

Chaotic Structure Test and Predictability Analysis on Traffic Time Series in the City of Istanbul

Ugur Yankaya, H. Murat Celik

Izmir Institute of Technology, Architecture Faculty, City and Regional Planning,

URLA / İZMİR

Tel: (0232) 750 70 54 / 750 70 30

uguryankaya@iyte.edu.tr, muratcelik@iyte.edu.tr

Serhan Ozdemir, Hakki Erhan Sevil

Izmir Institute of Technology, Engineering Faculty, Mechanical Engineering,

URLA / İZMİR

Tel: (0232) 750 67 54 / 750 67 73

serhanozdemir@iyte.edu.tr, erhansevil@iyte.edu.tr

Abstract

Empirical studies suggest that traffic flow generally exhibits irregular and complex behavior. Modeling of traffic flow characteristics is difficult and needs new techniques. In this study, we analyzed chaotic structure in traffic time series data collected from an urban arterial in Istanbul over a period of about 1 week. Nonlinear techniques (correlation dimension and metric entropy) are used to identify chaotic structure. After detecting chaotic characteristics, the predictability of time series data was examined. It is found that the traffic flow at the main road of Ikitelli – Mahmutbey location displayed a periodicity close to 24 hrs, and a 100 minute long prediction interval which is indicative of low dimensional chaos as found from the computed metric entropy. Traffic time series data included speed, occupancy rates, and volume at each lane on the main road of Ikitelli – Mahmutbey on the European side.

Keywords: Chaotic systems, Traffic flow, Time series analysis, Correlation dimension, Metric entropy

1. INTRODUCTION

Many systems may exhibit complex, multi-dimensional, and irregular behaviors. For example, a system that may be known as a stochastic system may represent different characteristics such as chaotic and fractal. Therefore, different techniques need to be used to expand our understanding about the systems. In the last decade, nonlinear techniques to model and predict such complex systems have been applied by many researchers in various fields of natural - social sciences and engineering. Also, these techniques may give better results rather than deterministic techniques.

Recent studies have shown that transportation systems may exhibit chaotic or fractal behaviors (Lan et al., 2003; Nair et al. 2001; Li and Shang 2007; Shang et al., 2007; Frazier and Kockelman, 2004; Shang, Li, and Kamae, 2005; Xu et al., 2008; Shang et

al., 2008). In spite of growing interest and voluminous empirical literature, there is no consensus that traffic flow exhibits fractal or chaotic behavior. Chaos theory implies system determinism. Transportation systems are affected from many factors such as human and weather factors. Therefore, the methods based on chaos theory may not be applied in the modeling of transportation systems, since these factors can be random rather than deterministic (Frazier and Kockelman, 2004).

Complex phenomena in general present a nonlinear behavior. Such systems are generally known as deterministic chaotic system. Traffic flow exhibits irregular and complex behavior (Nair et al., 2001; Shang, Li and Kamae, 2005). It is expected that nonlinear methods based on chaos theory (e.g. Lyapunov exponent, Correlation dimension, and Kolmogorov entropy) may provide better model performance for understanding behavioral structure of traffic and transportation systems. Nonlinear methods needs time series data set over a long time period. There is a lack of empirical studies in Turkey due to the lack of time series data for traffic flow characteristics. On the other hand, there is a growing interest about nonlinear analysis for traffic time series data in the literature. Therefore, gaining more contributions to the traffic flow from different cases and techniques needs to be explored. These contributions may provide information about predicting traffic flow dynamics and application in intelligent transportation systems. If we can understand the factors that produce chaos in a traffic flow, it can be possible to build new kinds of traffic models.

To determine the fractal or chaotic characteristics of a data set, different methods can be used. Power spectrum analysis (Shang et al. 2007 and Frazier and Kockelman 2004), Hurst exponent (Lan et al. 2003, Shang et al. 2007, Li and Shang 2007), correlation dimension (Lan et al. 2003, Frazier and Kockelman 2004, Shang et al. 2005, Shang et al. 2006), and the Lyapunov exponent (Lan et al. 2003, Shang et al. 2005, and Nair et al. 2001) are mostly used techniques. Since traffic flow has complex dimension, chaotic approaches are suitable for complex and multidimensional systems.

