

# Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach

Eleni I. Vlahogianni, Matthew G. Karlaftis \*, John C. Golias

*Department of Transportation Planning and Engineering, School of Civil Engineering,  
National Technical University of Athens, 5, Iroon Polytechniou Str.,  
Zografou Campus, Athens 157 73, Greece*

Received 18 July 2003; accepted 20 April 2005

---

## Abstract

Short-term forecasting of traffic parameters such as flow and occupancy is an essential element of modern Intelligent Transportation Systems research and practice. Although many different methodologies have been used for short-term predictions, literature suggests neural networks as one of the best alternatives for modeling and predicting traffic parameters. However, because of limited knowledge regarding a network's optimal structure given a specific dataset, researchers have to rely on time consuming and questionably efficient rules-of-thumb when developing them. This paper extends past research by providing an advanced, genetic algorithm based, multilayered structural optimization strategy that can assist both in the proper representation of traffic flow data with temporal and spatial characteristics as well as in the selection of the appropriate neural network structure. Further, it evaluates the performance of the developed network by applying it to both univariate and multivariate traffic flow data from an urban signalized arterial. The results show that the capabilities of a simple static neural network, with genetically optimized step size, momentum and number of hidden units, are very satisfactory when modeling both univariate and multivariate traffic data.

© 2005 Elsevier Ltd. All rights reserved.

**Keywords:** Traffic flow; Multivariate time series; Short-term predictions; Neural networks; Genetic optimization

---

---

\* Corresponding author. Tel.: +30 210 7721 280; fax: +30 210 7722 404.  
E-mail address: [mgk@central.ntua.gr](mailto:mgk@central.ntua.gr) (M.G. Karlaftis).

## 1. Introduction

Accurate real-time information provision and short-term predictions of traffic parameters such as volumes, travel speeds and occupancies, is a research topic that has attracted considerable interest in the literature. This is, at least in part, a result of the increasing penetration of Intelligent Transportation Systems (ITS) technologies in everyday life. ITS technologies, with Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS) as examples, attempt to deal with the traffic congestion and travel time problems facing commuters in many urban areas worldwide by better synchronizing traffic signals and by assisting drivers on selecting routes based on accurate real-time information on traffic conditions. The successful implementation and wide-spread acceptance of ITS technologies depends, to a large extent, on the quality and accuracy of the information given to the traffic management authorities and the drivers. As such, it comes as little surprise that the field of short-term traffic parameter modeling and prediction has attracted significant scientific interest which, because of the continuous increase in computing speed and the development of increasingly flexible methodological approaches, only increases with time. Since the early 1980's, researchers have used an extensive variety of specifications to model traffic characteristics and produce short-term forecasts with a large amount of literature concentrating on predictions from time-series models that range from ARIMA to dynamic generalized linear models and, especially in recent years, on neural networks.

The literature has shown that neural networks are one of the best alternatives for modeling and predicting traffic parameters possibly because they can approximate almost any function, regardless of its degree of nonlinearity and without prior knowledge of its functional form. However, because of the developmental nature of neural networks, a large degree of uncertainty is present when trying to select the optimal network structure and, to overcome this, researchers have to rely on very time consuming and questionably efficient rules-of-thumb. The present paper concentrates on the development of optimized neural network models to forecast traffic flow in highly congested urban signalized arterials. The focus is on providing an advanced, genetic algorithm based, multilayered structural optimization strategy that can assist both in the proper representation of traffic flow data with temporal and spatial characteristics and in the selection of the “most” appropriate neural network structure.

From a methodological standpoint, the current paper extends past research by introducing an optimization approach that provides the means for assuring that the neural network developed has the optimal structure and, consequently, the “best” behavior with respect to the traffic flow set it is required to model. The formulation uses multivariate data (i.e. data from multiple loop detectors) from urban signalized arterials, an area that, despite its great importance, has not been adequately researched in the past. The remainder of this paper is organized as follows. The next Section is a brief overview of the background in short-term traffic forecasting and further details the motivation behind the approach developed in this paper. The third Section describes the methodological and optimization framework to be used in the paper and the fourth Section briefly discusses the temporal and spatial representation of traffic data in neural structures. The fifth Section discusses the data used in this study and the sixth Section describes the model estimation process and presents the empirical results. Finally, the seventh Section summarizes and discusses the findings of the paper.

## 2. Background and motivation

Traffic flow can be considered as both a temporal and a spatial phenomenon. Its conceptual as well as practical dependence on time and space (distance from the prediction point) has been established in the traffic forecasting literature, especially in applications in highly congested and densely signalized urban environments. The temporal and spatial attributes of traffic flow suggest that it is a dynamic as well as a fast evolving phenomenon, especially in urban areas, where traffic volumes frequently approach congestion. As [Head \(1995\)](#) suggested, in the urban environment the presence of signalization gives traffic a spatio-temporal behavior that is more random and difficult to study than in freeways. Further, as [Stathopoulos and Karlaftis \(2001a\)](#) showed, traffic flow in urban signalized arterials has a certain temporal and spatial behavior that exhibits a randomness which escapes the traditional perception of periodicity (monthly, weekly, daily or even hourly periodicities) in traffic operations. In a later study, the same authors showed that this randomness is very irregular and cannot be aggregated even when examining it in 2-hour periods ([Stathopoulos and Karlaftis, 2003](#)).

To model and forecast short-term traffic flow, statistical time series analyses, in the form of the ARIMA family of models have been the most widely used approaches ([Levin and Tsao, 1980](#); [Hamed et al., 1995](#); [Kirby et al., 1997](#); [Williams et al., 1998](#); [Williams, 2001](#)). But, although the temporal variability of traffic flow appears to be a crucial characteristic in short-term forecasting, the traditional time series methodologies frequently seem unable to capture the rapid variations of flow in urban areas ([Head, 1995](#); [Abdulhai et al., 1999](#)).

[Angeline et al. \(1994\)](#) attempted to apply an ARIMA structure to model urban traffic flow but showed that these models suffer from their natural tendency to concentrate on the mean values of the series and, as such, suffer from an inadequacy in modeling the extreme ones. The use of other techniques such as non-parametric regression ([Smith and Demetsky, 1997](#); [Smith et al., 2002](#); [Clark, 2003](#)), local linear regression ([Sun et al., 2003](#)) and kalman filtering models ([Okutani and Stephanedes, 1984](#); [Whittaker et al., 1997](#); [Stathopoulos and Karlaftis, 2003](#)) have shown an improved behavior in modeling short-term traffic flows but, have not been tested in signalized arterials.

Regarding the spatial dependence of flow that involves its propagation through space and whether this influence can contribute to ameliorate the predictive performance of a model, the literature indicates a difficulty in modeling such behavior due to the resulting increase in model complexity. To account for this, some authors have turned to multivariate modeling where models are developed using data from the location of interest as well as downstream locations, yielding models that are better equipped to capture the contribution of correlated sequential points of interest to short-term flow forecasts ([Williams, 2001](#); [Stathopoulos and Karlaftis, 2003](#)).

Despite the existence of some statistical based studies that examine the spatial nature of traffic flow, short-term traffic forecasting has been investigated using almost exclusively neural networks ([Vythoulkas, 1993](#); [Clark et al., 1993](#); [Kwon and Stephanedes, 1994](#); [Lingras and Mountford, 2001](#); [Kirby et al., 1997](#); [Abdulhai et al., 1999](#); [Yasdi, 1999](#); [Zhang, 2000](#); [Chen et al., 2001](#); [Chen and Grant-Muller, 2001](#)). This is, at least in part because, as opposed to conventional statistical forecasting models, neural networks have certain inherent proprieties that can deal with several modeling issues concerning the implementation of a multivariate approach to modeling flow. First, they are empirical (data-driven), self-adaptive models that have the ability to capture the underlying relationships without the need of a priori assumptions regarding the problem exam-

ined (Zhang et al., 1998). And, second, because of their ability to learn from data, even if the underlying relationships are not apparent, their non-linear nature and their ability to generalize (after being trained to a sample data they can generate predictions from a part of the data that had not been used for training), makes them a useful tool for working with databases with noise, an often seen phenomenon when modeling time series from real time traffic forecasting applications.

The above characteristics indicate the theoretical ability of neural networks in modeling the temporal and spatial variability of traffic. The literature on neural networks has to exhibit a large number of neural network models ranging from purely static to highly dynamic structures. Many of these models have already been used in modeling traffic flow. Some examples are the simple structures of Multilayer Perceptrons (MLPs) (Clark et al., 1993; Vythoulkas, 1993; Smith and Demetsky, 1994; Chang and Su, 1995; Lee and Fambro, 1999; Gilmore and Abe, 1995; Dougherty and Cobbet, 1997; Ledoux, 1997; Innamaa, 2000; Florio and Mussone, 1996; Yun et al., 1998; Zhang, 2000; Chen et al., 2001), the radial basis MLPs (Lyons et al., 1996; Park and Rilett, 1998; Chen et al., 2001), the time-delayed neural networks (Lingras and Mountford, 2001; Yun et al., 1998; Yasdi, 1999; Abdulhai et al., 1999; Dia, 2001; Ishak et al., 2003) and the recurrent neural networks (Dia, 2001; Van Lint et al., 2002).

