

Local and Global Iterative Algorithms for Real-Time Short-term Traffic Flow Prediction

Eleni I. Vlahogianni, Ph.D and Matthew G. Karlaftis, Ph.D.
*National Technical University of Athens
Greece*

1. Introduction

When considering short-term prediction systems that operate in real-time and in an "intelligent" technology-based environment, the effectiveness depends, mostly, on predicting traffic information in a timely manner (Smith & Oswald, 2003). This implies that regardless of the traffic conditions met in real-time, a prediction system should not only be able to generate accurate single step ahead predictions of traffic flow, but also to operate on an iterative basis to produce reliable multiple steps ahead predictions in cases of data collection failures.

The bulk of research in short-term traffic flow prediction has concentrated on data-driven time-series models that construct the underlying rules of complex traffic datasets rather than working based on pre-determined mathematical rules; these models can be parametric - such as ARIMA models - or non-parametric (a review of the literature, methodologies and approaches used can be found in Vlahogianni et al., 2004). Among non-parametric approaches, neural networks and especially Multi-layer Feed-forward Perceptrons (MFLPs) have been shown to be most effective in forecasting traffic flow variables because of their propensity to account for a large range of traffic conditions and generate more accurate predictions than classical statistical forecasting algorithms (Smith & Oswald, 2003, Vlahogianni et al. 2004, Vlahogianni et al., 2005, Wang et al., 2006). Multi-layer Feed-forward Perceptrons (MFLPs) have memoryless neurons while input activation is a function only of the current input state and not of the past input-output relations and are very popular and widely used in time series traffic forecasting partly because of their ability to capture non-linear behavior in the data structures they model, regardless of their complexity (Hornik et al., 1989). They are usually trained under a pattern-based consideration; computations aim at distinguishing clusters that have different statistical properties (mean and variance).

However, the static operation of MLFP contradicts the temporal evolution of traffic flow. Short-term traffic flow prediction is significantly affected by the temporal evolution of traffic variables (Abdulhai et al., 2002, Stathopoulos & Karlaftis, 2003). The core research in traffic flow over the years has investigated a wide variety of nonlinear phenomena such as phase transitions (Kerner & Rehborn 1996), hysteretic effects (Hall et al. 1992, Zhang 1999), localized spatio-temporal behavior (Kerner 2002) and others. These phenomena - mainly observed in freeway bottlenecks - are difficult to predict and replicate in a simulation environment based on classical traffic flow theory (Kerner 2004). Moreover, recent efforts in

Source: Urban Transport and Hybrid Vehicles, Book edited by: Seref Soylu,
ISBN 978-953-307-100-8, pp. 192, September 2010, Sciyo, Croatia, downloaded from SCIYO.COM

studying the temporal evolution of traffic flow in signalized arterials have indicated that traffic volume patterns - a sequence of volume states that define the time window for prediction - exhibit significant temporal variability and demonstrated that short-term traffic flow has variable deterministic and nonlinear characteristics that can be related to the prevailing traffic flow conditions (Vlahogianni et al. 2006a, 2008).

Although traffic flow theory has formulated improved theoretical bases in order to account for such phenomena - for example the three-phase traffic flow theory (Kerner 2004) - these effects are not taken into consideration in the process of short-term predictions using data-driven algorithms. Moreover, as indicated in most intelligent transportation system architectures, the dynamic nature of prediction algorithms are imperative to confront probable malfunctions in the data collection system or excesses in computational time (Smith & Oswald 2003, Vlahogianni et al. 2006b).

Iterative predictions provide the means to generate information on traffic's anticipated state with acceptable accuracy for a significant time horizon in cases of system failure. The present paper focuses on providing a comparative study between local and global iterative prediction techniques applied to traffic volume and occupancy time-series. The remainder of the paper is structured as follows: The concept of iterative predictions along with the basic notions of the proposed local and global prediction techniques in traffic flow prediction is presented in the following section. Next, the characteristics of the implementation area as well as the results of the comparative study are summarized. Finally, the paper ends with some concluding remarks.

