Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow

Loukas Dimitriou\textsuperscript{a}, Theodore Tsekeris\textsuperscript{a,b}, Antony Stathopoulos\textsuperscript{a,\ast}

\textsuperscript{a} Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, 5 Iroon Polytechniou, 157 73 Athens, Greece
\textsuperscript{b} Centre for Planning and Economic Research, Athens, Greece

Received 6 November 2006; received in revised form 13 November 2007; accepted 14 November 2007

Abstract

This paper presents an adaptive hybrid fuzzy rule-based system (FRBS) approach for the modeling and short-term forecasting of traffic flow in urban arterial networks. Such an approach possesses the advantage of suitably addressing data imprecision and uncertainty, and it enables the incorporation of expert’s knowledge on local traffic conditions within the model structure. The model employs univariate and multivariate data structures and uses a Genetic Algorithm for the offline and online tuning of the FRBS membership functions according to the prevailing traffic conditions. The results obtained from the online application of the proposed FRBS are found to overperform those of the offline application and conventional statistical techniques, when modeling both univariate and multivariate traffic data corresponding to a real signalized urban arterial corridor. © 2007 Elsevier Ltd. All rights reserved.

Keywords: Traffic flow modeling; Short-term forecasting; Fuzzy rule-based systems; Global optimization; Urban networks

1. Introduction

Accurate forecasting of traffic conditions in the short run, i.e. within the next few minutes, constitutes a crucial element for the efficient deployment of advanced traffic management and information systems (ATMIS) to prevent or mitigate congestion problems in metropolitan areas. Short-term traffic forecasts can be used to describe future traffic state variables, such as traffic flows, support the operation of online adaptive and transit priority signal control and route guidance systems along urban arterial streets. Several methods have been developed and implemented for the short-term prediction of traffic flow, mostly in freeways and, at a smaller extent, in urban arterial networks.
Examples of such methods include the Kalman filtering (Kwon and Stephanedes, 1994) and extended versions of it (Gazis and Liu, 2003), the auto-regressive integrated moving average (ARIMA) model (Hamed et al., 1995) and various extensions of it, such as combination with Kohonen maps (KARIMA) (Van der Voo r t et al., 1996), the subset ARIMA model (Lee and Fambro, 1999) and the multivariate ARIMA or ARIMAX (Williams, 2001). Other studies have focussed on statistical pattern recognition techniques (Davis et al., 1990), several forms of non-parametric regression models (Smith and Demetsky, 1997; Smith et al., 2002; Sun et al., 2003), state-space methods (Stathopoulos and Karlaftis, 2003) and Bayesian network approaches (Sun et al., 2006).

Most of these statistical methods are principally based on some strong assumptions, such as those concerning stationarity and ‘known’ statistical distributions of traffic variables, e.g., the (multivariate) normal distribution, which are practically irrelevant to the operation of realistic urban networks (Stathopoulos and Karlaftis, 2001). In addition, such methods are not capable to adequately respond to problems of imprecision and uncertainty pertaining to complex and high-frequency datasets, particularly those related to abrupt unexpected events and the changing dynamics of urban traffic conditions (Smith et al., 2002; Vlahogianni et al., 2004).

Nevertheless, with the exception of the widely applied artificial neural network (ANN) models (Dougherty, 1995), computational techniques from the field of artificial intelligence (AI) have received a limited consideration as alternatives for the modeling and prediction of urban traffic flow. Such techniques can be regarded as more flexible, robust and appropriate for addressing issues of uncertainty and imprecision, in comparison to the conventional statistical approaches mentioned above. Moreover, it has become evident in the existing literature that a skillful combination of different AI techniques and other meta-heuristics or optimization methods from operations research can lead to a more efficient behavior and greater flexibility when dealing with real-world and large-scale problems.

Despite that there are not many implementations of hybrid AI techniques in the field of road traffic forecasting, the existing ones can be considered as promising. Relevant examples are nonetheless restricted to hybrid neural network models, including combined Genetic Algorithm (GA) and ANN models (Lingras and Mountford, 2001; Vlahogianni et al., 2005) and combined fuzzy logic and ANN models (Yin et al., 2002; Ishak and Aleksandru, 2003; Quek et al., 2006).

Furthermore, new interest in hybrid AI methods arises from the use of hybridized fuzzy logic-based models. Fuzzy logic provides a natural framework to include expert (interpretable) knowledge in modeling nonlinear stochastic processes and complex systems (Zimmermann, 1996). A fuzzy rule-based system (FRBS) is a rule-based system where fuzzy logic is used as a tool for representing different forms of knowledge about the problem at hand and modeling the relationships that exist between its variables. FRBS are able to combine both linguistic and numerical information in a seamless way, so that provide a flexible and robust description of processes with varying complexity. In contrast to other approximate methods, such as neural networks, that store knowledge implicitly (‘black-box’ approach), the knowledge in a FRBS remains explicit as the system behavior is described by local rules (Cordon et al., 2001). The specific type of FRBS that is used in the present study is described in Section 2.1.

Nonetheless, the rule base of fuzzy systems may be difficult to be determined. This is because experience or expert’s knowledge rules may not always involve practical rules-of-thumb concerning the appropriate selection of those values of system parameters that can provide optimized solutions. Stochastic global search (meta-heuristic) algorithms, such as Genetic Algorithms (GAs), can offer the ability to explore a large search space for desirable solutions only requiring a simple scalar performance measure. Thus, they can suitably calibrate the internal structure of fuzzy membership functions, which assign crisp values to fuzzy sets, and help obtain such a set of parameter values that describe the desired behavior of the system, in terms of an appropriate optimization criterion.