Interestingly, most studies have concentrated on developing purely static forms of neural networks such as the popular MLPs. The basic characteristic of such networks is that their neurons are memoryless; that is, the input activation is a function only of the current input state and not of the past input-output relations (Hush and Horne, 1993). Principe et al. (2000) indicate that static neural networks use statistical proprieties such as the mean and variance to distinguish each data cluster. This implies that, in cases of data clusters with the same mean and variance, MLPs are not statistically efficient (Principe et al., 2000).

The use of static MLPs seems to be contradictory, at least on a conceptual level, with the initial consideration of traffic's dynamic behavior since a static structure is required to learn from, and generalize to, data that encompass both temporal and spatial dimensions. Nevertheless, MLPs have been very popular and widely used in time series traffic forecasting, partly because of their ability to capture non-linear and non-stationary behavior in the data structures they model. This ability is a result of the hidden layer(s) embedded in their structure, which introduce non-linearity as an internal procedure (Haykin, 1994). The literature on traffic forecasting shows that MLPs can easily model multivariate phenomena (Clark et al., 1993; Van Lint et al., 2002; Ishak et al., 2003). Moreover, they have the ability to map any continuous function regardless of how complex it is ("Universal function approximation" property) (Hornik et al., 1989).

A crucial argument in favor of using static MLPs concerns the nature of dynamic neural structures. If we wish to construct a dynamic neural structure, then the incorporation of memory lines that are responsible for the model's temporal and dynamic internal behavior largely imposes very complex neural structures. Apart from complexity that makes the networks computationally demanding and slow, these models frequently exhibit an unstable behavior because of their need to be optimized in both space and time, implying the model's look-back time window (Principe et al., 2000).

In general, the problem of traffic flow prediction should be considered as a 3-dimensional optimization process; optimization in distance from the point of interest, in time and in space.

Optimization in distance from the point of interest refers to the spatial evolution of traffic flow; in case the distance from the point of interest is crucial in modeling flow series, then flow is studied from either a sequential or a network perspective. When studying traffic flow's evolution in time, prediction becomes more complicated. Assuming that flow has a temporal dependence to past values (long-term memory), time (in the traffic flow prediction setting) has spatial implications as well. In essence, we seek the look-back time window (past values) that carries information valuable for prediction. This look-back time window is characterized by the frequency of taking delayed or lagged information e.g. every one or two backwards steps (time), as well as the total length of looking back e.g. 5 or 6 steps ahead (space). From the above, it is apparent that when optimizing prediction algorithms, it must be done both in terms of the distance from the point of interest and in terms of the temporal and spatial aspect of the flow series. In the present paper, the optimization is restricted to structural and learning issues of the MLPs. The so called look-back time window is fixed in order to facilitate the solution of the problem.

The question then, when using the simpler static MLPs, is to find a way to obtain the (near) optimal performance of the prediction model. In general, the literature treats neural network prediction performance optimization from two different angles: (i) the external input feeding that deals with the manner in which input data is presented to the network and, (ii) the optimization of the network's internal structure with regards to the process of learning (optimizing the learning parameters) and the network's dimensions (meaning the complexity of the network regarding its degrees of freedom). In the first line of thought, the literature has to exhibit hybrid (that use a primary classification technique that feed into a neural network) structures such as principal component analysis MLPs (Ishak and Alecsandru, 2003), modular MLPs (Park and Rilett, 1998; Ishak and Alecsandru, 2003) and self-organizing maps MLPs (Chen et al., 2001). Moreover, Genetic Algorithms have been applied to the input space optimization (Lingras and Mountford, 2001; Abdulhai et al., 1999).

On the other hand, the lack of dedicated internal structural optimization efforts that optimize both the learning process and the model's dimensions indicates that, when using even the simple MLP structures, a model's optimal performance cannot be assured, if only by a very time consuming and open-ended trial-and-error process. This can be considered as a weak point regarding the implementation of neural networks in traffic forecasting, especially when considering that the optimization of a data-driven model with a large number of degrees of freedom is a very important step in the model development phase and can make a significant difference in predictive accuracy (Reed and Marks, 1998).

In short, using non-optimized MLPs in practice, especially in traffic forecasting where efficient training is a requirement for successful implementation, may be considered as problematic. This paper aims at improving the process of short-term traffic flow modeling through a methodological approach that optimizes the performance of static MLP structures using advanced error back-propagation techniques and genetic algorithms. The genetic algorithms will assist the learning optimization of the network and, as such, will play the role of meta-optimizers. The objective is to deal with, and evaluate, two fundamental issues in model development:

1. The predictive performance of an optimized static MLP that is trained with traffic data from highly congested urban signalized arterials that include both a temporal and a spatial dimension, and

2. The quality of traffic forecasts obtained from static MLP models whose learning process and model structure have been optimized using a combination of a local search method (such as the gradient descent method) with global optimization techniques (such as genetic algorithms). In particular, it is of interest to investigate whether the advantages of optimizing a static neural structure (such as the MLPs) outweigh the possible theoretical complexities of using Genetic Algorithms.

### 3. Multilayer feed-forward perceptrons (MLPs)

The MLP belongs to the feedforward neural networks that are usually trained using the error back-propagation learning rule. The concept of a back-propagation MLP can be thought of as a two passes procedure through the different layers of the network: a forward one, in which the weights are fixed and a backward pass where the weights are adjusted according to the error correction rule. Fig. 1 is a schematic representation of the process involved. The purpose of this paper is not to analyze the way back-propagation works but to discuss and optimize the basic issues affecting its performance in traffic flow prediction. (Detailed analysis of error back-propagation process can be found in Haykin (1994), Bishop (1995) and Principe et al. (2000).)

Training MLPs with error-back propagation can be seen as an optimization problem with respect to two basic issues: (i) the supervised learning procedure, usually meaning the adaptation of the weights (called learning optimization) and, (ii) the optimization of the dimensions of the network and the learning parameters such as the learning rate or step size (called network structure optimization).

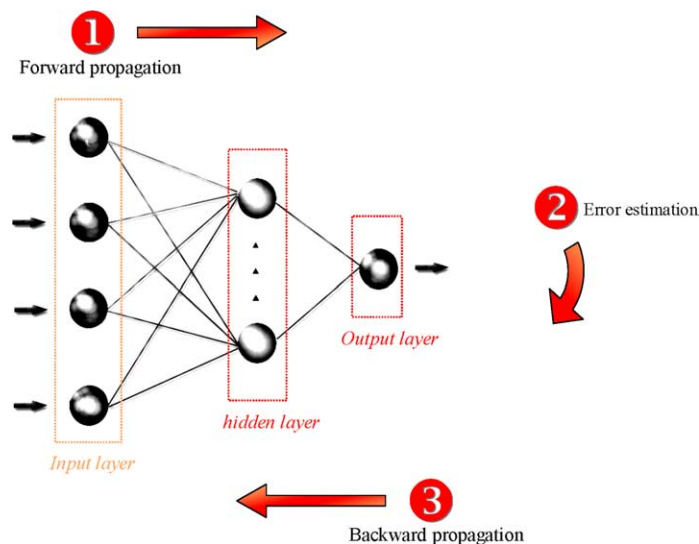


Fig. 1. Back-propagation multilayer perceptron with one hidden layer.



### 3.1. Learning optimization

The process of learning optimization deals mainly with the speed of convergence and the trapping to local minima. The error-back-propagation rule can be considered as an optimization problem regarding the searching for the optimum in the gradient descent space (the search in the gradient descent space is mathematically expressed by  $\nabla J(k)$ , where  $J(k)$  is the performance surface at the cycle  $k$ ). It is the most widely used learning algorithm due to its straightforward and simple implementation (Principe et al., 2000). Nevertheless, it suffers from two shortcomings; first, it is a local search method (each weight updating is done using information locally to that weight) that can be trapped to local minima (leading to a false training) and, second, that the performance surface may have many flat regions suggesting that the gradient-descent method will move slowly (if at all), as the weights are being modified proportionally to the gradient (leading to slow training) (Principe et al., 2000). The locality of the gradient descent method is considered the major reason that back-propagation is statistically efficient and is easy to implement in an extended number of applications (Bishop, 1995).

