Enhancing Predictions in Signalized Arterials with Information on Short-Term Traffic Flow Dynamics

Eleni I. Vlahogianni

Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Published online: 30 Apr 2009.

To cite this article: Eleni I. Vlahogianni (2009) Enhancing Predictions in Signalized Arterials with Information on Short-Term Traffic Flow Dynamics, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 13:2, 73-84

To link to this article: http://dx.doi.org/10.1080/15472450902858384

Please scroll down for article
Enhancing Predictions in Signalized Arterials with Information on Short-Term Traffic Flow Dynamics

ELENI I. VLAHOGIANNI
Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Short-term traffic flow predictions are an essential part of intelligent transportation systems. Previous research underlines the difficulty in systematically assessing the predictability of traffic flow near capacity or during congested conditions. In this article a neural network prediction scheme is proposed that is consistent with the pattern-based evolution of traffic flow and has the capability of exploiting past information to acquire knowledge on the traffic dynamics in order to enhance predictability. Findings indicate that pattern-based predictions are more accurate—in the traffic flow regimes considered—when compared to other local and global prediction techniques that operate under the time-series consideration. The pattern-based prediction scheme was also found to outperform the other methods tested in the knowledge of the anticipated traffic flow state in all traffic flow conditions considered.

Keywords Neural Networks; Traffic Flow; Short-term Prediction; Pattern-based Prediction

Short-term traffic flow predictions should be well-timed, accurate, and consistent with the manner traffic flow evolves over time. In general, prediction of traffic flow is accomplished by two distinct approaches: the model-based and the statistical or data-driven approaches. In model-based approaches assumptions are made regarding the existence of sufficient a priori information in the form of physical rules (for instance the conservation law), while data-driven approaches attempt to analyze a sequence of observations produced by the underlying mechanism directly; from the statistics or dynamics obtained from a sequence of observations the aim is to infer knowledge about the future evolution.

Data driven techniques are quite popular in traffic flow prediction because they provide the structural flexibility to account for the unpredictable effect of adaptive signalization, as well as complexity in time-series of traffic variables induced by nonrecurring events or the self-organization of traffic flow. Moreover, they have exhibited better predictive behavior than modeling approaches when applied to traffic conditions near congestion (Qiao et al., 2001; Smith et al., 2002; Smith and Oswald, 2003; Geroliminis and Skabardonis, 2006).

Although these approaches have given promising results, there are some conceptual and practical issues that fail at being addressed, such as the decreased accuracy during nonrecurring traffic flow conditions and their difficulty to adapt to the complex variable statistical behavior of traffic flow in signalized arterials (Hua and Fagri, 1993; Chang and Su, 1995; Head 1995, Ledoux, 1997; Yin et al., 1998; Stathopoulos and Karlaftis, 2003a and b; Kamarianakis et al., 2005). The present study extends past research on short-term traffic flow prediction by proposing a...
neural network prediction scheme that is consistent with the pattern-based evolution of traffic flow and has the capability of exploiting past information to acquire knowledge about traffic dynamics in order to enhance predictability.

**SHORT-TERM TRAFFIC FLOW PREDICTION: OBJECTIVES AND METHODS**

The problem of short-term traffic flow prediction using data-driven approaches decomposes into predicting one or more traffic flow variables (volume, speed, occupancy, and so on); given data in the form of a sequence of a variable \( x_n \) \( (n = 1, 2, \ldots, N) \), the next state \( x_{n+1} \) will be generated by generalizing on a mathematical rule. This implies that knowledge about the manner traffic flow evolves may be captured in the near past. These approaches to traffic flow forecasting are in general not concerned with the causality of several traffic-related phenomena—for example the effect of traffic control strategies in the flow evolution of signalized arterials—but rather with the irregularities found in the time-series and patterns of traffic variables—due to the effect of all causalities arising from traffic-related operations (signalization, incidents and so on).

