Engineering Applications of Artificial Intelligence 84 (2019) 51–65

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A bilevel approach to enhance prefixed traffic signal optimization ☆

Ricardo García-Ródenas, María L. López-García, María Teresa Sánchez-Rico, Julio Alberto López-Gómez *

Departamento de Matemáticas, Escuela Superior de Informática, University of Castilla-La Mancha, Spain

ARTICLE INFO

Keywords:
Traffic signal control
Time-of-day intervals
Evolutionary computation
Clustering
Multivariate time series analysis

ABSTRACT

The segmentation of multivariate temporal series has been studied in a wide range of applications. This study investigates a challenging segmentation problem on traffic engineering, namely, identification of time-of-day breakpoints for pre-fixed traffic signal timing plans. A large number of urban centres have traffic control strategies based on time-of-day intervals. We propose a bilevel optimization model to address simultaneously the segmentation problems and the traffic control problems over these time intervals.

Efficient memetic algorithms have been developed for the bilevel model based on the hybridization of the particle swarm optimization, genetic algorithms or simulated annealing with the Nelder–Mead method. Numerically the effectiveness of the algorithms using real and synthetic data sets is demonstrated.

We address the problem of automatically estimating the number of time-of-day segments that can be reliably discovered. We adapt the Bayesian Information Criterion, the PETE algorithm and a novel oriented-problem approach. The experiments show that this last method gives interpretable results about the number of reliably necessary segments from the traffic-engineering perspective.

The experimental results show that the proposed methodology provides an automatic method to determine the time-of-day segments and timing plans simultaneously.

1. Introduction

The study of the dynamic aspects of traffic is essential for a proper modelling of traffic and its related phenomena, like traffic congestion among others. Intelligent Transportation Systems (ITS) address this level of uncertainty through advanced monitoring systems of the traffic network in real time, which makes possible to determine the system state and to respond to unexpected situations. The implementation of these systems (traffic responsive) requires significant economic investment and complex maintenance processes, which means that at the present moment, the number of urban areas equipped with these systems remains small. Many studies and methods for traffic planning and control in urban networks are based on the assumption of time-of-day (TOD) intervals which determine dynamic congestion patterns. A TOD is a period of time where traffic dynamics can be considered stationary, so it is possible to assume that traffic in this interval follows a specific behaviour, which is defined through a pattern. Once TODs have been determined, timing plans are developed and optimized for each TOD using heuristic and metaheuristic algorithms. Typically, between three and five plans are run in a given day.

Traditionally, TOD determination and optimization plans are tasks that has been performed visually by traffic engineers making it challenging and subjective. Currently, there are two possibilities to address the problem of defining control strategies for traffic networks based on TOD intervals.

- The *sequential methodology* assumes a two-phase approach, in the first phase the determination of the TOD intervals is carried out while in the second phase an adequate control strategy is determined for each TOD.
- The *simultaneous methodology* addresses the above two phases simultaneously throw bilevel optimization.

Determining TOD intervals has been widely addressed in the literature using mainly cluster analysis techniques, among others. Nevertheless, the applied methods do not consider the so-called *time-domain-constraints* (see López García et al. (2014)). The omission of these constraints leads to a noisy and infeasible detection of TOD intervals as well as transitions between different TODs with a high cost, in terms of quality of modelling. These noisy TODs are clusters which do not follow an intuitive TOD scheme as the majority of clusters. For that

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.engappai.2019.05.017.

* Corresponding author.
E-mail addresses: Ricardo.Garcia@uclm.es (R. García-Ródenas), MariaLuz.Lopez@uclm.es (M.L. López-García), Teresanchezr@gmail.com (M.T. Sánchez-Rico), JulioAlberto.Lopez@uclm.es (J.A. López-Gómez).

https://doi.org/10.1016/j.engappai.2019.05.017
Received 25 November 2016; Received in revised form 4 February 2019; Accepted 30 May 2019
Available online xxxx
0952-1976/

reason, these clusters have to be re-assigned to other TODs, since if it is not done, the transition between these unfeasible clusters and the other ones will be highly expensive and the obtaining of a larger number of clusters will produce worst results. The re-assigning process of unclean clusters was made again by traffic engineers.

The time-domain-constrained data clustering problem tackles a clustering problem in which data are labelled with the time where they were gathered. Furthermore, the time-domain constraints impose that the obtained clusters need to be contiguous in time. That is, if two data are grouped in the same cluster then all data with a time label between both data must be assigned to the same cluster. Thus, time series segmentation is solved through the partition of a time series into homogeneous clusters which are close in terms of time, solving the problem of noisy clusters and minimizing the costs of transitions between clusters.

A second challenge for TOD determination and traffic light optimization is the development of a bilevel optimization model which addresses simultaneously the definition of control strategies. However, it implies a high computational cost, derived from the two hard optimization tasks of determining TODs and optimizing timing plans. In order to solve successfully the model, memetic algorithms which combine the advantages of metaheuristics and local search strategies must be developed, in order to build more accuracy and faster algorithms which achieve an optimum trade-off between exploration and exploitation. Hybridizations of metaheuristics and local search algorithms have been proven that outperform the results of them individually (see Shi and Eberhart (1998), Espinosa-Aranda et al. (2013), Li and Schonfeld (2015), and Sabar et al. (2017)).

Moreover, there is not in the literature an automated methodology for determining TOD breakpoints and optimizing traffic signal times in each segment. For this reason, in this paper the problem is formalized following a simultaneous methodology through an approach given in López García et al. (2014). The model is formed to simultaneously determine the TODs and the traffic control strategy. In the upper level the TODs breakpoints are determined by optimization and the lower level problem is represented by a traffic control problem. The integration of both levels avoids local optima in the TOD breakpoints pursued. This paper focuses on a control strategy based on timing plans for intersections but can be easily extended to methods using the whole traffic network.

The main contributions of the paper are:

- The determination of TODs breakpoints and the development of the optimal traffic control for each TOD following a simultaneous approach which considers time-domain constraint in a clustering problem has been dealt with. The versatility of the proposed bilevel model to be adopted in different traffic problems has been proved. Specifically, in this paper we have addressed the optimal traffic signal timing for each TOD. The results of the computational experiments show that, in the case of an isolated intersection, our bilevel proposal based on a simultaneous methodology reduces the waiting time for users about 3% in comparison with a sequential methodology.
- Memetic algorithms have been applied to the bilevel model. The hybridization of Particle Swarm Optimization (PSO) and the Nelder–Mead (NM) method, the hybridization of Simulated Annealing (SA) with NM, and of Genetic Algorithms (GA) with NM have been analysed. In the literature only proposals based on GA and SA have been tackled. The results of the computational experiments show the effectiveness of the memetic algorithms respect to the metaheuristics for the problems addressed.
- Despite the fact that the determination of number of cluster has been widely studied in the literature, automatic determination of the number of TODs has not been previously studied. This aspect has been considered in this paper through the Bayesian Information Criterion (BIC), the PETE algorithm and finally, an oriented-problem approach has been developed with promising results.

The article is organized as follows. Section 2 reviews traffic signal control and optimization based on TODs. In Section 3 the bilevel model is formulated. In Section 4 the memetic computing for its resolution is described. The automatic determination of the number of TODs is carried out in Section 5. In Section 6 the numerical experiments over real and synthetic data are carried out and finally the conclusions obtained are analysed.

2. Related work

During the last few decades traffic signalization has experienced a great evolution, from the first pre-fixed signals with fixed times to the real-time traffic signalization. A roughly taxonomy of traffic signal systems can be stated as follows (see van Katwijk et al. (2006), Gordon et al. (1985)).

1. **Fixed Time Systems.** These systems fix a predetermined control strategy on a set of time intervals given.
   (a) *Fixed Time Systems (Pre-timed)*. In fixed time systems, the timing plan is determined from historical data, fixing predetermined rates. The main disadvantage of this approach is the inability to adapt to demand fluctuations over the time.
   (b) *Plan Selection Systems*. They select the most appropriate timing plan from a repository of plans according to the received inputs from sensors which detect the current state of the traffic network.
2. **Traffic Responsive Systems.** They usually operate in real-time and make decisions according to current traffic conditions which are collected through a detection system.
   (a) *Actuated.* They operate in real-time applying a control strategy according to the current state of the traffic network. In this case, these systems are capable of adjusting the length of the current phase in response to flow and demand variation.
   (b) *Adaptive*. These systems represent the most complex and sophisticated traffic signal control systems. They operate in real-time through an optimization algorithm to choose the optimum timing plan. With regard to actuated systems, adaptive traffic signal control evaluates a set of possible control strategies in real-time incorporating a decision making in order to choose the best control mechanism. Moreover, these systems are classified according to the type of communication system on centralized or distributed.

Despite the fact that traffic adaptive systems with distributed processing are the most promising systems, the infrastructure in the most populated and important cities in the world are not ready for implementing this kind of system now, since an expensive technological infrastructure is required. Relevant studies of this type methodology are Abdoos et al. (2013), McKenney and White (2013) and Khamis and Gomaa (2014) that propose traffic responsive systems based on multi-agent systems for traffic signal optimization.

The above discussion motivates the study of determining TOD intervals and it is the why it is still a very current research topic. Moreover, Koonce et al. (2008) showed that non-adaptive systems, for example pre-fixed time systems, with adequate design and with regular updates obtain acceptable results in comparison with traffic responsive systems. For this reason it is crucial to have specialized tools capable of automating the planning process and making planning changes when new mobility patterns are detected. In addition, robust optimization techniques allow robust pre-timed systems which are less sensitive to traffic flow fluctuations (see Yin (2008), Lee et al. (2011)).

2.1. Reviewing the time-of-day identification problem

Traditionally, traffic engineers determined TOD breakpoints, developed and optimized traffic timing plans based on their expert knowledge and with the only historical data of traffic volumes in critical intersections. With the appearance of data mining and machine learning techniques, different approaches have been widely used in this process. The TOD determining problem can be addressed through cluster analysis or by means of segmentation of multivariate time series. The first approach deals with solving the so called *clustering with time-domain constraints* and the second with the *change-point detection problem*. The standard solution of change-point detection problem involves: (i) the number of change-points, (ii) their locations, and (iii) functions for determining curve fitting between successive change-points. In TOD determination problem the issue (i) is fixed and the fitted function related with issue (iii) is considered constant. Fu (2011) reviews the change-point detection problem but it does not report applications in traffic domain. Both approaches are considered to be equivalent. However, as far as we know, only cluster analysis techniques have been applied in current literature to solve it.

Hauser and Scherer (2001) provided a solution based on hierarchical clustering taking into account the volumes and occupancies of different intersections, determining different TOD intervals and optimizing plans for each of them.

Smith et al. (2002) proposed a new method based on hierarchical clustering, with the novelty of making a high-resolution definition of the state of the system, taking into account not only the flow or occupancy as traffic parameter but also the density. With these data, different plans were developed for each TOD. In order to address the unfeasible clusters problem, Park et al. (2003) proposed an approach based on clustering using a Genetic Algorithm (GA). To do that, the fitness function in the codification of genetic algorithm introduces a penalty with the purpose of avoiding unclean clusters.

Furthermore, Wong and Woon (2008) proposed a method based on $k-$means algorithm and silhouette index. This method uses the traffic parameters flow, speed and occupancy and it determines the TODs following an iterative scheme between a TOD determination stage and a traffic control stage. In each iteration the TODs are refined using the new information of the traffic control problem and the empty clusters are deleted. This method can be viewed as heuristic to solve the bilevel model which appear in the simultaneous methodology.

However, these works did not consider the time-domain-constraint into the clustering process, which produced the so-called *unclean* clusters. Although the approach of removing unclean clusters provide the minimization of transition costs, they did not consider the transition costs in the optimization tasks. To achieve it, Lee et al. (2011) proposed an approach based on a GA to explicitly consider the transition costs during the optimization tasks in a coordinated-actuated traffic signal system.

Moreover, local search strategies have also been applied to determine TOD intervals. One example is Park and Lee (2008), where the problem is solved using the optimal cycle length per time interval through a greedy search algorithm. The authors proposed this methodology due to the quickness of local search strategies. This work takes into account the transition cost but it does not consider explicitly the time-domain constraints.