This paper presents the development and implementation of a hybrid, meta-optimized FRBS, which is a fuzzy rule-based system, augmented with meta-heuristic optimization techniques, for the automatic tuning of its parameters. More specifically, the process used here employs a GA to tune different components of the fuzzy system through an adaptive rolling horizon framework. The proposed system involves a multivariate structure, in terms of employing time and space-lagged input data along a specific urban arterial corridor. Such a multivariate formulation is capable to address complex cases of data, such as those corresponding...
to congested urban traffic flow, characterized by increased fluctuation and transition regimes. Besides, it can successfully absorb the effect of missing or incorrect data and it involves an adaptable internal form that periodically adjusts with the way traffic evolves.

As far as the organization of the paper is concerned, Section 2 presents the FRBS and the optimization method used for online tuning the system parameters according to the prevailing traffic conditions. Section 3 provides information about the sources and usage of the study data. Section 4 describes the preliminary investigation of the FRBS, including the design of different system components. Section 5 analyzes the computational performance of the model, including its comparison with other, conventional statistical models used in the literature, and discusses several practical issues involved in its deployment. Section 6 summarizes and concludes the contribution of the present study.

2. The structure of the fuzzy rule-based system

2.1. General description of the FRBS

Fuzzy sets theory, introduced by Zadeh (1965), has provided an additional framework for the robust treatment of the uncertainty pertaining to various transportation systems and the interpretation of expert knowledge in their structure. Related applications range from strategic and operational transportation planning to the design and evaluation of traffic control strategies. Various models have been blossomed in the field of transportation adopting fuzzy sets theory, especially since early 1990s. The majority of those models refers to the development of alternative traffic control strategies, as they were first applied by Pappis and Mamdani (1977) to the signal control of a single junction and, then, by Chen et al. (1990) to the freeway traffic control. Other related applications include modeling of the route choice process (Lotan and Koutsopoulos, 1993) and other discrete choice behavior problems (Vythoulkas and Koutsopoulos, 2003), dynamic traffic assignment models (Liu et al., 2003), and various estimation and optimization problems involving variables with intrinsic uncertainty (see Teodorovic and Vukadinovic, 1998).

The key concept of fuzzy systems is that the variables used in them are expressed through a multi-valued logic to capture the uncertainty and vagueness of the different system states. The interactions between different fuzzy variables result in the state of the system. The interpretation of the knowledge on the interactions among variables is expressed through a set of IF–THEN rules. Fuzzy systems operating with these rules are referred to as fuzzy rule-based systems (FRBS) and they can be distinguished into two main types: (a) the Mamdani FRBS and (b) the Takagi–Sugeno–Kang FRBS (see Cordon et al., 2001). The present study employs a Mamdani-type FRBS, which was first introduced in the studies of Mamdani (1974) and Mamdani and Assilian (1975). The structure of such a fuzzy system is based on three stages, i.e., the fuzzification of the inputs, the inference system and the defuzzification of the outputs.

The FRBS performs a mapping between the inputs and outputs by taking into account the knowledge on the relationships among system variables, which are expressed in a fuzzy manner. In the first stage (fuzzification), the states of the system are expressed within the multi-valued logic. In contrast with the crisp representation of a value, the fuzzy reasoning describes values through fuzzy sets. Each fuzzy set $A$ represents the possibility (degree) of a value to be member of the set in 0–1 scale:

$$ A = \{ (x, \mu^A_x(x)) | x \in X \} $$

where $\mu^A_x(x)$ denotes the degree of membership of value $x$ to the fuzzy set $A$ and $X$ is the domain of $x$. The function that assigns crisp values to fuzzy sets is called membership function (MF). Fig. 1 illustrates a characteristic example of a MF for the representation of traffic flow through employing three different states, i.e., ‘Low’, ‘Medium’ and ‘High’, to describe the variable ‘Volume’.

The selection of the MFs (number and shape) is made according to the particular features of the problem. The second stage (inference system) is composed of two components: (a) the formation of the fuzzy rules and (b) the execution of the fuzzy operations. The set of fuzzy rules incorporates the expert knowledge pertaining to the system and constitutes the rule base (RB) or knowledge base (KB). The complete rule base connects possible input states to output states of the system variables. After all rules have been implemented, the aggregation of information leads to a fuzzy set (number) that characterizes the system output for the specific
conditions. The present study adopts the centre of gravity (CoG) defuzzification algorithm to produce a crisp output from this fuzzy number.

The benefits of system modeling and control using the above described Mamdani’s FRBS can be summarized into the following points:

- FRBS are able to incorporate all forms of information available about the nature of the system. Namely, qualitative expert knowledge information on the interactions of the system variables and states, and quantitative knowledge on the system variables (measurements).
- The model-free approach makes FRBS more tractable than classical control schemes, since the mathematical modeling of complex, large-scale systems is a demanding task.
- Under certain conditions, FRBS are universal approximators since they have the ability to approximate any function to the desired degree (Kosko, 1992).
- FRBS provide a robust framework for handling the uncertainty observed when controlling real-world systems (Cordon et al., 2001).

In the current context, the use of FRBS is appropriate for capturing the spatio-temporal structure of a series of measurements, such as those of urban traffic flow. Furthermore, due to the well-proven ability of fuzzy sets to represent human perceptions, the FRBS can enhance the understanding and short-term modeling of traffic flow through introducing expert knowledge about the uncertainty incorporated in the traffic data, which is associated with complex interactions between drivers’ behavior and road supply conditions under different levels of demand and service. Despite the advantages of employing FRBS in the specific problem, it is also noted that their usage necessitates expert judgement and knowledge of the particular local traffic conditions, in order to ensure that the setup of fuzzy rules and MFs will be plausible and computationally efficient, particularly for real-time application purposes as the present one. The next subsection presents a suitable process for the online calibration of the MFs of the FRBS.