A large number of efforts in short-term prediction of traffic variables ranging from ARIMA family of models, State-Space models, local linear approaches, nonparametric regression and neural networks can be traced in the literature; a comprehensive review can be found in Vlahogianni and colleagues’ 2004 study. Moreover, hybrid artificial approaches, such as mixtures of ARIMA and neural networks (Danech-Pajouh and Aron, 1991), self-organizing neural networks (Van Der Voort et al., 1996; Chen et al., 2001), fuzzy neural networks (Yin et al., 2002; Ishak and Alexandru, 2004), genetically optimized neural networks (Abdulhai et al., 2002; Vlahogianni et al., 2005, 2007a), and wavelet-based neural networks (Jiang and Adeli, 2005; Xie and Zhang, 2006) have gained interest due to their enhanced approximating capabilities and adaptable features (learning); these characteristics establish a methodological environment consistent with the observed variability in short-term traffic flow that helps to improve the average predictive accuracy. All the above form an interesting field of research, but are rather confusing regarding the objectives and methods applied to short-term traffic flow prediction.

Some remarks should be emphasized; first, the conceptual problem of short-term traffic flow prediction, as initially stated, is not a simple time series problem; findings have indicated that traffic flow exhibits discontinuities in its temporal evolution (Newell, 1992; Hall et al., 1992); therefore, the prediction method should be robust to pattern variability, rather than be based on data stream continuity.

Second, predicting the short-term traffic flow is not a simple choice between univariate and multivariate methods. It is a fact that traffic flow cannot be explicitly determined by the knowledge of a single variable for example traffic volume; the knowledge of occupancy is also critical in determining the traffic flow conditions (Daganzo, 1997). Moreover, it is also a fact that in each traffic flow regime (free-flow, synchronized flow, or congestion) each variable’s temporal evolution has a different effect in the transitional behavior of traffic flow (Kerner, 2004). Consequently, the focus should not be on developing complex algorithmic structures to account for the prediction of more than one variable, but accounting for traffic variables predictability with regards to the ability to accurately infer knowledge on the anticipated traffic conditions and transitional phenomena.

Third, the selection of the proper methodology to prediction is not a simple choice between linear or nonlinear methods, stochastic or not, and so on. In principle, traffic flow has a variable statistical behavior when studied in short-term. This complex behavior is magnified in signalized arterials; the effects of platoon dispersion and queue overflows are critical in the temporal evolution of link traffic flow (Newell, 1971). Recent findings have indicated variable statistical behavior and strongly nonlinear features near capacity, as well as a highly transitional behavior (Kerner, 2004; Vlahogianni et al., 2006). From a traffic engineering perspective, it is important to have stable prediction capabilities (in terms of accuracy) in all traffic flow conditions, especially during the formation of congestion; current intelligent traffic management systems are founded on this concept. From a methodological perspective, this behavior is difficult to be model by a single statistical prediction approach, as each model has the structural and learning propensity for representing a specific temporal statistical behavior (Kantz and Schreiber, 1997). Consequently, there is a need to determine the physics of short-term traffic flow with regards to transitions and regimes prior to developing prediction models.

The fourth remark has to do with purely methodological issues. It has been observed that neural networks have exhibited prominent predictive capabilities in short-term traffic prediction; this is also reflected to the various structures and learning algorithms tested on traffic datasets until now (for a review see Dougherty, 1995; Adeli, 2001). From this perspective, a distinction should be made between static and dynamic neural structures. Static neural networks (implying multilayer feed forward structures) are flexible nonlinear parameterized models; their parameters are calibrated through learning and adapted according to the available data. These structures are simple to train and have well-founded mathematical approximation capabilities successfully tested in various prediction problems (Principe et al., 1999). Their deficiency is that they operate as static pattern classifiers and cannot manipulate dynamic information appropriately.

This methodological inconsistency is surpassed using dynamic structures of neural networks that encompass a short-term memory mechanism that allows treating data as a temporal sequence and, thus, creating an internal nonstationary environment for prediction; however, there are certain trade-offs in using dynamic structures; their learning is complex and unstable. Moreover, they adopt a time-series approach to modeling traffic...
ENHANCING PREDICTIONS IN SIGNALIZED ARTERIALS

The integrated processes involved in the proposed methodology to traffic flow prediction.

flow and their robustness to the variable traffic flow dynamics is questionable.