In the literature, there are a large amount of chaos and fractal studies in the last decade for communication networks like the internet and other fields. Many of these are related to network traffic. In the case of transportation systems, traffic time series data has been analyzed for only one location. Also, the outcomes of the studies in the literature differ substantially. Frazier and Kockelman (2004) applied chaotic data analysis techniques to traffic flow data for Freeway in California. Traffic flow data were analyzed by Fourier power series, the correlation dimension, and the largest Lyapunov exponent. In their study, traffic flow exhibits chaotic characteristics. Lan et al. (2003) tested the presence of chaos for traffic flow time series data. Hurst exponent, Lyapunov exponent, correlation dimension, and Kolmogorov entropy were calculated from the automatic traffic count records (different time scale) for 20 selected stations out of Miami Freeway in the United States. They found that there is a strong evidence of chaotic structure, rather than random. Shang et al. (2005) applied non-linear time series modeling techniques such as correlation dimension and Lyapunov exponent in order to analyze the traffic data collected from the Beijing Xizhimen Highway in China. The results indicated that chaotic characteristics exist in the traffic system. Xu and Gao (2007) investigated the dynamical behavior of network traffic flow, which is based on the dynamical gravity model in the case of origin – destination trip network. They found chaos in the traffic network. Shang et al. (2006) tested the presence of chaos in the traffic time series by employing the correlation dimension method for the highway in Beijing, China. They did not find strong evidence for the evidence of chaotic behavior

in the dynamics of traffic time series. Shang et al. (2008) investigated multi-fractal behavior and long-range correlations in traffic time series for Highway data in China. It is realized that traffic time series has an almost mono-fractal behavior. Also, traffic speed fluctuation is turned out to be a multi-fractal. Shang et al. (2007) investigated the presence of fractal in the traffic data for Beijing Yuquanying Highway in China. The power spectrum, the empirical probability distribution function, the statistical moment scaling function, and the autocorrelation functions were used. Traffic time series data exhibited fractal behavior. Li and Shang (2007) studied to determine whether multi-fractal parameters could be used to classify of traffic flow for Beijing Yuquanying Highway. Their analysis was mainly based on Hurst exponent. They found that multi-fractal process is appropriate in the analysis of the traffic flow. Belomestny and Siegel (2003) investigated statistical properties of highway velocity time series data obtained from German highways. There is a strong evidence of a long-range dependence. Also, they found that time series data set has stochastic nature. Wang et al. (2005) tried to forecast the traffic flow in a short-term based on theory of chaos and to identify the chaos in traffic flow time series for Shanghai City in China. There is chaos in traffic flow. Also, approaches based on deterministic chaos can be effective in short-term traffic flow forecast. Nair et al. (2001) analyzed traffic flow data and characterized it as chaotic.

Traffic system management is an important tool to use transportation system capacity efficiently in recent years. The tool needs to predict condition of the system from a time to another time with real-time data. Therefore, it is known how the factors such as traffic signals directing the system will behave in the next position. There may be differences between prediction and observation. However, we should have prior information about the probable condition at $t+1$ time. What signal times should be important for effective usage of existing capacity. In addition, the determination of the flexibility of the system against unexpected situations is important. It is worthwhile Whether or not the data relate to the system at a time have reliable information about the condition of the system at the next time. This study examines system predictability considering traffic flow characteristics for continental crossing in Istanbul.

2. NONLINEAR TIME SERIES ANALYSIS

In general, dimension is the number of parameters is needed to identify of an object. The dimension of a line is 1, and for a plane it is 2. However, as in many examples in everyday life, all objects are not smooth as geometrically. Complex, non-integer dimensions are called fractal dimensions. Considering that the most of naturally observed phenomena, or scientific researches have nonlinear dynamics and due to the fact that non-linear dynamics (especially chaotic systems) have fractal dimensions, the importance of fractal dimension analysis methods have increased, especially in recent years.

In this study, the variation in non-linear dynamical quantities of Istanbul traffic data was analyzed. In short, the purpose of this study is to test the predictability of traffic flow data through nonlinear time series analysis techniques, and to provide accurate forecasting information for advanced signaling systems.

2.1. Correlation Dimension

Correlation dimension is used to determine the geometry and complexity of an attractor. Correlation dimension is calculated via correlation sum. The correlation sum of a certain point s_n in a vector space is simply the sum of the all points in a “epsilon” (ϵ) neighborhood. In this study, calculation methods, which were introduced by Kantz and Schreiber (Kantz and Schreiber, 1997) will be the used for nonlinear invariants. More information on nonlinear invariants and their calculation methods can be found in previous works of the authors (Sevil, 2006).

The correlation sum of a time series can be found by the following equation:

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=i+1}^N H(\epsilon - \|s_i - s_j\|) \quad (1)$$

Where N is the total number of points and H Heaviside is the step function.