2.2. Description of the FRBS tuning process and the MFs

FRBS can be augmented with suitable calibration techniques to improve the performance of the model. The calibration mechanism, typically known as tuning process, is used to complement expert’s decisions on both the fuzzification and rule base formation in the FRBS. The tuning process employed here, as it is
illustrated in Fig. 2, combines an optimization method, i.e. a Genetic Algorithm which is analytically described in Section 2.3, with the FRBS so that improve the value of some performance indicator.

In particular, the shape parameters of the MFs are calibrated to optimize the mapping from inputs to outputs, in terms of minimizing the measure of the mean absolute relative error (MARE) between the estimated (simulated) and the observed (actual) values at each time interval over a given study period. The process of tuning the FRBS membership functions can be generally expressed through the following minimization problem:

\[
\min_{\theta} \text{MARE} = \sum_{t} \frac{|F(Y_t, \theta) - y_t|}{y_t}
\]

subject to \(G(\theta) \leq Q\) (3)

where \(\theta\) represents the control variables (MF parameters), \(Y_t\) are the input variables at time \(t\), \(y_t\) are the observations at time \(t\), \(F(Y_t, \theta)\) are the estimates of the FRBS based on \(Y_t\) and \(\theta\), and \(G(\theta) \leq Q\) is the set of constraints concerning the control variables \(\theta\).

Several alternative types of MFs can be adopted, including triangular, trapezoidal or bell-shape. Amongst them, the triangular-shape function can be considered as providing a simpler representation of traffic flow, which eases the understanding and facilitates the clear distinction between different traffic states, during both the specification and calibration stages of the model. Fig. 3a shows a coding scheme with the parameters \(a^4_A, b^4_A\) and \(c^4_A\) to be estimated for tuning a triangular MF for the state \(S\) of variable \(A\), with \([w^4_{A,i}, w^4_{A,j}]\) being the range of the MF wherein the state \(S\) of variable \(A\) is defined, between the lower (left-hand-side) and upper (right-hand-side) boundaries \(w^4_{A,i}\) and \(w^4_{A,j}\), respectively. There are several alternative coding schemes for representing triangular MFs, such as the scheme shown in Fig. 3b. Nevertheless, the first approach shown in Fig. 3a is preferred here due to its ability to represent and handle more conveniently the given problem constraints in the model calibration process (see Sections 2.3 and 4.1).

---

**Fig. 2. Flow chart of FRBS tuning process.**

Please cite this article in press as: Dimitriou, L. et al., Adaptive hybrid fuzzy rule-based system approach ..., Transport. Res. Part C (2007), doi:10.1016/j.trc.2007.11.003
Due to the high complexity involved in the functional forms of FRBS (nonlinear and conditionally expressed functions), the adoption of classical, derivative-based methods of optimization cannot be considered as appropriate. For this reason, the present study employs an evolutionary computing approach (see Section 2.3) to calibrate the MFs of the FRBS, although other methods employing different stochastic search mechanisms, such as, Simulated Annealing and Tabu Search, could also be considered for this purpose (see Teodorovic and Vukadinovic, 1998).

2.3. The GA used to calibrate the FRBS

Evolutionary computing encompasses a set of nature-inspired AI search and optimization methods. These methods are based on an initial population of feasible candidate solutions and the use of stochastic mechanisms for exploring the search space and obtaining a solution trajectory which gradually approaches optimality. Among them, the most well-known are Genetic Algorithms (GAs), which rely on the mechanics of natural selection from a population of solutions and the Darwinian survival of the fittest (see Holland, 1975; Goldberg, 1989; Mitchell, 1998). The use of GAs has become particularly attractive for the gradient-free, stochastic global optimization of highly complex, unconstrained and constrained, problems. The following paragraphs provide a description of the basic features of the GA to be used in the current application, as they are shown in the corresponding box in Fig. 2.

The search mechanism of the GA relies on a stochastic process, which utilizes information about the performance of the system to be optimized in relation to values of different control parameters. In this process, an initial population of candidate solutions is tested and, subsequently, a new, genetically improved population is produced which contains candidates with a higher probability to reach the optimal solution. Each of the candidate solutions consists of a sequence of values of the control variables, and it takes the form of a string (chromosome), i.e., a binary coded value forming a string of 0s and 1s or real numbers in the decimal numbering system, called alleles. The GA stochastic process is composed of three genetic operators, as they are shown in Fig. 4, which refer to the reproduction, crossover and mutation.

![Fig. 3. Tuning variables of a triangular MF using different coding schemes.](image)

![Fig. 4. The operators of the present GA stochastic process.](image)
Reproduction: The convergence of GAs is based on the concept that chromosomes with higher performance are more likely to be close to optimum values. In this sense, the production of the new population should rely more on the most ‘successful’ individuals. Hence, this operator performs the reproduction of an intermediate population, referred to as parent population, from which a new, genetically improved population will be produced. There are various methods for the reproduction of the individuals of the parent population, such as the roulette wheel and tournament selection. In this study, the tournament selection method is adopted. In this method, the parent population is formed by choosing the most ‘powerful’ amongst a number of randomly chosen individuals from the old population. In any of the above methods, the Darwinian principle of survival of the fittest individuals is attained.

Crossover: After selecting the parent population, the exchange of genetic material (crossover) among the member individuals is the mechanism that leads to the production of a new improved population. The crossover is performed by randomly mating the individuals and exchanging parts of their chromosomes, according to a pre-specified rate (probability of two mates to crossover) and pattern. It has been proved that, under certain conditions, the crossover operation can lead to a global optimum solution (Goldberg, 1989). The crossover rate is usually chosen to be rather large (>50%). The present study uses a crossover rate equal to 80%. Finally, there are many patterns for crossover, such as the single-point crossover, n-point crossover and scattered crossover. The current application uses a scattered crossover pattern, wherein randomly selected parts of each chromosome are exchanged (see Fig. 4) to allow the transmission of genetic information among individuals.

