/10.1016/j.trc.2015.03.032

Network Signal Setting Design: meta-heuristic optimisation methods

Giulio E. Cantarella*, Stefano de Luca, Roberta Di Pace and Silvio Memoli
Department of Civil Engineering, University of Salerno Italy, EU

g.cantarella@unisa.it; sdeluca@unisa.it; rdipace@unisa.it; smemoli@unisa.it

* Corresponding author

ABSTRACT

This paper aims to investigate the application of meta-heuristic optimisation methods to Network Signal Setting Design. The adopted approaches are (i) three step optimisation, in which first the stage matrix (stage composition and sequence), the green timings at each single junction are optimised, then the node offsets are computed in three successive steps; (ii) two step optimisation, in which the stage matrix is defined at a first step, then the green timings and the node offsets are computed at a second step. In both approaches the stage matrix optimisation is carried out through explicit complete enumeration.

In the first approach multi- criteria optimisation is followed for single junction signal setting design (green timings), whilst the coordination (node offsets) is approached through mono-criterion optimisation, as well as for the synchronisation (green timings and offsets) in the second approach.

A new traffic flow model mixing CTM and PDM has been applied. This model allows to explicitly represent horizontal queuing phenomena as well as dispersion along a link. Some meta-heuristic algorithms (i.e. genetic algorithms, hill climbing and simulated annealing) are investigated in order to solve the two problems

The proposed strategies are applied to two different layouts (a two junction arterial vs. a four junction network) and their effectiveness is evaluated by comparing the obtained results with those from benchmark approaches implementing mono-criterion optimisation only.

Keywords: Network Signal Setting Design; Meta-heuristics; Cell Transmission model; Platoon Dispersion model.

1 INTRODUCTION AND MOTIVATION

Traffic lights are one of the most common ways to control a road junction network. The design of control variables can be formulated as an optimisation problem, often named Network Signal Setting Design (NSSD).

In accordance with literature, the main ongoing issues in connection with network signal setting design are related to:

- ✓ the decision variables involved in the optimisation (green timings, green scheduling, and node offsets);
- ✓ the number of objective functions (criteria) considered in the optimisation methods (mono-criterion vs. multi-criteria);
- ✓ the traffic flow model needed to compute delay indicators.

In terms of decision variables, two main approaches are adopted in literature: coordination and synchronisation. In the first case, optimisation may be achieved in two ways: (i) first the green scheduling, say the stage matrix is defined, then the green timings at each single junction and eventually the node offsets are computed in separated steps; (ii) the green scheduling and timings are computed together in the first step, the coordination is carried out in the second step.
In the case of synchronisation, the scheduling is carried out at a first step, say the stage matrix is defined, and the green timings and the coordination (node offsets) are computed at a second step for each junction in the network.
Two of the most straightforward software programs are: TRANSYT14[1]® (TRL, UK) and TRANSYT-7F® (FHWA, USA). Both are able to compute the signal timings, the offsets and the cycle length by combining a traffic flow model and a signal setting optimiser. Both may be used for coordination or synchronisation; in the former case Oscady Pro® (TRL, UK; Burrow, 1987) may be used for green timings and scheduling optimisation. TRANSYT14® generates several (but not all) significant stage sequences to be tested but the optimal solution is not endogenously generated, while TRANSYT-7F® is able to optimise the stage sequence for each single junction starting from the ring and

[1] Recently TRANSYT15® has been released.

barrier NEMA (i.e. National Electrical Manufacturers Association) phases. Further investigations, with respect to the stage sequence optimisation may be found in Park et al. (2000). They carried out a simulation framework made up of a mesoscopic flow simulator and Goldberg's genetic algorithm optimiser which, showed that the developed simulator provided better results than those obtained using the software TRANSYT-7F®. Other investigations may be found in Hadi and Wallace (1993; 1994). As regards the solution methods, in TRANSYT14®, the Hill-Climbing and the Simulated Annealing are implemented, whereas in TRANSYT-7F®, the Hill Climbing and some Genetic Algorithms may be used, but none of them are able to solve the multi-criteria optimisation (they optimise linear combinations of two criteria).

With regards to the latter case, some interesting contribution may be found in Ceylan (2006) which focused on the integration of a combined optimisation procedure (based on GAs and HC) with TRANSYT for the evaluation of a performance index and in Putha et al., (2012) which compared the GAs with the Ant Colony Optimisation, by applying the coordination approach at network level. Differently from previous authors that considered the PDM for traffic flow modelling, Lertworawanich, (2011) proposed the application of GAs for network traffic signal setting trough CTM for traffic flow modelling. All above cited authors considered the combination (by weighs) of objective functions as a proxy of multi-criteria optimisation.

Finally traffic models may be classified according to the level of detail adopted for representing the traffic systems. The first are microscopic models, which describe both the space-time behaviour of the systems' elements (i.e. vehicles and drivers) as well as their interactions at disaggregated level. The second are mesoscopic models, which represent traffic by groups of vehicles or a single vehicle, whereby the activities and interactions are not described at a detailed level. The final models are macroscopic flow models which describe the traffic at a high level of aggregation as a flow without distinguishing its constituent parts.

In the case of a signalised network, two main issues are to be adressed: (i) the modelling of the dispersion between interacting junctions, which is strictly related to the distance travelled on the connecting links and (ii) the spillback (i.e. the link blockage) and the merging and diverging modelling (i.e. the lane blockage).

All these phenomena may be observed in macroscopic traffic flow models. In general, the approach proposed by Robertson with the platoon dispersion model (PDM) in the 1969 is the most straightforward for modelling the dispersion of platoons by estimating vehicle arrivals at downstream locations based on an upstream vehicle departure profile and a desired traffic stream speed. However, this model shows a main weakness since it cannot describe the spillback phenomena and it does not model the effects of blocking back (i.e. horizontal queuing). Based on this consideration, recent enhancements in traffic models are related to the application of the cell transmission model (CTM) (Daganzo, 1994; Lo, 1999, 2001; Lo and Chow, 2004; Chow and Lo, 2007) where a link is represented (discretised) by 'cells' which experience a flow carried out as the minimum value between the (maximum) number of vehicles that can be 'sent' by the upstream cell and those that can be 'received' by the downstream cell. It should be noted that CTM cannot describe density variations.

Promising approaches leading for stochastic CTM have been proposed in Sumalee et al. (2011), in which the sending and receiving functions are reproduced by random parameters of the fundamental flow–density diagram. An approach for signalised arterials has been proposed in Jabari and Liu (2011), in which a variety of time headway distributions, that are dependent on traffic states, are simulated; more extensive studies regarding density estimation have been dealt with in enhanced L(agged)-CTM by Szeto (2008). Other researchers have focused on the lane blockage simulation (Li and Chang, 2010; Li, 2011) by considering the 'sub-cell concept'. Currently, in TRANSYT14® (Binning et al. 2008), the CTM may be adopted as an alternative to the PDM, for short distances whereas in the case of long distances the PDM is still preferred to the CTM.

On the basis of all the previous considerations, in this paper:

- the stage matrix (SM), defined by stage composition and stage sequence, can be obtained by enumerating (Ex.en) all possible stages; other methods such as Oscady, allows for green timings and scheduling optimisation; in this case the SM is obtained from the scheduling (see also Improta and Cantarella, 1984);
- the green timings (GT) and offsets (Off) are optimised;

- some meta-heuristic algorithms are analysed for the solution of the mono-criterion/multi-criteria optimisation and applied to two different layout configurations (a two junction arterial vs. a four junction network);
- finally, in terms of traffic flow modelling, an extension of the CTM is proposed to include the traffic dispersion (CT&PDM).

Summing up, the paper contributions are reported in bold in the following Table 1 where criteria (Crt) and objective functions (Of) are considered. A further comparison is shown with regards to the optimisation (Opt) Tools in terms of Benchmark (Ben), Proposal (*) and Future Perspectives (FP).

Table 1: Paper contributions, Benchmarks and Future perspective on Network signal setting design

<table>
<tr><th rowspan="2">Opt</th><th rowspan="2">var</th><th colspan="9">Strategy</th></tr>
<tr><th colspan="2">three step</th><th colspan="5">two step</th><th colspan="2">one step</th></tr>
<tr><td rowspan="3">Crt</td><td>SM</td><td>Ex.en</td><td>Ex.en</td><td>Ex.en.</td><td>Ex.en.</td><td>Ex.en</td><td rowspan="2">mono</td><td rowspan="2">multi</td><td rowspan="3">mono</td><td rowspan="3">multi</td></tr>
<tr><td>GT</td><td>mono</td><td>multi</td><td rowspan="2">mono</td><td rowspan="2">mono</td><td rowspan="2">multi</td></tr>
<tr><td>Off</td><td>mono</td><td>mono</td><td>mono</td><td>mono</td></tr>
<tr><td rowspan="3">Of</td><td>SM</td><td>ND</td><td>ND</td><td>CF</td><td>CF</td><td>CF</td><td rowspan="2">CF</td><td rowspan="2">CF & of</td><td rowspan="3">TD</td><td rowspan="3">TD & of</td></tr>
<tr><td>GT</td><td>CF</td><td>CF & of</td><td rowspan="2">TD</td><td rowspan="2">TD</td><td rowspan="2">TD & of</td></tr>
<tr><td>Off</td><td>TD</td><td>TD</td><td>TD</td><td>TD</td></tr>
<tr><td rowspan="4">Tools</td><td></td><td>Ben</td><td>*</td><td>Ben</td><td>*</td><td>FP</td><td>Ben</td><td>FP</td><td>FP</td><td>FP</td></tr>
<tr><td>SM</td><td>Ex.en</td><td>Ex.en</td><td>Ex.en</td><td>Ex.en</td><td rowspan="3">x</td><td rowspan="2">OSCADY (GT&S)</td><td rowspan="3">x</td><td rowspan="3">x</td><td rowspan="3">x</td></tr>
<tr><td>GT</td><td>OSCADY (GT)</td><td>UNISA (GT)</td><td rowspan="2">TRANSYT (Sy) PDM/CTM</td><td rowspan="2">UNISA (Sy) CT&PDM</td></tr>
<tr><td>Off</td><td>TRANSYT (C) PDM/CTM</td><td>UNISA (C) CT&PDM</td><td>TRANSYT (C) PDM/CTM</td></tr>
</table>

-C: coordination; Sy: Synchronisation; GT&S: Green timing and Scheduling
-CF: capacity factor; TD: total delay
- of: generic objective function
-PDM/CTM: platoon dispersion model or cell transmission model; CT&PDM: cell transmission and platoon dispersion model

The paper is organised as follows: the proposed strategies are described in section 2, after some preliminary considerations the main purposes of the paper are reported in 2.4 where the proposed traffic flow model is described, in 2.5 where the optimisation problem is shown and in 2.6 where the adopted algorithms are summarised; in section

3, some computational results are shown; finally, conclusions and research perspectives are discussed in section 4.

2 PROPOSED STRATEGIES

This paper dealt with the network signal setting design through the application of two approaches: three step optimisation and two step optimisation as defined in the previous section.

The basic notations, the constraints and the optimisation problems are described below.