A common alteration of the classic gradient descent method is the addition of the momentum term  $\alpha$ . The alteration consists of adding a term that shows that the weights are updated proportionally to the updating in the previous training cycle. This generally implies that the momentum suppresses oscillations and stabilizes the weight trajectory in areas where the gradient descent is small (Reed and Marks, 1998).

Another approach is to avoid using the same learning rate across the network. A rule-of-thumb is to use a learning rate that decreases when moving from the output to the input layer. Moreover, as literature shows, the learning rate should not be kept constant during the process of learning but must decrease when the adaptation is unstable (rattling) and increase in flat points of the performance surface (Bishop, 1995). A mathematical expression of the previous is called the adaptive step size method or delta-bar-delta rule:

$$\Delta\eta_{ij}(n+1) = \begin{cases} k & \text{if } S_{ij}(n-1)D_{ij}(n-1) > 0 \\ -b_{ij}(n) & \text{if } S_{ij}(n-1)D_{ij}(n-1) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $k$  is an additive constant,  $b_{ij}$  is a multiplicative constant,  $S_{ij}$  measures the average of the previous gradients and  $D_{ij}$  is the current gradient.  $S_{ij}$  is given by

$$S_{ij} = (1 - \lambda)\nabla w_i(n-1) - \lambda S_i(n-1) \quad (2)$$

The term  $\lambda$  in the last equation is a smoothing factor. The first part of Eq. (1) refers to the case of slow convergence ( $S_{ij}$  and  $D_{ij}$  have the same sign) and the second line of Eq. (1) is when the weight oscillates (Principe et al., 2000). What is rather obvious from the above is that the adaptive step size method has several internal parameters, such as  $k$ ,  $b$ ,  $\lambda$ , that have to be calibrated.

A more advanced searching technique is to use the conjugate gradients. The conjugate gradient method is a line search method (meaning searching methods that start from a search direction in the weight space and minimize along the line). The basic concept of conjugate gradient descent is to change the search direction for the optimal step size by weighting the previous direction to the minimum with the new direction:

$$\mathbf{s}_{ij}^{\text{new}} = -\nabla \mathbf{J}^{\text{new}} + \alpha \mathbf{s}^{\text{old}} \quad (3)$$

where  $\mathbf{s}$  is the direction vector,  $\mathbf{J}$  is the performance surface and  $\alpha$  is a trade-off parameter between the two directions (Principe et al., 2000).

### 3.2. Network structure optimization

The purpose of training a neural network is not only to obtain an exact representation of the training data, but also to build a model that represents the physical process that generated that data. The generalization ability of a neural network, i.e. its ability to produce predictions for data it has not been trained on, is not only influenced by the learning process but also by the structure of the network itself. Few degrees of freedom (weights) lead, in general, to a poor generalization ability. On the other hand, if the network has more than the necessary weights it can be caught in a situation where it has learnt the data so well that it has memorized it (overtraining) and thus does not have the ability to generalize. Although there are several techniques to stop training before the network begins to overtrain (such as the popular early stopping approach), a crucial role to achieving a good generalization is based on limiting the network structure size (usually referring to the structure of the hidden layer(s)).

The search for the optimal network structure is still largely treated by the literature as an extensive trial-and-error process. For example, there are some constructive and destructive hill climbing methods relying on the idea of starting from a minimum structure and adding processing units (or the opposite), but these methods are both extremely time consuming and only indicative of the “true” optimal network size (Angeline et al., 1994). The literature also seems to indicate that in the case of network structure optimization, the application of Genetic Algorithms as a search method for finding the near-optimal structure may prove to be the most efficient approach (Yao, 1999; Reed and Marks, 1998). Despite this, genetic optimization of neural structures has not been used, to the best of our knowledge, in short-term traffic flow modelling and forecasting even though it could be potentially very helpful in practical applications.

### 3.3. Optimization using genetic algorithms

As previously discussed, the essential concern in modelling MLPs is the specification of their optimal structure with respect to the number of hidden units and the learning rule. In a more microscopic view of optimizing these two issues, one can trace many network parameters such as the learning rate  $\eta$ , the momentum  $\alpha$  as well as the number of hidden units  $h$  that are not automatically adjusted but have to be estimated by the practitioner. The most commonly used approach is to adopt a trial-and-error process in order to discover an optimal value for the three variable parameters. However, most commonly used rules-of-thumb applied to finding the optimum number of hidden neurons (Masters, 1993; Sarle, 1998), the values of learning rate, momentum and the combination of the above (Principe et al., 2000), are very time consuming due to the vast dimensions of the solution space and may not necessarily lead to the optimal result. The requirement for searching for a near-optimal neural network structure implies the coordinated optimization of the network with respect to structural and learning issues along with a meta-optimization effort in a more detailed level (calibration of learning parameters).



The question then becomes whether a method exists that could fulfill the described optimization requirements. Goldberg (1989) explains that current approaches in optimization such as the calculus-based, the enumerative and the stochastic methods are either local-based search methods (meaning that they could provide a local optimum) or suffer from the implications of the “dimensionality curse” (because of the large solution space that has to be searched) and, as such, cannot be considered as robust. A promising optimization method that, at least in theory, is considered to have the natural propensity of overcoming the above obstacles is the Genetic Algorithms (GAs) approach. GAs are stochastic search algorithms that search over a population of possible solutions but differ from the most commonly encountered optimization methods in four ways (Goldberg, 1989):

- They work with a coding of the parameter set, not the parameters themselves.
- They search from a population of points, not a single point.
- They use payoff (objective function) information.
- They use probabilistic transition rules, not deterministic rules.

GAs are based on three mechanisms: the selection, the crossover and the mutation. Their concept is founded on the principles of genetics. Initially, there is a random population of potential solutions  $A$  encoded as chromosomes and each chromosome consists of genes (Fig. 2). During the selection, a chromosome from the current generation’s population is chosen to be included in the next generation’s population. Selection does not introduce new chromosomes but recreates the population’s chromosomes, increasing the number of fitter ones and decreasing the number of less fit ones (Goldberg, 1989). The basic concept is to give preference to better chromosomes allowing them to transfer their genes to the next generation. In this paper the *roulette* selection scheme is used; in short, the chance of a chromosome being selected is proportional to its fitness (Goldberg, 1989), suggesting that preference is given to the fittest chromosome (survival of the fittest).

Before entering the next generation’s population, selected chromosomes may pass through a process of crossover and mutation (depending upon the probability of crossover and mutation). Crossover is an operation that combines two chromosomes (parents) to produce a new chromosome (offspring; Fig. 3 explains the concept of crossover). The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the best characteristics from each of them.

Mutation is an operation that alters one or more gene values in a chromosome from its initial state. This can result in an entirely new set of gene values. With these new gene values, the genetic algorithm may be able to arrive at a better solution than was previously possible

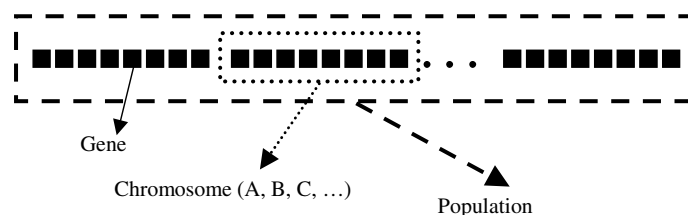


Fig. 2. Representation of a population.

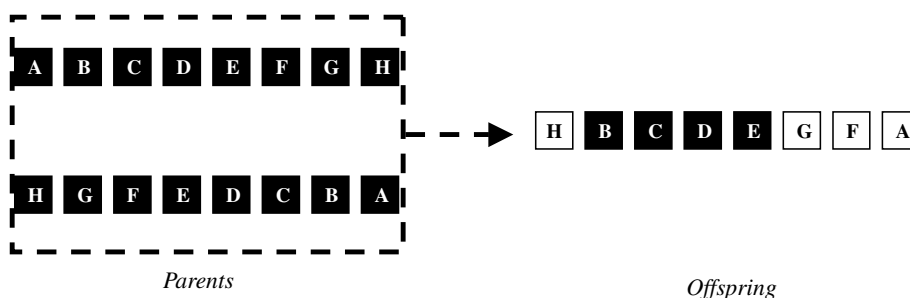


Fig. 3. The process of crossover.

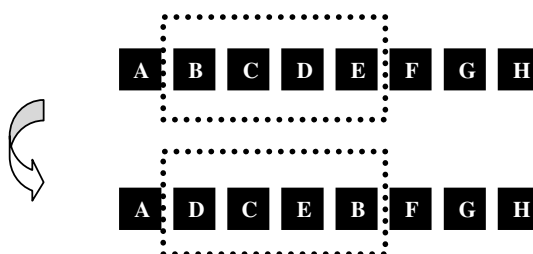


Fig. 4. The process of mutation.