Mutation: This operation provides a mechanism for preventing local convergence through randomly altering some allelic values according to a pre-specified (typically small, such as <5%) rate. A mutation rate equal to 1% is used in this study.

The coding scheme of the current GA is based on the representation of the control variables in the binary numbering system. In particular, a binary number is assigned to each control variable, i.e., \( a_s^A \), \( b_s^A \) and \( c_s^A \) parameters of MFs. After the decoding into decimal numbers, a manipulation of the values is performed in order to scale them and make them comply with the problem constraints, as it is described by the following inequalities:

\[
\begin{align*}
  w_{ij}^A & \leq a_s^A \\
  a_s^A + b_s^A + c_s^A & \leq w_{ij}^A \\
  a_s^A, b_s^A, c_s^A & \geq 0
\end{align*}
\]

A crucial factor of the performance of GAs is the population size. Large populations can increase the exploratory power of the search space, while they decrease the convergence speed, since they increase the number of evaluations of the objective function. There is always a trade-off between the number of individuals in the population and convergence speed. In this study, a population of 200 individuals is used in the offline application of the model. In the online application, where convergence speed matters, a smaller population size is used, that is equal to 50 individuals. The smaller population size used by the online model, in comparison to the population size used by the offline model (see Section 4.2), can be explained by the fewer information resources (training sample size) needed for fitting the former model than those required for fitting the latter model.

3. Description and analysis of the study data

An adaptive mechanism for representing and predicting traffic flow should ideally enable the incorporation of all available information obtained from various measurement sources in the structure of the model. Nonetheless, even in the case of networks simpler than those of urban arterials, such as freeways, the dimensionality of the problem can considerably increase with the number of measurement sections. Thus, in the current context of urban traffic operations, a procedure is followed in order to: (i) select a limited part of the network (a number of arterial sections) for demonstration analysis purposes, and (ii) determine the number of measurement sections, whose observed traffic flows are to be used as input in the modeling process.
The first task is carried out here through the simultaneous modeling of the traffic flow time series originated from 140 key locations equipped with loop detectors around the urban road network of the Greater Athens Area (GAA) in Greece. These data consist of traffic flow measurements collected at the end of every 90-s signalization cycle and aggregated at intervals of 3-min duration corresponding to the last 3 days of May 2000. Based on Tsekeris and Stathopoulos (2006), the method of principal component analysis (PCA) is adopted to identify common underlying sources of temporal variability in traffic flow, referred to as eigenflows, over the whole network. These eigenflows are then used to reconstruct total traffic variability within a lower-dimensional structure that preserves its important properties. The results of applying the PCA in the given network help identifying groups (clusters) of different locations across specific urban arterial corridors wherein the observed traffic flow has a large impact on the aggregate network traffic conditions during the period of the study. In addition, the estimation of the principal components of traffic variability facilitates the detection of those locations which mostly affect, i.e. are positively or negatively highly correlated with the other measured traffic flow time series in a particular corridor.

The selected study data consist of time series of traffic flow (smoothed volume) measurements obtained from loop detectors placed in tandem along the North–South direction of a major urban signalized three-lane arterial, i.e. Alexandras Avenue, in Athens, Greece. Based on the results of the PCA for the given network (for details, see Tsekeris and Stathopoulos (2006)), traffic data captured by loop detector L101 (see Fig. 5) were considered as a suitable candidate for our prediction analysis experiment. This is because the flow time series measured at that point were found to exhibit significant fluctuations and carry the largest impact on the flows traversing neighbouring locations during most parts of the day. These fluctuations may be attributed not only to the traffic conditions occurring at the specific location during the current and past time intervals, but also to those conditions prevailing at locations upstream of the study location. Namely, the modeled process may exhibit systematic dependence on the observations made at neighbouring locations.

The selection of the number of measurement sections in the specific corridor and, hence, the input variables to be considered in the analysis determine the appropriate model data structure. This structure relies on the spatial and temporal evolution of the traffic flow in the arterial corridor under study. Although standard analytical approaches, such as partial and auto-correlation analysis, analysis of variances (ANOVA) and PCA, are appropriate for guiding the modeling process to include a particular set of measurement sections, they

Fig. 5. Coded representation of the study area and the loop detector layout.
may lead to erroneous inferences with regard to specifying the model data structure, as they are primarily descriptive and phenomenological in nature. Such approaches are based on the statistical identification of common temporal and/or spatial correlation patterns among aggregate traffic flows measured at different sections in tandem along a linear freeway or arterial corridor, at specific time-of-day or day-of-week periods (e.g., see Stathopoulos and Karlaftis, 2003).

On the contrary, the use of a micro-simulation procedure for the detailed analysis of traffic conditions can provide a theoretically sound and interpretive approach for determining the causal temporal and spatial relationships among aggregate flows measured along a signalized urban arterial corridor. The micro-simulation procedure used here refers to the traffic software integrated system (TSIS) platform (Federal Highway Administration, 2001). The calibration of the micro-simulation model is based on real data concerning the measured traffic flows at the count locations of the study arterial.

The specific approach can identify the input variables (flows at the study section and upstream sections) by estimating the distribution of travel times of individual vehicles from each upstream measurement location to the study section, as a function of different traffic demand levels, as they are generated within the TSIS framework. The current procedure takes into account different traffic flow regimes at the operational level of each vehicle over different periods of the day and incorporates information on actual traffic signal control plans in the given arterial corridor. Sub-area coding and the loop detector layout are shown in Fig. 5.