Lee and Kim (2011) proposed a methodology based on the $k-$means algorithm and the VPLUSKO index – which is defined on the volume and occupancy – to determine TOD breakpoints. The VPLUSKO index helped to calibrate the model and adjust the breakpoints. However, this work do not address the problem of automatically determining the optimum number of TODs under any traffic conditions like the rest of reviewed studies.

Soft-computing techniques such as fuzzy logic and metaheuristics have also been applied to determine TOD breakpoints. In Angulo et al. (2011), authors presented a plan selection system based on two setups: off-line and on-line. In the off-line step, different mobility patterns were determined using historical data which were synthesized. In the on-line phase, the current traffic conditions are matched with one of the determined patterns in the off-line phase, where a pattern corresponds to a fuzzy prototype.

Ratrout (2011) proposed a process to determine optimal TODs based on $k-$means method. This work is a first approximation to address the inclusion of time-domain constraints in the TOD determination problem. Instead of the traffic volumes, this work proposes the use of $z-$ score and the time as features in the clustering process. The approach falls on a sequential methodology.

In last years, new technologies and sensors, like GPS, APC etc., have allowed to collect data in real-time in order to address different transport problems. The problem of determining TOD breakpoints has been applied to other transportation domains. Bie et al. (2015) applied GPS data for bus scheduling, determining TODs for bus lines. In this work, the authors used the dwell times at stops and inter-stop travel times like clustering indexes to partition data, using a hierarchical approach.

Various criteria for automatic model selection have been widely used to determine the number of clusters (in our problem the number of TODs) in the data. The Bayesian information criterion (BIC), Akaike's Information Criterion (AIC) and Minimum Description Length (MDL) are some notable examples of these criteria. However, despite of the profuse literature about this topic (Chiu et al., 2001; Xia and Chen, 2007) there is not an index which outperforms the rest in all application domains.

The TOD determination problem viewed as a segmentation problem also contributes with its own methods to determine the number of TODs such as the PETE method (Vasko and Toivonen, 2002) or the BIC's adaptation method (ΔBIC) for the segmentation of time series (Chen and Gopalakrishnan, 1998; Wang et al., 2008).

Briefly, Table 1 shows a summary of the works analysed in this section grouped by the approach used to solve determining TODs breakpoints, sequential or simultaneous.

2.2. Memetic-related works to traffic control systems

The traffic signal control methods are deployed on the basis of optimization procedures. Genetic algorithms (GA) have been widely and successfully applied (see Araghi et al. (2015), Chin et al. (2011), Teklu et al. (2007), Ceylan and Bell (2004) among others). The growing complexity of the traffic signal systems has motivated the use of new metaheuristics like in García-Nieto et al. (2012) where *particle swarm optimization* (PSO) is applied for determining cycle programmes of traffic lights in two large networks.

Nowadays, there are several enhancements of metaheuristics that improve its performance in different ways. One of these strategies are memetic algorithms (MA). The concept of MA was first coined by Moscato and Norman (1992). In that work, the authors developed an algorithm for the travelling salesman problem using local search heuristics with the purpose of improving the exploitation capability of population-based algorithms. Since then, this concept has been formalized, theoretically studied and applied to many complex optimization problems (see the reviews Chen et al. (2011) and Neri and Cotta (2012)).

Currently, MAs have shown their ability to achieve high performance and superior robustness across a wide range of problem domains. To mention a few examples, MA has been applied to power systems like in Hu et al. (2015), where a MA is used for feature selection in order to forecast mid-term interval loads, or aerodynamic, as in Qu et al. (2017), where it has been used to optimize aerodynamic shape or aircrafts and determine the optimal settings of shape parameters.

Recently, Sabar et al. (2017) applied MAs to an adaptive signal timing settings system. In this work, MA determine good quality signal timing settings within an acceptable amount of time. This task is crucial in real time systems and this motivation has been followed along this work.

Table 1
A review of traffic signal optimization based on time-of-day intervals.

Methodology	Technique	Problem	**Reference**
Sequential approach	Hierarchical clustering	Determining and optimizing TOD intervals on a three-intersection corridor.	Hauser and Scherer (2001)
		Developing timing plans for a three coordinated actuated intersection.	Smith et al. (2002)
	Genetic algorithm	Determining optimum TODs, removing unfeasible clusters in a three-intersection actuated and coordinated intersection.	Park et al. (2003)
	Quadratic sequential programming	Signal timing optimization on an isolated fixed-time four-stage signalized intersection	Yin (2008)
	Clustering k−means and statistical techniques	Determining and optimizing TOD intervals into a ten coordinated signalized intersection in Seoul.	Lee and Kim (2011)
	Clustering k−means	Determining optimum TOD intervals in a three pretimed coordinated signalized arterial intersection.	Ratrout (2011)
	Clustering k− means and fuzzy logic	Determining and optimizing TOD for traffic signal optimization	Angulo et al. (2011)
Simultaneous approach	Clustering k−means	Optimizing signal timing plans on extremely congested roads	Wong and Woon (2008)
	Greedy search	Determining optimum TOD intervals into coordinated actuated traffic signal arterial operations.	Park and Lee (2008)
	Genetic algorithm	Determining optimum TOD intervals into coordinated actuated traffic signal arterials.	Lee et al. (2011)
	Hierarchical clustering	Time series partition for determining TOD intervals to optimize bus lines	Bie et al. (2015)

3. Models to identify time-of-day breakpoints

The main objective of this section is to state the optimization models for the determination of TODs based on the time-domain-constrained data-clustering problem. The goal is to define time intervals in which traffic demand is approximately stationary and therefore the dynamic component within each interval can be considered negligible. This is a resolution strategy for addressing the management and the control of traffic for non-stationary demand (dynamic demand).

We will start addressing the sequential methodology in which TODs are determined in a first step and after that, the traffic control problem is solved in each TOD. Then, this approach will be extended to a simultaneous methodology in which both problems are solved on a single optimization model, in this case, the bilevel optimization model which is formalized in Section 3.2.

3.1. Sequential methodology

Urban traffic networks are mathematically modelled by a *directed graph* $\mathcal{G} = (N, A)$ in which the set of nodes N represents the *intersections* and the so called *centroids*. The centroids are dummy nodes which model city areas with generation/attraction of trips. The set of links A represent urban roads and the so-called *connectors*, which are dummy links joining the centroid nodes with the intersection nodes.

Fig. 1 shows a representation of a traffic network. This network consists of 4 centroids, 4 connectors and 11 links representing the streets of the urban area modelled. In addition to the network (supply) in these systems the *origin–destination matrix* is considered representing the demand between different centroids. In this example we consider four origin–destination pairs $A \to C, A \to D, B \to C, B \to D$. The index ω denotes one of these pairs and W the set of all origin–destination pairs. To model the variation of demand over time the *intensity of demand* $q_\omega(t)$ is introduced. These functions are not directly observable in the network.

New technologies allow the monitoring of traffic networks in real-time. These traffic control systems are located in a subset of links of the network, denoted by $\hat{A} \subset A$ and we will call this the set of *sensors*. The traffic parameters in the link $a \in \hat{A}$ are the *traffic flow* q_a (veh./hour), the *speed* v_a (km./hour) and the *density (or alternatively occupancy)* k_a (veh./km.), which is linked to the two above by the equation:

$$q_a = v_a k_a, \qquad a \in A. \tag{1}$$

For simplicity, we will focus on the flow observed in each arc of the traffic network. This methodology can be generalized in order to

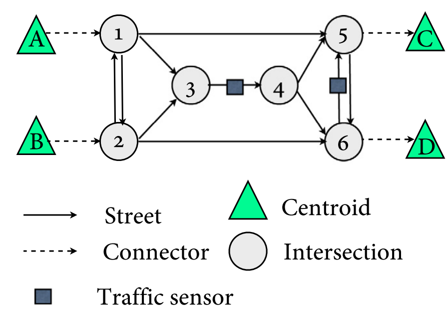

Fig. 1. Modelling of a traffic network.

consider the three types of traffic parameters on a simple way. García-Ródenas et al. (2017) proposes the use of geodesic distance over euclidean distance, since the first one allow to incorporate the three types of traffic parameters and take into account the correlations between the link counts.

Let us a data set D, where $D = \{(t_j, \mathbf{q}_j)\}_{j=1}^{N}$. This data set is composed by tuples which in turn consist of $\mathbf{q}_j \in \mathbb{R}^d$ which is a vector of link flows. Indeed, flow data $\mathbf{q}_j$ are labelled with an index which represent the time instant t_j for which the data was gathered. Thus, it is possible to assume the data are time ordered. Henceforth, it is mathematically expressed as follows, $t_j < t_{j+1}$ where $j = 1, \dots, N-1$.

- **Step 1: TOD determination problem**

TOD determination problem has been dealt in the literature as a clustering problem. Cluster analysis tries to find TODs with a certain level of internal homogeneity and heterogeneity between different clusters. The scheme most widely used addresses the problem of minimizing the variability into each group or cluster. The most popular criterion followed to do that consists of minimizing the *sum squared error (SSE)*, whose mathematical formulation is the following: Supposing that it is wanted to partition N traffic observations in clusters $(C_1, C_2, \dots, C_K)$. Each cluster C_k is defined by the so-called *centroid*:

$$\bar{\mathbf{q}}_k = \frac{1}{|C_k|} \sum_{j \in C_k} \mathbf{q}_j. \tag{2}$$

where $|C_k|$ is the number of objects in cluster C_k and represents the mean flow vector in each TOD k.

The TOD determination problem used in the literature can be formulated according to:

$$\begin{aligned}
&\text{Minimize} \quad SSE(\mathbf{w}) = \sum_{j=1}^{N}\sum_{k=1}^{K} w_{jk}\|\mathbf{q}_j - \bar{\mathbf{q}}_k\|^2 \\
&\text{Subject to:} \quad \sum_{k=1}^{K} w_{jk} = 1; \qquad j = 1, \cdots, N \\
&\qquad\qquad\quad \sum_{j=1}^{N} w_{jk} \geq 1; \qquad k = 1, \cdots, K \\
&\qquad\qquad\quad w_{jk} \in \{0, 1\}
\end{aligned} \tag{3}$$

where the objective function is the sum squared error within each cluster and the binary variable w_{jk} takes the value 1 if j observation is assigned to cluster k and 0 otherwise. The first constraint imposes that all objects are assigned to some cluster and the second one requires that there is not empty clusters. The k-means algorithm can be considered as a greedy algorithm to solve the optimization problem (3).