2.1 VARIABLES AND CONSTRAINTS

Assuming that the green scheduling is described by the stage matrix (i.e. the stage matrix composition and sequence), let

c be the cycle length, assumed known or as a decision variable (common to all junctions);

for each junction (not explicitly indicated)

t_j be the duration of stage j as a decision variable;

t_{ar}, be the so-called all red period at the end of each stage to allow the safe clearance of the junction, assumed known (and constant for simplicity's sake);

Δ be the approach-stage incidence matrix (or stage matrix for short), with entries δ_{kj} =1 if approach k receives green during stage j and 0 otherwise, assumed known;

l_k be the lost time for approach k, assumed known;

$g_k = \Sigma_j \, \delta_{kj} \, t_j - t_{ar} - l_k$ be the effective green for approach k;

$r_k = c - g_k$ be the effective red for approach k;

y_k be the arrival flow for approach k, assumed known;

s_k be the saturation flow for approach k, assumed known;

$(s_k \cdot g_k) / (c \cdot y_k)$ be the capacity factor for approach k;

and for each junction in the network

ϕ_i be the node offset as the time shift between the start of the plan for the junction i and the start of the reference plan, say the plan of the junction number 1, $\phi_1 = 0$. Given such a reference value, all the other m-1 nodes, where m is the number of junctions in the network, are independent variables.

As well known, computing total delay through a traffic flow model needs the link offset between each pair of adjacent junctions, $\phi_{ij} = -\phi_{ji}$.
Let the junction network be represented by an undirected graph with a node for each junction and an edge for each pair of adjacent junctions (the actually traffic directions are irrelevant). According to this representation

- if such a network is loop less, all the m-1 link offsets are independent (as many as the independent node offsets) and may be used as decision variables; arterials are a special case of such kind of networks;
- on the other hand, if the network contains k independent loops, the number of independent link offsets will be equal to m – k; in this case it is better to use the m-1 independent node offsets as optimisation variables.

Some constraints were introduced in order to guarantee:

stage durations being non-negative

$$t_j \geq 0 \qquad \forall\, j$$

effective green being non-negative

$$g_k \geq 0 \qquad \forall\, k$$

this constraint is usually guaranteed by the non-negative stage duration, but for a too short cycle length with regard to the values of all-red period length and lost times, say

$$\Sigma_j \, MAX_k \, (\delta_{kj} \, l_k + t_{ar}) \geq c$$

consistency among the stage durations and the cycle length

$$\Sigma_j \, t_j = C$$

the minimum value of the effective green timing

$$g_k \geq g_{min} \qquad \forall k$$

A further constraint was included in order to guarantee that the capacity factor must be greater than 1 (or any other value)

$$((s_k \cdot g_k) / (c \cdot y_k) - 1) \geq 0 \quad \forall k$$

Such a constraint may be added only after having checked that the maximum junction capacity factor for each approach k in the junction i is greater than 1, otherwise a solution may not exist whatever the objective function is.

Finally let assume

$$c \geq \phi_i \geq 0.$$

2.2 OBJECTIVE FUNCTIONS

At a single junction, the objective functions in the optimisation problems were:

- the junction capacity factor computed as

$$CF = MAX_k\ (s_k \cdot g_k) / (c \cdot y_k) \tag{1}$$

- the total delay computed
  - for non-interacting approaches (isolated or external junctions) by the two terms Webster's formula (Webster, 1958) as

$$TD = \Sigma_k\ y_k \cdot (0.45 \cdot c \cdot (1 - g_k /c)^2 / (1 - y_k / s_k)$$

$$+ y_k \cdot 0.45 / (s_k \cdot g_k / c) \cdot ((g_k / c) \cdot (s_k / y_k) - 1))) \tag{2}$$

  - for the interacting approaches by evaluating vehicles queuing interval by interval and considering input as the flow obtained by cyclic flow profiles. A more detailed expression consistent with the traffic flow modelling will be described in subsection 2.4.

Further objective functions could be considered such as the queue length, the number of stops etc.

2.3 STAGE COMPOSITION AND SEQUENCE

A set of feasible stages is built up by assuming that all the approaches in a stage are mutually compatible and each stage is maximal. Given the set of stages, that is the stage composition, some of them (not necessary all) produce a feasible sequence if all approaches have green at least in one stage. Commonly if an approach has green in more than one stage, those stages are adjacent (that is relevant for sequence with at least four stages).

All possible sequences can be grouped in equivalent classes; each class contains all sequences leading to the same green timings, the same TDs and CFs. Optimal offsets generally change in a consistent way among all the sequences in an equivalent class, thus:

- when only two stages are available {(1,2)} two possible sequences {(1,2); (2,1)} are equivalent and only one equivalent class exists;
- when three stages are available {1,2,3}, two equivalent classes exist, namely, the class of sequences generated by translation {(2,3,1); (3,1,2)} and the class of sequences generated by mirroring {(3,2,1); (2,1,3); (1,3,2)}.

As dealt with in a future paper, this analysis may be further extended when four on more stages are available.

Let m be the number of junctions in the network and ni the number of equivalent classes for each junction i, $\prod_{i=1:m} ni^{mi}$ combinations of classes can be generated and must be tested to achieve optimum.

In this paper an explicit enumeration will be carried out. The implicit enumeration will be addressed in future papers.

2.4 TRAFFIC FLOW MODEL

The most straightforward continuous traffic flow model is the first order model developed concurrently by Lighthill & Whitham (1955) and Richards (1956), based on the assumption that the number of vehicles is conserved between any two points if there are no entrances (sources) or exits (sinks). This produces a continuous model known as the Lighthill-Whitham-Richards (LWR). To solve the LWR space continuous problem the Cell Transmission Model was introduced (Daganzo; 1994). Since CTM assumes the same speed for all the vehicles on a road, it cannot fully predict realistic traffic flow behaviour as the platoons keep the same density when moving from the upstream stop-line section to the downstream section, and all vehicles travel at the same free flow speed.

One of the aim of this paper was to implement a traffic flow model which overcame the limitations of PDM and CTM. In particular major drawbacks of the PDM are in queuing simulation since queues are assumed vertical; on the other hand the CTM include the horizontal queuing at the cost of not considering the platoon dispersion. Finally, the proposed model allows horizontal queues and platoon dispersion modelling

2.4.1 CELL TRANSMISSION MODEL

In a cell-transmission model, a link is represented by a sequence of small sections (cells) and a record of the cell contents (number of vehicles) as time goes on is kept. This record is updated at closely spaced instants (clock ticks) by calculating the number of vehicles that crosses the boundary separating each pair of adjoining cells during the corresponding clock interval.

The main drawback of the CTM is that the density is equally distributed (such as constant) over the cell. The number of vehicles in the cell depends on its spatial extension and smaller is the cell length (thus smaller is the clock tick), lower is the error in density estimation made by finite approximation. On the other hand, particularly in case of large network, the number of cells (and computational complexity of the problem) will increase.

Thus in the proposed model the clock tick is around 1 second and in order to improve the density approximation as discrete variable, an additional equation based on speed-density relationship, was included.

In the CTM formulation, let

n_i, be the number of vehicles on the cell i;

Q_i, be the maximum flow rate in cell i

d_{i+1}, be the wave speed coefficient of cell i+1

N_{i+1}, be the maximum number of vehicles present in the cell i+1.

The average flow $Y_i(t)$ on the link i from clock tick t to clock tick t+1 is the result of a comparison between the maximum number of vehicles that can be “sent” by the cell directly upstream of the boundary

$$S_i(t) = \min\{Q_i , n_i\} \tag{3}$$

and those that can be “received” by the downstream cell

$$R_i(t) = \min \{Q_i , d_i[N_i - n_i]\} \tag{4}$$

Hence the flow $Y_i(t)$ can be written as:

$$Y_i(t) = \min \{S_i(t), R_{i+1}(t)\} \quad (5)$$

2.4.2 CELL TRANSMISSION AND PLATOON DISPERSION MODEL (CT&PDM)

As shown above, in the CTM equations, due to the assumption that all vehicles travel at the same speed (keeping as such the same density) for getting in the downstream section, no platoon dispersion may be detected. To overcome this shortfall, the flow propagation has been modelled by employing for each cell the well-known Drake speed-density relationship in which the corresponding flow, $X_i(t)$, is computed as follows:

$$X_i(t) = k_i(t)\, v_0 \exp [-0.5 (k_i(t) / k_m)^2] \quad (6)$$

where:

$k_i(t) = [n_i(t) + n_{i+1}(t)] / 2L$: is the density of cell i and cell i+1 at time t, being L the length of the cell;

v_0 : is the free-flow-speed;

k_m : is the traffic density at maximum flow,

The flow $Y_i(t)$ may be so formulated as:

$$Y_i(t) = \min \{S_i(t), R_{i+1}(t), X_i(t)\} \quad (7)$$

2.4.3 TOTAL DELAY COMPUTATION

Let the period of analysis T be a multiple of the cycle duration in so far as the quantities relating to a cycle were repeated in the subsequent cycles. The interval of duration T was divided into intervals of equal duration $\Delta\tau$[2]. For the computation of the total delay the cumulated input flows, $Ciy_k^j(t)$, and those output, $Coy_k^j(t)$, on the stop line on approach k of junction j, in the subsequent sub-intervals t were compared.

The Deterministic Total Delay (DTD_k^j) cumulated in the interval [0, T] for approach k of junction j was then given by the following expression:

$$DTD_k^j = \Sigma_{t=1..T/\Delta\tau} (Ciy_k^j(t \times \Delta\tau) - Coy_k^j(t \times \Delta\tau))\, \Delta\tau \quad (8)$$

Thus delay experienced on an interacting approach is a function of the offsets between the timing plans. In fact, such a delay depends on the output flow in the downstream

[2] in our application a clock tick of 1 second were adopted.

junction which is obtained by starting from the input flow in the upstream junction through the phenomenon of dispersion.

Let s_k^j be the saturation flow on approach k of junction j, the Stochastic and Oversaturation component of Total Delay $SOTD_k^j$ on approach k of junction j is computed using the following expression

$$SOTD_k^j = \{ [(y'^j_k - s_k^j)^2 + (4y'^j_k/T)]^{0.5} + (y'^j_k - s_k^j)\}\ T/4 \quad (9)$$

and considering the average of the values of the cyclic flow profile along the connecting link arriving at approach k of considered junction j, y'^j_k as input flow.

An example of the model is shown in following with regard to cyclic flow profile (see Figure 6, Figure 7 and Figure 8) and with regard to the flow progression over network links (see Figure 9).

2.5 OPTIMISATION PROBLEMS

The paper dealt with the application of two optimisation strategies:

A. three step optimisation; in this case the scheduling, the green timings A(1) at each single junction and the offsets A(2) were computed in three separate steps;

B. two step optimisation; the decision variables might be computed in two steps.

In particular, with regards to the latter this could be applied by following one of two feasible alternative combinations of variables

B.1) in the first combination (Sc&Gt-Off), the scheduling (Sc) and green timings (Gt) were computed at a first step, and the offsets (Off) were computed at a second step;

B.2) in the second combination (Sc-S), the scheduling was carried out at a first step by defining the stage matrix for each junction, then the green timings and the offsets (e.g. the synchronisation approach, S) were computed at a second step.

In this paper, the A and B.2 strategies were adopted. In Figure 1 a general framework of such strategies is shown. The stage- sequence incidence matrix for each single junction was generated externally and then iteratively embedded in the NSSD until the optimum solution was reached.

The explicit enumeration was ended when the iteration $k = \prod_{i=1:m} ni^{mi}$.