2. Iterative predictions of traffic flow

Most predictions systems are dependent on data transmission. This suggests that continuous flow of volume and occupancy data is necessary to operate efficiently. However, it is common for most real-time traffic data collection systems to experience failures (Stathopoulos & Karlaftis 2003). For this, a real-time prediction system should be able to generate predictions for multiple steps ahead to ensure its operation in cases of data collection failures. The multiple steps ahead prediction problem can be formulated based on two conceptual approaches:

- *Direct approach:* Given a time-series of a variable - for example volume $V(t), V(t-1), \dots$, a model is constructed to produce the state of a variable at $V(t+h)$ steps ahead, where $h > 1$.
- *Iterative or recursive approach:* Given a time-series of a variable a single step ahead model is constructed to produce a prediction $\hat{V}(t)$ at time t that is then fed backwards to the network and is used as new input data in order to produce the next step $\hat{V}(t+1)$ prediction at $t+1$:

$$\hat{V}(t+1) = \{\hat{V}(t), V(t), V(t-1), \dots\} \quad (1)$$

The direct approach to prediction has been implemented using neural networks in Vlahogianni et al. (2005). The reliability of direct prediction models is suspect because the model is forced to predict further ahead (Sauer 1993). This is the main argument in using iterative models in multiple steps ahead prediction. On the other hand, iterative predictions have the disadvantage of using the predicted value as input that is probably corrupted. Literature in short-term prediction of chaotic time-series indicates that the iterative approach to multiple steps ahead prediction is more accurate than direct one (Casdagli 1992).

Iterative approaches to neural network modeling have previously been applied to traffic flow predictions and results indicate that predictability decreases when the model uses previous predictions as inputs in the process of short-term predictions (Zhang 2000). This can be probably attributed to the fact that short-term traffic flow dynamics are not smooth and could not be captured by the simple structures of MLFP implemented so far (Vlahogianni et al. 2005). Two distinct categories of prediction techniques are further investigated; the local linear prediction techniques and the global neural networks. The next sub-sections provide the basics of the above techniques.

2.1 Local Weighted Linear model

The locally weighted linear (LWL) model finds similar local dynamics and produces a good estimate of the future state of a system, while neural networks are global approaches that attempt to build a single complex model for the entire range of behaviors identified in the time series. LWL belongs to the category of memory-based local prediction models. Locality in prediction is achieved by utilizing only the nearby states of the variable to be predicted. In brief, taken the case of predicting the next state of traffic volume $V(t+\tau)$ where τ is the embedding delay that matches the prediction step, the algorithm searches the k nearest neighbors of $\mathbf{V}(t)$ i.e. the k nearest states $\mathbf{V}(t')$ ($t' < t$) that minimize $\|\mathbf{V}(t) - \mathbf{V}(t')\|$, where $\|\cdot\|$ a metric (for example Euclidean) (Casdagli 1992). The local linear predictor results by fitting a linear polynomial to the pairs of $\mathbf{V}(t)$, $\mathbf{V}(t'+\tau)$; the contribution of each of the k nearest neighbors is weighted by a weighting function (kernel) $K = e^{-\frac{d}{h}}$ - d 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 (Kononov 2007).

2.2 Temporal neural networks

The global iterative model to be used as an alternative theoretical approach to the proposed prediction framework is the temporal structures of neural networks. These networks are an extension of the static MLFPs that engage a memory mechanism to reconstruct the time-series under prediction in the m -dimensional Phase Space. The memory can be limited to the input layer or can extend to the entire neural networks - for example the hidden layer - and aims at creating a State-Space reconstruction process and converts the time series data of volume and occupancy in vectors; for example $\vec{V}_t = \{V_{t-\tau}, \dots, V_{t-(m-1)\tau}\}$, where τ is the time delay of and m is the dimension the reconstructed volume in the State-Space. Networks that encompass tap delay memory structures only to the input layer of the form $\mathbf{y}(n) = [y(n), \dots, y(n-m+1)]^T$ are known as time-delayed neural networks (TDNN), whereas networks that incorporate more complex memory mechanisms - for example a Gamma memory (de Vries & Principe 1992) - are known as time-lagged neural networks (TLNN).