The micro-simulation results demonstrated a significant dependence of traffic conditions at loop location L101 on the conditions of the immediately upstream (about 1200 m away) loop location L103, with an average travel time between the two locations equal to 166 s. The travel times estimated under different demand levels are mostly less than 180 s, although inbound and outbound flows from the intervening intersections (access points) between the two locations render the dependence pattern highly stochastic on a cycle-by-cycle basis. Hence, given that the prediction mechanism is used here to provide short-term (3-min-ahead) forecasts, traffic flows originated before the immediately previous time interval and from locations upstream of section L103 are excluded from the set of input variables.

Based on these results, two models with different data structures are used in the short-term traffic flow prediction process. On the one hand, the univariate model, which provides future flow $F_{t+1}$ at the section of the examined loop (L101) based on a single input, namely, the current information about flow $F_t$ on that location. On the other hand, the multivariate model, which provides future flow $F_{t+1}$ based on two inputs, namely, the current state of flow at the examined loop L101 section and the current state of flow at the loop L103 section upstream. The system output refers to the one-step-ahead predicted flow $F_{t+1}$ at measurement location L101. The available real-time data consist of 480 3-min traffic flow measurements for each of the two locations (i.e. L101 and L103) and each of the three days of the study period, yielding a total of $2 \times 3 \times 480 = 2880$ measurements. The following section describes the empirical model building and calibration processes.

4. Design of the FRBS components

4.1. Interpretation of expert’s knowledge

In the current application, a parsimonious and comprehensible representation of the system structure is adopted to investigate the model performance. Specifically, each variable value is represented by three fuzzy sets, which correspond to three different states of the traffic flow: Low, Moderate and Heavy. Each fuzzy set is expressed in terms of three values: $a^s_m, b^s_m$ and $c^s_m$ (see Fig. 3a). Several constraints are added to ensure that each fuzzy set maintains its validity through the calibration process, on the basis of historical observations of the local traffic conditions and the road capacity characteristics of the study arterial sections. First, the maximum observed traffic volume on both measurement sections is considered as equal to 160 vehicles per 3-min interval (veh/interval). However, the maximum allowed traffic flow is defined as 200 veh/interval in order to allow the model accounts for extreme traffic conditions. Namely,

$$a^s_m, b^s_m, c^s_m \geq 0 \quad \forall s \in S, m \in M, \text{ and}$$

$$a^s_m + b^s_m + c^s_m \leq 200 \quad \forall s \in S, m \in M$$

Please cite this article in press as: Dimitriou, L. et al., Adaptive hybrid fuzzy rule-based system approach ..., Transport. Res. Part C (2007), doi:10.1016/j.trc.2007.11.003
where \( m \) denotes the fuzzy variable in the universe of discourse \( M \) and \( s \in S \) is the alternative state represented by the MF with parameters \( a^s_m, b^s_m \) and \( c^s_m \) using the fuzzy variable \( m \). The range of the values of each MF is defined as: (a) \([0, 100]\) for the Low state, (b) \([50, 150]\) for the Moderate state and (c) \([100, 200]\) for the Heavy state, using the following relationships:

\[
\begin{align*}
0 & \leq a^m_s + b^m_s + c^m_s \leq 100 \quad \forall s \in S_{\text{low}}, \ m \in M \\
50 & \geq a^m_s \quad \forall s \in S_{\text{moderate}}, \ m \in M \\
150 & \leq a^m_s + b^m_s + c^m_s \leq 150 \quad \forall s \in S_{\text{moderate}}, \ m \in M \\
100 & \geq a^m_s \quad \forall s \in S_{\text{heavy}}, \ m \in M \\
200 & \leq a^m_s + b^m_s + c^m_s \leq 200 \quad \forall s \in S_{\text{heavy}}, \ m \in M
\end{align*}
\]

Fig. 6 illustrates the normalized fuzzy partition process used here to facilitate the initial setting of the fuzzy sets. Based on this partition scheme, a gradual transition from one state to another is achieved. The RB is constructed under the knowledge that the spatio-temporal propagation of traffic flow takes place smoothly and no sudden shock occurs. Hence, the RB for the univariate model comprises of 3 rules, as they are described in Table 1. For the multivariate model, the RB takes additionally into account the knowledge that traffic forecasts will be more dependent on the current state of the flow at the examined loop section (L101), while the information provided from the upstream loop section (L103) is used to add spatial correlation effects in the model. In particular, the RB of the multivariate model comprises of nine rules, which are analytically presented in Table 2.

### 4.2. The tuning of the membership function parameters

In the case of the univariate FRBS model, the aim of the process is to tune the shape of three MFs that represent one input and of three MFs that represent one output. Provided that each MF is expressed in terms
of three values (control variables), i.e. $a^3_1$, $b^3_1$ and $c^3_1$, the process aims at estimating 18 values (6 MFs × 3 values) of MFs that minimize the MARE (three values for each of the three MFs of the input variable and three values for each of the three MFs of the output variable). In the case of the multivariate model, the calibration process involves two inputs and one output, which leads to the estimation of 27 values (9 MFs × 3 values) of MFs that minimize the MARE.

An appropriate training dataset (see below) is selected and used for the calibration of all models before using them for the out-of-sample prediction of traffic flow. Except of the univariate and the multivariate models, the FRBS is also distinguished here into the offline and the online prediction models. The offline model estimates the FRBS parameter values used in the prediction process based on the calibration procedure performed once using an adequately sized training dataset. Following the practical rule of thumb (2/3 of the data used for training – 1/3 for testing), the training dataset used to calibrate the offline model corresponds to traffic flow measurements collected on May 29 (Monday) and 30 (Tuesday), including $2 \times 480 = 960$ intervals, while the out-of-sample prediction is performed for the third day of May 31 (Wednesday), i.e. for 480 intervals.