(Fig. 4 is a schematic representation of the above concept). Mutation is an important part of the genetic search as it helps to prevent the population against any permanent fixation of genes in chromosomes (Mitchell, 1998). Its purpose is to maintain diversity within the population and inhibit premature convergence. Mutation and selection (without crossover) create parallel, noise-tolerant, hill-climbing algorithms (Winter et al., 1996).

Both crossover and mutation occurs during evolution according to different types of predetermined probabilities. These probabilities are estimated based on certain rule-of-thumb. For example, crossover should usually be selected at high values (a good starting value is 0.9) (Gen and Cheng, 2000). Moreover, the mutation rate is chosen as the inverse of the number of chromosomes (population) and much lower than the crossover probability to avoid permutation (Eiben and Smith, 2003).

### 3.4. A combined approach to learning and structure optimization of neural networks

As has become apparent from the preceding discussion, the optimization of neural network structures is a complicated process. First, it involves the search over both the learning and the appropriate model structure surfaces of solutions. Second, it should be a trade-off between adopting a local aspect in searching for the optimum weight matrix and employing a more global meta-optimizing technique in order to fully calibrate the microscopic aspects of learning and generalization. In this framework, the present paper tracks the performance of different Multi-layer feed-forward Perceptron structures that have genetic algorithms embedded in their structure and are applied to traffic flow forecasting. More precisely:

- MLPs are trained with classic error-back-propagation and GAs are used for step size  $\eta$ , momentum  $\alpha$ , and are used for optimizing the processing units  $h$ .
- MLPs are trained with classic back-propagation with adaptive step size and GAs are used for optimizing the processing units  $h$ .
- MLPs are trained with scaled conjugate gradient descent and GAs are used for optimizing the processing units  $h$ .

#### 4. Temporal and spatial representation of data in MLPs

The basic question when trying to predict time series traffic data in MLPs, is finding a way of incorporating both its temporal and its spatial characteristics. In the absence of internal memory that could account for the temporal behavior, MLPs have to be presented externally with data that encompass such characteristics. For example, consider the series of flow data which varies as a function of time ( $V(t)$ ). In order to predict the value of  $V$  at a given time  $t$ , the network must be trained using pairs of input-output values, where the input values could be time-lagged events of  $V$ , such as  $V_{t-1}, V_{t-2}, \dots, V_{t-n}$ . After training, the network should be capable of predicting an observed  $V'_t$  using as inputs  $V'_{t-1}, V'_{t-2}, \dots, V'_{t-n}$ . In this framework, the MLP acts like a function approximator (Bishop, 1995); essentially, the model tries to take advantage of possible statistical correlation between time lagged flow events and represent the underlying function that connects them. This approach can be considered as a univariate time series non-linear prediction model. Moreover, assuming that the MLP is presented with time lagged data but also with data with spatial attributes (traffic volume from detectors located upstream of the measurement point), the resulting MLP can be considered as a multivariate non-linear time series prediction model.

#### 5. The data

In the present study we use data from a major urban signalized arterial in Athens, Greece, to model and predict traffic flow. Alexandras Avenue (Fig. 5) is located on the boundaries of the core of Athens and connects two highly congested urban arterials (Kifissias Av. and Patission Av.). The North-South direction is congested during large parts of the day. The total length of the area

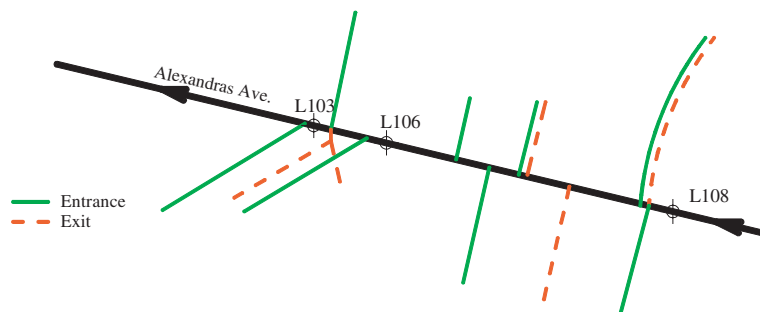


Fig. 5. Graphical representation of the control locations in Alexandras Avenue.

of study is 1.5 km. The loop detection system is organized as can be seen in Fig. 5. Loop detectors L108 (1.2 km from L103), L106 (0.3 km from L103) and L103 capture data at some of the most crucial intersections; further analysis indicated that flow exhibits great fluctuation during most parts of the day in L103. For that reason, and in order to test the robustness of neural networks in time series data with significant variations, loop L103 was selected as the target loop for forecasting.

Due to the complexity of the area of study (many entrances and exits that are not controlled by loop detectors) as well as the densely located traffic signalization system, a more macroscopic data-driven approach is selected for modeling traffic flow. It was also impossible to retrieve coordinated traffic information with the traffic signalization plans; as such, information regarding the signalization was discarded from the modeling process. The only essential and accurate knowledge about the way traffic evolves in the area stems from traffic volume; this will be used to predict traffic in the area.

The available data consists of traffic flow measurements in 3 min intervals between January and May of 2000. Although previous studies using the same data have demonstrated that different time periods of the day possess different autoregressive and cross-correlation characteristics (Stathopoulos and Karlaftis, 2001b), this paper uses a more unified treatment of the data by ignoring the time period and attempting to predict a flow series using a single optimized model.

Moreover, based on earlier analyses of this data in the same area, Stathopoulos and Karlaftis (2001b) reported that the data exhibit a positive autocorrelation with relatively short-term lags and a high correlation between sequential loop detectors both for short and longer term lags; as a result, two different data sets were developed: one including univariate traffic flow data (implying that only time lagged flow data from the loop of interest, loop L103, are used), and one with multivariate flow data (implying that time lagged data from both the loop of interest and from previous ones are used). Based on the previous considerations, the resulting data set consists of 7300 measurements of traffic flow at 3 min intervals in three specific control locations, loop detectors L103, L106 and L108 (the data used this study are available from ftp://data:data@147.102.154.49).

## 6. Empirical findings

The model development was divided into two stages: the application of the proposed optimization techniques to the univariate data of traffic flow data and the application of the same techniques to the multivariate set. Both sets included 7300 3-min traffic flow measurements. In order to model the univariate set of data, an MLP with one-hidden layer was developed using 3 inputs (the 3 time lagged periods of flow from L103) and 5 output units (flow for L103 for time intervals  $t + 1$ ,  $t + 2$ ,  $t + 3$ ,  $t + 4$ ,  $t + 5$ ). These time lags resulted from the research of Stathopoulos and Karlaftis (2001b) based on a spectral and cross-spectral analysis. In the multivariate setting, the same type of network was used but this time it included 13 inputs (3 time lagged periods of flow from L103, and 15 min time lagged data from L106 and L108) and 5 output units (flow for L103 for time intervals  $t + 1$ ,  $t + 2$ ,  $t + 3$ ,  $t + 4$ ,  $t + 5$ ). The selection of the time lags for L106 and L108 was based on the idea of having enough information on the spatial evolution of traffic over the arterial examined. The specification of the structure of the developed MLP model, that is the

dimensions of the input, the hidden and the output space, the activation function, the maximum number of cycles of training and the error criterion, are summarized in Table 1. In order to train and predict with neural networks, the data set was separated in three subsets; the training set (60% of the available data), the cross-validation set (10%), and the training set (30%). The training procedure is tested using the cross-validation set; the prediction performance of the developed models is exhibited in the test set.

The three optimization approaches followed are described in Table 2. More specifically, the first involves a combination of a genetically optimized hidden layer with a genetically optimized gradient descent learning rule with momentum. The second is a mixture of a genetically optimized hidden layer with a genetically optimized gradient descent learning rule with adaptive step size and momentum. The final optimization approach combines the scaled conjugate gradient descent with genetically optimized hidden layer. The genetic encoding is binary, while the chromosome consists of the parameters to be meta-optimized. For example, in the first approach it seeks to optimize the values of the learning rate  $\eta$  the momentum  $\alpha$ , and the number of hidden units  $h$ . The crossover and the mutation are 0.9 and 0.02 respectively; population size and generation are 50 and 100 respectively.

Table 3 summarizes the results from the test set for all the optimization strategies for both the univariate and the multivariate data sets. The measures of forecasting performance are the Mean Absolute Error (MAE), the Mean Relative Percent Error (MRPE) and the correlation coefficient  $\rho$  between actual and predicted flow series (Washington et al., 2003).