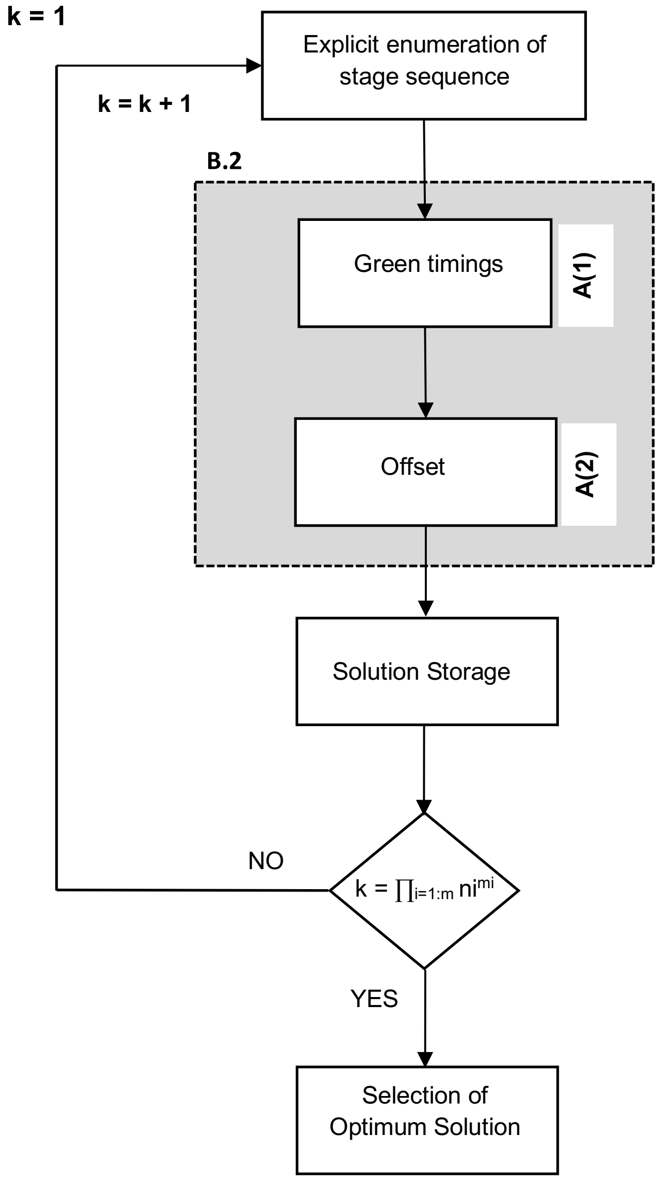


Figure 1: General Framework of A and B.2 optimisation strategies (for each junction)

In three step optimisation (A) once all the sequences are explicitly enumerated, multi-criteria Genetic Algorithms based on total delay minimisation and capacity factor maximisation were applied for green timings at each single junction; the minimisation of the network total delay, with respect to the independent link offsets (loop less network) or node offsets (network with loops) was carried out by applying a Hill-

Climbing algorithm; the scheduling optimisation was computed through explicit enumeration of all the possible sequences.

In the two step optimisation (B.2), the green timings and the offsets were computed simultaneously (Sc-S), by considering mono-criterion optimisation based on total delay minimisation. The adopted algorithm was the Simulated Annealing.

In Table 2 and Table 3 each strategy is summarised with respect to the optimisation method and the objective function; in particular, bold cases refer to applications described in section 3.

Table 2: Overview of three step strategy (A) w.r.t. Optimisation methods/objective functions

Strategy	**Variables** **GT** **Optimisation method**	**Variables** **Offsets** **Optimisation method**	**Variables** **GT** **Objective functions**	**Variables** **Offsets** **Objective functions**
A	Mono	Mono	TD / CF	**TD**
	Multi	**Mono**	**TD & CF**	

Table 3: Overview of two step strategy (B.2) w.r.t. Optimisation methods/objective functions

Strategy	**Variables** **GT** / **Offsets** **Optimisation method**	**Variables** **GT** / **Offsets** **Objective functions**
B.2	**Mono**	**TD**
	Multi	TD & o.f.*

*o.f.: It represents a generic objective function

2.6 SOLUTION ALGORITHMS

Some main issues should be addressed to solve the NSSD: i) the optimisation approach, mono/multi-criteria functions; ii) the variables in the single junction optimisation or in the network optimisation; iii) the network layout (arterial as network loop less and network with at least one loop).

Based on all the previous considerations, meta-heuristic algorithms are generally adopted since they allowed for the approximation of the optimal solution of difficult combinatorial problems. In general, these approaches have evolved through interactions and analogies derived from biological, physical, computer and decision making sciences (Genetic Algorithms, GAs, Hill-Climbing, HC, Simulated Annealing, SA and other not used in this paper). To get a fully operational algorithms from a meta-heuristic requires the specifications of several functions and/or parameters whose meaning depends on the meta-heuristic itself.

As better described in sub 2.5. we cope with different optimisation problems; for each of them the selection of an algorithm was based on the trade-off between the effectiveness of the algorithm related to the space solution exploration, the parameters to be set and the computational effort depending on the algorithm complexity. Some considerations about benefits and drawbacks of the algorithms used in this paper will be described in following.

In SA the number of the parameters to be set is lower than the GAs and in terms of computational effort, the performance is generally better. Furthermore, SA can easily handle changes in the objective function, while with GAs the selection of the solution based on the fitness function may complicate the search for the optimum. On the other hand for highly dimension space of solution Gas are generally more effective than SA, for their flexibility due to the two genetic operators, crossover and mutation at the cost of a higher number of parameters to be set (for unfamiliar readers a more detailed description of the GAs is given in Appendix A)

GAs, in particular Non Dominated Sorting Genetic Algorithms II (NSGA-II), guarantee the exploration of a wider range of a solutions space, allowing for the computation of an approximation of the entire Pareto front. This provides a higher resilience in multi-criteria optimisation problems.

The HC algorithm uses a small amount of memory with respect to other approaches and it is characterised by lower processing time. One of the main drawbacks of the HC algorithm is the risk of hitting a local minimum/maximum. When the algorithm finds a local minimum/maximum value, which is lower/higher than the others in the neighbourhood, it freezes. However, this does not exclude the presence of far global minima with lower values. The HC search has the same drawbacks of a greedy search. It moves quickly to the best node by also using a 'short-sighted' strategy. For this reason its application is generally associated with optimisation problems without large range of a solutions space.

In order to clearly summarise the adopted approaches as well as highlighting the correspondence between the problems described previously and the algorithms, two further columns are added to Table 2 and Table 3 leading to Table 4 for three step optimisation (A) and Table 5 for two step optimisation (B.2).

Table 4: Overview of three step strategy(A) w.r.t. Optimisation methods/objective functions including Algorithms

Strategy	layout	Variables GT	Variables Offsets	Variables GT	Variables Offsets	Variables GT	Variables Offsets
		Optimisation method		Objective functions		Algorithms	
A	Two junction arterial	Mono	Mono	TD / CF	TD	HC	HC
		Multi	**Mono**	**TD & CF**		**GAs**	**HC**
	Four junction network	Mono	Mono	TD / CF		HC	HC
		Multi	**Mono**	**TD & CF**		**GAs**	**HC**

Table 5: Overview of two step strategy (B.2) w.r.t. Optimisation methods/objective functions including Algorithms

Strategy	layout	Variables GT	Variables Offsets	Variables GT	Variables Offsets	Variables GT	Variables Offsets
		Optimisation method		Objective functions		Algorithms	
B.2	Two junction arterial	Mono		TD	TD	HC	
		Multi		TD & o.f.*		GAs	
	Four junction network	**Mono**		**TD**		**SA**	
		Mono		TD & o.f.*		GAs	

*o.f.: It represents a generic objective function

3 NUMERICAL APPLICATIONS

In this section, the numerical results obtained by the proposed approaches and benchmark tools are compared; in particular, different layouts, a two junction arterial and a four junction network, different meta-heuristics and different traffic flow models are analysed.

The adopted strategies are shown in Table 6, in particular:

- the approaches are the UNISA, as described in section 2, vs. the benchmark approach, OSCADY PRO® /TRANSY14® - TRL software;
- the traffic flow model (TM) is the cell transmission and platoon dispersion model (CT&PDM), described in subsection 2.4 vs TRANSYT14® traffic flow models, i.e. the Platoon Dispersion Model (PDM) or the Cell Transmission Model (CTM)[3];
- -the optimisation strategies are three step, A, and two step, B.2;
- -the algorithms are HC,GAs and SA.

[3] TRANSYT14 may be applied with either traffic models (alternatively but not simultaneously).

This is a post-peer-review, pre-copyedit version of an article published in Transportation Research Part C The final authenticated version is available online at: http://dx.doi.org/10.1016/j.trc.2015.03.032

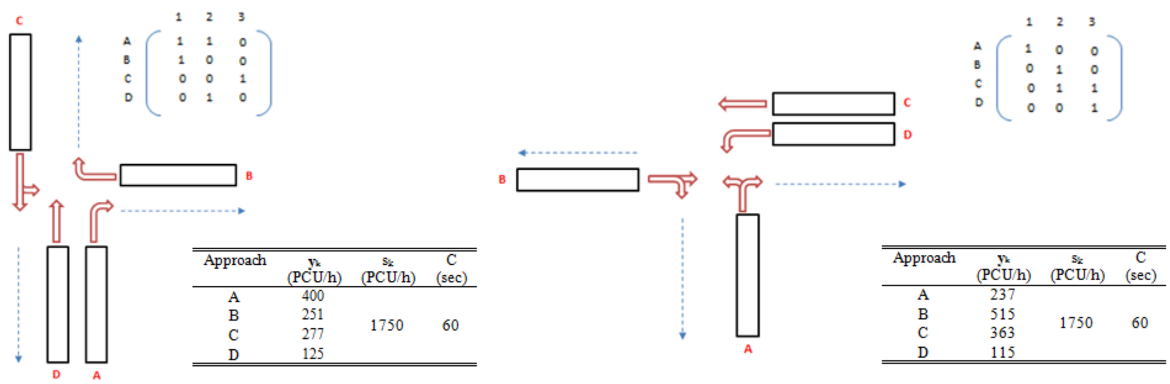
Table 6: Adopted strategies

Layout	Strategy	UNISA	TM	TRL (OSCADY& TRANSYT[4])	TM
Two junction arterial	A	UNISA-GA&HC	CT&PDM	OSCADY&TRANSYT-HC	PDM or CTM
Four junction network	A	UNISA-GA&HC		OSCADY&TRANSYT-HC	
	B.2	UNISA-SA		TRANSYT-SA	

The implementation of such strategies was worked out using a MATLAB (Release 2013b) code provided by the authors. All the applications were run on a PC having an Intel(R) Xeon(R) CPU E5-1603, clocked at 2.8GHz and with 4GB of RAM.

3.1 TWO JUNCTION ARTERIAL

In this section the performances of three step method (A) applied to a two junction arterial (as a very simple example of a loop less network), are described. The layout together with the stage matrix and the main input data are shown in Figure 2. Once defined the connecting link between the two junctions their interaction is simulated through the traffic flow model CT&PDM.



$$\begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline A & 1 & 1 & 0 \\ B & 1 & 0 & 0 \\ C & 0 & 0 & 1 \\ D & 0 & 1 & 0 \end{array}$$

Approach	y_k (PCU/h)	s_k (PCU/h)	C (sec)
A	400	1750	60
B	251		
C	277		
D	125		

$$\begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline A & 1 & 0 & 0 \\ B & 0 & 1 & 0 \\ C & 0 & 1 & 1 \\ D & 0 & 0 & 1 \end{array}$$

Approach	y_k (PCU/h)	s_k (PCU/h)	C (sec)
A	237	1750	60
B	515		
C	363		
D	115		

Figure 2: Layout, Stage matrix, characteristics for junction 1 (left) and junction 2 (right)

Firstly UNISA-GA&HC CT&PDM was implemented, through the application of multi-criteria GAs to achieve single junction green timings (see Table 7 and Table 8) and mono-criterion HC to achieve optimal offsets (see Table 10 and schematic representation in Figure 3), by simulating traffic flow with CT&PDM. Successively a comparison, in terms of offsets and total delay, with OSCADY&TRANSYT–HC (PDM)

[4] The authors are awake that the total delay is analytically computed via different ways in UNISA and TRANSYT14.

and OSCADY&TRANSYT–HC (CTM) was carried out (see Table 10). In OSCADY PRO® a mono criterion optimisation, based on the maximisation of single junction capacity factor, was used to achieve single junction timings. In TRANSYT14® HC a mono-criterion optimisation, based on the minimisation of the arterial total delay, to get optimal offsets, was adopted, simulating alternatively traffic flows by PDM or CTM. Graphical outputs are shown in Figure 4.