On the contrary, the online model periodically updates (adapts) these parameter values by processing real-time traffic flow information. The calibration process of the online prediction model uses a regularly updated training dataset corresponding to the traffic flow measurements of the past 60 intervals prior to the prediction interval. The adoption of such a rolling horizon (from $F_{t-60}$ to $F_t$) that covers a period of 3 h was empirically considered here as appropriate for the one-step-ahead prediction ($F_{t+1}$) of traffic flow at the given arterial corridor. For comparison purposes, all models were tested for the period spanning the whole day of May 31, i.e. for a set of 480 intervals.

The stopping criterion of the calibration process of both the offline and online models is defined as the convergence of the GA when no significant improvement (<0.001 of the MARE) in the value of the objective function is achieved after a series of 50 generations. In the case of the offline multivariate model, the convergence of the GA (see Fig. 7a) results in a MARE $\approx 0.18$. In the case of the online multivariate model, a fast convergence of the GA (see Fig. 7b) is observed within the first 30 s of running time, resulting in a significantly smaller MARE = 0.02, in relation to the offline model. Thus, the selection of such a rolling horizon duration as that of 3 h for the online updating of the MF parameters does not practically affect the real-time convergence process of the GA, while it can help FRBS being adequately responsive to the evolution of prevailing traffic conditions.

The reduction of the MARE for the online model can be attributed to the smaller size of the training dataset used in the calibration process, in comparison to the offline model. The calibration process of the univariate models (not shown in the figures) is also found to rapidly converge at approximately similar levels of the MARE. Fig. 8 illustrates the space–time evolution (spread) of the three states of traffic flow, as they are obtained from the results of the calibration of the membership functions corresponding to the
two inputs and the one output (forecast) and indicated with the different colours, using the online multivariate model for a typical, in terms of the morning traffic level, period of the day at the given arterial sections.

Fig. 7. Convergence of the calibration process based on the Genetic Algorithm for (a) the offline multivariate model using the training dataset and (b) the online multivariate model for a typical period of the day.

Fig. 8. Space–time representation of the results of the calibrated membership functions corresponding to the two inputs, $L_{101}(t)$ and $L_{103}(t)$, and the output forecast, $L_{101}(t + 1)$, of the online multivariate model for a typical period of the day.

two inputs and the one output (forecast) and indicated with the different colours, using the online multivariate model for a typical, in terms of the morning traffic level, period of the day at the given arterial sections.
5. Results and discussion

5.1. Description of the ARIMA and Kalman filter models

In addition to the univariate and the multivariate offline and online FRBS, two conventional, commonly used statistical approaches are also implemented and evaluated here, for comparison purposes, for the one-step-ahead prediction of traffic flow. The first approach refers to an ARIMA model. In order to implement this model, the time series are first-order differenced to obtain stationarity at the 99% significance level based on the Augmented Dickey Fuller (ADF) test. Then, traffic flow \( y_t \) at time \( t \) is estimated by using the following form of an ARIMA \((p,d,q)\) model:

\[
\phi(L)\Delta^d y_t = \theta(L)e_t
\]

where \( e_t \sim \text{i.i.d. } N(0,\sigma^2) \), \( \Delta = y_t - y_{t-1} \), \( d \) denotes the order of integration, \( L \) is the lag operator such that \( L^p y_t = y_{t-p} \), and

\[
\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p
\]

\[
\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q
\]

The ARIMA vector of model parameters \( \theta = (\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2) \) is estimated by using past observations of traffic flow \( y_1, y_2, \ldots, y_t \) and, then, maximizing the log-likelihood function \( L(\theta) = \log f_{y_t, y_{t-1}, \ldots, y_1} \). After following the standard procedure for model selection by examining the auto-correlation and partial auto-correlation functions (ACF and PACF, respectively), and testing the performance of various configurations from the ARIMA family of models, the ARIMA \((1,1,1)\) model without constant term and \( \phi_1 = 0.0468, \theta_1 = 0.7260 \) (statistically significant at the conventional levels of confidence) is selected in order to produce the out-of-sample predictions. The dataset used for both model selection and training consists of traffic flow measurements corresponding to the two previous days of the study period, as in the case of calibrating the offline FRBS.

The second statistical approach concerns an adaptive version of the autoregressive moving average (ARMA) model, which is expressed in a state-space form that can be recursively estimated with a Kalman filter (KF). The KF follows an online adaptive calibration process using updated traffic flow information from the previous time interval. Although the KF also relies on normality assumptions regarding the nature of the prediction errors, it can be still considered as an effective predictor. Namely, even in cases where the normality assumption is dropped, KF is an optimal estimator within the class of all linear estimators, in the sense that it minimizes the mean squared error (MSE) (Harvey, 1989).

In the current application, the prediction mechanism of the traffic flow series \( y_t \) at some time point \( t \) is expressed within a state-space form and a time-invariant ARMA \((1,1)\) process without a constant term is used, as follows:

\[
y_t = x_t + \epsilon_t
\]

\[
a_{t+1} = \phi x_t + \theta \epsilon_t, \quad t = 1, \ldots, n
\]

where \( \phi \) and \( \theta \) are the model parameters, \( \epsilon_t \sim (0, \sigma^2 I) \) in Eqs. (14) and (15) denotes the same white noise process, with \( I \) being the unitary matrix, and \( x_t \) is the state variable, with \( a_1 = y_{t-1} \) and an initial value \( a_1 = 0 \). When new observations \( y_t \) become available, the KF recursively estimates the system state in order to produce optimal forecasts for \( y_{t+1} \). Similar expressions of various models from the ARMA family in state space form can be found in Harvey (1989) and Hamilton (1994). Finally, the parameters vector \( \theta = (\phi, \theta, \sigma^2) \) of the model has been estimated offline by maximizing the log-likelihood function \( L(\theta) = \log f_{y_t, y_{t-1}, \ldots, y_1} \), given a training set \( y_1, y_2, \ldots, y_t \), which consists of traffic flow measurements corresponding to the two previous days of the study period, as in the case of calibrating the ARIMA model and the offline FRBS.