Fig. 2 illustrates the determining TOD problem. In the figure on the left, a possible solution for the TOD problem solving (3) is shown. In this figure, it is shown that the transition between clusters C_1 and C_2 and between clusters C_3 y C_4 is chaotic and produce higher transition costs between traffic control strategies. This solution does not fulfil the temporal constraints, formulated by:

$$\left.\begin{array}{l} \text{If } t_j < t_i < t_{j'}, \\ j, j' \in C_k \end{array}\right\} \Rightarrow i \in C_k \tag{4}$$

We introduce decision variable $\mathbf{s}$ (instead of $\mathbf{w}$) in order to incorporate the temporal constraints (4) into the cluster analysis. Assume that $T = [a, b]$ is the whole time period in which data has been extracted and henceforth, this is the period that must be partitioned. Suppose the limit points $a = s_0 < s_1 < s_2 < \cdots < s_{K-1} < s_K = b$ (see Fig. 2) as decision variables. Thus, it is possible to describe the set

$$C_k(\mathbf{s}) = \left\{ j \in \{1, \ldots, N\} \ / s_{k-1} \leq t_j < s_k \right\}; \quad k = 1, \ldots, K \tag{5}$$

The centroid has been considered as the mean flow in the kth TOD and it is shown in Eq. (6):

$$\bar{\mathbf{q}}_k(\mathbf{s}) = \frac{1}{|C_k(\mathbf{s})|} \sum_{j \in C_k(\mathbf{s}')} \mathbf{q}_j \tag{6}$$

where $|\cdot|$ is the cardinal of a set. According to the previous formalization, the optimization model is stated as:

$$\begin{aligned}
&\text{Minimize} \quad SSE(\mathbf{s}) = \sum_{k=1}^{K} \sum_{j \in C_k(\mathbf{s}')} \|\mathbf{q}_j - \bar{\mathbf{q}}_k(\mathbf{s}')\|^2 \\
&\text{Subject to:} \quad s_{k-1} < s_k; \qquad k = 1, \ldots, K \\
&\qquad\qquad\quad s_0 = a, s_K = b
\end{aligned} \tag{7}$$

- **Step 2: Traffic control problem** Once the set of TODs has been determined, the mean flow $\bar{\mathbf{q}}_k$ is computed and it is assumed this is the traffic pattern which operates in each TOD period. The traffic control problem find an optimal strategy, which will be defined by a parameter vector Θ_k, which will optimize a certain criterion, like waiting time, etc. In a general way, these kind of problems are formulated by:

$$\underset{\Theta_k}{\text{Minimize}} L(\bar{\mathbf{q}}_k; \Theta_k); \qquad k = 1, \ldots, K \tag{8}$$

3.2. Simultaneous methodology

These methodologies try to find the optimum traffic control strategy in terms of accuracy and efficiency. In order to do that, they determine the TOD intervals where pre-time (which corresponds with the first step in the sequential methodology) and traffic signal timing plans (step two in the sequential methodology) will be applied. The simultaneous methodology proposes the following bilevel optimization model which integrates both steps:

$$\begin{aligned}
&\textbf{TOD determination problem} \qquad \underset{\mathbf{s}}{\text{Minimize}}\, \widetilde{J}(\mathbf{s}) := \sum_k \widetilde{J}_k(\mathbf{s}). && (9)\\
&\qquad\qquad \text{Subject to: } s_{k-1} < s_k; \qquad k = 1, \ldots, K. && (10)\\
&\qquad\qquad s_0 = a, s_K = b. && (11)\\
&\textbf{Traffic control problem on TOD } k \quad J_k(\mathbf{s}) = \underset{\Theta_k}{\text{Minimize}} \sum_{j \in C_k(\mathbf{s})} L(\mathbf{q}_j; t_j; \Theta_k);. && (12)
\end{aligned}$$

The decision variables in this case are the control strategies Θ_k and the period $[s_{k-1}, s_k]$ where the strategy will be applied. Besides, the objective function in this problem change the goal of finding time intervals with minimum variations in flow intensity to finding intervals where optimal traffic control strategies are obtained. Another important feature is that traffic control problem is not posed exclusively on $\bar{\mathbf{q}}_k$ but also on the set of observations belonged to each TOD. Note that $C_k(\mathbf{s})$ may have a short and insufficient number of observations or even it can be empty, leading to make the traffic control problem (12) bad posed. To deal with this problem, we assume that $\widetilde{J}(\mathbf{s}) = +\infty$. As may be seen from the previous formalization, the problem (12) depends on the application that is being tackled. Note that the formulation is general enough in order to tackle the use of simulation models to analyse the traffic network. If we note $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_K) \in \mathbb{R}^{Q \times K}$, it is possible to consider that flow depends on the traffic control strategies implemented in the system $\mathbf{q}(\Theta)$ and simulation models allow to compute $\mathbf{q}(\Theta)$ as a quality index $L(\mathbf{q}_j(\Theta); t_j; \Theta_k)$ which it is being considered.

In this work we consider a combined approach of TOD determination and optimal traffic signal timing. Without loss of generality, we consider an isolated fixed-time signalled intersection. The parameter Θ_k represents the optimal timing plan for the TOD k, and the criterion $L(\mathbf{q}_j; t_j; \Theta_k)$ for period j is the total delay time in the intersection in that period j. In that case the lower level problem cannot be explicitly solved and the problem have a bilevel nature.

In this problem the control variables are $\Theta_k := (\mathtt{C}^k, \mathbf{g}_k)$ where $\mathtt{C}^k$ is the cycle length (seconds) and $\mathbf{g}_k = (\cdots, g_{ki}, \ldots)$ denotes the vector of effective green time for each line group i at TOD k.

We use the delay equation in the Highway Capacity Manual (HCM) (Council et al., 2010) to estimate the delay per vehicle for period $j \in C_k$

$$d(\mathbf{q}_j, t_j; \mathtt{C}^k, \mathbf{g}_k) := \frac{\sum_{i=1}^{n} \left[\frac{\mathtt{C}^k (1-\lambda_i^k)^2 q_{ji}}{2(1-\lambda_i^k \min(1, x_i^k))} + 900 T_j q_{ij} \left(x_i^k - 1 + \sqrt{(x_i^k - 1)^2 + \frac{4x_i^k}{c_i^k T_j}} \right) \right]}{\sum_{i=1}^{n} q_{ji}} \tag{13}$$

where

- n: is the number of lane groups.
- λ_i^k: is the effective green split per lane group i at TOD k, i.e. $\lambda_i^k = \frac{g_{ki}}{\mathtt{C}^{\mathtt{t}}}$.
- s_i: is the saturation flow for lane group i (veh/h).
- x_i^k: represents the degree of saturation in line i at TOD k, i.e. $x_i^k = \frac{q_{ij}}{\lambda_i^k s_i}$.
- T_j: is the duration of the analysis period, i.e $T_j = t_j - t_{j-1}$ and t_0 the starting time.
- c_i^k : is the capacity for lane group i (veh/h), i.e $c_i^k = \lambda_i^k s_i$

The total delay time at period j can be computed as

$$L(\mathbf{q}_j, t_j; \mathtt{C}^k, \mathbf{g}_k) := d(\mathbf{q}_j, t_j; \mathtt{C}^k, \mathbf{g}_k) T_j \left[\sum_{i=1}^{n} q_{ji} \right] \tag{14}$$

Therefore, the optimization model of signal timing can be written as:

$$\widetilde{J}_k(\mathbf{s}) := \underset{\mathtt{C}^k, \mathbf{g}_k}{\text{Minimize}} \sum_{j \in C_k(\mathbf{s})} L(\mathbf{q}_j, t_j; \mathtt{C}^k, \mathbf{g}_k) \tag{15}$$

subject to linear constraints on $\mathtt{C}^k$ and $\mathbf{g}_k$

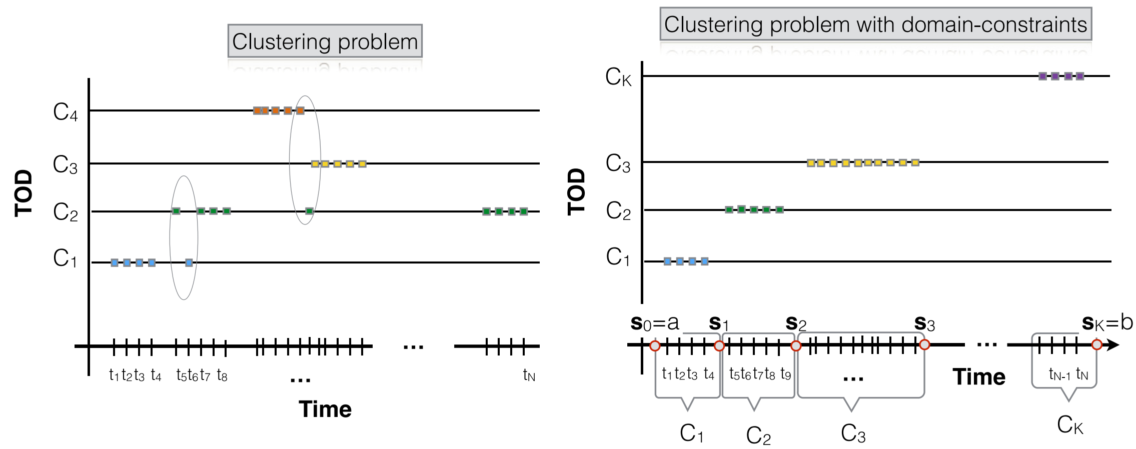

Fig. 2. Clustering problem with and without time-domain constraints.

3.3. Reformulation of simultaneous TOD determination problem using unconstrained optimization

Briefly, by summarizing all the formalization proposed in this article, the simultaneous TOD determination problem given in Eqs. (9)–(12) can be written as follows:

$$\begin{array}{ll} \text{minimize} & \widetilde{J}(\mathbf{s}) \\ \text{subject to} & s_{k-1} < s_k; \qquad k = 1, \dots, K \end{array} \tag{16}$$

The bilevel optimization problem (16) proposed in this work includes two optimization tasks or levels. In the first or upper level, clustering with time-domain constraint is made. The vector $\mathbf{s} = (s_1, \dots, s_{K-1}) \in \mathbb{R}^{K-1}$ defines the decision variables which will be used in the upper level problem, since when this vector is arranged, the evaluation of $\widetilde{J}(\mathbf{s})$ leads to K independent traffic control problems. We assume that optimization tools are available to obtain the optimum solution to traffic control problem Θ_k for the $\mathbf{s}$ variable included in the upper level.

According to this definition, the previous model can be reformulated as an unconstrained optimization model. To do that we define

$$\hat{J}(\mathbf{s}) := \begin{cases} \widetilde{J}(\mathrm{sort}(\mathbf{s})) & \mathbf{s} \in [a,b]^{K-1} \\ +\infty & \mathbf{s} \notin [a,b]^{K-1} \end{cases} \tag{17}$$

where $\mathrm{sort}(\mathbf{s})$ function put in order the elements of $\mathbf{s} \in \mathbb{R}^{K-1}$ from the bottom to the top and $[a,b]^{K-1}$ is $[a,b] \times \cdots \times [a,b]$ a hypercube in $\mathbb{R}^{-1}$K. In this way, (16) can be reoriented as an unconstrained optimization problem as it is shown in Eq. (18):

$$\begin{array}{ll} \text{minimize} & \hat{J}(\mathbf{s}) \\ \text{subject to} & \mathbf{s} \in \mathbb{R}^{K-1} \end{array} \tag{18}$$

The main advantage of this reformulation is that it allows the use of plenty developed algorithms to unconstrained optimization. It must be underlined that this reformulation can be also applied to step one in the sequential methodology, since it is enough to identify $\widetilde{J}(\mathbf{s}) = SSE(\mathbf{s})$.

4. Memetic algorithms for the bilevel TOD determination problem

Then, the problem shown in Eq. (18) will be named as the bilevel TOD determination problem.

In order to harness the advantages of local search, Espinosa-Aranda et al. (2013) developed a memetic algorithm to introduce local search into meta-heuristic algorithms based on population with the purpose of improving the solution in promising regions.

These memetic algorithms are based on a *population* of individuals or candidates. Each one of them constitutes a solution in the solution space. The population-based algorithms explore successfully the search space due to the use of a population allows to avoid local optima.

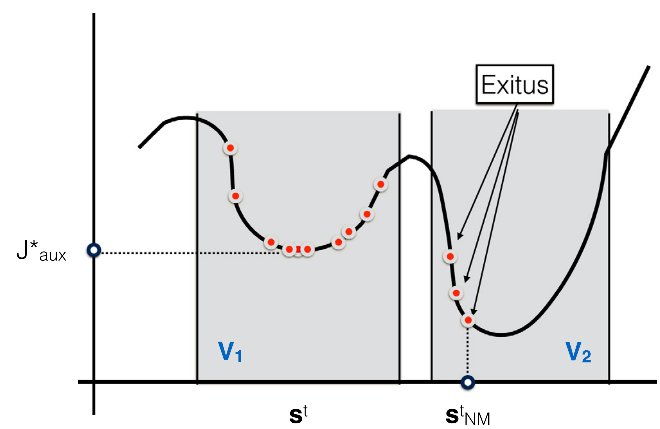

Fig. 3. Illustration of hybridization strategy.

However, these algorithms do not encourage a exploitation phase. For this, the proposed MA is an instance of this class of algorithms and it provides a trade-off between accuracy and computational cost. The proposed algorithm is shown in Table 2. The algorithm has two important parameters which must be fitted n_c and n_r. The parameter n_c, is used in exploration phase and it takes into account the number of successive improvements achieved by the global optimization algorithm. In this case, the local exploitation method will start from the best solution achieved.