All applications were carried out considering a fixed value of the cycle length computed by following the Webster indication. According to the procedure defined in subsection 2.3, let 2 be the number of junctions and 2 the equivalent classes, 4 combinations of sensitive sequences on arterial total delay were tried as shown in Table 9. The analysis was carried out in OSCADY&TRANSYT-HC (PDM) and as shown in the table, the optimal sequence corresponds to (1,2,3) for junction 1 and (3,2,1) for junction 2. Table 7 and Table 8 show how the results carried out are affected by the trade-off between total delay (minimisation) and capacity factor (maximisation) induced by the multi-criteria procedure. In general, a different effect may be reached by considering the capacity factor as a constraint and the total delay as an objective function, to be minimised; in this latter case, in fact, a dominance effect of the total delay is expected in terms of a performance indicator.

Table 7: Multi-criteria optimisation based on Total delay minimisation and Capacity Factor maximisation at junction 1 (Pop size = 20; Mutation rate = 0.8; Crossover rate = 0.0001). UNISA GA&HC

Stream	Green timings [s]	Total Delay [PCU-hr/hr]	Capacity Factor
A	36	1.32	2.20
B	23	1.55	2.03
C	24	1.69	2.11
D	13	0.99	2.04

Table 8: Multi-criteria optimisation based on Total delay minimisation and Capacity Factor maximisation at junction 2 (Pop size = 20; Mutation rate = 0.8; Crossover rate = 0.0001). UNISA GA&HC

Stream	Green timings [s]	Total Delay [PCU-hr/hr]	Capacity Factor
A	19	1.57	1,81
B	30	2.89	1.40
C	41	0.97	2.80
D	11	2.64	1.48

Table 9: Stage sequences vs. TD (fixed length between junctions equal to 500 m); equivalent sequences are in curly brackets

Stage sequence junction 1*	Stage sequence Junction 2**	TD [PCU hr/hr]
(1,2,3); {(2,3,1); (3,1,2)}	(1,2,3); {(2,3,1); (3,1,2)}	15.18
(1,2,3); {(2,3,1); (3,1,2)}	**(3,2,1); {(1,3,2); (2,1,3)}**	**13.83**

(3,2,1); {(1,3,2); (2,1,3)}	(1,2,3); {(2,3,1); (3,1,2)}	21.57
(3,2,1); {(1,3,2); (2,1,3)}	(3,2,1); {(1,3,2); (2,1,3)}	17.01

* for junction 1 (stage 1: A,B/stage 2: A,D/stage 3: C);**for junction 2 (stage 1: A, stage 2: B,C/stage 3: C,D).

For each single junction, the green timings were obtained through the GAs application, and using the optimal stage sequences reported above, a sensitivity analyses of TD with respect to distance between two junctions was carried out. Starting from the stages duration obtained through the GAs implementations, once fixed four layouts based on connecting link lengths variability (i.e. from junction 1 to junction 2: 500 m; 400 m; 300 m; 200 m), the arterial total delay and the offset between the signal timing plans of the interacting junctions were carried out through traffic flow modelling. The considered stage sequences will be consistent with results above obtained (see Table 9) for a given distance between junctions (L=500m).

Table 10: Numerical results for OSCADY&TRANSYT-HC (PDM), OSCADY&TRANSYT-HC (CTM) and UNISA-GA&HC (CT&PDM) w.r.t. length variations

Length	**OSCADY&TRANSYT HC (PDM)**			**OSCADY&TRANSYT HC (CTM)**			**UNISA GA&HC (CT&PDM)**		
[m]	**Offset [sec]**	**TD [PCU hr/hr]**	**DOS* [%]**	**Offset [sec]**	**TD [PCU-hr/hr]**	**DOS* [%]**	**Offset [sec]**	**TD [PCU-hr/hr]**	**DOS* [%]**
500	55	13.83	71	47	15.84	80	45	10.12	67
400	44	12.77		39	14.41		34	11.51	
300	34	11.38		24	14.96		22	11.90	
200	22	9.85		13	26.67		10	23.97	

* DOS: is the maximum degree of saturation between the two junctions

With reference to the results shown in Table 10 two main observations should be made. The first one is related to the estimation of the maximum degree of saturation between the junctions which was lower in UNISA-GA&HC (CT&PDM) than OSCADY&TRANSYT-HC (PDM/CTM). That was justified by the different structure of the objective function adopted in the single junction signal setting. Indeed, as above described, in UNISA-GA&HC a trade-off between total delay (to be minimised) and capacity factor (to be maximised) was looked for.

The second was related to the total delay/offset estimation with respect to the length variation; as shown in Figure 4a, offset almost linearly increases with length for all the optimisation strategies. This trend might be explained by a decrease in the vehicles cruise time which forced a temporary close start of a downstream signal plan compared to the upstream one. With regards to the total delay (see Figure 4b) it was noted that such an indicator in OSCADY&TRANSYT-HC (PDM) decreased against link length for distances closer than 300 m, since no horizontal queues were modelled. On the other hand similar trends were observed in OSCADY&TRANSYT-HC (CTM) and UNISA-GA&HC (CT&PDM) and lower total delays were obtained from the latter since a dispersion function was applied. In conclusion, the proposed traffic flow model seemed to overcome the limitations both of PDM (in modelling horizontal queues) and CTM (in modelling platoon dispersion).

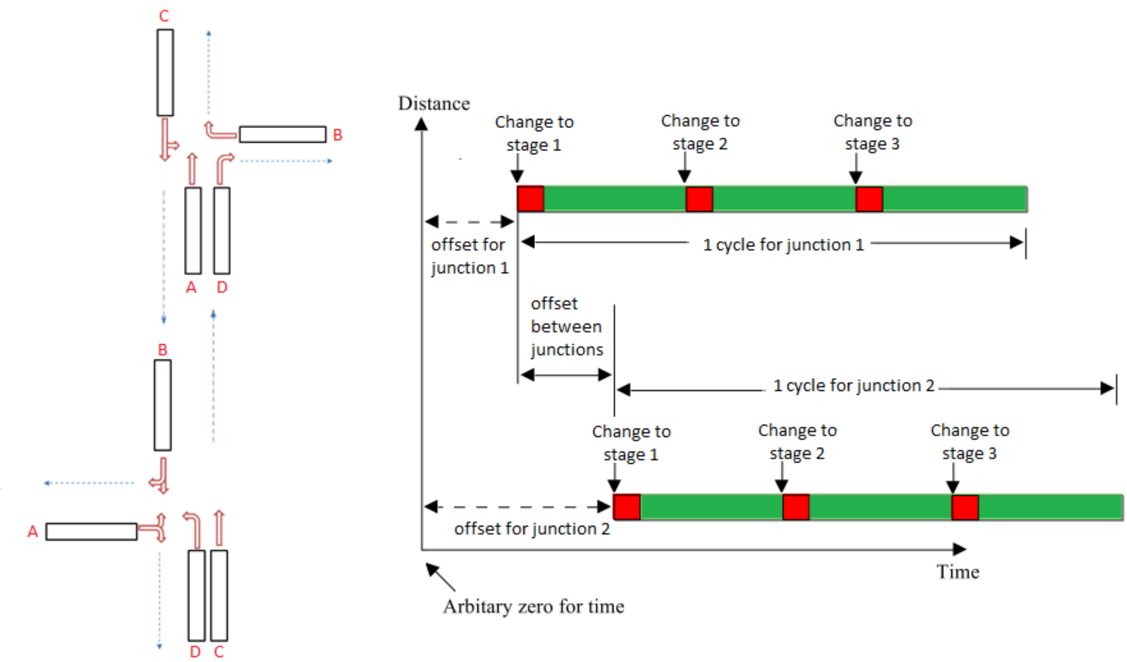
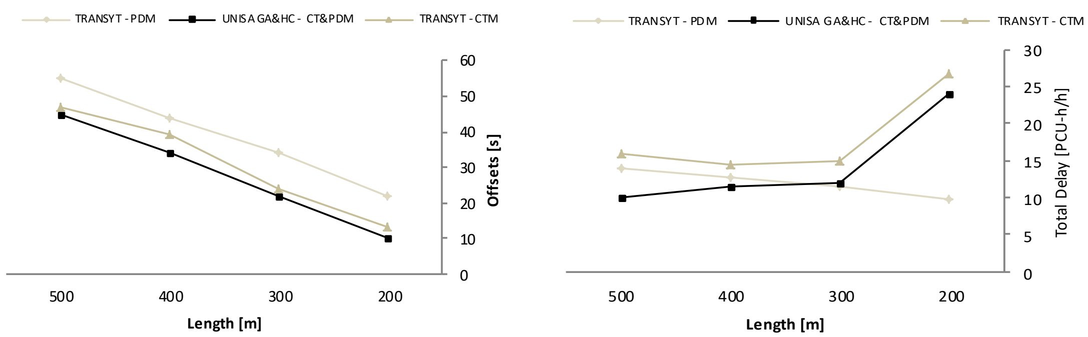


Figure 3: Schematic representation of the offsets (green and red bars are not proportional to actual values)

Figure 4: a) Optimal link offset w.r.t. length; b) Total Delay w.r.t. length



The results obtained in the two junction arterial showed the effectiveness of the adopted model (CT&PDM) in simulating traffic flows approaching signalised junctions further than queue forming phenomena along the connecting link.

3.2 FOUR JUNCTION NETWORK

In this section, the design strategies and optimisation algorithms were then analysed in more depth through an application to a network with four interacting signalised junctions forming a loop (shown in Figure 5 as a square for simplicity's sake).

Firstly, three step strategy (A) through UNISA-GA&HC (CT&PDM) was carried out. The NSGA-II algorithm was applied to obtain single junction timings while optimal link offsets were achieved through a HC algorithm.

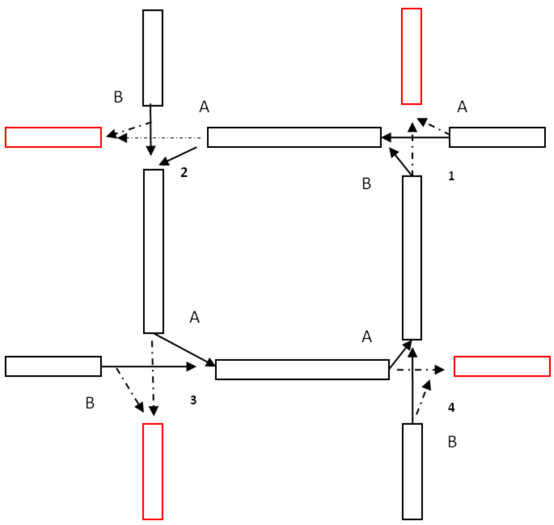


Figure 5: Four junction network layout

The results, collected in terms of link offsets, network total delay and network degree of saturation, DOS, were then subsequently compared with those obtained by TRANSYT14® - TRL implementation, where mono-criterion optimisations both to achieve single junction timings (obtained by internally implementing OSCADY PRO® - TRL) and optimum link offsets (here a HC was adopted, selected from a choice set HC/ SA) were applied.