In equation (1), the variable s represents the vectors that are obtained through embedding the scalar values to a phase space. The embedding method is nothing but creating m dimensional vectors from one dimensional data by using method of delay, which can be calculated as

$$s_n = (s_n, s_{n-\nu}, \dots, s_{n-(m-2)\nu}, s_{n-(m-1)\nu}) \quad (2)$$

where s_n is the vector, m is the dimension of the vector and s_i represent the real time series data. In equation (2), the parameter ν is the optimal lag and can be obtained either using autocorrelation function or average mutual information function. In this study, average mutual information function is used for determining optimal lag ν .

Correlation sum, basically, is equal to the total number of neighbor points (s_i, s_j) which are closer than a distance ϵ . Considering infinite number of data points with ϵ approaching to zero, it is expected that correlation sum has an exponential relation with the parameter ϵ

$$C(\epsilon) \propto \epsilon^D \quad (3)$$

From equation (3), a dimension value D based on the behavior of a correlation sum, can be defined (Eqn 4). This dimension is called correlation dimension and it is quantitative value for time series.

$$D = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log C(N, \epsilon)}{\log \epsilon} \quad (4)$$

2.2. Metric Entropy (Kolmogorov – Sinai Entropy)

Metric entropy is quantitative measure used to characterize the chaotic behavior of the systems in multi-dimensional phase space. Metric entropy is proportional to loss of information of a dynamical system for a certain time. In other words, metric entropy represents the predictability of a given system, and it originates from information theory. Considering the observation of a system as a source of information, metric entropy gives a quantity about its predictability while entire past of the system has been

observed. The lower limit of informational metric entropy can be defined by using time series analysis.

In the literature, the relation between metric entropy and the predictability of a system, and the real life implementation of this relation was presented (Sevil and Ozdemir, 2006). Metric entropy of a time series is equal to

$$K_2 = \lim_{m \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \log \frac{C(m, \varepsilon)}{C(m+1, \varepsilon)} \quad (5)$$

In equation (5), the subscript 2 represents that the metric entropy value that is calculated by time series analysis is the lower limit of the metric entropy value determined from information theory.

3. TRAFFIC FLOW DATA and DATA ANALYSIS

In this study, we use the traffic flow data observed on the urban arterial in European site in Istanbul over a period of about 1 week, from 11/24/2008 to 11/30/2008 (24 h for each day). Our testing data come from the automatic traffic count records called as RTMs for each lane of the selected roads. The data were obtained from Istanbul Metropolitan Municipality, Department of Transportation and Directorate of Traffic. The raw data includes speed, volume, and occupancy rates. They are collected every ten minutes for each lane. We analyzed mainly speed data in the concept of this study. Some studies in the literature analyzed different time scales of 1 minute, 3 minute, 5 minute, 10 minute. In Istanbul, traffic counts are only recorded in the time scales of 10 and 30 minutes. In the concept of the study, we only analyze traffic speed data at second lane (s2) collected from RTMS22 reading the traffic flow crossing from Ikitelli to Mahmutbey (Figure 1 / RTMS22).

Figure 2 represents the time series plot of the speed data observed over a period of about 1 week. The fluctuations can only be seen in speed data. The detecting chaotic structure is difficult. Therefore, we need additional methods identifying behavior of time series data.