Fig. 3 illustrates the basis of this algorithm. When an algorithm fall into a local minimum environment V_1, the value of the objective function is difficult to enhance and new improvements are not obtained until the algorithm scapes successfully from this region in the search space. The role of n_c parameter measures the number of successes (successive improvements) and it is an indirect measurement to check if the algorithm has get out from the V_1 neighbourhood to other neighbourhood V_2 which contains better solutions. Once this fact has been detected (the algorithm has changed its application environment to other more promising) is appropriate to apply a method with good capabilities of local search, it means, a good convergence to the local minimum.

The method used in the exploitation phase is the Nelder–Mead (NM) simplex algorithm. This computational scheme was introduced in Nelder and Mead (1965) to unconstrained optimization problems. Some advantages of this method are that gives notably improvements in first iterations and it is very useful in nonlinear optimization problems for which derivates may not be known.

Table 2
Memetic algorithm for the bilevel TOD determination problem.

Step 1.	(*Initialization*). Initialize the number of iterations (N), the global optimizer parameters and randomly generate an initial population of solutions. Initialize the number of iterations n_c, n_r associated with the global optimizer and NM respectively. Set the counters $t = 1$ and $n = 0$ and let $\widehat{J}^*_{aux} = +\infty$.
Step 2.	(*Exploration stage*). Apply one iteration of global optimization algorithm to the current population. Let $\mathbf{s}^t$ be the current solution and $\widehat{J}^*$ its objective value. If $\widehat{J}^*_{aux} > \widehat{J}^*$, then let $n = n + 1$ and $\widehat{J}^*_{aux} = \widehat{J}^*$.
Step 3.	(*Exploitation stage*). If $n = n_c$ apply n_r iterations of NM algorithm by initializing the method from $\mathbf{s}^t$. Let $\mathbf{s}^t_{NM}$ be the solution found, then replace the best solution of the population by $\mathbf{s}^t = \mathbf{s}^t_{NM}$ and take $n = 0$.
Step 4.	(*Stopping criterion*). If the current iteration is $t = N$, Stop; otherwise set $t = t + 1$ and go to Step 2. Output: The best TODs defined by $(\mathbf{s}^N_1, \mathbf{s}^N_2, \ldots, \mathbf{s}^N_{K-1})$ and its optimal value $\widehat{J}^*$.

This paper uses a hybridization of a standard PSO (Kennedy and Eberhart, 1995; Shi and Eberhart, 1998; Clerc and Kennedy, 2002), GA (Goldberg, 1989) and SA (Kirpatrick et al. (1983) and Černý (1985)) with a NM algorithm. The first algorithm so-called SPSO-NM has been successfully employed in a timetabling train problem (Espinosa-Aranda et al., 2015).

4.1. A Simulated Annealing (SA*) for the bilevel TOD determination problem

Simulated Annealing (SA) is a popular local search meta-heuristic based on the metaphor of a technique for heating and controlled cooling of a material to increase the size of its crystals, reducing their defects. A key point of this algorithm is that it allows hill-climbing moves in order to find a global optimum, thus providing a way of not being trapped in local optima.

The general idea of the algorithm is quite easy. The algorithm starts with an initial solution $\mathbf{s}$. This solution will be compared in each iteration of the algorithm with a new generated solution $\mathbf{s}'$. It can be generated randomly or using some specific rule in a neighbourhood $V(\mathbf{s})$ of the current solution s.

A realization of SA is effected by choosing the neighbourhood function, since its definition has a great impact on the performance of the algorithm. In this paper, the next specific neighbourhood function is proposed.

$$V(\mathbf{s}) := \left\{\mathbf{s}' \in [a,b]^{K-1} / s'_k = s_k;\ \text{for all } k \neq k' \text{ with } k' \in \{1,\ldots,K-1\}\right\} \tag{19}$$

Another issue which must be addressed is the choice of the generation probability function which determines the probability of generating a new solution. In this case, a uniform distribution has been used and the probabilities are proportional to the size of the neighbourhood $V(\mathbf{s})$. Thus, the generation of a new solution $\mathbf{s}' \in V(\mathbf{s})$ will be carried out in three steps:

$$\begin{aligned} &\mathbf{s}' = \mathbf{s} \\ &\text{Choose a random number } k' \in \{1,\ldots,K-1\} \\ &s'_{k'} = a + Rand() \cdot (b-a). \end{aligned} \tag{20}$$

Fig. 4 shows the neighbourhoods $V(\mathbf{s})$ considered by SA and its variant SA* when $K = 3$. The essential advantage of SA* is that allows wide movements, being able to avoid local minima. For this reason, it is an excellent candidate as an algorithm to exploration stage in hybridization.

The probability of accepting the new generated solution $\mathbf{s}'$ is defined by the next rule.

$$\mathbb{P}(\text{Accept } \mathbf{s}' \text{ as next solution}) = \begin{cases} \exp[-(\widehat{J}(\mathbf{s}') - \widehat{J}(\mathbf{s}))/\mathcal{T}_n] & \text{If } \widehat{J}(\mathbf{s}') - \widehat{J}(\mathbf{s}) > 0 \\ 1 & \text{If } \widehat{J}(\mathbf{s}') - \widehat{J}(\mathbf{s}) \leq 0 \end{cases}$$

Following this scheme, best solutions are always accepted, although there is a fraction of non-improving solutions which are accepted. It is made with the purpose of avoiding local optima and can be compared with the mutation process in a genetic algorithm. The fraction of worst solutions which are accepted depends on a temperature parameter $\mathcal{T}_n$

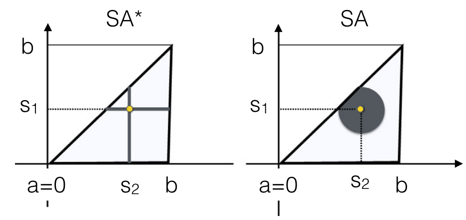

Fig. 4. Illustration of neighbourhoods for SA and SA* and $K = 3$.

which is particular of SA algorithm. The use of the neighbourhood function (19) and the probability function (20) in the basic SA originates the so-called SA* and its pseudocode is shown in Table 3.

5. Determination of the optimal number of TODs

The bilevel TOD determination problem assumes that the number of TOD intervals K is known. This section looks at four methods for determining the optimal number of TOD intervals. The first two are based on the widely used Bayesian Information Criterion and employ all the registered time series, while the third algorithm is based on Vasko and Toivonen (2002) and the fourth is proposed in this work and it is problem-oriented approach. Those last two methods operate with the average observed values.

5.1. Bayesian Information Criterion (BIC)

The BIC is a likelihood criterion for model comparison that penalizes models with additional parameters. The BIC is defined mathematically as:

$$BIC(K) = -2\log L_K(\mathbf{q}_1,\ldots,\mathbf{q}_n) + \lambda m_K \log(n) \tag{21}$$

where $\{\mathbf{q}_1,\ldots,\mathbf{q}_n\}$ is the complete data to be modelled. The first term represents the maximum log likelihood of the data under the model with m_K parameters. The second term $\lambda m_K \log(n)$ is responsible for penalizing the candidate models according to their number of parameters m_K and λ is the penalty weight ($\lambda = 1$ according to the BIC theory). The optimum model corresponds to the one for which the value of BIC, given by Eq. (21) is minimum. If we assume that the observations in a TOD are drawn from a full-covariance Gaussians $\{\mathbf{q}_j\}_{j\in C_k} \sim N(\mu_k, \Sigma_k)$, the BIC for the $K-$TODs solution is defined

$$BIC(K) = \sum_{k=1}^{K} n_k \log(|\Sigma_k|) + K\left(d + \frac{d(d+1)}{2}\right)\log(n) \tag{22}$$

in which n_k denotes the number of articles in cluster C_k and d is the dimension of the flow vector space.

A two-stage process presented by Chiu et al. (2001) for determining the optimum number of clusters. The procedure is described below, following the work of Xia and Chen (2007).

The two-stage process first examines the BIC for all potential clustering solutions. The goal is to find the smallest number of clusters that have the lowest BIC, because the BIC decreases first and then increases as the number of clusters increases, the BIC for each K (clustering solution) is calculated. Beginning from $K = 1$, the first

Table 3
SA* algorithm for the bilevel TOD determination problem.

Step 1.	*(Initialization).* Initialize the algorithm's parameters: number of iterations (N), initial solution $\mathbf{s}$, temperature cooling schedule $\{\mathcal{T}_n\}$, and initial temperature $\mathcal{T} = \mathcal{T}_0$. Select a repetition schedule, $\{M_n\}$, that defines the number of iterations executed at each temperature. Set the temperature change counter $n = 0$ and repetition counter m=0.
Step 2.	Generate a new solution $s' \in V(\mathbf{s})$ through Eq. (20)
Step 3.	Compute $\Delta = \widehat{J}(\mathbf{s}') - \widehat{J}(\mathbf{s})$. If $\Delta \leq 0$ then $\mathbf{s} = \mathbf{s}'$, $\widehat{J}^* = \widehat{J}(\mathbf{s}')$ and $\mathbf{s}^* = \mathbf{s}'$. Otherwise $\Delta > 0$, set $\mathbf{s} = \mathbf{s}'$ with probability $\exp[-\Delta/\mathcal{T}_n]$. Take $m = m + 1$.
Step 4.	If $m = M_n$ then $n = n + 1$ and $m = 0$.
Step 5.	*(Stopping criterion).* If the current number of iterations is $t = N$, Stop; otherwise let $t = t + 1$ and go to Step 1. Output: The best TOD intervals of $[a, b]$ defined by $\mathbf{s}^*$ and its objective value $\widehat{J}^*$.

$K = \widehat{K}$ value that satisfies $BIC(\widehat{K}) < BIC(\widehat{K} + 1)$ is chosen as a coarse estimate of the number of clusters. In the second stage, the ratio of changes in dispersion measurement is used to determine the optimum number of clusters based on the coarse estimate obtained in the first stage. The ratio of changes in dispersion measurement is defined as $R(K) = s_{K-1}/s_K$ for $K = 2, \ldots, \hat{K}$, in which s_{K-1} denotes the change in dispersion measurement if K clusters are merged into $K - 1$ clusters.

The parameter s_K can be computed as $s_K = l_K - l_{K+1}$, in which $l_K = \sum_{k=1}^{K} n_k \log(|\Sigma_k|)$. This second stage is based on the understanding that a significant increase in $R(K)$ will be observed when two clusters that should not be merged are merged. The $R(K)$ value for each $K (K = 2, \ldots, \hat{K})$ is calculated, and the two largest $R(K)$ values are identified as $K = K_1$ (the largest) and $K = K_2$ (the second largest). Xia and Chen (2007) uses an empirical threshold value of $R(K_1)/R(K_2) = 1.15$; that is, if $R(K_1)/R(K_2) > 1.15$, K is set to K_1; otherwise, K is set to $\max(K_1, K_2)$.

5.2. ΔBIC

In segmenting an audio stream the BIC has widely used (Chen and Gopalakrishnan, 1998; Wang et al., 2008). It can be shown (Chen and Gopalakrishnan, 1998) that if the expression

$$\Delta BIC(\mathbf{s}_k(K)) = (n_k + n_{k+1})\log(|\Sigma_k \cup \Sigma_{k+1}|) - n_k \log(|\Sigma_k|) \\ - n_{k+1}\log(|\Sigma_{k+1}|) \\ - \left(d + \frac{d(d+1)}{2}\right)\log(n_k + n_{k+1}) \text{ with } k = 1, \ldots, K-1. \quad (23)$$

is positive, the time $\mathbf{s}_k$ is a good candidate for a segment boundary. Note that $\Sigma_k \cup \Sigma_{k+1}$ represents the variance–covariance matrix of the observations $\{\mathbf{q}_j\}_{j \in C_k(\mathbf{s}) \cup C_{k+1}(\mathbf{s})}$. It is possible to apply this criterion to the K–solution to determine if its border points $\mathbf{s}_k$ are significant. The criterion used is to choose the solution with significant points of maximum cardinality K.