The training of TLNN for iterative predictions feeds back the prediction at time $t+1$ and utilizes it as an input for the generation of next prediction step $t+2$. For global models, this simple method is plagued by the accumulation of errors and the model can quickly diverge from the desired behavior (Principe et al. 2000). For this reason the training in the specific iterative neural network model is conducted via the temporal back-propagation algorithm known as Back-propagation to time (BPTT) (Webros 1990). BPTT are trajectory learning algorithms that create virtual networks by unfolding in time the original topologies on which back-propagation training is applied; the main objective is to improve the weights so

as to minimize the total error over the whole time interval (Webros 1990). More specifically, all weights are duplicated spatially for an arbitrary number of time steps τ ; as such, each node that sends activation to the next has τ number of copies as well. For a training cycle n , the weight update is given by the following equation (Haykin 1999):

$$\mathbf{w}_{ji}(n+1) = \mathbf{w}_{ji}(n) + \eta \delta_j(n) \mathbf{x}_i(n) \quad (2)$$

where, $\mathbf{w}_{ji}(n+1)$ and $\mathbf{w}_{ji}(n)$ are the weights of the i -th synapse of the neuron j at training cycle $n+1$ and n respectively, η is the learning rate, $\mathbf{x}_i(n)$ ($i=1,2,\dots,n$) is the input vector and $\delta_j(n)$ is given by:

$$\delta_j(n) = \begin{cases} e_j(n) \phi'(v_j(n)), & j \text{ neuron in the output layer} \\ \phi'(v_j(n)) \sum_{r \in A} \Delta_r^T(n) \mathbf{w}_{rj}, & j \text{ neuron in the hidden layer} \end{cases} \quad (3)$$

where, $e_j(n)$ is the network's error, ϕ is the nonlinear activation function. Moreover, if A is a set of all neurons whose inputs are fed by the j neuron in the hidden layer in a forward manner, then $v_j(n) = \sum_{i=1}^m \mathbf{w}_{ji} \mathbf{x}_i(n) + b_j$ is the induced local field of neuron r that belongs to the

A and $\Delta_r^T(n) = [\delta_r(n), \delta_r(n+1), \dots, \delta_r(n+m)]^T$ is the local gradient vector.

Regardless of being static or dynamic, neural networks suffer from a usually manual trial-and-error process of optimization mainly with respect to their structure (number of hidden units) and learning parameters. On the other hand, automation in the optimization process of a neural network is most critical in complex neural structures. Recently, genetic algorithms have gained significant interest as they can be integrated to the neural network training to search for the optimal architecture without outside interference, thus eliminating the tedious trial and error work of manually finding an optimal network. Genetic algorithms are based on three distinct operations: selection, cross-over and mutation (Mitchell 1998); these operations run sequentially in order for a fitness criterion (in the specific case the minimization of the cross-validation error) to converge.

3. Implementation and results

The specific iterative prediction approach is implemented using 90-seconds volume and occupancy data collected along a signalized arterial in the centre of Athens (Greece) (Fig 1). Previous research on the area has demonstrated that traffic flow in the specific area - defined by the joint consideration of volume and occupancy - has a variable deterministic and nonlinear temporal evolution (Vlahogianni et al. 2008). The subsequent reconstruction of the series of volume and occupancy in the Phase-space using the mutual information criterion (Fraser & Swinney 1986) and the false nearest algorithm (Kennel et al. 1992) indicate that both volume and occupancy can be determined by a vector with time delay that equals to 1 and dimension that equals to 5.

Based on the above results, the recursive prediction techniques are set to have a look-back temporal window of 1-hour to retrieve and evaluate the traffic flow patterns. Moreover, for the recursive TDNN training, the neural network is set to be unfolded in an $m=5$ dimensional space in order to produce predictions. The depth of the Gamma of the TLNN was also calibrated to reflect the above reconstruction. Apart from the memory mechanism,

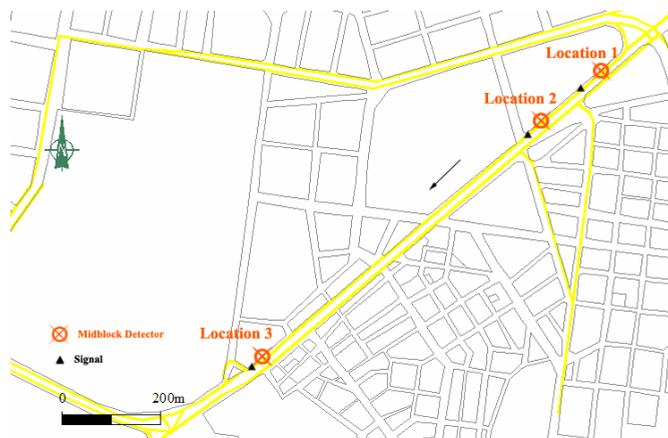


Fig. 1. Representation of the set of arterial links under study.

the difference between TDNN and TLNN implemented is that the second network extends the memory mechanism to the hidden layer too, in order to provide a fully non-stationary environment for the temporal processing of volume and occupancy series. The specifications regarding data separation as well as the genetic algorithm optimization are depicted on Table 1.