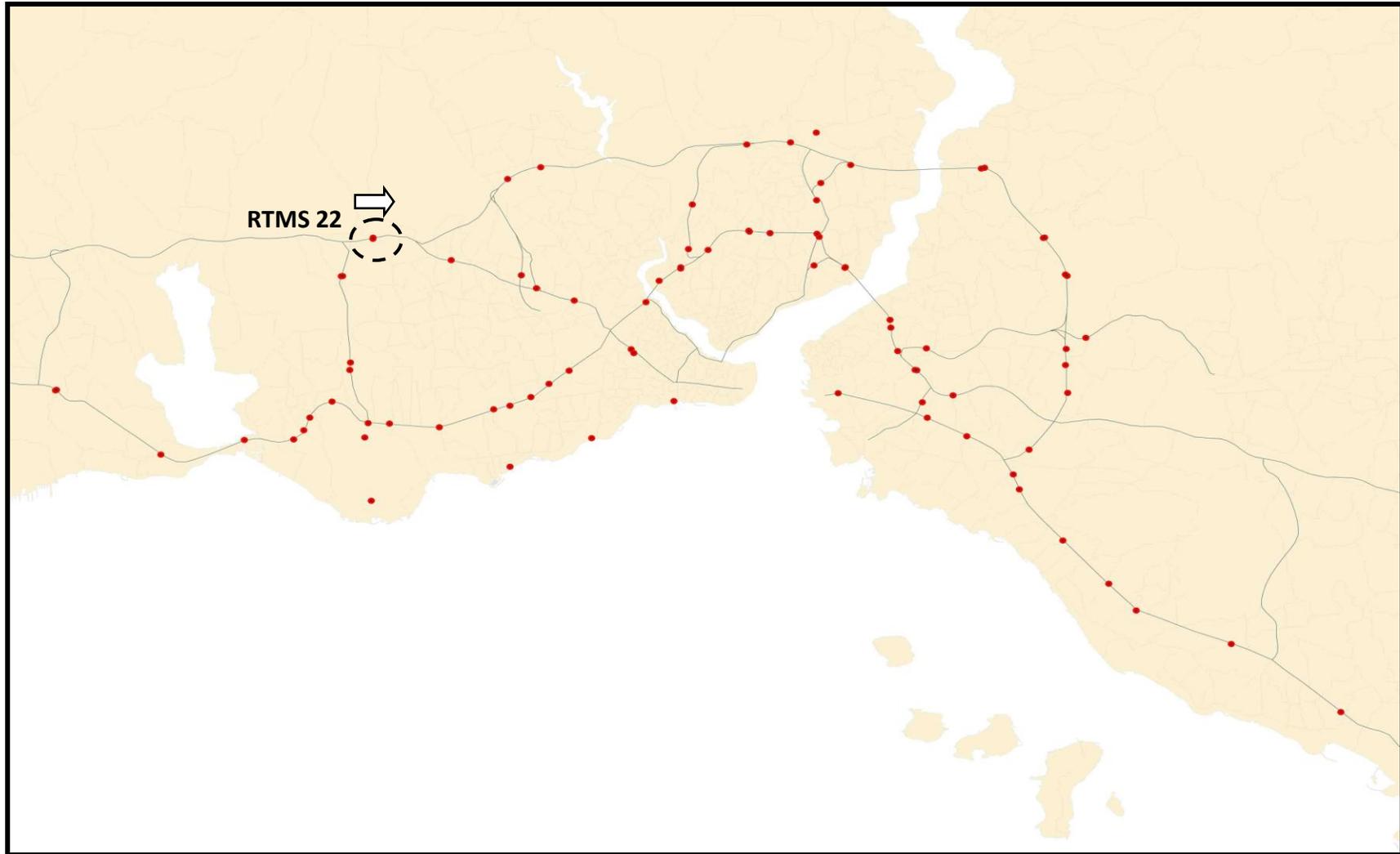


Figure 1: The sensor locations on the main roads in Istanbul

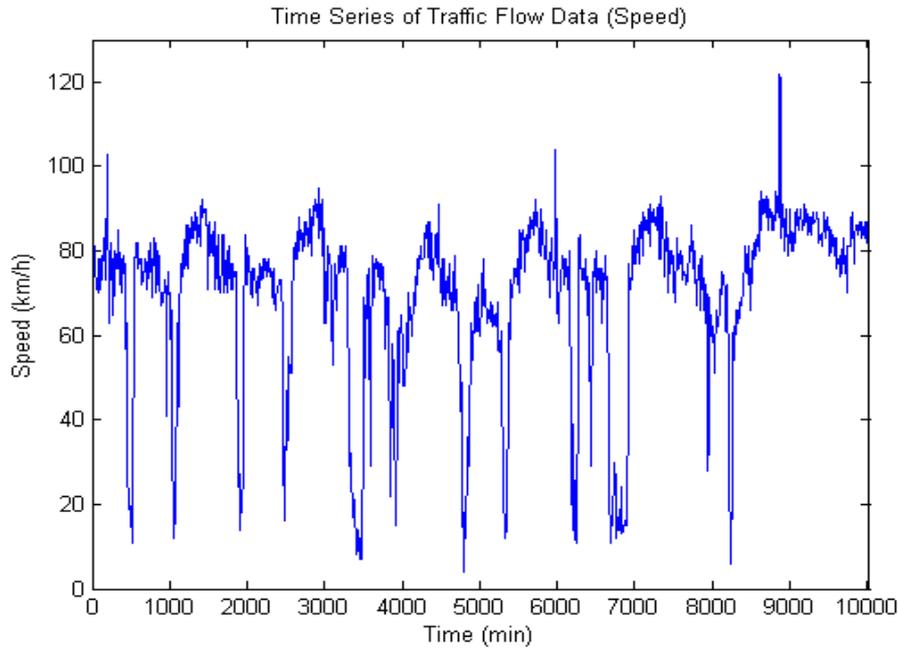


Figure 2: Traffic Flow data (speed) of the direction through Ikitelli – Mahmutbey (rtms22/s2)