As reported earlier, there have already been some efforts for predicting traffic flow with ARIMA and State-space modeling (Stathopoulos and Karlaftis, 2003). Since these studies used the same dataset, as well as the same input parameters both in the univariate and the multivariate modeling, provides for establishing a solid comparison between the proposed methodology and

Table 1  
Model specifications

	Univariate model	Multivariate model
Input space	3	13
Output space	5	5
Hidden layers	1	1
Activation function	tanh	tanh
Train stopping	Cross-validation (early stopping)	Cross-validation (early stopping)
Max number of epochs	2000	2000
Error criterion	Mean square error (MSE)	Mean square error (MSE)

Table 2  
Optimization strategies used

Weight adaptation	Learning genetically optimized parameters	Structure
Gradient descent	Step size $\eta$ and momentum $\alpha$	Genetically optimized hidden layer
Gradient descent	Momentum $\alpha$ and parameters $k$ , $b$ , $\lambda$	
Scaled conjugate gradient descent	–	

Table 3  
Summary of the results

Data	Model	Performance evaluation	Prediction				
			$t + 1$	$t + 2$	$t + 3$	$t + 4$	$t + 5$
Univariate	Gradient descent	MAE	6	11	12	13	14
		MRPE	8.54	16.76	18.74	20.36	21.71
		$r$	0.95	0.84	0.80	0.77	0.75
	Adaptive step size	MAE	6	11	12	13	14
		MRPE	8.49	16.88	18.67	20.49	21.83
		$r$	0.95	0.84	0.80	0.77	0.75
	Scaled conjugate gradient descent	MAE	7	11	12	13	14
		MRPE	8.71	17.01	18.86	20.66	22.02
		$r$	0.95	0.84	0.80	0.77	0.74
	ARIMA	MRPE	16	n/a	n/a	n/a	n/a
Multivariate	Gradient descent	MAE	7	11	11	12	13
		MRPE	8.39	16.66	18.31	19.99	20.39
		$r$	0.95	0.85	0.82	0.79	0.78
	State-space	MRPE	13	n/a	n/a	n/a	n/a

other commonly used predictive approaches. However, the multiple steps ahead forecasting has not been dealt with in the previous studies. Moreover, a comparison of the performance of the optimized networks with the one of a conventional non-optimized MLP that uses a trial-and-error procedure for selecting the proper number of hidden units  $h$  was undertaken.

### 6.1. Univariate time-series

In general, the results show that the advanced techniques of gradient descent did not improve the performance of the MLP significantly. All three learning strategies resulted in a network with similar performance (according to mean absolute error, mean percent error and  $r$  values) both for one step and five steps ahead. A careful look at the one-step ahead predictions and particularly the time series (Figs. 6–8) and scatter plots (Figs. 9–11) show that the model has been trained well and presents a good ability to generalize. Figs. 6–8, which are the time series plots for all the univariate models developed, show that all models capture the underlying relationships in the data.

Moreover, from both time series and scatter plots (Figs. 9–11), the generalization ability seems to be sufficient in the boundary flow values. The one-step ahead predictions are more than sufficient, especially when considering that the correlation coefficient  $\rho$  is 0.95. The overall performance of the univariate model can be considered as satisfactory taking into account that  $\rho$  drops from 0.95 (one step ahead) to 0.74 (5 steps ahead), without retraining the network for subsequent steps.

Regarding the learning convergence, which is a way of evaluating the contribution of each of the three learning optimization strategies, two interesting results can be extracted. As seen in Figs. 12–14 which demonstrate the manner in which Mean Square Error (MSE) changes over the number of epochs or training cycles, genetically optimized gradient descent with momentum and the conjugate gradient descent are faster in reaching the lower levels of MSE (values under 0,02).



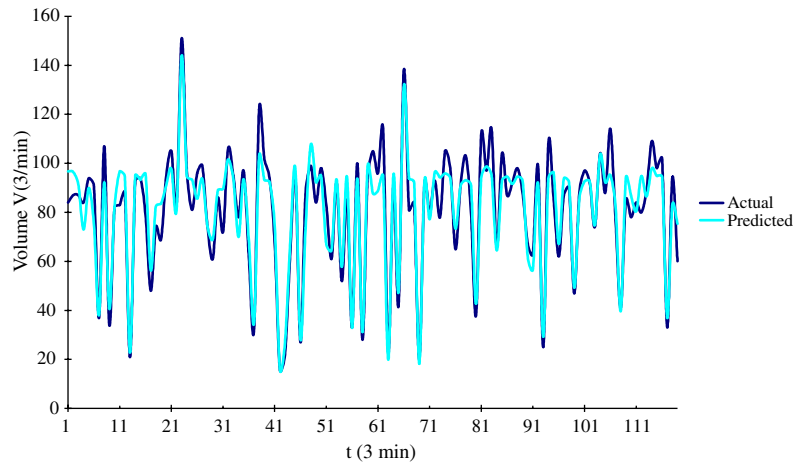


Fig. 6. Time series of univariate model with genetically optimized gradient descent with momentum learning rule.

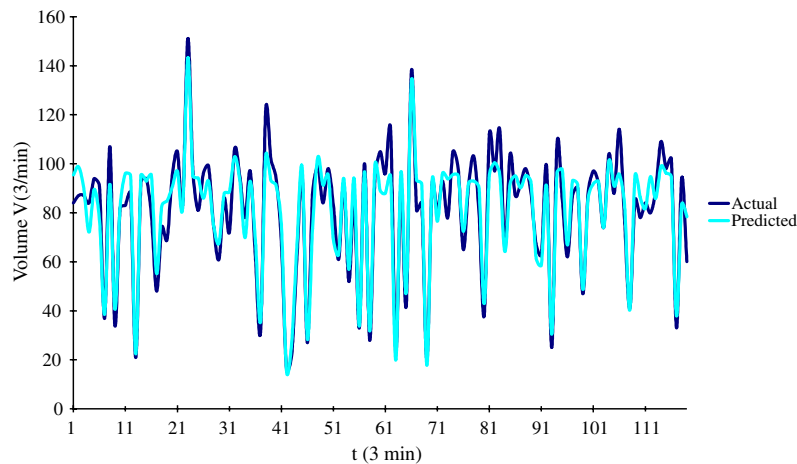


Fig. 7. Time series of univariate model with genetically optimized gradient descent with momentum and adaptive step sizes learning rule.

The first two strategies result in a learning curve that fluctuates much before stabilizing to low values of MSE (Figs. 12–14). On the other hand, the Scaled Conjugate Gradient Descent approach reaches low values of MSE in a smoother way (Fig. 14). Moreover, only the genetically optimized gradient descent with momentum and the genetically optimized gradient descent with adaptive step sizes manage to train in less than 2000 epochs of training. Conjugate gradient is much slower in reaching a result, as expected, without performing any better than the rest of the models. Practically, all three alterations of the gradient descent algorithm have similar performance. Yet, the first is preferable as it has a simpler structure than the other two, it is faster than conjugate gradient and, more importantly, needs less optimization effort than adaptive step sizes.

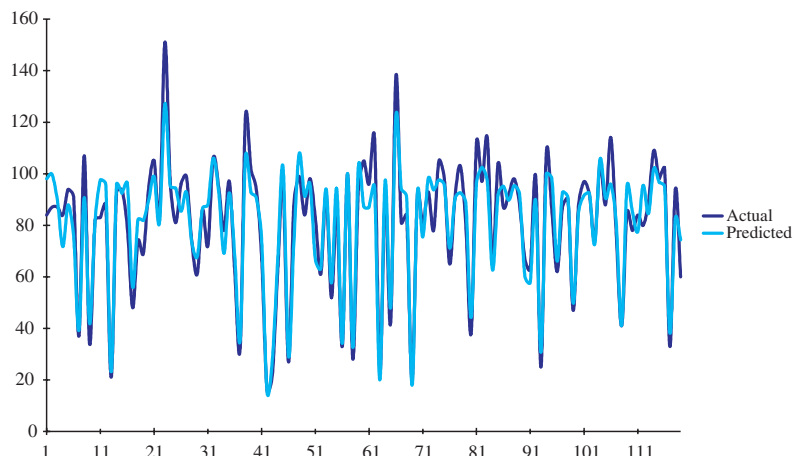


Fig. 8. Time series of univariate model with scale conjugate gradient descent learning rule.

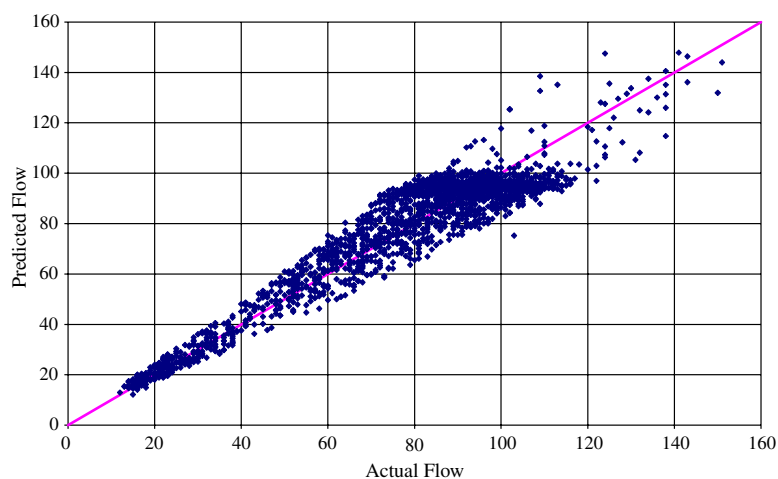


Fig. 9. Scatter plot of univariate model with genetically optimized gradient descent with momentum learning rule.