Parameters	Specifications
Datasets: TR-CV-TE *	60%-20%-20%
Levels	1 hidden layer
Optimization	Genetic algorithm
Back-propagation	Genetic algorithm
Chromosome	$h \in [5,25]$, $\gamma \in [0.01 - 0.1]$, $\mu \in [0.5 - 0.9]$ **
Fitness function	Mean square error (cross-validation set)
Selection	Roulette
Cross-over	Two point ($p=0.9$)
Mutation	Probability $p=0.09$

* Training - Cross-validation - Testing

** h : neurons in hidden layer, γ : learning rate, μ : momentum

Table 1. Data and neural network specifications for iterative short-term volume and occupancy prediction.

The results of the comparative study are summarized in Table 2. As can be observed the TLNN performs significantly better - with regards to the mean relative percent error of prediction - than the local weighted linear model under the iterative prediction framework in both volume and occupancy. When compared to the iterative predictions of a TDNN, it is observed that TDNN performs comparable to the TLNN. However, as the same does not apply to the case of occupancy; further statistical investigation is conducted to the series of volume and occupancy in order to explain the behavior of the models regarding occupancy predictions. Results from a simple LM ARCH (Eagle 1982) that tests the null hypothesis of no ARCH effect lying in the data series of volume and occupancy shows that occupancy exhibits higher time-varying volatility than volume that is difficult to be captured.

Iterative Models	Mean Relative Percent Error (%)	
	Volume	Occupancy
LWL	21	30
TDNN	14	26
TLNN	13	22

Table 2. Prediction Results (Mean Relative Percent Error) of the comparative study.

Fig 2 and Fig 3 depict the relationship of the actual and the predicted values of volume and occupancy equally. A systematic error is observed in the predictions of volume using the local prediction model. Moreover, there seems to be a difficulty in predicting high volume values as observed in Fig. 2. As for the occupancy predictions, there seem to be much more scattered than the ones of volume; R2 values are lower than the ones of volume.

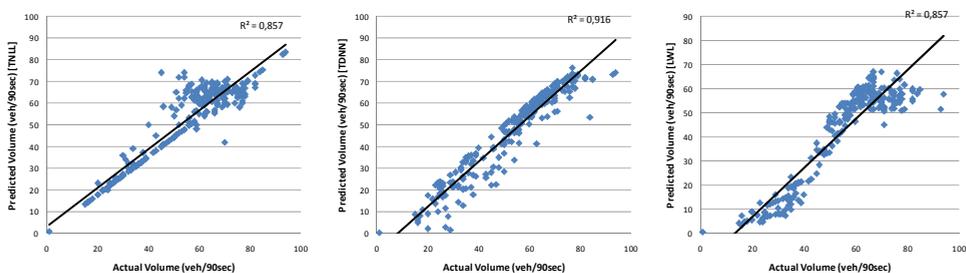


Fig. 2. Actual versus predicted values of traffic volume for the three iterative prediction techniques evaluated.

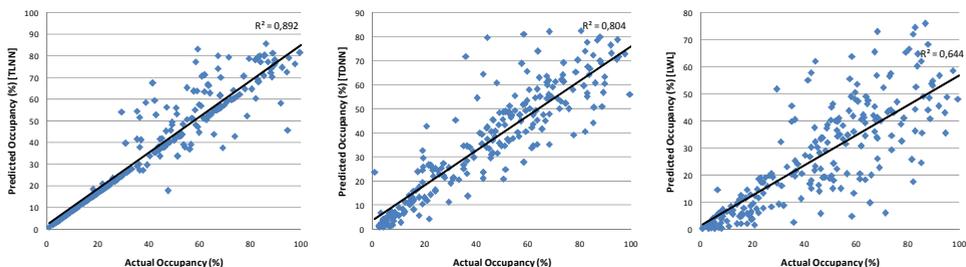


Fig. 3. Actual versus predicted values of occupancy for the three iterative prediction techniques evaluated.