5.3. A modified PETE algorithm

Vasko and Toivonen (2002) present the so-called PETE algorithm to determine the number of time segments. This method generates a p–value for each increase in the number of segments. In this paper a modification of this method is adapted in order to reduce its computational cost. Let $e(\mathbf{s}^K)$ be the segmentation error of the solution $\mathbf{s}^K$. By using a Monte Carlo simulation the random error of adding a new segment in the partition is calculated as follows: A random segment $k \in \{1, \ldots, K\}$ is selected, then the observations $C_k(\mathbf{s}^K)$ are randomly ordered and the segment is randomly partitioned. The new error denoted by $e_j(\mathbf{s}^K)$ is calculated. Drawing from the random sample $\{e_j(\mathbf{s}^K)\}$ the p–quantile e_p of the random error is calculated, if it satisfies

$$e(\mathbf{s}^{K+1}) < e_p \quad (24)$$

then the $K + 1$-solution is selected and the procedure is repeated.

5.4. An approach oriented to the TOD determination problem

The desired objective is a suitable model of the dynamic mechanisms which operate in the traffic network. In this context the selection of a high number of TODs has as a consequence a greater analytic cost but gives more satisfactory results. Henceforth a trade-off between computational cost and accuracy must be achieved. To that end a natural method is to require a number of clusters K in which the mean relative error $\bar{e}$ in describing the time series as a set of stationary flows does not exceed a value $\tilde{e}$ given by the planner, that is

$$K^* = \operatorname*{Arg\,minimize}_{K} \left\{ K : \bar{e}(\mathbf{s}^K) < \tilde{e} \right\}. \quad (25)$$

This value $\bar{e}$ has a physical interpretation and thus allows a priori determination.

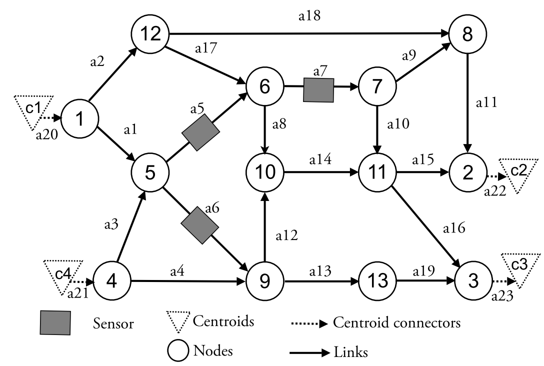

Fig. 5. Nguyen–Dupuis network.

6. Computational experiments

The objectives of these computational tests are:

1. To analyse the performance of the proposed MAs applied to the TOD determination problem. This goal has been analysed in Experiment 1.
2. To evaluate the proposed methodology in a real case. The purpose is to compare a sequential methodology in which, in a first stage the number of TODs is determined and later the traffic signal control for each TOD is optimized, with the simultaneous methodology. The numerical results are collected in Experiment 2.
3. The objective in Experiment 3 is to analyse the above four indexes to determine the optimal number of TODs over real data.

Experiment 1: A comparison of the performance of different MAs

The first data set was generated through a simulation experiment, using the dynamic traffic load model developed in Sánchez-Rico et al. (2014); the Nguyen–Dupuis network shown in Fig. 5 has been used.

The results are shown in Fig. 6. These results consist of 4 daily traffic patterns, considering traffic flow or density over a set of three sensors located in the network. Eight synthetic data sets are considered and $K = 5$ TODs is set.

The combined model given in Section 3.2 has a bilevel structure, in which the evaluation of the objective function requires solving K optimization problems. This has motivated the need for efficient resolution algorithms. In Experiment 1, the standard PSO, SA, SA*, GA

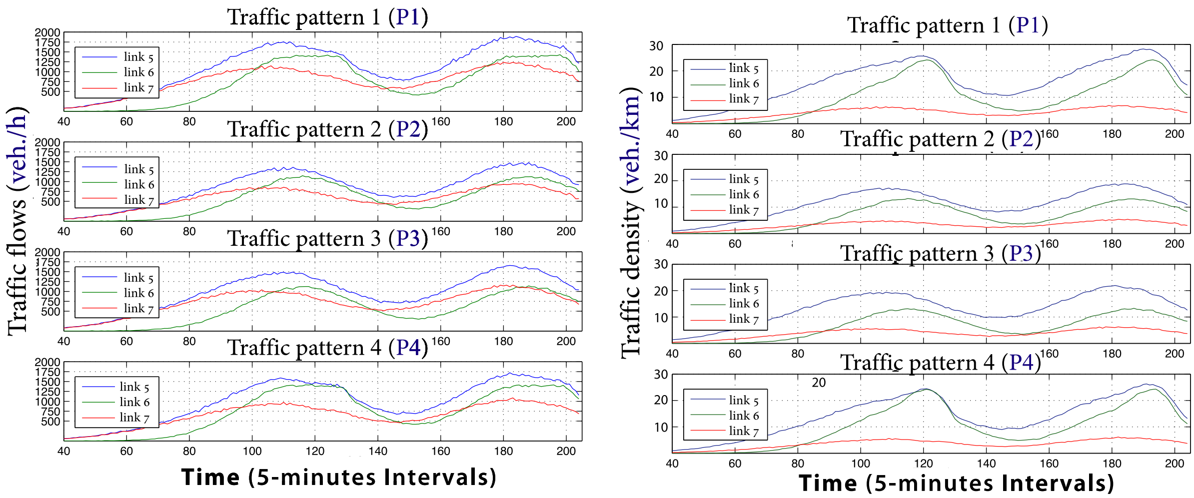

Fig. 6. Synthetic data set based on flows and densities.

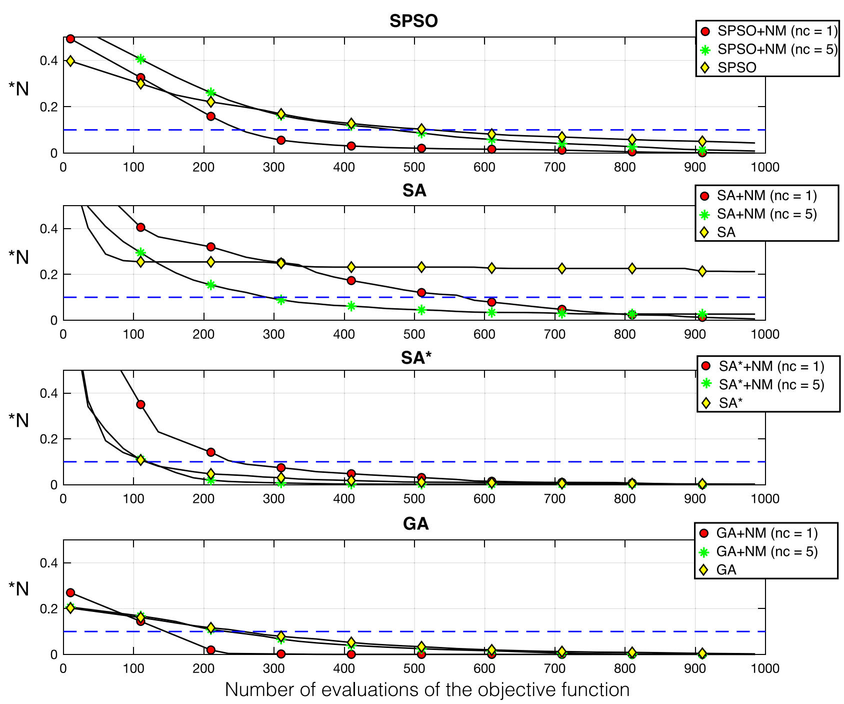

Fig. 7. Average performance of the MA in density-based tests.

algorithms and their hybridizations with the NM have been tested on the problem (7). Note that the works (Park et al., 2004; Lee et al., 2011) use GA algorithms for solving this problem and the GA can be considered as the baseline test.

Metaheuristic algorithms depend on a large amount of parameters and the adjustment of these parameters is one of the main challenges of their application. The parameters used in GA, SA and SA* takes their values from Matlab functions `ga` and `simulannealbnd`. In PSO algorithm, inertia weight has taken the value $w = 1/(2 * log(2)) = 0.721$, and the learning parameters $c_1 = 0.5 + log(2) = 1.1913$ and $c_2 = c_1$. Furthermore, two neighbourhood topologies have been considered, the *gbest swarm* in PSO and the *lbest swarm* in its hybridization, taking an specific number of particles (neighbour count) $np = 3$. Finally, gbest swarm topology has been considered to the crude PSO because it converges more quickly than lbest topology. Population size in GA and PSO has been fixed to 40 individuals/particles.

Due to the number of parameters and the endless possibilities to compare and adjust them, we are focus on the effect of the parameter n_c in the hybridization proposed in the paper. The selected parameters for the hybridization are $n_c \in \{1, 5\}$ and $n_r = 100$. The objective of this test is not to find the best algorithm, since it depends on the kind of traffic network, the number of link counts, the TODs number and the settings chosen in the algorithms. In this paper, the main objective is to test if the proposed MAs improve the performance of baseline algorithms.

As the algorithms have a probabilistic nature each instance was run 10 times over the eight test problems. In order to visualize the results obtained, the average change in the 10 runs and for all the test problems is considered, and in addition the value of the objective function has been standardized as

$$\mathbf{Z}^* = \frac{\hat{J}(\mathbf{s}) - J_{\mathrm{MIN}}}{J_{\mathrm{MAX}}} \tag{26}$$

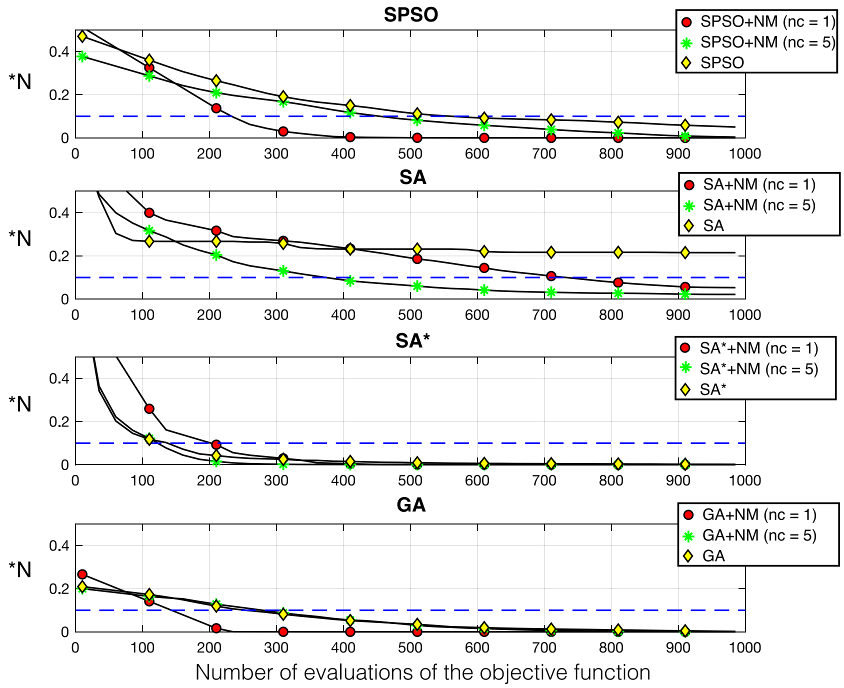

Fig. 8. Average performance of the MA in flow-based tests.

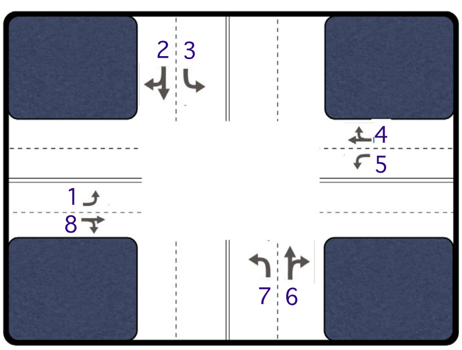

Fig. 9. A four-leg intersection.

Table 4
Performance of the MAs in flow-based and density-based tests.