PATTERN-BASED SHORT TERM TRAFFIC FLOW PREDICTION

The methodology is based on neural network structures that learn to generalize in specific temporal behavior with respect to the temporal dynamics of traffic flow patterns. In this approach, predictability of short-term traffic flow is addressed by attaining the basic pattern of traffic flow evolution. The term basic pattern describes the information a prediction model has to acquire from the past in order to predict a variable’s future state. Moreover, the predictability will also be treated separately with respect to the various traffic flow conditions determined by the volume-occupancy relationship; for this, the joint study of the temporal behavior of volume and occupancy is considered.

As can be observed in Figure 1 that demonstrates the processes integrated to the proposed prediction framework, the methodology after reconstructing the basic pattern from volume and occupancy series and identifying the statistical characteristics of their joint temporal evolution, it clusters the identified volume and occupancy patterns provided by Eq. (1). More specifically, recurrent behavior is quantified based on the closeness of volume and occupancy trajectories in the Phase-Space (Marwan et al., 2007):

\[
CR_{i,j}^{m} = (\epsilon_i - \|\vec{V}_i - \vec{O}_j\|), \vec{V}_i \in R^m, \vec{O}_j \in R^m, i, j = 1 \ldots N
\]

where \( N \) is the number of states of volume \( V \) and occupancy \( O \) and \( \epsilon_i \) is the threshold of distances \( r_{i,j} \). Eq. (2) that results to a matrix of distances known as the Cross-Recurrence Plot is the basis for the proposed analysis known as the Cross-Recurrence Quantitative Analysis (CRQA) (Zbilut et al., 1998). The cross-recurrence analysis has a demonstrated ability to detect very subtle patterns in time series without the need of a priori assumptions on nonstationarity and nonlinearity (Zbilut, 2004).

Further quantification of the recurrent behavior can lead to quantify the deterministic structure (%DET), as well as the nonlinearity \( (L_{\text{max}}) \) of a dynamical system (Marwan et al., 2007). Table 1 shows the definitions and mathematical formulation of these measures.

In brief, the concepts of determinism and nonlinearity are attributes of the pattern-based evolution of traffic flow; traffic flow, when studied in the form of series of volume and occupancy, may evolve in patterns that are recurrent (in time) and ‘close’ (high values of %DET), or recurrent but isolated (stochastic—significantly lower values of %DET). Moreover, nonlinearity

\[
\vec{V}(t) = \vec{V}_i = (V_i, V_{i+\tau}, \ldots, V_{(m-1)\tau+i})
\]

\[
\vec{O}(t) = \vec{O}_i = (O_i, O_{i+\tau}, \ldots, O_{(m-1)\tau+i})
\]

\[
\vec{V}_i = \begin{bmatrix} V_{i1} \\ V_{i2} \\ \vdots \\ V_{im} \end{bmatrix}, \vec{O}_j = \begin{bmatrix} O_{j1} \\ O_{j2} \\ \vdots \\ O_{jm} \end{bmatrix}, \tau = 1 \\
\|\vec{V}_i - \vec{O}_j\| = \sqrt{\sum_{l=1}^{m} (V_{il} - O_{jl})^2}
\]

Table 1 Crossrecurrence measures mathematical definitions.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrence %REC</td>
<td>Percent of recurrent points in the recurrence matrix [ \frac{1}{N-k} \sum_{l=1}^{N-k} I(P(l)) ]</td>
</tr>
<tr>
<td>Determinism %DET</td>
<td>Percent of recurrent points in the recurrence matrix that forms diagonal lines [ \frac{1}{N} \sum_{l=1}^{N} (P(l) \text{ is the frequency distribution of the lengths } l \text{ of the diagonal lines}) ]</td>
</tr>
<tr>
<td>Nonlinearity ( L_{\text{max}} )</td>
<td>Maximum length of the diagonal lines [ \max {</td>
</tr>
</tbody>
</table>
depicted by the evolution of $L_{\text{max}}$ (maximum of the diagonal recurrence lines in the time window $W$) has been shown to be equivalent of the inverse of the largest positive Lyapunov Exponent (Gao and Cai, 2000); this means that low values of $L_{\text{max}}$ are indicative of a divergence from the near past evolution of traffic flow (compared to the near past pattern evolution). These two characteristics of traffic flow evolution have been associated to the occurrence of transitional queue conditions defined by an analytic LWR traffic flow model (Vlahogianni et al., 2007b).