In order to investigate the performance of the iterative models during the formation of congested conditions, two distinct time periods are selected for further studying the time series of the actual and predicted volume and occupancy with regards to different methodologies. These two periods depict the onset of the morning (Figure 4) and the afternoon peak (Fig. 5).

As can be observed, although iterative TLNN exhibited improved mean relative accuracy when compared to the iterative TDNN, both models seem to capture the temporal evolution of the two traffic variables under study. In the case of afternoon peak where the series of volume exhibit a oscillating behavior - in contrast to the trend observed in volume and occupancy during the onset of the morning peak, both neural network models either over-

estimate of under-estimate the anticipated values of traffic volume. As for the LWL model, predictions as depicted in the time series of the actual versus the predictive values of traffic volume and occupancy can be considered as unsuccessful.

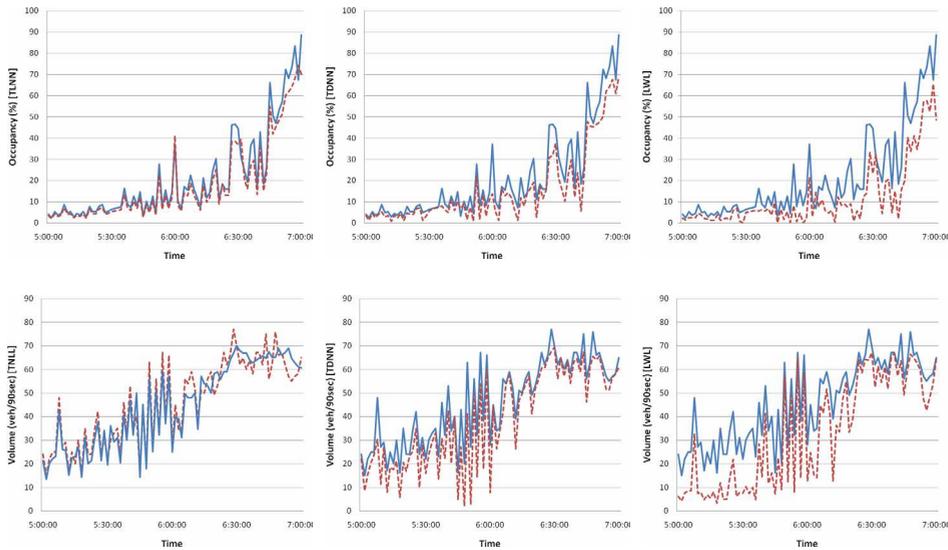


Fig. 4. Time-series of actual and predicted (dashed line) values of traffic volume (vh/90sec) for the onset of the morning peak.

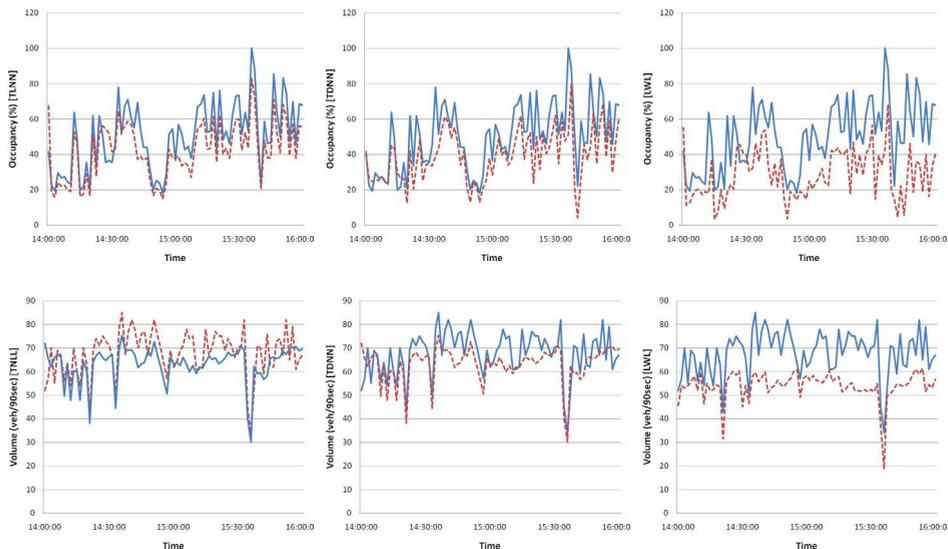


Fig. 5. Time-series of actual and predicted (dashed line) values of traffic volume (vh/90sec) for the onset of the afternoon peak.