Algorithm	Flow-based tests average		Density-based tests average	
	500 eval	1000 eval	500 eval	1000 eval
SPSO	1201E−01	5020E−02	1078E−01	4340E−02
SPSO + NM (nc = 1)	9176E−04	1764E−08	2220E−02	1636E−04
SPSO + NM (nc = 5)	8870E−02	3700E−03	9420E−02	8800E−03
SA	2316E−01	2150E−01	2317E−01	2121E−01
SA + NM (nc = 1)	1978E−01	5310E−02	1321E−01	5600E−03
SA + NM (nc = 5)	6610E−02	2170E−02	4770E−02	2640E−02
SA*	1070E−02	1800E−03	1190E−02	2600E−03
SA* + NM (nc = 1)	2033E−04	2790E−05	3530E−02	4766E−05
SA* + NM (nc = 5)	9064E−04	8471E−04	2300E−03	1200E−03
GA	3900E−02	2000E−03	3720E−02	2300E−03
GA + NM (nc = 1)	7145E−05	3945E−05	1993E−04	1840E−04
GA + NM (nc = 5)	3750E−02	4867E−04	2820E−02	5541E−04

Table 5
Results obtained with the Nelder Mead algorithm.

Problem	Flows-based tests		Densities-based tests	
	Average Z^*	Success rate (%)	Average Z^*	Success rate (%)
$P1$	0.425	7.0	0.694	9.0
$P2$	0.574	9.0	0.502	2.0
$P3$	0.397	11.0	0.600	2.0
$P4$	0.750	7.0	0.507	2.0

where J_{MAX} and J_{MIN} are respectively the maximum and minimum found for all the algorithms in a given problem. This standardization means that the 8 test problems are equally important.

The evolution of the algorithms is shown in Figs. 7 and 8, while Table 4 shows the results of each algorithm when 500 and 1000 evaluations are computed. It is observed that the hybridization speeds up the original algorithm. In addition it can be seen that the SA^*+NM and $GA+NM$ algorithms with $n_c = 1$ outperforms the other algorithms.

In the previous experiment, it has been shown that the hybridization of metaheuristic algorithms with NM method allow to accelerate the convergence of the base method. Besides, it is observed that all methods converge to the same solution in each of the ten executions completed (if the mean value in 10 tests is $\mathbf{Z}^* \to 0$, henceforth, each one individually also converges to zero). The acceleration in the base procedure is due to the NM algorithm. In this article, it is intended to answer the question about if it would be better to apply exclusively NM algorithm. In the next test, the previous 8 problems have been solved 100 times, departing from different random points in each execution. A success in the experiment is considered if the reached value satisfies $\mathbf{Z}^* < 0.001$. Table 5 shows the mean value obtained $\mathbf{Z}^*$ and the success percentage. The results indicate that in 800 runs, NM has converged into a local minima in 751 times. It demonstrates that MA algorithms are more robust and allow to alleviate the shortcomings of NM method.

Experiment 2: A real case study

This example is taken from Yin (2008) and consists of a real-world intersection between 164th Street SW and Alderwood Mall Parkway in the City of Lynwood, Washington. The flows were recorded in March–April 2005, 36 observed flow patterns were retrieved for the PM peaks (4:30–6:30 p.m.) between Tuesday and Thursday. Based on these data two test problems were designed, in the first the 36

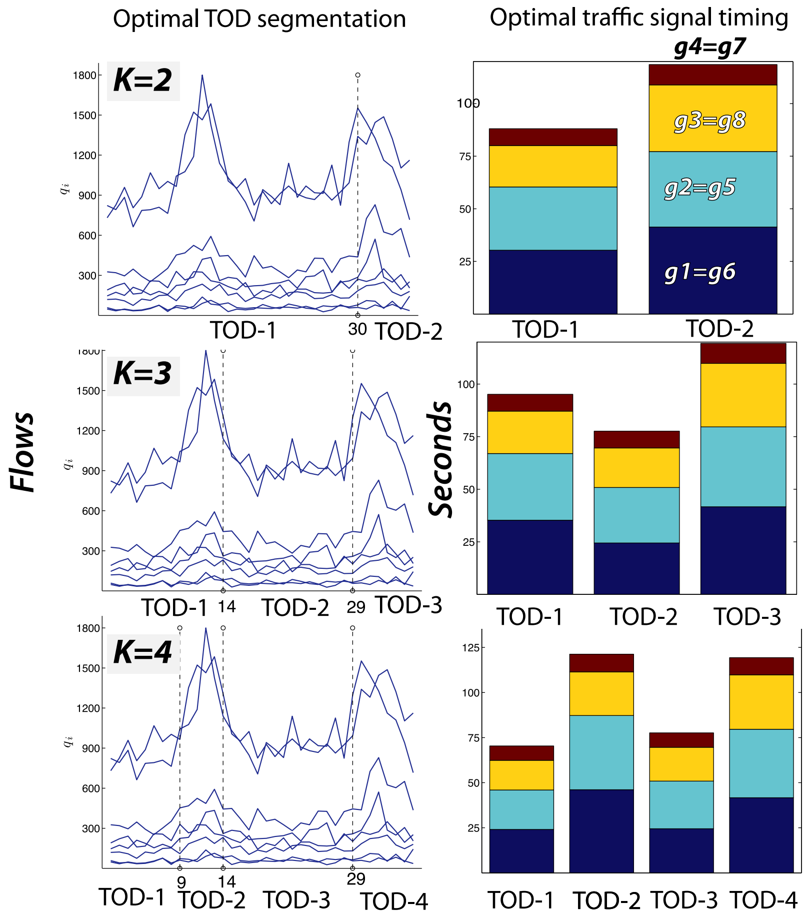

Fig. 10. Test 2 solution through simultaneous methodology.

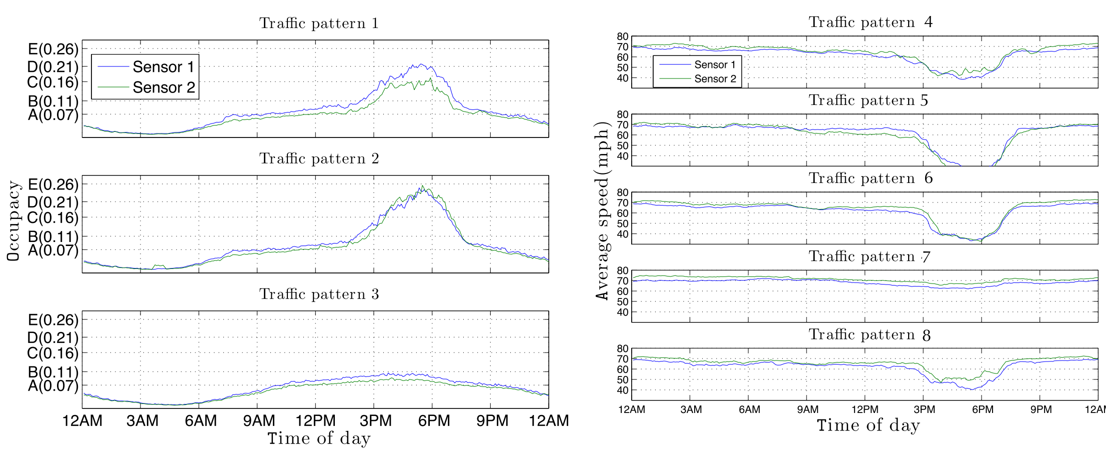

Fig. 11. Real data set. Average daily occupancy and speed profiles.

patterns are considered as consecutive over a day (named Test 1). The second (named Test 2) is obtained by modifying the original data to introduce a dependence on the current time of day, in particular the flows of the jth-period have been multiplied by the weighting factor $f_j = 0.85 + 0.65\exp(-0.15 * (j-12)^2) + 0.65\exp(-0.15 * (j-32)^2)$ with $j = 1, \dots, 36$.

The intersection is shown in Fig. 9 and the saturation flow rates s_i for groups $i = 1, \dots, 8$ are $1900, 3800, 3800, 1900, 1900, 3800, 3800, 1900$. A specific lead–lag phasing sequence is used in the example and the resulting constraints for the traffic-signal timing problem are:

$$g_1 + g_2 + g_3 + g_4 + L = \mathtt{C} \tag{27}$$

$$g_1 = g_6, g_2 = g_5, g_3 = g_8, g_4 = g_7 \tag{28}$$

$$g_1 \geq g_{\min}, i = 1, 2, \dots, 8 \tag{29}$$

$$\mathtt{C}_{\min} \leq \mathtt{C} \leq \mathtt{C}_{\max} \tag{30}$$

where L is the total lost time per cycle, 14 s used in the example; $g_{\min}$ is the minimum green time, 8 s used, and $\mathtt{C}_{\min}$ and $\mathtt{C}_{\max}$ are the minimum and maximum cycle length, specified as 50 s and 140 s, respectively. When using Eq. (14) to calculate the total delay, the duration of each time period $j = 1, \dots, 36$ is set as $T_j = 0.25h$.

Over the two test problems a *sequential* and a *simultaneous* methodology is applied. The sequential methodology is the one followed in

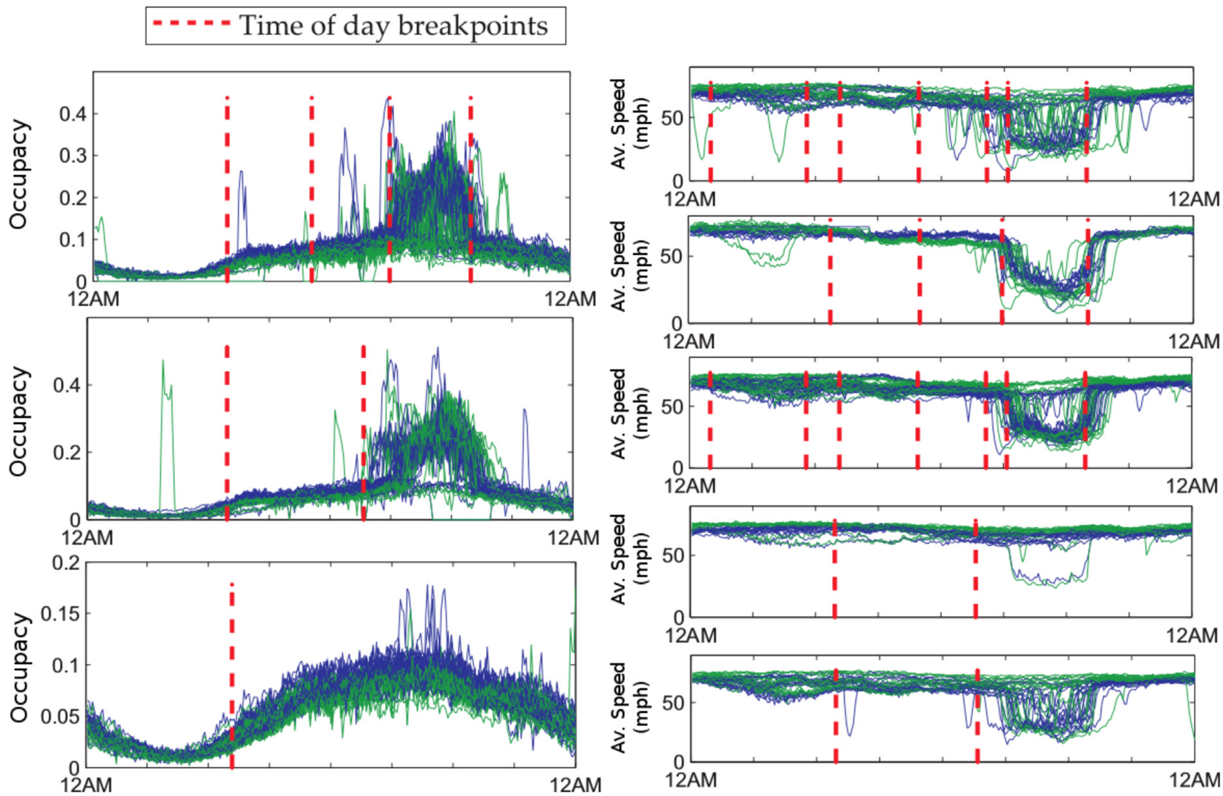

Fig. 12. Solution of the TOD problem using the BIC rule.