5.2. Analysis of results and comparison of the models

Fig. 9a and b illustrates the time series graphs of the observations and predictions resulted from the Kalman filter and the ARIMA model. Fig. 10a and b illustrates the corresponding time series graphs obtained from the

offline and the online FRBS when using a univariate data structure. Correspondingly, Fig. 11a and b illustrates the time series graphs obtained from the offline and the online FRBS when using a multivariate data structure. Each graph also includes a zoom-in window to allow a closer view of the temporal pattern evolution of the two series.

The measures of the mean absolute relative error (MARE), the mean squared relative error (MSRE) and the $R^2$ goodness-of-fit statistic between the actual and the predicted flow series are considered as appropriate and used here for measuring the performance of the short-term traffic flow prediction models. Table 3 indicates the increased accuracy of the online FRBS, in comparison to the other models, in terms of the performance measures of the MARE, MSRE and $R^2$. The prediction accuracy gain resulted from the use of the FRBS, especially that with the online multivariate setting, in comparison to the use of the other models, indicates that the study traffic data encompass complex patterns in their evolution that cannot be adequately modeled by the conventional statistical approaches, such as the Kalman filter and ARIMA models.

As it is shown from the graphical analysis of the time series patterns, the FRBS, particularly the online version of it, captures the general trend of the traffic flow time series and timely reflects a portion of the

**Fig. 9.** Observations vs. predictions resulted from (a) the Kalman filter and (b) the ARIMA model.
upward and downward shifts in this trend. Both the Kalman filter and the ARIMA model (see Fig. 9) provide a more responsive, in terms of the predicted magnitude of the shifts, but time-delayed representation of the fluctuation pattern, resulting in higher prediction errors, in terms of all the performance measures, in comparison to the FRBS (see Table 3). This accuracy loss of the ARIMA and Kalman filtering models can be attributed to the fact that the unpredictable part of their standard specification forms, as they are described in Section 5.1, cannot adequately address real-time stochastic changes in the traffic flow generation process.

On the other hand, the fuzzy rule-based representation of traffic data in the FRBS enables the predicted flows to retain a more smoothed pattern, which is less sensitive to the high-frequency variations of traffic flow, in contrast with the pattern obtained from the standard statistical models. The forecasts produced by the online FRBS (see Fig. 11), especially that with the multivariate setting, are found to be more responsive to the actual traffic flow variations and, hence, more accurate, in comparison to the forecasts produced by the offline FRBS (see Fig. 10). This is because the multivariate model incorporates information on both the time and spatial dependence effects of the examined location, in comparison to its univariate counterpart.

Fig. 10. Observations vs. predictions resulted from (a) the offline univariate FRBS and (b) the online univariate FRBS.
In addition, the online model can better handle the stochasticity and nonlinearity of the spatio-temporal evolution of the urban traffic process, in comparison to the offline model, without being dependent on the fluctuations of traffic conditions in the study arterial corridor over different time-of-day periods. This is explained by the real-time adaptive tuning of the system parameters, according to the prevailing traffic conditions. The existing outcome signifies the importance of adopting such unified approaches as the online multivariate FRBS.

Table 3
Results of the performance measures of different prediction models

<table>
<thead>
<tr>
<th>Model</th>
<th>ARIMA</th>
<th>Kalman</th>
<th>Univariate Offline FRBS</th>
<th>Multivariate Offline FRBS</th>
<th>Univariate Online FRBS</th>
<th>Multivariate Online FRBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$ (%)</td>
<td>76.727</td>
<td>79.948</td>
<td>77.468</td>
<td>80.020</td>
<td>83.458</td>
<td>84.233</td>
</tr>
</tbody>
</table>

In addition, the online model can better handle the stochasticity and nonlinearity of the spatio-temporal evolution of the urban traffic process, in comparison to the offline model, without being dependent on the fluctuations of traffic conditions in the study arterial corridor over different time-of-day periods. This is explained by the real-time adaptive tuning of the system parameters, according to the prevailing traffic conditions. The existing outcome signifies the importance of adopting such unified approaches as the online multivariate FRBS.
FRBS for the short-term traffic flow modeling and forecasting through the online treatment of traffic data with a single model that circumvents the effect of time period.

The scatter diagram presented in Fig. 12 verifies the close systematic relationship between the observed and predicted flows based on the use of the online multivariate FRBS model. In particular, the values of the intercept and slope of the best-fit line, which it almost coincides with the line of equivalence, are about equal to zero and unity, respectively, with $R^2 = 84.23\%$. Moreover, the diagram shows that this systematic relationship holds for both the low and increased traffic flow conditions, which correspond to the time periods from midnight to 6:30 a.m. and from 6:30 a.m. to midnight, respectively, and they are illustrated by the two distinct (lower and upper) data clusters. The small number of data points between the two clusters indicates the short duration of the transition period from low to increased traffic flow conditions.