The effect of the optimization of step size, momentum and hidden layer can be easily understood by comparing the performance of the classical error-back-propagation MLP with the optimized one. In order to establish a comparison, a classical error-back-propagation MLP was trained using a trial-and-error procedure concerning the hidden layer, by starting from an initial value of 10 hidden units and adding 1 unit at every new training until the network reaches the lowest MSE in the cross-validation and several rules-of-thumb to estimate step-size and momentum (details of these rules can be found in Principe et al., 2000; Swingler, 2001). This MLP reached its best predictive performance at 95 hidden units when trained with univariate time series data, 40 more than the optimized MLP. As has been well established in the literature, larger networks imply more degrees of freedom which, in turn, results in significantly higher computational effort.

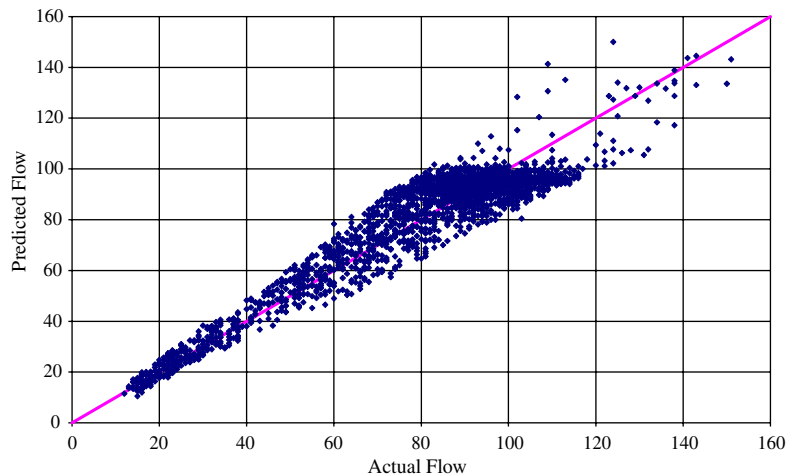


Fig. 10. Scatter plot of univariate model with genetically optimized gradient descent with momentum and adaptive step sizes learning rule.

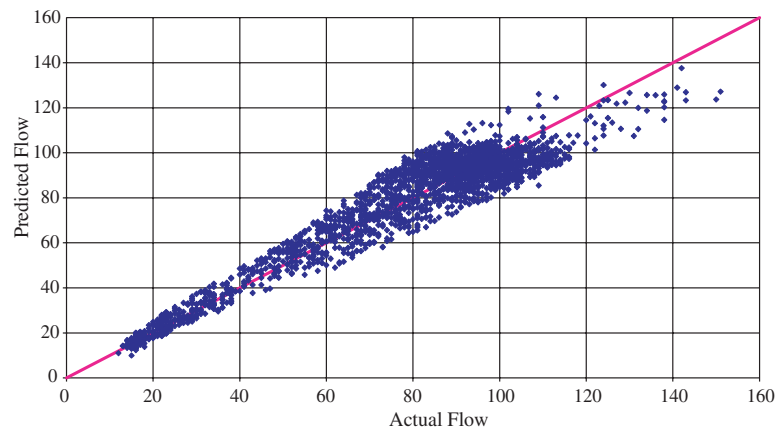


Fig. 11. Scatter plot of univariate model with conjugate gradient descent learning rule.

It is interesting to examine the qualitative implications of the results in the univariate thinking. It is, in general, risky for a forecasting model to predict 15 min ahead using only 9 min of past information. However, the forecasting power is acceptable for only 2 steps ahead, as the accuracy of the predicted flow drops rapidly after 6 min ahead prediction. The accuracy of the predicted flow drops rapidly after 6 min ahead prediction. The accuracy of the neural networks in the two steps ahead forecasting is better than the accuracy given for one step ahead by ARIMA models (Table 3). These results are indicative of a certain kind of robustness of neural networks. Although there is no statistical information on the long-term memory of the series, meaning the probability that the next states will follow the pattern of the three previous ones, neural networks have shown that they can solidly retrieve patterns of flow and generate predictions for more than one step ahead.

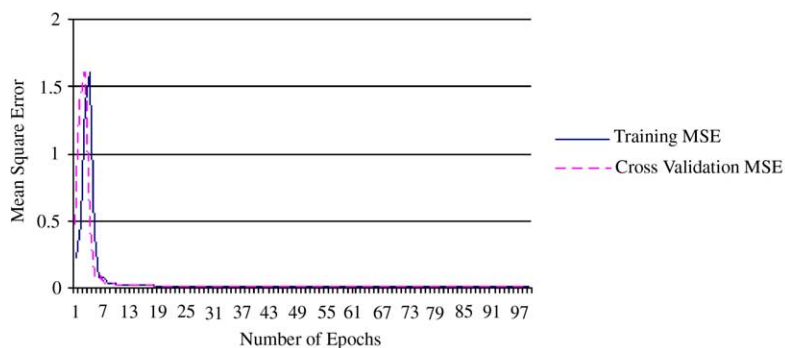


Fig. 12. Learning curve of genetically optimized gradient descent trained neural networks with momentum learning rule.

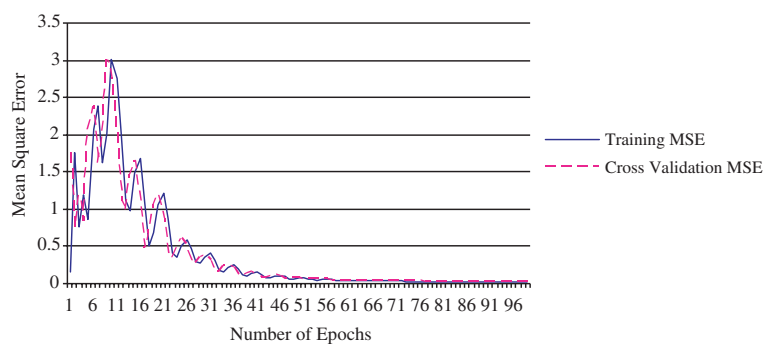


Fig. 13. Learning curve of genetically optimized gradient descent trained neural networks with momentum and adaptive step sizes learning rule.

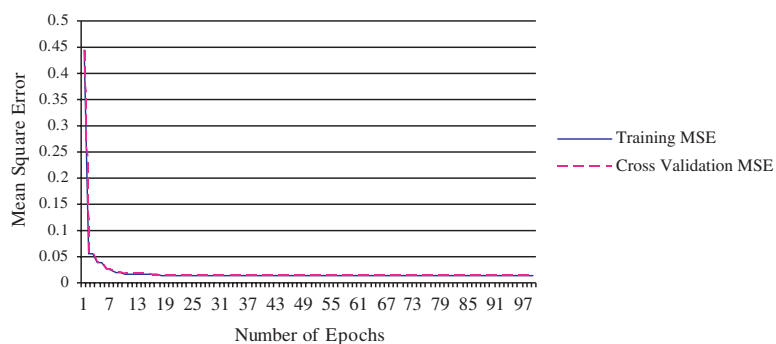


Fig. 14. Learning curve of scaled conjugate gradient descent learning rule.

## 6.2. Multivariate time-series

From the traffic engineering perspective, results taken from the multivariate neural networks are very interesting. The statistical evaluation of the predictive performance shows a slight improvement when using multivariate data in the 5 steps ahead prediction.

Figs. 15 and 16, which are the time series and scatter plots of the actual versus the predicted flow for the Multivariate Model, show that the performance of the MLP was similar to the Univariate Model.