Further comparisons were subsequently carried out with reference to two step strategy (B.2) where single junction green timings were designed simultaneously with link offsets considering the minimisation of network total delay. The SA algorithm, as described in sub-section A.3, was applied.

In terms of flow simulation modelling, a detailed comparison was carried out among TRANSYT14® - TRL (PDM/CTM) and UNISA (CT&PDM). Such a match focuses on the model's capability in capturing both dispersion and traffic bottleneck phenomena when length variations over the link composing the network occurred (once given fixed flows).

3.2.1 THREE STEP OPTIMISATION (A):

Once the stage matrix for each junction was fixed as well as the value of the cycle length computed by the Webster indication (equal to 90 sec), a multi-criteria optimisation (minimisation of total delay and maximisation of capacity factor) performed by NSGA-II algorithm was performed to achieve single junction timings. All information related to the network, in terms of flow and saturation flow for each approach, are summarised in Table 11.

Table 11: Flow and saturation flow for each approach

Junction	Stage	y_k [veic/h]	s_k [veic/h
1	A	400	1750
	B	245	
2	A	390	
	B	227	
3	A	422	
	B	286	
4	A	375	
	B	191	

Results carried out for each junction with regard to the optimisation criteria are shown in the following tables (from Table 12 to Table 15). As in the case of two junction arterial application, the multi-criteria optimisation showed, in terms of performance indicators, the trade-off effect between the two criteria considered in the optimisation; in fact, any solution might be expected to be not dominant with respect to both criteria at the same time.

Table 12: Multi-criteria optimisation based on Total delay minimisation and Capacity Factor maximisation at junction 1 (Pop size = 20; Mutation rate = 0.8; Crossover rate = 0.0001). UNISA-GA&HC

Stream	Total Delay [PCU-hr/hr]	Capacity Factor	Effective green
A	1.27	1.66	30
B	2.35	2.15	55

Table 13: Multi-criteria optimisation based on Total delay minimisation and Capacity Factor maximisation at junction 2 (Pop size = 20; Mutation rate = 0.8; Crossover rate = 0.0001). UNISA-GA&HC

Stream	Total Delay [PCU-hr/hr]	Capacity Factor	Effective green
A	1.75	2.32	39
B	1.91	2.47	46

Table 14: Multi-criteria optimisation based on Total delay minimisation and Capacity Factor maximisation at junction 3 (Pop size = 20; Mutation rate = 0.8; Crossover rate = 0.0001). UNISA-GA&HC

Stream	Total Delay [PCU-hr/hr]	Capacity Factor	Effective Green
A	1.49	1.88	32

B	2.30	2.68	53

Table 15: Multi-criteria optimisation based on Total delay minimisation and Capacity Factor maximisation at junction 4 (Pop size = 20; Mutation rate = 0.8; Crossover rate = 0.0001). UNISA-GA&HC

Stream	Total Delay [PCU-hr/hr]	Capacity Factor	Effective Green
A	1.53	2.38	39
B	1.79	2.44	46

Starting from the timings obtained by the application of the GAs' implementations, once five layouts (varying the connecting links length) were fixed (see Table 16), the network total delay, the degree of saturation and the link offsets between the signal plans of the interacting junctions were carried out through HC algorithm and modelling network flow propagation by CT&PDM. The results were then compared with those obtained by OSCADY, to achieve single junction timings, and TRANSYT-HC (PDM/CTM) to obtain optimal link offsets (see from Table 17 to Table 19).

Table 16: Lengths combination and ID

L1 [m]	L2 [m]	L3 [m]	L4[m]	Lengths combination ID
500	200	600	200	I
400	190	500	180	II
300	170	350	160	III

As highlighted in Table 17, even though comparable results in the computation of network total delay were identified for lengths combination I and II, there was a significant discrepancy for lengths combination III, among the results carried out in OSCADY&TRANSYT-HC (PDM), OSCADY&TRANSYT-HC (CTM) and UNISA-GA&HC (CT&PDM). De facto, as the distances decreased, the Network Total Delay did not change significantly in OSCADY&TRANSYT-HC (PDM), while it increased considerably in UNISA-GA&HC (CT&PDM) and OSCADY&TRANSYT-HC (CTM). Such a result may be related to the higher reliability of CT&PDM (in UNISA-GA&HC) and CTM (in OSCADY&TRANSYT-HC) in modelling queues in terms of its spatial extensions. Therefore, when short links were considered and traffic was more likely to be restricted by downstream traffic and traffic signals, CT&PDM and CTM could be considered as more reliable compared to PDM which employed vertical queuing.

Table 17: Numerical results for OSCADY&TRANSYT-HC (PDM), OSCADY&TRANSYT-HC (CTM) and UNISA-GA&HC (CT&PDM) w.r.t. length variations

Lengths combination	OSCADY&TRANSYT HC (PDM)		OSCADY&TRANSYT HC (CTM)		UNISA GA&HC (CT&PDM)	
[ID]]	TD [PCU-hr/hr]	DOS [%]	TD [PCU-hr/hr]	DOS [%]	TD [PCU-hr/hr]	DOS [%]
I	10.14	65	13.11	72	8.48	60

II	12.53	12.81	12.83
III	7.53	22.41	19.34

Table 18: Optimised values of the offsets achieved by OSCADY&TRANSYT-HC - (PDM) and by UNISA-GA&HC (CT&PDM) w.r.t. length variations

Lengths comb	OSCADY&TRANSYT-HC (PDM)				UNISA-GA&HC (CT&PDM)			
[ID]	Φ1-2 [sec]	Φ2-3 [sec]	Φ3-4 [sec]	Φ4-1 [sec]	Φ1-2 [sec]	Φ2-3 [sec]	Φ3-4 [sec]	Φ4-1 [sec]
I	77	24	0	79	54	31	58	37
II	80	20	3	77	49	20	54	57
III	35	26	45	74	52	3	46	79

Table 19: Optimised values of the offsets achieved by OSCADY&TRANSYT-HC - (CTM) and by UNISA-GA&HC (CT&PDM) w.r.t. length variations

Lengths comb	OSCADY&TRANSYT-HC (CTM)				UNISA-GA&HC (CT&PDM)			
[ID]	Φ1-2 [sec]	Φ2-3 [sec]	Φ3-4 [sec]	Φ4-1 [sec]	Φ1-2 [sec]	Φ2-3 [sec]	Φ3-4 [sec]	Φ4-1 [sec]
I	19	51	45	65	54	31	58	37
II	8	32	50	0	49	20	54	57
III	71	33	63	13	52	3	46	79

Furthermore, as already shown in the case of two junction arterial, due to the multi-criteria optimisation, the capacity factor obtained in UNISA-GA&HC (CT&PDM) was better than those obtained by the benchmark strategies.

With respect to the layout of ID I, cyclic flow profiles (CFPs) obtained through UNISA-GA&HC (CT&PDM) are shown below. In particular Figure 6 refers to results obtained at upstream of link 1-2, Figure 7 refers to results obtained in a middle section of the link 1-2, and Figure 8 at downstream of link 1-2.

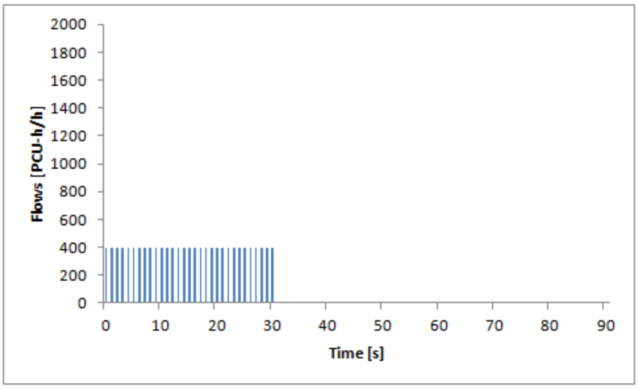


Figure 6: UNISA-GA&HC (CT&PDM) CFP at the beginning of link 1-2 (signalised junction 1)

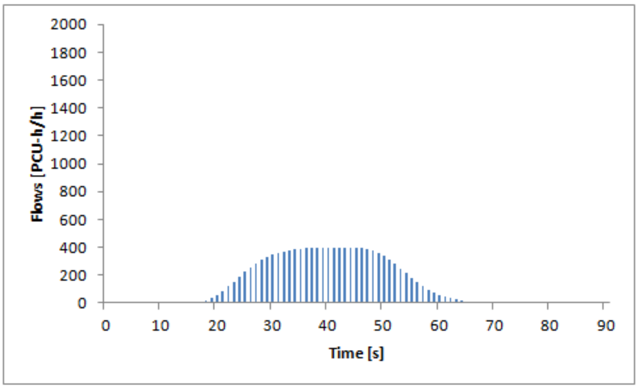


Figure 7: UNISA-GA&HC (CT&PDM) CFP in the middle section of link 1-2

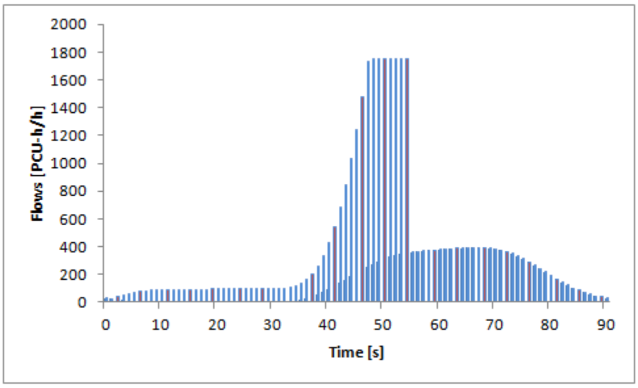


Figure 8: UNISA-GA&HC (CT&PDM) CFP at the end of link 1-2 (signalised junction 2)

3.2.2 TWO STEP OPTIMISATION (B.2)

The network parameters (i.e. the stage matrix for each junction, the input flows and the three combinations of links length) were fixed as equal to those adopted in previous implementations. The network total delay, the degree of saturation, the link offsets between the signal plans of the interacting junctions and the green timings (see from Table 20 to Table 24) were carried out by adopting UNISA-SA (CT&PDM), thus the green timings and link offsets are simultaneously optimised through a SA algorithm and modelling network flow propagation by CT&PDM The results were then compared with those obtained by TRANSYT-SA (PDM/CTM)

As it was previously stated (in three step optimisation (A)), the adopted approaches provided different queuing models which directly affected the obtained results. As shown in Table 20, in the case of higher distances (length combinations I and II), the total delay was overestimated (12.04 PCU-hr/hr and 14.25 PCU-hr/hr) in TRANSYT-SA (CTM) whereas lower values were produced by TRANSYT-SA (PDM) (9.04 PCU-hr/hr and 7.21 PCU-hr/hr) and UNISA-SA (CT&PDM) (6.71 PCU-hr/hr and 9.41 PCU-hr/hr); at lower distances (length combination III), the value computed by TRANSYT-SA (PDM) (6.85 PCU-hr/hr) seemed to be not sensitive to the reduction of distances while similar results (15.38 PCU-hr/hr vs. 14.16 PCU-hr/hr) were shown in TRANSYT-SA (CTM) and UNISA-SA (CT&PDM) since they took account of the spatial extension of queues. These results confirmed the suitability of the proposed model for simulating both dispersion and queue blockage phenomena.