4. Conclusions

Modern intelligent transportation systems require prediction algorithms that are adaptable and self-optimized in terms of the prevailing traffic flow conditions. Neural networks have been for long considered a prominent approach short-term prediction of traffic variables. The present paper extends past research by focusing on purely temporal structures of neural networks that provide iteratively short-term traffic flow predictions. A comparative study is conducted between local prediction techniques and neural networks with respect to the predictive accuracy. Results indicate that the global neural networks techniques outperform the local predictors, both when considering the mean behavior of the models and their behavior in critical traffic flow conditions, such as the onset of the morning and afternoon peak in signalized arterials. The optimal accuracy is attained by the TLNN that is the most complex temporal neural network among those tested.

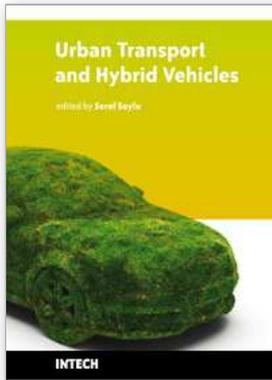
From a conceptual standpoint, the TLNN implemented is fully compatible with the complex non-stationary features of traffic flow. From a methodological standpoint a central consideration should be kept in mind; as the aim is mainly at the real-time implementation, the extensive computational time to train and optimize such networks should be considered. It is evident that a retraining strategy is needed in order for the neural structures to incorporate and learn newly observed traffic flow events. Although the last is not required during the entire real-time operation of the model, research should be focus on the manner the accuracy of iterative predictions decreases over time, as well as the formulation of a mathematical or empirical criterion to evaluate the time neural networks should be retrained.

5. References

- B. Abdulhai, H. Porwal and W. Recker, "Short-term traffic flow prediction using neuro-genetic algorithms," *Intelligent Transportation Systems Journal*, vol.7, no. 1, pp. 3-41, Jan. 2002.
- M. Casdagli, "Chaos and deterministic versus stochastic non-linear modeling", *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 54, No. 2, pp. 303-328, 1992.
- R.F. Engle, Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987- 1008, 1982.
- J. D. Farmer and J. J. Sidorowich, "Predicting Chaotic Time-Series", *Phys. Rev. Lett.*, vol. 59, issue 8, pp. 845 - 848, Aug 1987.
- A. M., Fraser and H. L., Swinney, (1986). Independent coordinates for strange attractors from mutual information, *Physical Review A*, 33 (2), pp. 1134-1140.
- F. L Hall, V. F. Hurdle and J. H. Banks, "Synthesis of Occupancy (Or Density) Relationships on Freeways," *Transportation Research Record* 1365, TRB, National Research Council, Washington, DC, pp. 12-18, 1992.
- S. Haykin, *Neural Networks: A comprehensive foundation*, Prentice Hall Upper Saddle River, NJ, 1999.
- K. Hornik, M. Stinchcombe and H. White, "Multilayer feed-forward networks are universal approximators," *Neural Networks*, vol.2, no. 5, pp. 359-366, Jul. 1989.