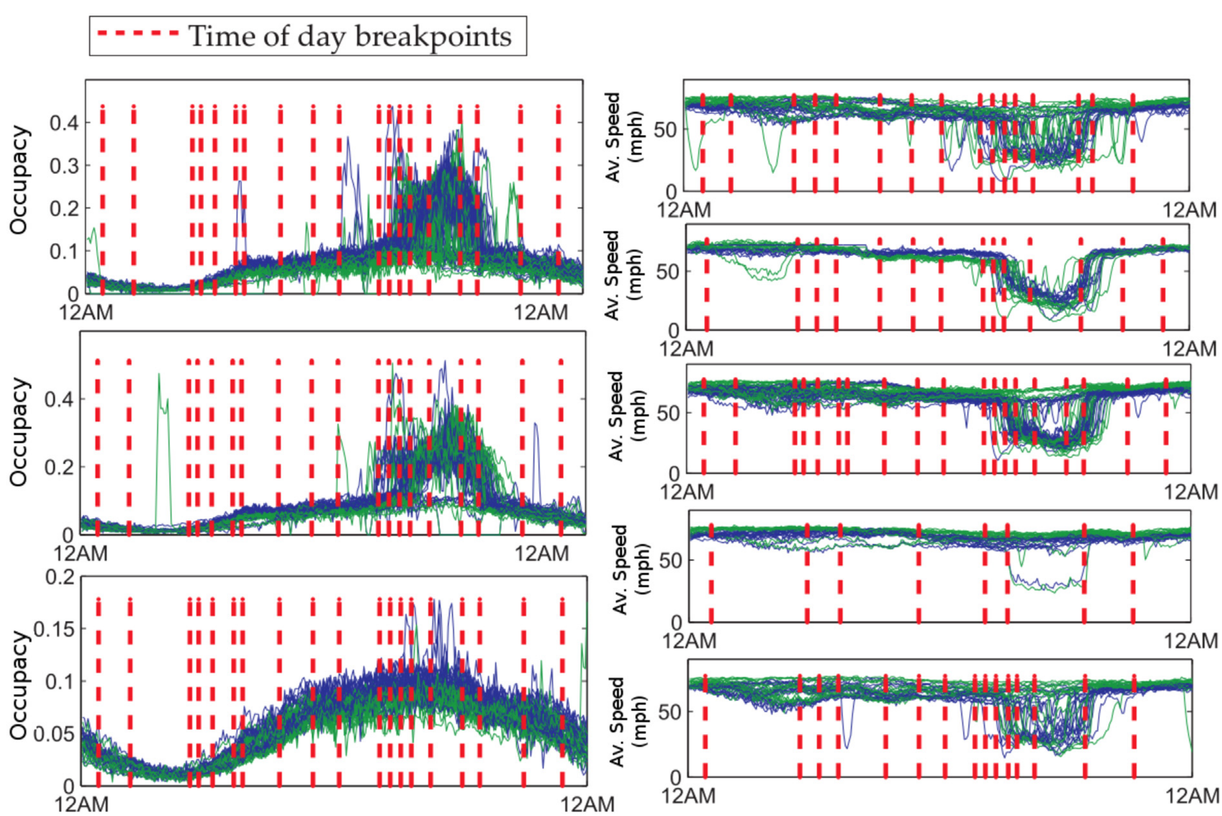

Fig. 13. Solution of the TOD problems using the ΔBIC rule.

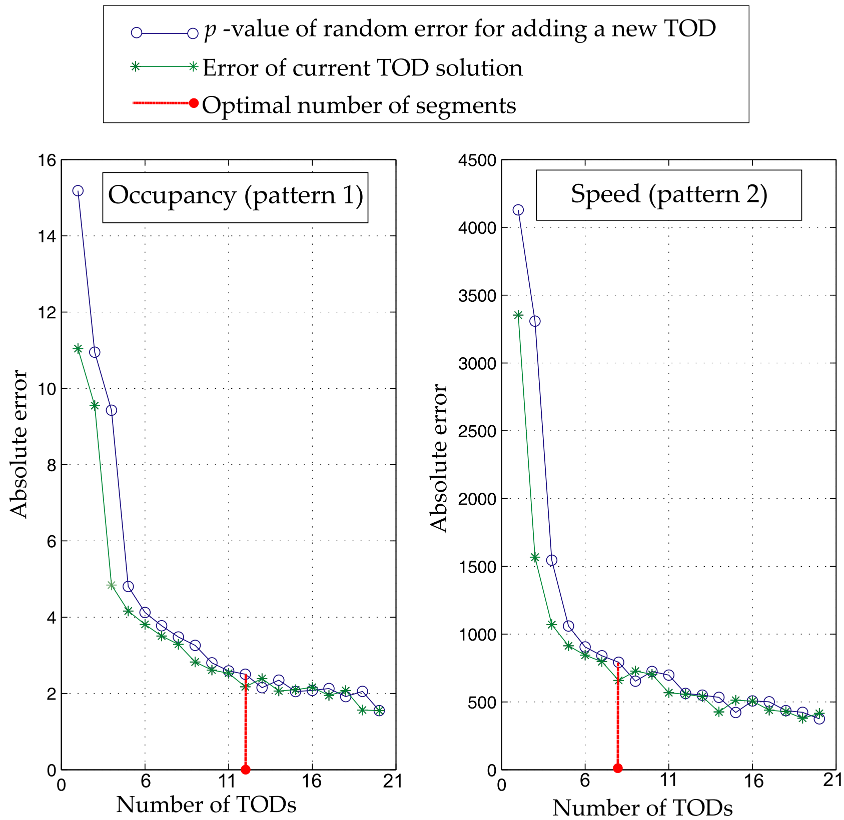

Fig. 14. PETE algorithm behaviour.

practice, first the TODs are determined and later the optimal traffic signal timing is calculated for each TOD. The simultaneous methodology is the one described in this paper and it directly minimizes total time in the intersection. Both methods require an algorithm to solve the problems (15). He and Hou (2012) use ant colony and a genetic algorithm to solve this problem. These algorithms present two disadvantages, the first is its high computational cost due to the large number of solutions and the second is, because the SA algorithm is used for the bilevel model, conflicts may appear between the accuracy of resolution of the lower level problem and the objective function value as shown in García and Marín (2002). This numerical experiment employs an interior-point algorithm (implemented in the MATLAB function `fmincon`). In our numerical tests, there is evidence that this option is faster and more efficient than the GA algorithm implemented in the MATLAB function ga. The results obtained are shown in Table 6. The first column shows the number of TODs considered, the second column the methodology used and the third the time required to obtain the best solution. The fourth column shows $\widetilde{J}_{\text{Delay}} = \sum_k \widetilde{J}_k(\mathbf{s})$ where $J_k(\mathbf{s})$ is calculated by using Eq. (14) and the fifth shows the value $\widetilde{J}_{\text{TOD}}$ where $J_k(\mathbf{s})$ is calculated by Eq. (7). The results obtained agree with expectation. Each methodology achieves better results for the objective function it is trying to minimize. The simultaneous methodology minimizes the total waiting time while the sequential gives TODs with lower variability. It is seen, however, that the sequential approach is very efficient and is capable of obtaining similar solutions to the simultaneous method. Both methodologies allow the systems to be recalibrated automatically. One conclusion is this, considering that traffic regulation in different time intervals is advantageous. The reduction in the waiting time by considering $K = 4$ instead $K = 2$ is about 3.1%. To go into this question more deeply, Fig. 10 shows the solutions obtained by employing the simultaneous methodology in Test 2. It depicts the TODs obtained for several values of K and how the optimal traffic signal timing is strongly dependent on the number of TODs. It shows that the cycle amplitude C is dependent on the level of congestion of the TOD.

Table 6
Comparison between the sequential and simultaneous methodologies.

K	Approach	CPU time (s)	$\widetilde{J}_{\text{Delay}}$ (s)	$\widetilde{J}_{\text{TOD}}$
Test 1				
2	Sequential	0.1	2 051 660.5	6036.3
	Simultaneous	212.1	2 021 479.3	6123.2
3	Sequential	1.7	1 992 496.1	5500.9
	Simultaneous	170.7	1 989 842.6	5509.4
4	Sequential	1.3	1 979 965.9	5242.8
	Simultaneous	169.0	1 977 440.8	5451.0
Test 2				
K	Approach	CPU time (s)	$\widetilde{J}_{\text{Delay}}$ (s)	$\widetilde{J}_{\text{TOD}}$
2	Sequential	0.2	2 750 230.8	10 340.7
	Simultaneous	145.9	2 737 846.5	10 759.0
3	Sequential	1.4	2 708 806.7	9540.2
	Simultaneous	152.4	2 703 892.0	10 038.9
4	Sequential	1.2	2 655 422.0	7326.6
	Simultaneous	192.9	2 653 730.4	7340.1

Experiment 3: Determination of the optimal number of TODs

In this section the algorithms described in Section 5 to identify the optimal number of TODs are tested. In order to assess the methods, real traffic data collected by the California Freeway Performance Measurement System (PeMS) is used. The PeMS collects the traffic data in real time from over 25,000 individual detectors across all major metropolitan areas of the State of California. In order to test the methodology proposed in this article, observations over 100 days in 2013 have been collected from two dual loop detector stations in the

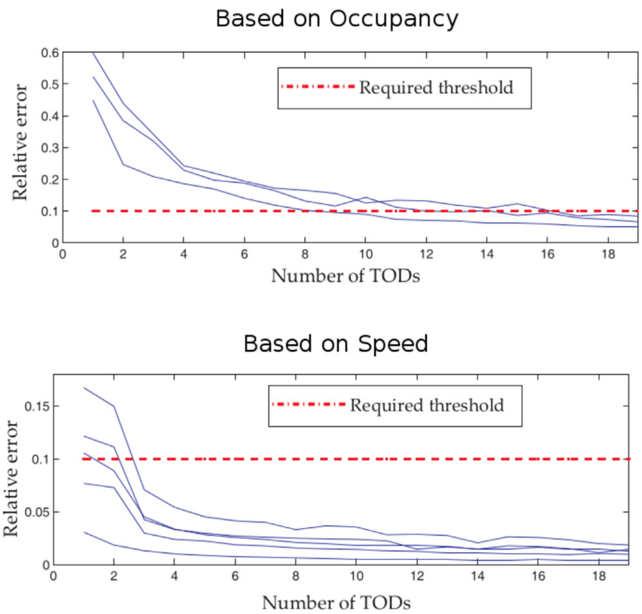

Fig. 15. OP algorithm behaviour.

Bay Area of California. The data have been previously classified into different traffic patterns on different days using the methodology given in García-Ródenas et al. (2017). The traffic profiles based on speed and occupancy are shown in Fig. 11. The data used are available at http://bit.ly/1hsTEjO.

For each of the 8 traffic patterns the TODs have been identified for $K = 1, \ldots, 20$, and by using the BIC, ΔBIC, modified PETE and the oriented-problem (OP) methods, the optimal number of TODs have been selected. The p value employed in the modified PETE method was $p = 0.05$ and the mean relative error $\tilde{e} = 0.1$ for the OP. The optimal number of TOD intervals resulting from each algorithm is shown in Table 7. The observed results indicate that the BIC is the most parsimonious method since it determines a reduced number of TODs, the ΔBIC method is highly sensitive and establishes a higher number of TODs. On the other hand it seems that the modified PETE algorithm finds an acceptable number of TODs. The OP method employs the same threshold for both types of data (occupancy or speed) but the variability level is different for each, which means that for the occupancy a high number of TODs is established but this is not the case for the speed.

To gain additional insight into the results the different procedures are displayed. The clustering results using the BIC and the ΔBIC criteria are shown in Figs. 12 and 13. The first column is associated with the occupancy and the second is based on the speed; finally the vertical lines separate the different TODs according to their algorithm. As can be observed in the figures the ΔBIC criterion produces a non-significant number of TODs which may not be helpful for traffic control. These methods tackle the existing variability between days.

Unlike the two previous methods, the modified PETE and the OP algorithms work with a average traffic profile. Fig. 14 illustrates the way the modified PETE algorithm works. The first graph corresponds to pattern 1 (occupancy) and the optimal number of time segments for this particular case is 12. It is possible to observe in the graph that the stopping criterion of the algorithm is when the $p-$ value of the mean error of adding a new random segment is less than the error obtained by increasing one TOD in the current solution.