Having assumed that traffic flow evolves in patterns with variable statistical characteristics with respect to determinism and nonlinearity, a clustering approach can be implemented in order to determine the prevailing characteristics of traffic flow patterns and identify traffic flow regimes on a volume-occupancy relationship. The implementation of the specific approach in signalized arterials has shown that traffic flow determined by the joint consideration of volume and occupancy evolves in patterns whose evolution can be clustered into four traffic flow regimes (Vlahogianni et al., 2008). The specific regimes exhibited in Figure 2 consist of a free-flow regime (Regime I) governed by weak determinism, as well as a congested regime (Regime IV) where traffic evolves in patterns of highly deterministic structure, in both regimes, non-linearity is weak. Moreover, in synchronized flow (traffic conditions near capacity), traffic flow exhibits strong nonlinearity (chaotic behavior); this implies that in synchronized flow where vehicles tend to keep similar individual speeds, small changes in vehicle characteristics—for example, speed, acceleration, braking, and so on—may result in significant changes in traffic flow evolution (Kerner, 2004). This finding is also reported in Daganzo (1997) that underlined that instability and chaotic-like behavior are observed near capacity. Interestingly, near capacity there exist two different regimes: Regime II that describes low deterministic traffic flow evolution and Regime III that encompass highly deterministic traffic flow pattern evolution. The above finding points to considering traffic flow evolution in synchronized flow as a piecewise deterministic nonlinear process (Regime III) that is disrupted by stochastic (random) events (Regime II).

Neural Network Predictors Based on Short-Term Temporal Patterns of Flow

The present article focuses on the multilayer feed-forward neural networks (MLP) that have one hidden layer that are trained with the supervised back-propagation rule. Each neuron is fed with data from the previous level in the form of a weighted sum, transforms this sum and propagates it to the net level. Let $w_{ij}$ be the weight of the connection between unit $k$ and $j$ the activation in the $p$th training example of the $k$th neuron will be:

$$ y_k^p = \Phi_k(s_k^p) $$

Where $s_k^p = \sum w_{kj} \cdot y_j^p + \theta_k$ is the weighted sum of the separate outputs of each of the connected units plus a bias or offset $\theta_k$ and $\Phi_k$ is a sigmoid function: $\Phi(s_k) = \frac{e^{sk} - e^{-sk}}{e^{sk} + e^{-sk}}$. The output will be:

$$ y_k(t + 1) = \Phi_k(s_k(t)) = \Phi_k \left( \sum_j w_{jk}(t) \cdot y_j(t) + \theta_k(t) \right) \quad (4) $$

Training consists of correcting the weights of the network with a portion of the error between predicted $y_k^p$ and actual values $d_k^p$ by:

$$ \Delta w_{jk}(t + 1) = \gamma \delta_k^p y_j^p + \mu \cdot \Delta w_{jk}(t) \quad (5) $$

Where, $\gamma$ is the learning rate, $\mu$ is the momentum term that is a constant that determines the effect of previous weight change and $\delta_k^p$ is the error determined by the delta rule (details can be found in Principe et al., 1999).

The above is a description of the manner static neural networks work. In the case of temporal processing, the MLP may engage a memory mechanism that reconstructs that time-series under prediction in the $m$-dimensional Phase Space. The simplest form of memory is the time-delayed that converts data in a vector of the form of Eq. (1). Memory mechanisms can be also extended to the hidden layer; in this case, the neural network known as time-lagged neural network (TLNN) preserves past internal states in order to improve on predictions.

The training of a TLNN can be done by a temporal form of Back-propagation algorithm. The basic equation that provide the activation $y_i(n)$ of the units $i, j$ at training cycle $n$ and the correction of the error $e_i(n)$ are:

$$ y_i(n) = (1 - m)y_i(n - 1) + m \left[ \sum_j w_{ij} x_j(n) \right] \quad (6) $$

$$ e_i(n) = (1 - m)e_i(n + 1) + \sum_j w_{ij} \delta_j(n) \quad (7) $$
where \( m \) is the depth of the memory and \( \delta_j(n) \) the correction of the weights according the delta rule. Weights are updated by:

\[
\frac{dE}{dm} = \sum_{n} \left[ -y(n-1) + \sum_{j} w_{ij}x_j(n) \right] e_i(n) \quad (8)
\]

These networks have a slower convergence than the static ones due to the temporal back-propagation algorithm, as the correction of the weights should be conducted for every temporal state of the network (Principe et al., 1999).