Table 20: Numerical results for TRANSYT-SA (PDM), TRANSYT-SA (CTM) and UNISA-SA (CT&PDM) w.r.t. length variations

Lengths combination	TRANSYT-SA (PDM)		TRANSYT-SA (CTM)		UNISA-SA (CT&PDM)	
[id]	TD [PCU-hr/hr]	DOS [%]	TD [PCU-hr/hr]	DOS [%]	TD [PCU-hr/hr]	DOS [%]
I	9.04	53	12.04	70	6.71	49
II	7.21	49	14.25	64	9.41	62
III	6.85	50	15.38	74	14.16	66

Table 21: Optimised values of the offsets achieved by TRANSYT-SA (PDM) and by UNISA-SA (CT&PDM) w.r.t. length variations

Lengths combination	TRANSYT-SA (PDM)				UNISA-SA (CT&PDM)			
[id]	Φ1-2 [sec]	Φ2-3 [sec]	Φ3-4 [sec]	Φ4-1 [sec]	Φ1-2 [sec]	Φ2-3 [sec]	Φ3-4 [sec]	Φ4-1 [sec]
I	28	21	69	62	54	13	72	41
II	43	20	53	64	49	22	44	65
III	45	19	51	65	63	3	39	75

Table 22: Optimised values of the offsets achieved by TRANSYT-SA (CTM) and by UNISA-SA (CT&PDM) w.r.t. length variations

Lengths combination	**TRANSYT-SA (CTM)**				**UNISA-SA (CT&PDM)**			
[id]	**Φ1-2 [sec]**	**Φ2-3 [sec]**	**Φ3-4 [sec]**	**Φ4-1 [sec]**	**Φ1-2 [sec]**	**Φ2-3 [sec]**	**Φ3-4 [sec]**	**Φ4-1 [sec]**
I	18	14	23	35	54	13	72	41
II	38	34	55	53	49	22	44	65
III	56	12	44	68	63	3	39	75

Table 23: Green timings achieved by TRANSYT-SA (PDM) and by UNISA-SA (CT&PDM) w.r.t. length variations

Lengths combination	**TRANSYT-SA (PDM)**				**UNISA-SA (CT&PDM)**			
[id]	**g_1A* [sec]**	**g_2A [sec]**	**g_3A [sec]**	**g_4A [sec]**	**g_1A* [sec]**	**g_2A [sec]**	**g_3A [sec]**	**g_4A [sec]**
I	43	33	58	38	37	44	42	45
II	40	41	43	42	33	38	44	33
III	39	37	42	40	38	36	41	31

*It refers to effective green duration of junction i and stage j (i.e. g_{ij})

Table 24: Green timings achieved by TRANSYT-SA (CTM) and by UNISA-SA (CT&PDM) w.r.t. length variations

Lengths combination	**TRANSYT-SA (CTM)**				**UNISA-SA (CT&PDM)**			
[id]	**g_1A* [sec]**	**g_2A [sec]**	**g_3A [sec]**	**g_4A [sec]**	**g_1A* [sec]**	**g_2A [sec]**	**g_3A [sec]**	**g_4A [sec]**
I	44	56	67	35	37	44	42	45
II	51	51	68	65	33	38	44	33
III	46	54	63	26	38	36	41	31

*It refers to effective green duration of junction i and stage j (i.e. g_{ij})

This is a post-peer-review, pre-copyedit version of an article published in Transportation Research Part C The final authenticated version is available online at: http://dx.doi.org/10.1016/j.trc.2015.03.032

4 CONCLUSIONS AND RESEARCH PERSPECTIVES

This paper focuses on the Network Signal Setting Design. The main purposes of the paper are

- the green timings and offsets optimisation (by embedded procedure) and the stage sequence optimisation (by explicit enumerative approach);
- the application of some meta-heuristic algorithms for the solution of the mono-criterion/multi-criteria optimisation;
- the traffic flow modelling by a combined cell transmission and platoon dispersion model.

An overview of all optimisation strategies is shown in the following Table 25 (repeating Table 4 and Table 5 for the reader’s convenience) in which there are the strategies herein applied to two different layouts (i.e. two junction arterial vs. four junction network). In particular, in bold are the strategies (i.e. three step optimisation (A)) applied to the two junction arterial, and both strategy (i.e. two step (B.2) and three step optimisation (A)) applied to the four junction network.

The obtained results have been compared with those obtained by benchmark software (i.e. OSCADY PRO® and TRANSYT14® - TRL). The proposed strategies may be considered effective with respect to the optimisation methods and the multi-criteria optimisation should be considered a suitable approach.

Table 25: Overview of two step (B.2) and three step strategy (A) w.r.t. Optimisation methods/objective functions including Algorithms

Strategy	**layout**	**Variables**		**Variables**		**Variables**	
		GT	**Offsets**	**GT**	**Offsets**	**GT**	**Offsets**
		Optimisation method		**Objective functions**		**Algorithms**	
B.2	Two junction arterial	Mono		TD	**TD**	HC	
		Multi		TD & o.f.*		GAs	
	Four junction network	**Mono**		**TD**		**SA**	
		Mono		TD & o.f.*		GAs	
A	**Two junction arterial**	Mono	Mono	TD / CF	**TD**	HC	HC
		Multi**	**Mono**	**TD & CF**		**GAs**	**HC**
	Four junction network	Mono	Mono	TD / CF		HC	HC
		Multi**	**Mono**	**TD & CF**		**GAs**	**HC**

*o.f.: It represents a generic objective function

Furthermore, the adopted traffic flow model (CT&PDM) aims to provide enhancements for effective traffic modelling among interacting junctions in which both dispersion and queue blockage phenomena are represented, thus, it combines the skills of CTM and PDM.

The "Time-distance diagrams" in Figure 9 are adopted to show the optimal green timings and offsets achieved through the implementation of UNISA-SA (CT&PDM). For brevity's sake results refer only to the case highlighted in grey in the above Table 25 and for distances L_1=500 m, L_2=200 m, L_3=600 m and L_4=200 m. The graphs provide a representation of the progression of traffic flow from stop-line to stop-line along a specified path (in this case from junction 1 to junction 4) by plotting traffic flow over time and distance. The progression bands are shown through a chromatic scale which indicates the consistency of the flows (considered as veh/clock tick. One clock tick is 1 second) for each cell adopted for the partition of the link length. According to space discretisation of CT&PDM, the cell length must be at least the maximum distance travelled by a vehicle in one clock tick that is the product of the free flow speed and the time measured in clock tick (in our case the free flow speed is 40km/h, the clock tick is 1/3600 h, thus the cell length is around 1/90 km). The number of cells for each link was computed once the cell length and the length of each stretch were known: L1= 5 x 10^{-1} km, thus the number of cells is 45; L2= 2 x 10^{-1} km, thus the number of cells is 18; L3= 6 x 10^{-1} km, thus the number of cells is 54; L4= 2 x 10^{-1} km, thus the number of cells is 18.

As regards Figure 9, It is worthy of interest to note that for the links with higher lengths (1-2 of length L1 and 3-4 of length L3), the platoon dispersion generates a non-uniform arrival profile, differently from the links 2-3 of length L2 and 4-1 of length L4 where the low distances among the junctions prevent the phenomenon of dispersion, inducing a uniform arrival profile.

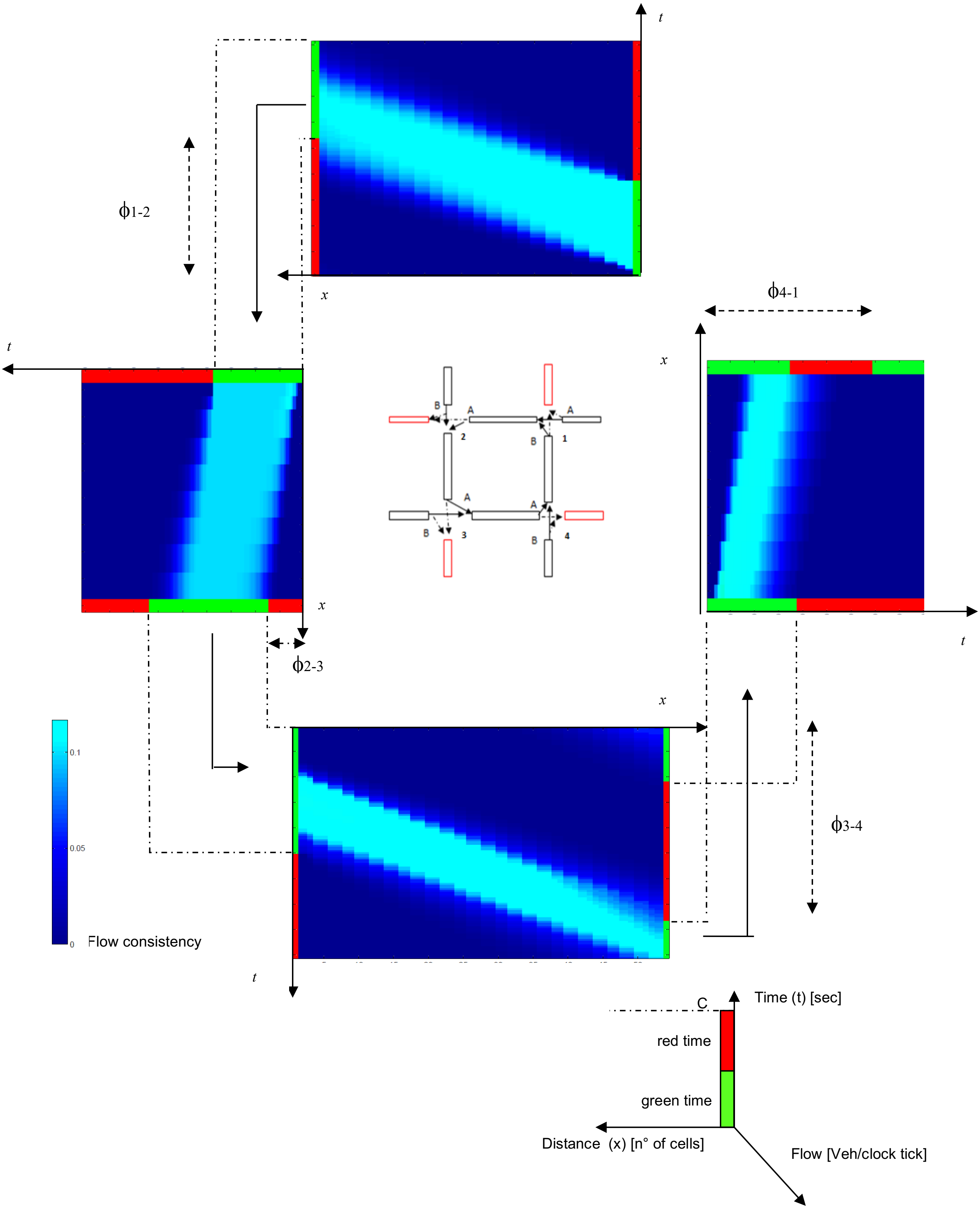


Figure 9: Time Distance Diagram in UNISA-SA (CT&PDM) (Id I, L1=500 m, L2=200 m, L3=600 m and L1=200 m) all over the network layout

The final comparison between the proposed methodology and the benchmarks was carried out in terms of running times. In Table 26 it is shown that no meaningful differences are observed in the case of the simpler layout (i.e. two junction arterial) and when the optimisation algorithm can be easily handled (i.e. HC algorithm, in the case of three step optimisation (A)). Although, some differences can be identified in two step optimisation (B.2), where a more complex algorithm (i.e. SA) is applied to achieve optimal results at a network level; in addition, in this latter case, an increasing level of interaction among decision variables and traffic flows could arise and may strongly affect the optimisation running time. In more detail, the highest value (185 s) is observed in TRANSYT-SA (CTM) and the lowest (74 s) is observed in TRANSYT-SA (PDM); notwithstanding that, the application of PDM (adopted by TRANSYT14® - TRL) may not be reliable when blocking phenomena occur (thus the lowest running time is imputed to the simpler adopted approach). Finally, the UNISA-SA (CT&PDM) running time is 95 s whereas that of TRANSYT-SA (CTM) is 185 s, thus, the application of the proposed approach is justified not only in terms of effectiveness but also in terms of efficiency.