Fig. 15 shows the results obtained from the OP method; the first graph shows the occupancy-based tests and the second the speed-based tests. In both graphs the red line represents the required threshold for the mean relative error value that must not be exceeded when selecting the number of clusters. This method has resulted in practice in

Table 7
Optimal number of TOD intervals by selection criterion and the problem.

Criterion	BIC	ΔBIC	PETE	OP
Problem	(Based on occupancy)			
Pattern 1	5	20	12	13
Pattern 2	3	20	12	18
Pattern 3	2	20	10	9
Problem	(Based on speed)			
Pattern 1	8	17	10	3
Pattern 2	5	15	8	4
Pattern 3	8	20	12	4
Pattern 4	3	9	12	2
Pattern 5	3	16	16	2

more consistent and comprehensive partitions, for example the optimal number of partitions for the occupancy problem varies between 10 and 18 while for the speed problem it varies between 2 and 4. The criterion of fixing a maximum mean relative error is easily interpreted by planners. This makes the method more easy to apply than the previous one.

7. Conclusions

In this paper a methodology for automatically updating pre-timed traffic control systems is set out. For this purpose a bilevel model has been formulated, which simultaneously includes the problem of determining the time-of-day breakpoints and the traffic control problem for each time interval. The proposed model has been solved by the use of a class of MAs. The effectiveness of the algorithms SA*+NM and GA+NM has been demonstrated on a collection of synthetic and real problems and they outperform GA, SA+NM and SPSO+NM methods. Furthermore, this feature allows us to apply SA*+NM or GA+NM algorithms to bilevel TOD determination problems.

The sequential and the simultaneous methodologies have been illustrated over a real problem. As might be expected the simultaneous method achieves better results than the sequential but the computational cost is higher. It can be seen that the sequential methodology, which is used in practice, achieves satisfactory results.

Finally, the automatic determination of the optimal number of TODs has been studied by employing the most promising methods, based on BIC, ΔBIC, a modified PETE algorithm and a OP approach. It is observed numerically that the criteria based on the BIC is conservative and produce a reduced number of TODs in comparison with the other methods. The index ΔBIC produces a large number of TODs. The oriented-problem method has an interpretation which allows its easy application and achieves fair values. The modified PETE method also achieves solutions in accordance with what is expected.

Acknowledgements

The research of García-Ródenas, López-García and López-Gómez was supported by Projects TRA2014-52530-C3-2-P and TRA2016-76914-C3-2-P of the Spanish-FEDER EU Ministry of Economy and Competitiveness.

The third author wishes to thank the Mexican government for its financial support through CONACYT.

References

Abdoos, M., Mozayani, N., Bazzan, A.L., 2013. Holonic multi-agent system for traffic signals control. Eng. Appl. Artif. Intell. 26, 1575–1587.
Angulo, E., Romero, F., García, R., Serrano-Guerrero, J., Olivas, J., 2011. An adaptive approach to enhanced traffic signal optimization by using soft-computing techniques. Expert Syst. Appl. 38 (3), 2235–2247.
Araghi, S., Khosravi, A., Creighton, D., 2015. A review on computational intelligence methods for controlling traffic signal timing. Expert Syst. Appl. 42 (3), 1538–1550.

Bie, Y., Gong, X., Liu, Z., 2015. Time of day intervals partition for bus schedule using gps data. Transp. Res. C 60, 443–456.
Černý, V., 1985. Thermodynamical approach to the traveling salesman problem an efficient simulation algorithm. J. Optim. Theory Appl. 45 (1), 41–51.
Ceylan, H., Bell, M.G., 2004. Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. Transp. Res. Part B 38 (4), 329–342.
Chen, S.S., Gopalakrishnan, P.S., 1998. Speaker, environment and channel change detection and clustering via the bayesian information criterion. In: Proc. DARPA Broadcast News Transcription and Understanding Workshop. pp. 127–132.
Chen, X., Ong, Y.S., Lim, M.H., Tan, K.C., 2011. A multi-facet survey on memetic computation. IEEE Trans. Evol. Comput. 15 (5), 591–607.
Chin, Y.K., Yong, K.C., Bolong, N., Yang, S.S., Teo, K.T.K., 2011. Multiple intersections traffic signal timing optimization with genetic algorithm. In: 2011 IEEE International Conference on Control System, Computing and Engineering, Penang, Malaysia, November (2011) 25-27. pp. 454–459.
Chiu, T., Fang, D., Chen, J., Wang, Y., Jeris, C., 2001. A robust and scalable clustering algorithm for mixed type attributes in large database environment. In: KDD '01 Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 263–268.
Clerc, M., Kennedy, J., 2002. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. IEEE Trans. Evol. Comput. 6 (1), 58–73.
Council, N.R., Board, T.R., of Sciences, N.A., 2010. HCM 2010, Highway Capacity Manual Concepts. HCM 2010, Highway Capacity Manual. National Academy of Sciences.
Espinosa-Aranda, J.-L., García-Ródenas, R., Angulo, E., 2013. A framework for derivative free algorithm hybridization. In: Tomassini, M., Antonioni, A., Daolio, F., Buesser, P. (Eds.), Adaptive and Natural Computing Algorithms. In: Lecture Notes in Computer Science, vol. 7824, Springer Berlin Heidelberg, pp. 80–89.
Espinosa-Aranda, J., García-Ródenas, R., Ramírez-Flores, M., López-García, M., Angulo, E., 2015. High-speed railway scheduling based on user preferences. European J. Oper. Res. 246 (3), 772–786.
Fu, T.-C., 2011. A review on time series data mining. Eng. Appl. Artif. Intell. 24 (1), 164–181.
García, R., Marín, A., 2002. Parking capacity and pricing in park'n ride trips a continuous equilibrium network design problem. Ann. Oper. Res. 116 (1–4), 153–178.
García-Nieto, J., Alba, E., Olivera, A.C., 2012. Swarm intelligence for traffic light scheduling application to real urban areas. Eng. Appl. Artif. Intell. 25 (2), 274–283, Special Section Local Search Algorithms for Real-World Scheduling and Planning.
García-Ródenas, R., López-García, M., Sánchez-Rico, M., 2017. An approach to dynamical classification of daily traffic patterns. Comput.-Aided Civ. Infrastruct. Eng. 32 (3), 191–212.
Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning, first ed. Addison-Wesley Longman Publishing Co. Inc., Boston, MA, USA.
Gordon, R., Reiss, R., Haenel, H., Case, E., French, R., Mohaddes, A., Wolcott, R., 1985. Traffic Control Systems HandBook. Implementation Package. U.S. Department of Transportation, Federal Highway Administration.
Hauser, T., Scherer, W., 2001. Data mining tools for real-time traffic signal decision support & maintenance. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Vol. 3. pp. 1471–1477.
He, J., Hou, Z., 2012. Ant colony algorithm for traffic signal timing optimization. Adv. Eng. Softw. 43 (1), 14–18.
Hu, Z., Bao, Y., Chiong, R., Xiong, T., 2015. Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. Energy 84, 419–431.
van Katwijk, R., Schutter, B.D., Hellendoorn, J., 2006. Traffic adaptive control of a single intersection A taxonomy of approaches. Proc. Vol. 39 (12), 227–232, Symposium on Control in Transportation Systems.
Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: IEEE International Conference on Neural Networks - Conference Proceedings, Vol. 4. pp. 1942–1948.
Khamis, M.A., Gomaa, W., 2014. Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework. Eng. Appl. Artif. Intell. 29, 134–151.
Kirpatrick, S., Gelatt, C.D., Vechi, M.P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680.
Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beaird, S., Braud, C., Bonneson, J., Tarnoff, P., Urbanik, T., 2008. Traffic Signal Timing Manual. United States Department of Transportation, Federal Highway Administration, Report No. FHWA-HOP-08-024.
Lee, I.-K., Kim, Y.-C., 2011. Optimization and evaluation for time-of-day signal timing of arterial traffic. Proc. Eastern Asia Soc. Transp. Stud. 2011, 338.
Lee, J., Kim, J., Park, B., 2011. A genetic algorithm-based procedure for determining optimal time-of-day break points for coordinated actuated traffic signal systems. KSCE J. Civ. Eng. 15 (1), 197–203.
Li, Z., Schonfeld, P., 2015. Hybrid simulated annealing and genetic algorithm for optimizing arterial signal timings under oversaturated traffic conditions. J. Adv. Transp. 49 (1), 153–170.
López García, M., García-Ródenas, R., Gómez, A., 2014. Hybrid meta-heuristic optimization algorithms for time-domain-constrained data clustering. Appl. Soft Comput. J. 23, 319–332.
McKenney, D., White, T., 2013. Distributed and adaptive traffic signal control within a realistic traffic simulation. Eng. Appl. Artif. Intell. 26 (1), 574–583.
Moscato, P., Norman, M.G., 1992. A memetic approach for the traveling salesman problem implementation of a computational ecology for combinatorial optimization on message-passing systems. In: Proceedings of the International Conference on Parallel Computing and Transputer Applications. IOS Press, pp. 177–186.
Nelder, J., Mead, R., 1965. A simplex method for function minimization. Comput. J. 308–313.
Neri, F., Cotta, C., 2012. Memetic algorithms and memetic computing optimization a literature review. Swarm Evol. Comput. 2, 1–14.
Park, B.B., Lee, J., 2008. A procedure for determining time-of-day break points for coordinated actuated traffic signal systems. KSCE J. Civ. Eng. 12 (1), 37–44.
Park, B., Lee, D.-H., Yun, I., 2003. Enhancement of time of day based traffic signal control. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Vol. 4. pp. 3619–3624.
Park, B., Santra, P., Yun, I., Lee, D.-H., 2004. Optimization of time-of-day breakpoints for better traffic signal control. Transp. Res. Rec. 1867, 217–223.
Qu, X., Zhang, R., Liu, B., Li, H., 2017. An improved tlbo based memetic algorithm for aerodynamic shape optimization. Eng. Appl. Artif. Intell. 57, 1–15.
Ratrout, N., 2011. Subtractive clustering-based k-means technique for determining optimum time-of-day breakpoints. J. Comput. Civ. Eng. 25 (5), 380–387.
Sabar, N.R., Kieu, L.M., Chung, E., Tsubota, T., de Almeida, P.E.M., 2017. A memetic algorithm for real world multi-intersection traffic signal optimisation problems. Eng. Appl. Artif. Intell. 63, 45–53.
Sánchez-Rico, T., García-Ródenas, R., Espinosa-Aranda, J., 2014. A Monte Carlo approach to simulate the stochastic demand in a continuous dynamic traffic network loading problem. IEEE Trans. Intell. Transp. Syst. 15, 1362–1373.
Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. In: Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., the 1998 IEEE International Conference on. pp. 69–73.
Smith, B., Scherer, W., Hauser, T., Park, B., 2002. Data-driven methodology for signal timing plan development a computational approach. Comput.-Aided Civ. Infrastruct. Eng. 17 (6), 387–395.
Teklu, F., Sumalee, A., Watling, D., 2007. A genetic algorithm approach for optimizing traffic control signals considering routing. Comput.-Aided Civ. Infrastruct. Eng. 22 (1), 31–43.
Vasko, K., Toivonen, H., 2002. Estimating the number of segments in time series data using permutation tests. In: Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. pp. 466–473.
Wang, D., Vogt, R., Mason, M., Sridharan, S., 2008. Automatic audio segmentation using the generalized likelihood ratio. In: Signal Processing and Communication Systems, 2008. ICSPCS 2008. 2nd International Conference on. pp. 1–5.
Wong, Y., Woon, W., 2008. An iterative approach to enhanced traffic signal optimization. Expert Syst. Appl. 34 (4), 2885–2890.
Xia, J., Chen, M., 2007. A nested clustering technique for freeway operating condition classification. Comput.-Aided Civ. Infrastruct. Eng. 22 (6), 430–437.
Yin, Y., 2008. Robust optimal traffic signal timing. Transp. Res. Part B 42 (10), 911–924.