Empirical evidence shows that neural networks and especially the static MLP structures are quite useful and flexible in approximating clear statistical dynamics (Arbib, 1995). Adding this to their structural and learning simplicity make static MLPs quite appealing to short-term traffic flow prediction and to the specific pattern-based consideration; however, issues of parameters’ optimality arise when determining the structure and learning parameters of neural networks. For this, a set of genetically optimized static MLPs for each different temporal behavior of traffic flow—as identified by crossrecurrence and cluster analysis—are developed. The networks are fed with patterns of traffic variables and the training is independent of the continuity in data or in traffic flow’s dynamics.

Genetic Algorithms for Neural Network Optimization

The optimization of neural networks can make a significant difference in predictive accuracy and training time (Reed and Marks, 1998). Optimized structures are considered those that have the optimum internal structure and learning parameters. In the specific implementation by structure, the number of units in the hidden layer and the values of learning rate and momentum in the back-propagation algorithm are implied. Genetic algorithms (GAs) are a prominent way of optimizing neural networks (Yao, 1999).

In brief, GA optimization is based on a solution or else a chromosome that consists of genes carrying information on the number of units in the hidden layer (\( h \)), the learning rate (\( \gamma \)) and the momentum (\( \mu \)). A population of chromosomes is created \( x_a \) (\( a = 50 \)) (100 generations). Furthermore, a chromosome \( x_i \) is evaluated on the basis of the mean square error between predicted and actual values in the crossvalidation set. After evaluating all the population’s chromosomes, a new population is created (generation: \( t + 1 \)); the chance of a chromosome getting selected is proportional to its fitness (or rank) (roulette selection). The differentiation in population between consecutive generations occurs due to the genetic operators of crossover and mutation (Mitchell, 1998). For new chromosomes (children) to be produces, a pair of parents is selected. Following, two points are randomly selected (probability 0.9) to interchange between parents to produce two new offspring (two point crossover). Moreover, genes in chromosomes are altered (probability 0.09) from their initial state in order to maintain diversity in the population. The process of cross-over and mutation is repeated until a generation at \( (t + n) \) a certain criterion of convergence is fulfilled (for example a threshold of the mean square error between predicted and actual values).

IMPLEMENTATION AND FINDINGS

The Data

The available data comes from an extended dataset of point measurements of volume and occupancy per 90 seconds (the average cycle length) from 140 arterial links in the center of Athens (Greece). Arterial links are controlled by midblock detectors located 90–120 meters from the stop line. For the specific analysis volume and occupancy data are extracted from three three-lane signalized arterials (Figure 3). These arterials serve the bulk of demand during the peak hours and experience significant inflows and outflows, as well as uncontrolled demand (midblock or side street traffic) inducing complexity to the distribution of traffic flow along the arterial links; however, simulation studies conducted in the area have shown that the signalization plans serve efficiently the demand and arrivals are uniformly distributed across cycle. Consequently, despite the midblock traffic, roadways maintain smooth operation that is suitable for further estimation and prediction implementations.

The descriptive statistics of the volume and occupancy series are seen on Table 2. As can be observed, the LM ARCH test (Eagle, 1982) conducted on both time-series indicated that volume and especially occupancy show signs of volatility. Previous research on the same dataset has focused on developing short-term prediction models for traffic volume prediction. All approaches were data-driven due to the uncertainty regarding signalization’s synchronization with the traffic flow measurements, as well as the possible influence of certain nonrecurrent

![Figure 3](https://example.com/figure3.png)  
**Figure 3** Map of the central business district of Athens (Greece). The signalized arterials utilized in the specific study are marked with dashed lines.
Table 2  Descriptive statistics and test for the presence of ARCH effect (volatility) for the series of volume and occupancy.