Table 26: Overview of computational time w.r.t. strategy and layout configurations

Strategy	Two junction arterial		Four junction network	
	Id	Running time [s]	Id	Running time [s]
B.2	-	-	UNISA-SA (CT&PDM)	95
			TRANSYT-SA (PDM)	74
			TRANSYT-SA (CTM)	185
A	UNISA-GA&HC (CT&PDM)	8	UNISA-GA&HC (CT&PDM)	14
	OSCADY&TRANSYT-HC (PDM)	8	OSCADY&TRANSYT-HC (PDM)	15
	OSCADY&TRANSYT-HC (CTM)	11	OSCADY&TRANSYT-HC (CTM)	18

In future papers other traffic flow models will be investigated, in particular more explicitly addressing the within day dynamics.

Several future research directions seem worthy of interest. The first one is related to multi-criteria network optimisation; in fact, at network level, the capacity factor is not univocally defined: it may be intended as the lowest (to be maximised) achieved among all the junctions or as the area throughout (to be maximised). In this last case, the network multi-criteria optimisation may be based on a wide area criterion.

Second research direction may be identified within the one step optimisation in which the stage sequence (scheduling), the green timings and the offsets are computed at the same time (including multi-criteria optimisation).

Finally, addressing the network signal setting design as a part of transportation supply design (with equilibrium assignment) could be worthy of interest.

ACKNOWLEDGMENTS

The paper is dedicated to Professor Matthew G. Karlaftis recently passed away, in memory of his effort and contribution in transportation research.

Authors wish to thank anonymous reviewers for their helpful comments. This research has been partially supported by the University of Salerno, Italy, EU under local grant n. ORSA132945 financial year 2013, Campania Region PON- Ricerca e Competitività 2000-2013 grant n. 04a2_G "SMART TUNNEL", and PhD program on transportation (Ph.D. School in Environmental Engineering).

REFERENCES

Baker, J.E. (1985) Adaptive selection methods for genetic algorithms, Proceedings of the 3rd International Conference on Genetic Algorithms and Applications, In: Grefenstette, J.J. (ed.), New Jersey, Lawrence Erlbaum: Hillsdale, pp 100-111.

Binning, J. C., Crabtree, M. R., & Burtenshaw, G. L. (2010) Transyt 14 user guide. Transport Road Laboratory Report nr AG48. APPLICATION GUIDE 65 (Issue F).

Burrow, I.J. (1987) OSCADY : a computer program to model capacities, queues and delays at isolated traffic signal junctions: Traffic Management Division, Traffic Group, Transport and Road Research Laboratory, Crowthorne, Berkshire.

Ceylan H. (2006) Developing Combined Genetic Algorithm-Hill-Climbing Optimization Method for Area Traffic Control. Journal of Transportation Engineering, 132(8), pp 663-671.

Chow, A. H., & Lo, H. K. (2007) Sensitivity analysis of signal control with physical queuing: Delay derivatives and an application. Transportation Research Part B: Methodological, 41(4), pp 462-477.

Daganzo, C.F. (1994) The cell-transmission model. Part 2: Network traffic, University of California, Berkeley, California.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. (2002) A Fast and Elitist Multi-Objective Genetic Algorithm : NSGA II. IEEE Trans. On Evolutionary Computation, vol.6, pp 182-197.

Hadi, M. A., & Wallace, C. E. (1993) Hybrid genetic algorithm to optimize signal phasing and timing Transpn.Res.Rec., 1421, pp 94-103.

Holland, J.H. (1975) Adaptation in Natural and Artificial System. The University of Michigan Press.

Improta, G. & Cantarella, G. E. (1984). Control system design for an individual signalized junction. Transportation Research Part B: Methodological, 18(2), pp 147-167.

Jabari, S. E., & Liu, H. X. (2012). A stochastic model of traffic flow: Theoretical foundations. Transportation Research Part B: Methodological, 46(1), pp 156-174.

Lertworawanich, P., Kuwahara, M., Miska, M., (2011) A new Multiobjective Signal Optimization for Oversaturated Networks. IEEE Transactions on Intelligent Transportation Systems, vol.12, no.4.

Li, Z. (2011). Modelling Arterial Signal Optimization with Enhanced Cell Transmission Formulations. ASCE Journal of Transportation Engineering, 137, pp 445-454.

Li, Z., Chang, G.,L., (2010) Modelling arterial signal optimization with enhanced cell transmission formulations. Transportation Research Part B, pp 2-19.

Lighthill, M. J., Whitham, G. B. (1955) On kinematic waves. II. A theory of traffic flow on long crowded roads, Proc. Roy. Soc. London., pp 317–345.

Lo, H. K., & Chow, A. H. (2004) Control strategies for oversaturated traffic. Journal of Transportation Engineering, 130(4), pp 466-478.

Lo, H. K. (2001) A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transportation Science, 35(2), pp 148-164.

Lo, H.K., (1999). A novel traffic signal control formulation. Transportation Research 33A (6), pp 433-448.

Park, B., B., Messer, J., C., Urbanik II, T., (2000) Enhanced Genetic Algorithm for Signal-Timing Optimization of Oversaturated Intersections. Transportation Research record 1727, no.00.1661.

Putha, R., Quadrifoglio, L., & Zechman, E. (2012) Comparing ant colony optimization and genetic algorithm approaches for solving traffic signal coordination under oversaturation conditions. Computer-Aided Civil and Infrastructure Engineering, 27(1), pp 14-28.

Richards, P.I., (1956) Shockwaves on the highway, Operations Research, Vol. 4, pp 42-51.

Robertson, D.I., (1969) TRANSYT: a traffic network study tool. RRL Report LR 253, Road Research Laboratory, England.

Sumalee, A., Zhong, R. X., Pan, T. L., & Szeto, W. Y. (2011) Stochastic cell transmission model (SCTM): a stochastic dynamic traffic model for traffic state surveillance and assignment. Transportation Research Part B: Methodological, 45(3), pp 507-533.

Szeto, W. Y. (2008) Enhanced lagged cell-transmission model for dynamic traffic assignment. Transportation Research Record: Journal of the Transportation Research Board, 2085(1), pp 76-85.

This is a post-peer-review, pre-copyedit version of an article published in Transportation Research Part C The final authenticated version is available online at: http://dx.doi.org/10.1016/j.trc.2015.03.032

APPENDIX A

A.1 GENETIC ALGORITHMS

GAs seek the optimal solution by simulating the evolution of a "population" of solutions (individuals), by mimicking the basic principle of bacteria evolution. Each solution is described by a vector of decision variables called a chromosome made up of genes. In our case, each gene is representative of the stage's duration; the number of genes in each chromosome depends on the number of stages computed by the use of the relationship between approaches and stages. Optimisation was carried out through an iterative process which was representative of a reproductive cycle of the individuals (solutions) in the population (see Figure 10).

GAs are based on the genetic operators application such as crossover and mutation; while the former promotes the ability of the algorithm to explore the wider areas in order to search for solutions, the latter, without departing from the areas which are searching for solutions, introduces some (random) diversifications of the same type. Based on previous considerations the crossover can be called the explorative operator and the mutation can be called the exploiter operator.

For all these reasons, GAs are characterised by a higher effectiveness in approximating the global solutions in optimisation problems as the iterative process of the algorithm allows the GAs to explore large areas of optimal solutions.

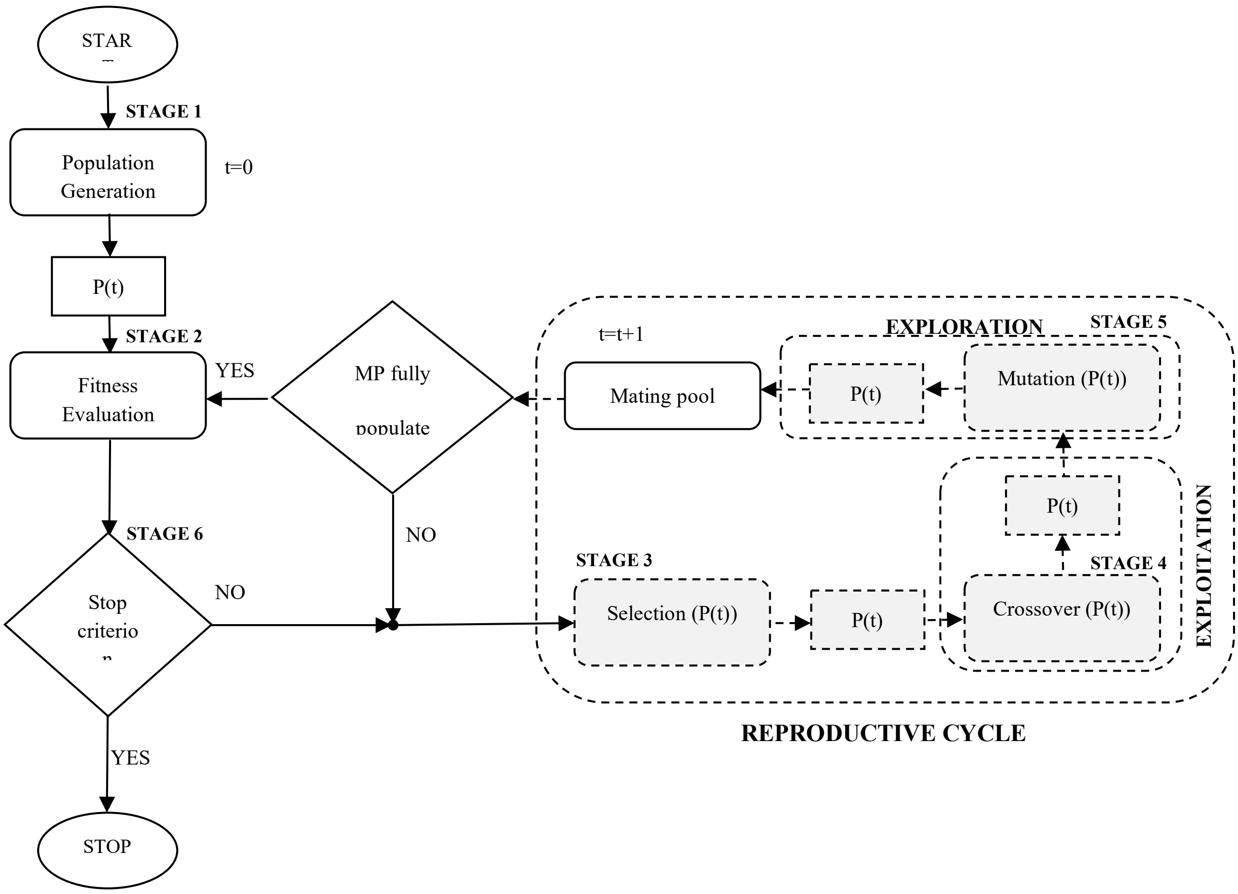


Figure 10: General Framework of GA

Stage [1] Population generation and Stage [2] Fitness Evaluation

At the beginning (t=0) the population, P(t), (in terms of each chromosome in the population) is randomly generated. At the successive step, for each solution, i, the fitness functions, ff_i, are calculated, so that the better the fitness function value of a solution, the greater its reproduction probability. In fact, the fitness function is strictly related to the probability, p_i, of being selected to be a parent; the probability is computed by normalising the fitness function, ff_i, of each chromosome, i, with regard to the fitness functions of all chromosomes i.e.

$$p_i = ff_i / \sum_j ff_j \qquad (10)$$

The procedure of probability (of being selected to be parent) computation is representative of the analogy between natural selection and the reproductive cycle (Holland, 1975).