Moreover, the network produced from the genetic optimization of the step size, momentum and hidden layer is similar with the one produced by the univariate time series. In general, this shows that the same network can efficiently model up to 13 input units. However, from the traffic

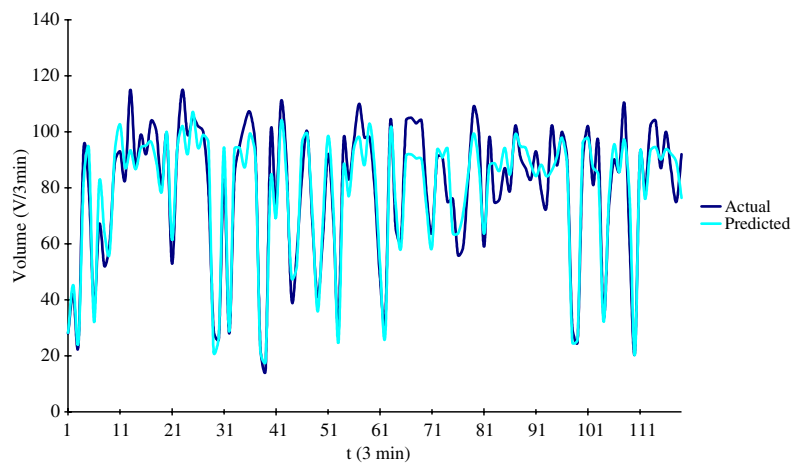


Fig. 15. Time series of multivariate model with genetically optimized gradient descent with momentum learning rule.

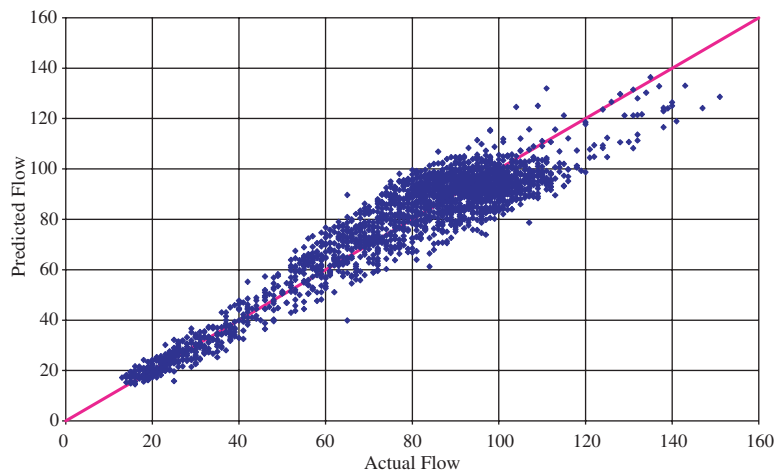


Fig. 16. Scatter plot of multivariate model with genetically optimized gradient descent with momentum learning rule.

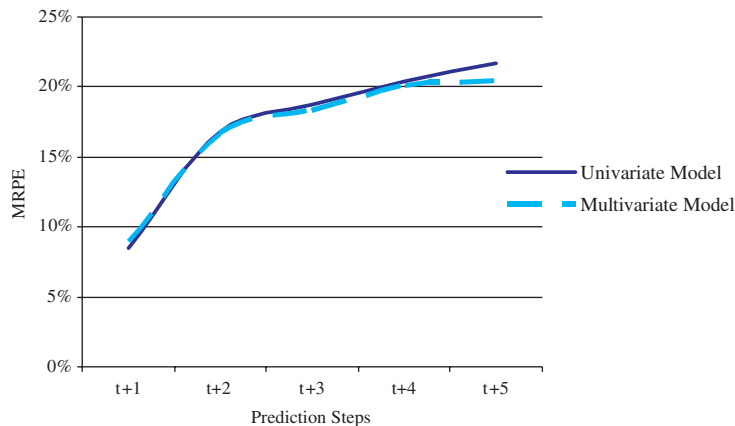


Fig. 17. Mean relative percent error progression of multivariate versus univariate model as a function of prediction step.

engineer's perspective, the main reason for anticipating improved predictions using a multivariate data set is that, in contrast with the Univariate Model which has a 3 step look-back time window, the multivariate one had a total of 5 steps back that implies a more comprehensive data set regarding the information about the evolution of traffic. This is depicted in Fig. 17, where a comparison between the Univariate and the Multivariate Model regarding the trend of MRPE is shown.

It can be seen that the Multivariate Model exhibits a better performance as the predictive horizon increases. Furthermore, and in order to test for the robustness of the MLP developed, the output space was increased to 20 steps ahead keeping the input space unchanged (15 min look-back time window). The results were not satisfactory for 20 steps ahead prediction (MRPE 31,74% and MAE 17 vehicles/3 min). Moreover, neural networks provide two step ahead predictions with equal accuracy compared to State-Space modeling (one step ahead prediction) in the multivariate setting; neural networks exhibited an enhanced accuracy both in the univariate and the multivariate setting.

## 7. Discussion and conclusions

Short-term prediction of traffic parameters such as flow is a crucial element of current ITS structures, yet complicated to formulate mathematically. Its difficulty lies in the nature of traffic flow which is a dynamically evolving phenomenon through both time and space. The present paper proposed an optimization framework for modeling this dynamic behavior through static forms of MLPs. The general idea has two parts; first, to assure the optimal internal performance of the model (with respect to the learning procedure and the network's hidden layer dimensions) and, second, to find a way of incorporating the temporal and spatial characteristics of traffic flow.

From a traffic perspective, although all three models developed capture the trend in the low flow regions, the genetically optimized back-propagation with momentum MLP seems to capture the essence of flow in time intervals where flow fluctuates much between mean values and extreme



ones. In essence, this implies that the genetically optimized network has a much better ability to capture high flow values which is of primary importance to traffic predictions and a major shortcoming of other tested approaches such as the ARIMA models. This result is promising especially in modeling traffic flow taken from highly congested urban signalized arterials.

Regarding the incorporation of a combination of temporal and spatial characteristics, the use of the multivariate data set was found helpful in the cases of multiple steps ahead forecasting. It appears that there is an improvement in the results when increasing the time window from 9 to 15 min ahead and using previous loop detector data. When the predictive time window is 6 min ahead, one can trace a shift in the trend of the Mean Relative Percent Error. While this is indicative of a possible improvement when using multivariate data, these results have to be evaluated with caution and certainly warrant further study. Moreover, the neural network approach gave improved predictions up to 2 steps ahead compared to the univariate ARIMA and multivariate State-Space models for one step ahead predictions (developed in earlier studies with the same data). This seems to indicate that the proposed models are quite robust.

From the results presented in the previous section, two important findings arise concerning model development; first, a non-linear static system was developed that represents a dynamic behavior and, second, this system was optimized with respect to its internal behavior. However, the answer to the question of why one should try to employ static models to capture the temporal and spatial variability of traffic flow, while the literature on neural networks has to exhibit several explicitly temporal structures, is more profound than simple statistical performance evaluation. It is highly related to the notion of optimization regarding the degree of complexity of the neural network. The conceptual notion of optimization in such networks involves the basic trade-off between generalization and accuracy, which is essentially one of finding a way of balancing good generalization and optimal accuracy levels, without increasing the complexity of the model regarding its internal structure (number of hidden layers, number of hidden units, memory lines etc.).

Regarding the use of an advanced optimization technique such as the Genetic Algorithms to calibrate several network parameters, this paper has achieved a combined optimization of two different structural parameters. The greater gain is that this optimization approach provides more means of assuring that the neural network developed has the optimal structure and, consequently, the “best” behavior with respect to the traffic flow set it is required to model. It should be emphasized that the proposed methodology provides means for calibrating certain essential modeling issues of neural networks without attaching to the procedure any topological or traffic characteristics. This is important, as it brings about the ability of the methodology to generalize in different implementation areas (both in terms of geometry and of traffic characteristics). When using data-driven approaches, the scope is to construct a solid methodological framework that, with the proper calibration, could efficiently predict traffic. It is evident that no empirical methodology can result in general conclusions. However, a well-organized and comprehensive approach (both in concept and in implementation) can claim methodological generality.

Moreover, the proposed optimization strategy tested in this paper has exhibited a fine combination between local and global optimization procedures. It first took advantage of the local nature of gradient descent in order to search for the minimum in the network performance surface, but it also exploited the power of the global concept of genetic algorithms in searching in a multivariate space of solutions with a combined optimization of learning parameters and number of hidden units. The overall contribution of genetic algorithms in the forecasting ability of the neural

network models developed as both optimization and meta-optimization techniques, led to a more confined structure of MLPs which is consistent with the data asked to model. Several confusing and, at times, inadequate rules-of-thumb concerning the selection of the value of the learning parameters such as the step size and momentum and the number of hidden units, can thus be avoided. The resulting MLP structure is closer to the simplest and optimal structure that could learn the specific traffic flow data.