<table>
<thead>
<tr>
<th></th>
<th>Free-Flow</th>
<th>Critical Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Max</td>
<td>115</td>
<td>100</td>
</tr>
<tr>
<td>Mean</td>
<td>58</td>
<td>24.60</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>16.8</td>
<td>18.54</td>
</tr>
<tr>
<td>Variance</td>
<td>282.93</td>
<td>343.62</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.86865</td>
<td>1.60</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.45</td>
<td>2.69</td>
</tr>
<tr>
<td>LM ARCH (5 lags)</td>
<td>6199.6**</td>
<td>8461.1**</td>
</tr>
</tbody>
</table>

**Rejection at 1% level of significance.

Table 3  Data and MLP specifications for pattern-based prediction.

<table>
<thead>
<tr>
<th></th>
<th>Regime I</th>
<th>Regime II</th>
<th>Regime III</th>
<th>Regime IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Data</td>
<td>12%</td>
<td>16%</td>
<td>54%</td>
<td>18%</td>
</tr>
<tr>
<td>TR–CV–TE</td>
<td>60%-20%-20%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure Levels</td>
<td>4-7-1</td>
<td>4-9-1</td>
<td>4-15-1</td>
<td>4-12-1</td>
</tr>
<tr>
<td>Optimization</td>
<td>Genetic algorithm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning</td>
<td>Genetic algorithm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetic algorithm optimization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromosome</td>
<td>$h \in [5, 25]$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitness function</td>
<td>Mean square error (cross-validation set)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection</td>
<td>Roulette</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-over</td>
<td>Two point ($p = 0.9$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutation</td>
<td>Probability $p = 0.09$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$h$: neurons in hidden layer.
$\gamma$: learning rate.
$\mu$:momentum.

events (incidents, restrictions to traffic flow enforced by police, and so on). It has been shown that static MLPs that have and genetically optimized internal structure and learning (Vlahogianni et al., 2005) outperformed State-Space and ARIMA models (Stathopoulos and Karlaftis, 2003a). The time-delayed (memory in the input layer) structures of MLP with genetically optimized memory implemented later in the same dataset showed improvement in accuracy, but reduced performance in transitional conditions (for example onset of congestion) (Vlahogianni et al., 2007a).

Pattern-Based Short-Term Predictions

Static neural network structures are developed one for each regime and each network is trained to generalize to specific dynamics regarding the deterministic and nonlinear behavior of traffic flow patterns. The neural networks are presented with patterns of volume and occupancy and are trained to produce one-step ahead volume and occupancy predictions. For this, the available data are categorized into traffic regimes—as determined on Figure 2—and then is further separated into three subsets: the training set, the crossvalidation set (for testing the training) and the testing set that will be used to evaluate the predictive performance of the neural networks. Networks are simple static Multilayer Feed-forward Perceptrons with one hidden layer that genetically optimized with respect to the number of hidden units. Training is done by the back-propagation algorithm; the learning rate and momentum are also been genetically optimized. The data, structural, learning, and optimization specifications of the networks are summarized in Table 3. Accuracy is evaluated using the Mean Absolute Error (MAE) and the Mean Relative Percent Error (MRPE) (Washington et al., 2003):

\[
\text{MAE} = \frac{1}{P} \sum_{p=1}^{P} |e^p| \tag{9}
\]

\[
\text{MRE} = \frac{1}{P} \sum_{p=1}^{P} \frac{|e^p|}{d^p_o} \tag{10}
\]

where $P$ is the number of data and $e^p = (d^p_o y^p_o)$ with $d^p_o$ the actual $y^p_o$ the predicted value.

Prediction results for volume and occupancy are summarized in Table 4. As can be observed, in all regimes, MLP predict traffic volume with similar levels of accuracy. The same applies to the case of occupancy. A $t$-student test to all predictions (volume and occupancy) in each regime is conducted in order to prove that there is no significant difference between predicted and actual values of traffic volume. Moreover, the tolerance intervals for volume and occupancy predictions are estimated in each regime; taking the worst case met in prediction errors (Regime III), with 95% confidence, at least 90% of the population MRPE lie between 0% and 18% for volume and 0% and
Table 4  Predictions of volume and occupancy in the test set using pattern-based MLPs.

