Available online at www.sciencedirect.com
ScienceDirect
Transportation Research Part C 14 (2006) 351–367
TRANSPORTATION RESEARCH PART C
www.elsevier.com/locate/trc

Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume

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 Street, Athens 15773, Greece

Received 5 April 2005; received in revised form 2 August 2006; accepted 12 September 2006

Abstract

Short-term traffic volume data are characterized by rapid and intense fluctuations with frequent shifts to congestion. Currently, research in short-term traffic forecasting deals with these phenomena either by smoothing them or by accounting for them by nonlinear models. But, these approaches lead to inefficient predictions particularly when the data exhibit intense oscillations or frequent shifts to boundary conditions (congestion). This paper offers a set of tools and methods to assess on underlying statistical properties of short-term traffic volume data, a topic that has largely been overlooked in traffic forecasting literature. Results indicate that the statistical characteristics of traffic volume can be identified from prevailing traffic conditions; for example, volume data exhibit frequent shifts from deterministic to stochastic structures as well as transitions between cyclic and strongly nonlinear behaviors. These findings could be valuable in the implementation of a variable prediction strategy according to the statistical characteristics of the prevailing traffic volume states.

Keywords: Traffic volume; Signalized arterials; Short-term prediction; Nonlinearity; Non-stationarity; State-space reconstruction; Recurrence plots; Recurrence quantification analysis

1. Introduction

Current research in traffic volume prediction based on data-driven approaches indicates that applied stochastic linear modeling such as the ARIMA fails at predicting shifts to extreme volume values (Davis et al., 1991; Hamed et al., 1995; Williams et al., 1998; Williams, 2001; Stathopoulos and Karlaftis, 2003), which could probably suggest that the process underlying traffic is more complicated than can be captured by a single linear statistical algorithm; moreover, there are suggestions that volume's temporal patterns vary

* Corresponding author. Tel.: +30 2107721280; fax: +30 2107722404.
E-mail addresses: elenivl@central.ntua.gr (E.I. Vlahogianni), mgk@central.ntua.gr (M.G. Karlaftis), igolias@central.ntua.gr (J.C. Golias).

0968-090X/$ - see front matter
doi:10.1016/j.trc.2006.09.002

according to prevailing traffic conditions (Smith et al., 2002). Despite literature emphasizing the need to produce accurate short-term predictions of the basic traffic variables such as volume, occupancy and speed, traffic is modeled by using either static linear approaches (ARIMA) that are restricted by a number of assumptions (Davis et al., 1991; Hamed et al., 1995; Kirby et al., 1997; Williams et al., 1998; Williams, 2001), or by nonlinear or non parametric models as, for example, neural networks, that treat traffic datasets in a continuous way clearly overlooking the effects of transitional behavior and extreme values (Kwon and Stephanedes, 1994; Smith and Demetsky, 1997; Park et al., 1998; Yun et al., 1998; Abdulhai et al., 2002; van Lint et al., 2002; Smith and Oswald, 2003; Ishak and Alecsandru, 2004; Vlahogianni et al., 2005).

It is of interest to note that in the literature, although different approaches make different assumptions regarding the underlying statistical properties, these assumptions are rarely – if ever – investigated a priori, but rather the appropriateness of the approaches is judged a posteriori based on the prediction success. Consequently, a key question to be addressed before selecting a modeling strategy is whether traffic's apparent irregularity (non-periodicity) is due to a distorted linear or a fundamentally nonlinear process. To answer this question, the statistical features of traffic volume such as detecting the degree and the power of non-stationarity and nonlinearity in the available time-series data should be addressed. It is also necessary to uncover the temporal evolution of the series of traffic volume (the temporal pattern of traffic volume). For example, assuming that volume exhibits dependence on past measurements (long term memory), a traffic volume state at time t is described by a series of past values at $t-1$, $t-2,\ldots$. The study of sequential states of volume in a time window, for example an hour, will provide the way traffic patterns evolve in time.

The scope of this paper is twofold: first, to identify the overall nonlinearity and non-stationarity in univariate series of traffic volume; second, to address the short-term evolution of these properties. The remainder of this paper is as follows; the next two sections discuss statistical methods for detecting nonlinearity and non-stationarity in series of short-term traffic volume. The following section is devoted to presenting a methodological approach for the detection of statistical properties of short-term evolution of traffic volume. Next, the implementation area and the empirical findings are presented and the final section summarizes the findings of the paper and offers some concluding remarks.

2. Detecting nonlinearity

Nonlinear approaches are said to be a prominent alternative to linear time series analysis of traffic volume. This is because, first, linear methods cannot account for all the irregular phenomena observed in traffic databases and, second, if a nonlinear process underlies the time series, the use of linear rules to describe it is conceptually false and can lead to largely erroneous results (Schreiber, 1999). The latter line of thought is not bidirectional, as modeling the irregularities of a traffic series by nonlinear methods does not indicate the existence of nonlinearities either in time series or in the underlying process generating them. As such, the ability to identify the nature of the series is essential in improving the understanding of the process involved and in providing an, as accurate as possible, approximation of complex traffic data structures.

The test of surrogate data is one of the most popular tests for nonlinearity. The underlying concept is the following (Schreiber, 1999): first, a null hypothesis of a linear Gaussian process that creates the time series is established; then, the *surrogate data* is constructed and, finally, given a selected statistic (e.g., prediction error or time reversibility) its performance is measured upon the real time series in order to reject or accept the primary established hypothesis (Paluš et al., 1995). An efficient way of reconstructing surrogates is the Fourier Transformation (Theiler et al., 1992). Consider a series of volume $\{V_t\}$ of N values taken every t, where $t = t_0, t_1, \ldots, t_{N-1} = 0$. A discrete Fourier transformation F of the original data is computed:

$$V(f) = F\{V(t)\} = \sum_{n=0}^{N-1} V(t_n)\mathrm{e}^{2\pi\mathrm{i}fn\Delta t} = A(f)\mathrm{e}^{\mathrm{i}\phi(f)} \tag{1}$$

where $A(f)$ is the amplitude and $\phi(f)$ is the phase. $V(f)$ is calculated in discrete frequency $f = -N\Delta f/2, \ldots, -\Delta f$, $0, \Delta f, \ldots, N\Delta f/2$, where $\Delta f = 1/(N\Delta t)$. Phases are randomized (rotation of phases in every f by a ω chosen uniformly in the range $[0, 2\pi)$):

$$\widetilde{V}(f) = A(f)\mathrm{e}^{\mathrm{i}[\phi(f)+\varphi(f)]} \tag{2}$$