## References

- Abdulhai, B., Porwal, H., Recker, W., 1999. Short-term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks. Transportation Research Board, 1999.
- Angeline, P.J., Saunders, G.M., Pollack, J.B., 1994. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks* 5 (1), 54–65.
- Bishop, Ch.M., 1995. *Neural Networks For Pattern Recognition*. Oxford university press, ISBN 0-19-853864-2.
- Chang, G.-L., Su, C.-C., 1995. Predicting Intersection Queue With Neural Network Models. *Transportation Research Part C* 3 (3), 175–191.
- Chen, H., Grant-Muller, S., 2001. Use of Sequential Learning for Short-Term Traffic flow Forecasting. *Transportation Research Part C* 9, 319–336.
- Chen, H., Grant-Muller, S., Mussone, L., Montgomery, F., 2001. A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting. *Neural Computing and Applications* 10, 277–286.
- Clark, S.D., Dougherty, M.S., Kirby, H.R., 1993. The use of neural networks and time series models for short-term traffic forecasting: a comparative study. *PTRC 21st Summer Annual Meeting*.
- Clark, S., 2003. Traffic prediction using multivariate nonparametric regression. *Journal of Transportation Engineering* 129 (2), 161–168.
- Dia, H., 2001. An Object-Oriented Neural Network Approach to Short-Term Traffic Forecasting. *European Journal of Operational Research* 131, 253–261.
- Dougherty, M.S., Cobbet, M.R., 1997. Short-Term Inter-urban Traffic Forecasts using Neural Networks. *International Journal of Forecasting* 13, 21–31.
- Eiben, A.E., Smith, J.E., 2003. *Introduction to Evolutionary Computing*. Springer, ISBN 3-540-40184-9.
- Florio, L., Mussone, L., 1996. Neural Network Models for Classification and Forecasting of Freeway Traffic Flow Stability. *Control Engineering Practice* 4 (2), 153–164.
- Gen, M., Cheng, R., 2000. *Genetic Algorithms and Engineering Optimization*. Wiley, ISBN 0-471-31531-1.
- Gilmore, J.E., Abe, N., 1995. Neural network models for traffic control and congestion prediction. *IVHS Journal* 2 (3), 231–252.
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, ISBN 0201157675.
- Hamed, M.M., Al-Masaeid, H.R., Bani Said, Z.M., 1995. Short-term prediction of traffic volume in urban arterials. *ASCE Journal of Transportation Engineering* 121 (3), 249–254.
- Haykin, S., 1994. *NN: a comprehensive foundation*. Macmillan, NY.
- Head, L.K., 1995. Event-based Short-term Traffic Flow Prediction Model. *TRB* 1510, 45–52.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feed-forward networks are universal approximators. *Neural Networks* 2, 359–366.
- Hush, D.R., Horne, B.G., 1993. Progress in supervised Neural Networks: What's new since Lippman? *IEEE Signal Progress Magazine*.
- Innamaa, S., 2000. Short-term prediction of traffic situation using MLP-Neural Networks. In: *7th World Congress on Intelligent Transportation Systems*, Turin, Italy.
- Ishak, S., Alecsandru, C., 2003. Optimizing traffic prediction performance of Neural networks under various topological, input and traffic condition settings. In: *Transportation Research Board 82nd Annual Meeting*, Washington, DC.

- Ishak, S., Kotha, P., Alecsandru, C., 2003. Optimization of dynamic neural networks performance for short-term traffic forecasting. In: *Transportation Research Board 82nd Annual Meeting*, Washington, DC.
- Kirby, H., Dougherty, M., Watson, S., 1997. Should we use neural networks or statistical models for short term motorway forecasting. *International Journal of Forecasting* 13, 45–50.
- Kwon, E., Stephanedes, Y.J., 1994. Comparative Evaluation Of Adaptive And Neural-Network Exit Demand Prediction For Freeway Control. *Transportation Research Record*, 1446.
- Ledoux, C., 1997. An Urban Traffic Flow Model Integrating Neural Networks. *Transportation Research Part C* 5 (5), 287–300.
- Lee, S., Fambro, D.B., 1999. Application of subset autoregressive moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record* 1678, 179–188.
- Levin, M., Tsao, Y.-D., 1980. On forecasting freeway occupancies and volumes. *Transportation Research Record* 773, 47–49.
- Lingras, P., Mountford, P., 2001. Time Delay Neural Networks Designed Using Genetic Algorithms for Short-term Inter-city Traffic Forecasting, IEA/AIE 2001, LNAI 2070, pp. 290–299.
- Lyons, G.D., McDonald, M., Hounsell, N.B., Williams, B., Cheese, J., Radia, B., 1996. Urban traffic management: the viability of short-term congestion forecasting using artificial neural networks. In: *24th European Transport Forum*, PTRC.
- Masters, T., 1993. *Practical Neural Networks Recipes in C++*. Academic Press, Inc. Hancourt Brace Jovanovich, Publishers.
- Mitchell, M., 1998. *An introduction to genetic algorithms*. MIT Press, ISBN 0262631857.
- Okutani, I., Stephanedes, Y.J., 1984. Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research, Part B* 18 (1), 1–11.
- Park, D., Rilett, L.R., 1998. Forecasting multiple-period freeway link travel times using modular neural networks. *Transportation Research Record* 1617, 163–170.
- Principe, J.C., Euliano, N.R., Lefebvre, C.W., 2000. *Neural And Adaptive Systems: Fundamentals Through Simulations*. John Wiley and Sons, Inc.
- Reed, R.D., Marks II, R.J., 1998. *Neural Smthing. Supervised learning in feedforward artificial neural networks*. MIT Press, Cambridge, Massachusetts.
- Sarle, W.S. (Ed.), 1998. Frequently Asked Questions about Neural Networks. Hypertext document: <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Smith, B.L., Demetsky, M.J., 1997. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering* 123 (4), 261–266.
- Smith, B.L., Williams, B.M., Oswald, K.R., 2002. Comparison of parametric and non-parametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* 10 (4), 303–321.
- Smith, B.L., Demetsky, M.J., 1994. Short-term Traffic Flow Prediction: Neural Network Approach. *TRB* 1453, 98–104.
- Stathopoulos, A., Karlaftis, M.G., 2001a. Temporal and spatial variations of real-time traffic data in urban areas. *Transportation Research Record Journal of the Transportation Research Board* 1768, 135–140.
- Stathopoulos, A., Karlaftis, M.G., 2001b. Spectral and cross-spectral analysis of traffic flow. In: *IEEE 4th Conference in Intelligent Transportation Systems*, San Francisco, California, USA, August, 2001.
- Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state-space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies* 11 (2), 121–135.
- Sun, H., Liu, H.X., Xiao, H., He, R.R., Ran, B., 2003. Short-term trtraffic forecasting using the local linear regression model. In: *Transportation Research Board 82nd Annual Meeting*, Washington, DC.
- Swingler, K., 2001. *Applying Neural Networks: A practical guide*. Morgan Kaufman Publishers, Inc, ISBN 0-12-679170-8.
- Van Lint, J.W.C., Hoogendoorn, S.P., Van Zuylen, H.J., 2002. Freeway travel time prediction with state-space neural networks: modeling state-space dynamics with recurrent neural networks. In: *Transportation Research Board 81st Annual Meeting*, Washington, DC.
- Vythoulkas, P.C., 1993. Alternative approaches to short-term traffic forecasting for use in driver information systems. In: Berkley, C.F. Dagango (Ed.), *Transportation and Traffic Theory, Proceedings of the 12th International Symposium on Traffic Flow Theory and Transportation*. Elsevier.

- Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC Press, London, ISBN 1-58488-030-9.
- Whittaker, J., Garside, S., Lindevel, K., 1997. Tracking and predicting network traffic process. *International Journal of Forecasting* 13, 51–61.
- Williams, B.M., 2001. Multivariate Vehicular Traffic Flow Prediction: An Evaluation of ARIMAX Modeling. *Transportation Research Record* 1776, 194–200.
- Williams, B.M., Durvasula, P.K., Brown, D.E., 1998. Urban Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *Transportation Research Record* 1644, 132–144.
- Winter, G., Periaux, J., Galan, M., Cuesta, P., 1996. *Genetic Algorithms in Engineering and Computer Science*. John Wiley & Sons, ISBN 047195859X.
- Yao, X., 1999. Evolving artificial neural networks. *Proceedings of the IEEE* 87 (9), 1423–1447 (Won the 2001 IEEE Donald G. Fink Prize Paper Award).
- Yasdi, R., 1999. Prediction of Road Traffic using a Neural Network Approach. *Neural Computation and Application* 8, 135–142.
- Yun, S.-Y., Namkoong, S., Rho, J.-H., Shin, S.-W., Choi, J.-U., 1998. A Performance Evaluation of Neural Network Models in Traffic Volume Forecasting. *Mathematical and Computer Modelling* 27 (9–11), 293–310.
- Zhang, G., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of art. *International Journal Of Forecasting* 14, 35–62.
- Zhang, H.M., 2000. Recursive Prediction of Traffic Conditions With Neural Networks. *Journal of Transportation Engineering* (November/December), 472–481.