<table>
<thead>
<tr>
<th>Regime</th>
<th>Volume MAE (vh/90sec)</th>
<th>MRPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regime I</td>
<td>3 ± 1</td>
<td>9.2 ± 3.7</td>
</tr>
<tr>
<td>Regime II</td>
<td>5 ± 2</td>
<td>8.5 ± 4.3</td>
</tr>
<tr>
<td>Regime III</td>
<td>5 ± 2</td>
<td>8.5 ± 5.2</td>
</tr>
<tr>
<td>Regime IV</td>
<td>5 ± 2</td>
<td>7.8 ± 4.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regime</th>
<th>Occupancy MAE (%)</th>
<th>MRPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regime I</td>
<td>2.34 ± 1.9</td>
<td>14.70 ± 6.5</td>
</tr>
<tr>
<td>Regime II</td>
<td>3.05 ± 2.1</td>
<td>14.20 ± 6.5</td>
</tr>
<tr>
<td>Regime III</td>
<td>3.36 ± 2.4</td>
<td>13.40 ± 6.9</td>
</tr>
<tr>
<td>Regime IV</td>
<td>7.30 ± 3.72</td>
<td>13.10 ± 6.0</td>
</tr>
</tbody>
</table>

29% for occupancy. Figure 4 and Figure 5 exhibit the scatter plots between actual and predicted values of volume and occupancy respectively: $r^2$ values are depicted for each regime. No systematic error is observed leading to assume that networks have approximated successfully the underlying behavior of volume. Occupancy predictions are, in general, worse than the ones of volume. This could be due to the inherent volatile behavior of occupancy that is enhanced comparing to the one observed in volume series (Table 2).

Comparing Pattern-Based Prediction with Time-Series Techniques

A comparative study is further conducted in order to evaluate the effectiveness of the proposed pattern-based prediction scheme—in terms of prediction accuracy—compared to the most commonly used time-series prediction techniques in previous researches. Both global and local time-series models are implemented. Global approaches attempt to build a single complex model for the entire range of behaviors identified in the time series. On the opposite, there exist local prediction techniques that are more suitable for capturing the dynamics of a specific regime.
that seek similarities locally in the dynamics in order to produce a good estimate of the future state of a system. From the global approaches, an ARIMA model and a TLNN are implemented. ARIMA models have been extensively applied to traffic flow data (Williams, 2001; Stathopoulos and Karlaftis, 2003a). As for the TLNN, the specific structure has been selected as it represents the state-of-the-art in time-series prediction using neural networks (Principe et al., 1999). As explained previously, these networks are MLPs that engage memory structures both in the input and the hidden layers.

Moreover, from the local approaches, the local weighted linear model that belongs to the category of memory-based local prediction models is implemented. Its local character of approximation lies in that the prediction is generated using only nearby states; these states define the neighborhood of prediction. To predict the next state of a traffic variable, for example the traffic volume \(V(t + \tau)\) where \(\tau\) is the embedding delay that matches the prediction step, we search the \(k\) nearest neighbors of \(V(t)\), that is, the \(k\) nearest states \(V(t') (t' < t)\) that minimize \(\|V(t) - V(t')\|\), where \(\cdot\) is a metric (for example Euclidean) (Farmer and Sidorowich, 1987). The local linear predictor results by fitting a linear polynomial to the pairs of \(V(t'), V(t' + \tau)\). In locally weighted linear prediction the contribution of each of the \(k\) nearest neighbors is weighted by a function (kernel) \(K = e^{-\frac{\text{distance}}{h}}\), where distance is the distance of a neighbor from the reference state and \(h\) is the distance of the reference state and its furthest neighbor, among the \(k\) nearest neighbors considered (Konov, 2007).

The above approaches are based on time series thinking, in contrast to pattern-based logic where patterns rather than time series are needed to train the neural networks. Moreover, all approaches utilize the same time window of past information for prediction; that is take traffic states four steps in the past to generate traffic volume’s and occupancy’s next state.

**Figure 5** Plots of predicted versus actual values of occupancy for all traffic flow regimes.
Results are depicted in Table 5. As can be observed, pattern-based prediction outperforms both TLNN and ARIMA models in all traffic flow regimes. The local weighted linear model seems to perform better that the ARIMA in the one step ahead prediction of both volume and occupancy; nevertheless, neural networks are found to be the optimal solution for short-term traffic flow prediction.