- M. B., Kennel, R., Brown and H. D. I., Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physics Reviews A*, 45, 3403, 1992.
- B. S. Kerner and H. Rehborn, "Experimental properties of complexity in traffic flow," *Phys. Rev. E* vol.53, no. 5, pp. R4275 - R4278 May 1996.
- B. S. Kerner, "Empirical macroscopic features of spatial-temporal traffic patterns at highway bottlenecks," *Physical Review E*, vol.65, no. 4, pp. 046138 1-30, Apr. 2002.
- B. S. Kerner, *The Physics Of Traffic: Empirical Freeway Pattern Features, Engineering Applications, And Theory*. Springer Verlag, New York., 2004.
- E. Kononov, *Visual Recurrence Analysis* version 5, May 2007.
- M. Mitchell. *An introduction to genetic algorithms*. MIT Press, ISBN: 0262631857, 1998.
- N. Y. Nikolaev and H. Iba, *Adaptive Learning of Polynomial Networks: Genetic Programming, Backpropagation and Bayesian Methods*, Springer, NY, 2006.
- J. C. Principe, N. R. Euliano and C. W. Lefebvre, *Neural and Adaptive Systems: Fundamentals Through Simulations*. John Wiley and Sons, Inc, 2000.
- T. Sauer, Time series prediction by using delay coordinate embedding, in: Weigend A.S., Gershenfeld N.A. (Eds.), *Time series prediction: Forecasting the future and understanding the past*, Addison Wesley, Reading MA, 1993.
- B. L. Smith and R. K. Oswald, "Meeting Real-Time Requirements with Imprecise Computations: A Case Study in Traffic Flow Forecasting," *Computer Aided Civil and Infrastructure Engineering*. vol. 18, no. 3, pp. 201-213, May 2003.
- A. Stathopoulos and M. G., Karlaftis, "A multivariate state-space approach for urban traffic flow modelling and prediction," *Transportation Research Part C*, 11(2), 121-135, April 2003.
- E. I. Vlahogianni, J. C. Golias and M. G. Karlaftis, "Short-Term Traffic Forecasting: Overview of Objectives and Methods," *Transport Reviews*, vol. 24, no. 5, pp. 533-557, Sept. 2004.
- E. I. Vlahogianni, M. G. Karlaftis and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research C*, vol.13, no. 3, pp. 211-234, June 2005.
- E. I. Vlahogianni, M. G. Karlaftis and J. C. Golias, "Statistical Methods for Detecting Nonlinearity and Non-Stationarity in Univariate Short-term Time-series of Traffic Volume," *Transportation Research Part C*,14(5), 351-367, 2006.
- E. I. Vlahogianni, M. G. Karlaftis, J. C. Golias, and N. D. Kourbelis, *Pattern-based Short-term Urban Traffic Predictor*. Proceedings of the 9th International IEEE Conference on Intelligent Transportation Systems, September, Toronto, Canada, 2006.
- E. I. Vlahogianni, M. G. Karlaftis and J. C. Golias, *Statistical Identification of Temporal Patterns in Short-term Urban Traffic Flow*, *Computer-Aided Civil and Infrastructure Engineering*, accepted for publication, 2008.
- B. de Vries and J. C. Principe, The gamma model—A new neural-net model for temporal processing, *Neural Networks*, 5, 565-576, 1992.
- Y. Wang, M. Papageorgiou, and A. Messmer, "A Real-Time Freeway Network Traffic Surveillance Tool," *IEEE Transactions on Control Systems Technology*, vol.14, no. 1, pp. 18-31, Jan. 2006.

- P. J. Webros, "Backpropagation Through time: What it does and How to do it", IEEE Proceedings Vol 78, No. 10, 1550-1567 Oct. 1990.
- H. M. Zhang, "A mathematical theory of traffic hysteresis," Transportation Research Part B, vol.33, no. 1, pp. 1-23. Feb. 1999.
- H. M. Zhang, "Recursive Prediction of Traffic Conditions With Neural Networks," Journal of Transportation Engineering, 126(6), 472-481, 2000.



Urban Transport and Hybrid Vehicles

Edited by Seref Soylu

ISBN 978-953-307-100-8

Hard cover, 192 pages

Publisher Sciyo

Published online 18, August, 2010

Published in print edition August, 2010

This book is the result of valuable contributions from many researchers who work on both technical and nontechnical sides of the field to be remedy for typical road transport problems. Many research results are merged together to make this book a guide for industry, academia and policy makers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Eleni Vlahogianni and Matthew Karlaftis (2010). Local and Global Iterative Algorithms for Real-Time Short-Term Traffic Prediction, Urban Transport and Hybrid Vehicles, Seref Soylu (Ed.), ISBN: 978-953-307-100-8, InTech, Available from: <http://www.intechopen.com/books/urban-transport-and-hybrid-vehicles/local-and-global-iterative-algorithms-for-real-time-urban-traffic-prediction>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.