Some more considerations need to be made in terms of fitness function estimation depending on the optimisation problem (e.g. mono-criterion optimisation vs. multi-criteria optimisation).

Stage [3] Selection

Based on the p_i, each solution is submitted to the selection procedure. Different approaches can be adopted for the selection and in our case the roulette wheel was adopted. With regard to each chromosome the p_i is computed, these probabilities are used by “composing” a roulette. The selection operator, generates a random value, depending on the membership range on the roulette wheel of each number, the corresponding solution is selected to be a parent for the successive generation.

Each time a solution is selected, a copy of the same is made and included in the mating pool (MP) until the same is full, i.e. when the mating pool size is equal to the size of the population.

Stage [4] Crossover and Stage [5] Mutation

After reproduction, each chromosome may be modified by two further genetic operators. This process refers to the crossover, whereby two new individuals are generated by mating with two already existing individuals, and through the mutation, an individual is randomly changed. By applying the two operators several times, a new population of solutions is obtained. This iterative procedure is repeated until some conditions (e.g. number of iterations or of improvement of the best solution) are satisfied.

The main parameters to fully specify GAs are the population size, the crossover probability (or rate, PC), and the mutation probability (or rate, PM). In terms of the crossover rate and the mutation rate, it can be observed that in order to balance the exploration and the exploitation effects introduced by two operators, the value rate of each one is usually heuristically chosen in order to respect the order relation: PM<PC.

Stage [6] Stop criterion

The iterative procedure is stopped when some criteria are satisfied, such as the fixed number of generations is reached or a solution is found that satisfies optimisation criteria where usually a threshold is introduced to evaluate the difference between the current solution and the solution reached in a previous generation. Convergence is reached when next generations do not lead to the further improvement of the fitness function.

GAs can be applied both to mono-criterion and multi-criteria optimisation problems. In terms of the algorithm procedure, this has an effect on computing the fitness function. In the first case, the general description is sufficient to be able to understand the procedure of the algorithm whereas in the case of the multi-criteria, further considerations are needed.

In particular, GAs for multi-criteria optimisation were implemented following the Non dominated Sorting Genetic Algorithm II (NSGA-II; Deb and Pratap, 2002). In accordance with this method, the selection (of being parents) of the chromosomes was made on the basis of the ranking of the solutions. Before starting the selection step, for each solution a value of rank was associated to the number of times in which the considered solution dominates the others. Using the obtained values, the rank based fitness assignment could be applied, in particular, to the fitness function, ff_i, for a given chromosome i, calculated by the linear ranking (Baker, 1985) approach as follows:

$$ff_i = 2 - SP + 2\,(SP - 1)\,(rank_i - 1) / N \tag{11}$$

where

SP is the selective pressure fixed to 1.5,

N is the population size,

$rank_i$, corresponds to the solution ranking and it was computed from the chromosome dominance hierarchy. Finally, an additional criterion was introduced for selecting among solutions with the same ranks. Each solution was attached to a value of crowding distance given by the Euclidean distance between the vector of the fitness functions of the solution and the vector of the best fitness function values, each one was defined as the best value among all solutions or in some cases the reference fitness function values were defined with regard to the considered criteria. If two or more solutions had the same rank value, selection at successive stages was based on the best value of the crowding distance.

All the defined multi-criteria problems were solved by applying NSGA-II.

A.2 Hill Climbing (HC) Algorithm

Hill Climbing is a neighbourhood-based meta-heuristic algorithm, without memory, which is deterministic in its basic version. The name originates from its ability to generate a succession of solutions by exploring the objective function surface which, if plotted, could be thought of as a series of hills and valleys in a multiple-dimensional world.

In this method, starting from an initial solution, successive iterations in the neighbourhood are performed until the current solution cannot be improved further. The algorithm stops when a local minimum/maximum is reached.

Different stochastic variants of this method are proposed in an attempt to endow it with a diversification strategy. For instance, the method can be applied by starting from multiple initial solutions which are randomly generated (as in the case of Shot-Gun Hill Climbing), or by varying the structure of the surroundings during the iterations.

In this paper, a basic Hill Climbing algorithm was applied in order to minimise the Total Delay at the downstream stop line (as a result of connecting link flow simulation). In order to reduce the risk of being trapped in a poor local optimum, a list of both small and large incremental offset alterations was set up (such increments are listed as percentages of the cycle time). Thereby, low increments made it possible to find an approximate local minimum of the Total Delay whilst high increments avoided getting trapped in that minimum.

A.3 Simulated Annealing (SA) Algorithm

The Simulated Annealing algorithm is a neighbourhood based meta-heuristic, which is inspired by the statistical mechanics to find solutions for both discrete and continuous optimisation problems. In particular, it takes the cue from the metallurgical phenomenon of annealing, in which a solid is brought to melting and then is slowly cooled to crystallize in a perfect lattice. Metropolis, in 1953, proposed for the first time a method to calculate the distribution of a particulates system in equilibrium temperature using a method of computer simulation. In this method, assuming that the system is in a configuration q having energy E (q), a new state r, having energy E (r), is generated, moving a particle from its initial position. The new configuration is then compared with the older one. If E (r) < E (q) the new state is accepted, if E (r) > E (q) it is not rejected, but rather, it is accepted with a probability P_a equal to:

$$P_a = \exp(-(E(r) - E(q)) / KT \quad (12)$$

where K is the Boltzmann constant and T is a control parameter, which by analogy with the original application is called temperature.

According to this method, there is, hence, a non-zero probability of reaching states of higher energy and so energy barriers that separate the global minima from the local minima can be climbed over. Note that the exponential function expresses the ratio between the probability of being in the leader configuration and the probability of being in q.

Kirkpatrick used the scheme of SA for combinatorial optimisation problems. To do this, he replaced the energy with a cost function and the states of a physical particles system with the solutions of a minimization problem.

The implementation of the SA algorithm requires the specification of some elements such as: a generator of random changes in solution; a fixed number of iterations before decreasing the temperature (L_k); an annealing schedule, i.e. an initial temperature and the rules for lowering it as the search progresses. Several cooling schedules are covered, including exponential, linear and temperature cycling. The simplest and most common temperature decrement rule is the Kirkpatrick geometric cooling scheme:

$$T_k = \alpha T_{k-1} \quad (13)$$

where α is a constant close to, but smaller than 1, usually assumed equal to 0.95.

In this paper the SA technique working on a single objective combinatorial problem, i.e. the minimisation of network total delay, was applied.

In Figure 11, the algorithm stages are described.

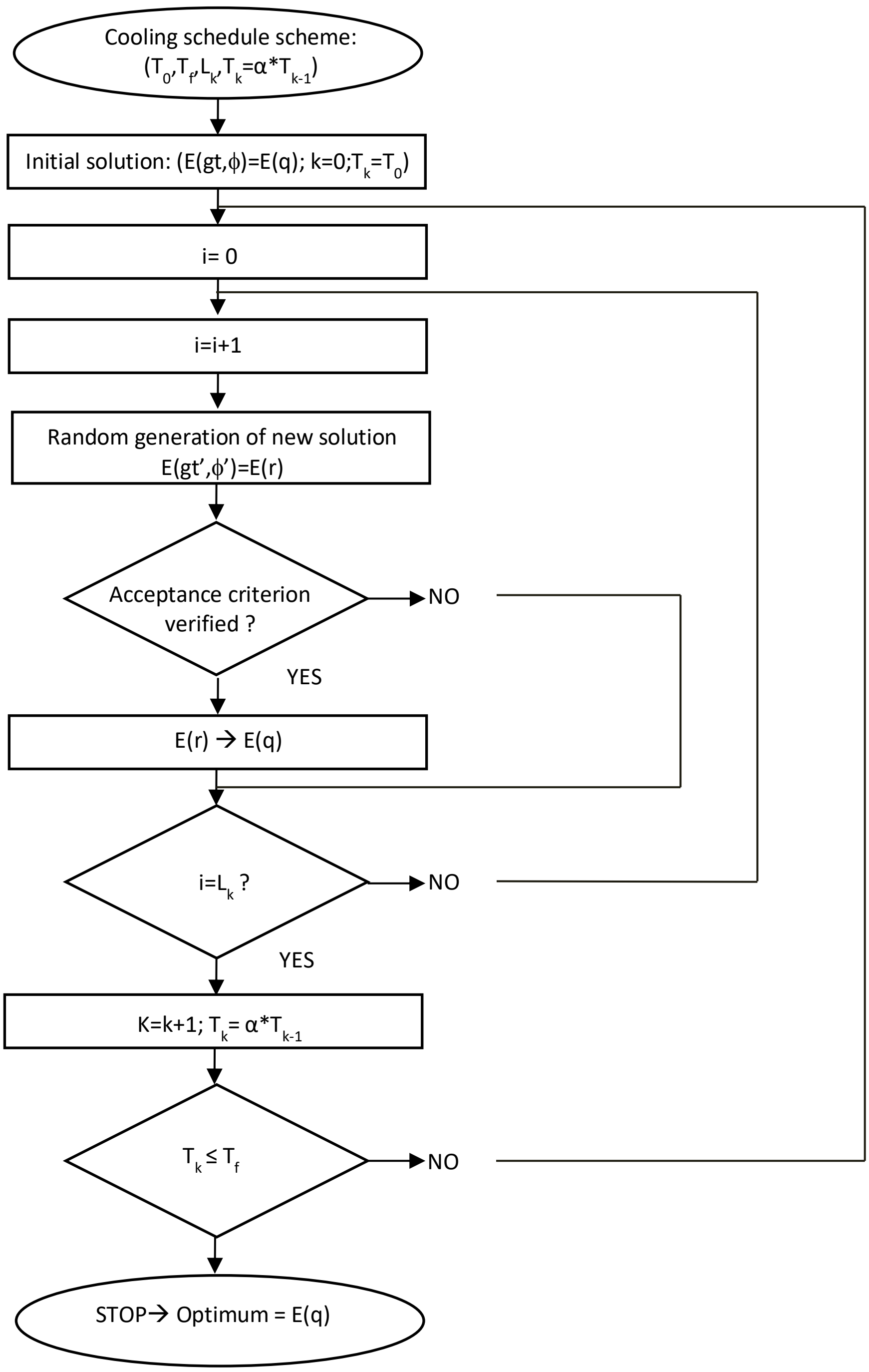


Figure 11: General description of the Simulated Annealing algorithm

Stage [1]

Let q_0 be an initial configuration which returns an Energy state (solution) $E(q_0)$, let T_0 be the initial temperature and T_f be the final temperature.

Stage [2]

For each decreasing value of T, T_k, the following steps were considered:

- ✓ By applying a random alteration of the current problem variables (i.e. green timings, green timings and offsets, Φ, among junctions) a feasible (energy state) new solution E(r) is generated. Such a solution is then compared with the best one archived E(q) (i.e. current solution):

$$\Delta E = E(r) - E(q) \quad (14)$$

- ✓ If $\Delta E \leq 0$, the new solution is accepted and so is archived as the new current solution, else, the new solution is accepted with a probability (12) and substitutes the current solution if such a criterion is satisfied;
- ✓ If the thermal equilibrium is not achieved i.e. if the number of iterations at current temperature is less than L_k the algorithm goes to step II a).

Stage [3]

If $T_k > T_f$, the temperature is further decreased and the algorithm goes to stage II.