Figures 6 depicts the actual versus predicted relationship of volume and occupancy for all three prediction strategies. As can be observed, the predicted relationship by the TLNN is problematic in high values of volume and at congestion. The same applies to the case of predictions provided by the ARIMA model. Regarding predictions of traffic flow using the local weighted linear predictor, Figure 6 shows that the predictor cannot capture the variability in traffic flow states in synchronized flow. The above results can be visualized in Figure 7 where the series of actual versus predicted values of volume and occupancy for the morning peak of a typical weekday are depicted. In the same figure the regime propagation during the same time period is provided. As can be observed, ARIMA predictions exhibit the worst fit to actual values. Regarding the rest of the methodologies evaluated, in the case of volume, TLNN predictions are more sensitive to sudden shifts of volume values. Moreover, in the case of occupancy, the pattern-based neural scheme is more accurate during shifts to high values as, well as during congested conditions.

In order to quantify the quality of prediction with respect to the information gain brought to the knowledge of the anticipated traffic flow state—meaning pairs of volume and occupancy at the...
Figure 7  Time-series of regimes and actual versus predicted values of volume (left column) and occupancy (right column) during the morning peak of a typical weekday.
next time interval—the mutual information between a method’s prediction and the actual traffic flow state is calculated in each traffic flow regime. The mutual information $I(x_i, x_j)$ between variables $x_i$ and $x_j$ measures the expected information gained about $x_j$, after observing the value of the variable $x_i$.

$$I(x_i, x_j) = \sum_{x_i, x_j \in X} P(x_i, x_j) \log \frac{P(x_i, x_j)}{P(x_i)P(x_j)} \quad (11)$$

The mutual information between a method’s prediction and the actual value of a traffic state can tell us how close their relationship is. Results are demonstrated on Table 6. As can be observed, numerical results match the previous qualitative observations made on Figure 6. That is, the proposed pattern-based neural network scheme is capable of predicting more accurately the anticipated state of traffic flow when compared to methods that operate under time-series thinking. The previous observation applies to all traffic flow regimes considered in the specific study. Interestingly, all models exhibit decreased performance in synchronized flow conditions (regimes II and III); this is probably due to the highly nonlinear characteristics of traffic flow, as well as the frequent transition between extreme deterministic and stochastic structure (Vlahogianni et al. 2008).

**CONCLUSIONS**

Accurate and timely short-term traffic flow predictions are an essential component of intelligent transportation systems. Various methods to predicting traffic flow variables such as volume, speed, travel time, and so on have been previously implemented; however, methods as presented cannot infer knowledge from predictions on the anticipated traffic flow regimes and transitions. Moreover, the prediction accuracy in critical traffic conditions, such as the onset of congestion, is questionable.

In this article, a methodological framework for developing accurate and consistent to traffic flow evolution predictors is presented; predictions are based on an advanced nonlinear methodology to quantify the statistical characteristics of patterns and identify regimes of traffic flow, as well as an artificial intelligence approach to pattern-based prediction. Findings indicate that predictions under the proposed framework are improved comparing to the ones under time-series thinking provided by a dynamic neural network, an ARIMA, as well as a local weighted linear model. The proposed pattern-based prediction scheme was also found to outperform the other methods tested in the knowledge of the anticipated traffic flow states in all traffic flow conditions.

The proposed approach encapsulates certain advantages; first, it improves on accuracy of short-term traffic flow predictions. Second, it is consistent with the dynamics of traffic flow; thus, it can claim applicability to all traffic flow regimes and can be adaptable to the changes in traffic flow evolution. Third, it is dependent on only volume and occupancy measurements and, for this, it can be considered as transferable. Finally, the proposed prediction scheme has limited dependence on the near past that makes it fast and easy to implement in real-time applications.

**REFERENCES**


Ishak, S., and Alecsandru, C. (2004). Optimizing traffic prediction performance of neural networks under various topological, input and

**Table 6  Method significance with respect to the information gain brought by each method to the knowledge of the anticipated traffic flow state (numbers correspond to the mutual information of a prediction method with the actual traffic state).**