5.3. Complexity and applicability issues

The present results demonstrate that the online multivariate FRBS can provide a more consistent framework for describing and predicting the dynamic evolution of urban traffic flow, at least in the short run, i.e. within the next 3 min, in comparison to the other modeling methodologies. The specific model can timely capture upward and downward shifts of the trend of the traffic flow time series, although the predicted flows are found to stay away from boundary conditions in cases of large rapid fluctuations of traffic conditions. This fact can be possibly attributed to the relatively wide representation of the evolution of traffic flow by using only three distinct states, i.e. those of Low, Moderate and Heavy traffic conditions, which it does not arguably allow making discernible more complex traffic regimes, such as those corresponding to transitory and boundary states. More specifically, the flows resulted from the multivariate FRBS do not exceed the 100 veh/interval (see Fig. 11), while the flows resulted from the online and offline univariate FRBS, which are generally less responsive to traffic variations, do not exceed the 95 veh/interval and 90 veh/interval, respectively (see Fig. 10).

Nonetheless, the present FRBS model primarily aims at filtering the abrupt variations of traffic flow rather than perfectly replicating them, which is rather impracticable for the case of urban signalized arterial networks. The ability of the model to satisfactorily reproduce changes in the temporal trend of traffic flow at different count locations can support traffic surveillance and control actions at more aggregate geographical levels (of paths or parts of the network), in contrast with the disaggregate levels of an individual link or intersection.

Furthermore, a more detailed representation of all possible traffic flow regimes, at a very high level of resolution, would increase the computational complexity of the FRBS. In particular, the complexity of the FRBS, in terms of the dimensions of the problem, is related to the size of two components. The first component refers...
to the size of the RB and, specifically, the number of rules that will be used in order to adequately describe the interrelationships among different states of different variables. Despite that no formal evidence exists regarding the optimal size of the RB, it is straightforward to derive that an increase in the size of the RB leads to a more comprehensive description of the system.

The second component refers to the size of the modeled system and, specifically, the number of variables used. The complexity concerning the problem dimensions in this case can be expressed as a function of the number of the modeled variables (inputs and outputs), the number of the different states of each variable that will be assigned to MFs and the number of the characteristic parameters of each MF used. The size of the problem with regard to the second component can be generally calculated as follows:

\[ O(F) = \sum_{U} \sum_{V} \left[ \Omega_{U,V}^{F} \Theta_{V}^{F} \right] \]  

(16)

where \( O(F) \) is the size of the problem concerning the calibration of the FRBS \( F \), \( U \) is the number of variables (inputs and outputs) used to model the system, i.e. the network traffic volumes in the current study, \( \Omega_{U,V}^{F} \) is the number of MFs of type \( V \) (here, all MFs are of triangular type) employed for describing different states of the \( U \) variables and \( \Theta_{V}^{F} \) is the number of the parameters of each MF of type \( V \).

Provided that the present application refers to an urban arterial corridor, the selected structure of the current FRBS can be regarded as providing a satisfactory trade-off between plausible generalization and prediction accuracy that overperforms the accuracy of the other methods. As it can be identified from Eq. (16), the size of the problem is significantly increased, in a multiplicative order, with respect to the system variables. Nevertheless, the proposed modeling and prediction framework can be well used in more general situations involving area-wide implementation, e.g., for traffic monitoring and control at the network level. In such cases, advanced decentralized computational techniques, such as those of parallel and distributed computing strategies, should be favorably employed to enable the optimal utilization of forecasts for operational purposes.

In particular, the online FRBS forecasts can be used to enhance the real-time deployment of dynamic traffic management systems (DTMS) with distributed signal control architecture and advanced traveler information systems (ATIS) along congested urban networks. Due to the ability of the FRBS to incorporate and handle expert knowledge about the uncertainty in the network traffic environment, it can provide useful insight into the interpretation and characterization of current and future road traffic conditions. More specifically, the proposed modeling framework offers the ability to manage large amounts of traffic flow data and translate them to different traffic states over different time-of-the-day periods. Furthermore, it can facilitate the visual representation of the evolving states of traffic flow (see Fig. 8) and the provision of qualitative information about the near-future traffic congestion levels (e.g., low, medium, high) through variable message sign (VMS) displays. These benefits could potentially be extended to encompass other traveler information services, such as the support of automatic incident detection and rerouting information strategies.

6. Conclusions

The present paper described the development and empirical application of an advanced hybrid AI technique, which is an adaptive hybrid fuzzy rule-based system (FRBS), for modeling and predicting the dynamic evolution of traffic flow in an urban arterial. The proposed FRBS enables the optimal tuning of the parameters involved in the membership functions through employing an appropriate meta-heuristic optimization method, i.e., a Genetic Algorithm. The proposed modeling framework encompasses a suitable combination between expert’s local knowledge and global optimization procedures. More specifically, on the one hand, traffic engineering judgement can be appropriately incorporated in the fuzzy rule base according to the knowledge and experience on local traffic conditions. On the other hand, this knowledge is consistently and timely optimized through the online tuning of the system parameters. Moreover, such a framework enables the incorporation of both the temporal and spatial characteristics of traffic flow in a unified manner, while it eliminates the effects of time period and extreme values on the forecasting performance.

The online adaptive FRBS model and, particularly, the one with a multivariate data structure, was found to be capable to result in short-term traffic forecasts of increased accuracy, as it considerably overperforms the
offline model and, especially, two other, commonly used statistical approaches, i.e., the Kalman filter and the ARIMA model. The results indicate the potential of the usage of the online FRBS as a short-term traffic forecasting tool for supporting the decentralized deployment of adaptive traffic signal control and information systems in congested urban arterials. The proposed methodology provides an efficient way to calibrate certain components of the FRBS without attatching to the procedure any topological or operational characteristics. Thus, the online FRBS can be easily calibrated and used in different network settings of diverse geometrical layouts, traffic regimes and control strategies.

References


Please cite this article in press as: Dimitriou, L. et al., Adaptive hybrid fuzzy rule-based system approach ..., Transport. Res. Part C (2007), doi:10.1016/j.trc.2007.11.003