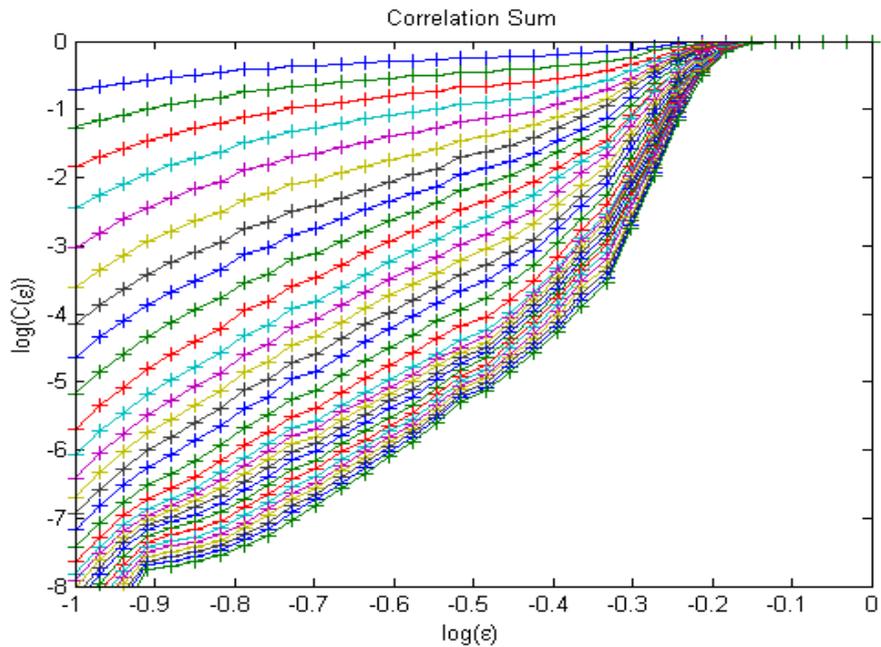


Figure 3: Correlation Sum of Traffic Flow data

Primarily, the scalar data has been normalized and transformed into phase space using method of delay-reconstruction. Then, for different embedding dimensions ($m=1-30$), correlation sum values were calculated (Figure 3). As shown in equation (4), the slope of correlation sum graph gives the correlation dimension plot (Figure 4 - Left). Due to the limit property, linear part (plateau) of the correlation dimension plot indicates the final dimension which is a

characteristic value for the system. In that region, the correlation dimension values can be represented due to changing embedding dimension and after a certain point the value will be constant. (Figure 4-Right, $D \sim 3.17$).

For scalar time series, the correlation dimension is a value that used to characterize the series. For instance, using correlation dimension, two traffic data in different times or different locations can be compared and be commented about their behaviors. Additionally, if correlation dimension values are rated (Eqn. 5), metric entropy graph is obtained (Fig. 5 - left). Likewise, the linear part can be plotted with respect to variable embedding dimensions, the steady state value indicates the metric entropy ($K \sim 0.1$) of the particular system.

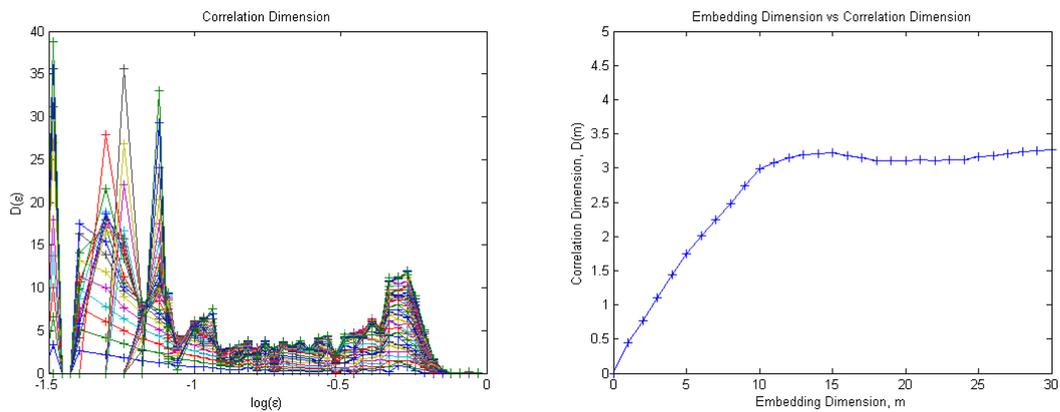


Figure 4: Correlation Dimension and Correlation Dimension variation plots for traffic flow data

Predictability of traffic data can be found by taking the inverse of metric entropy. For the particular implementation of this study, predictability is obtained as 10 ($\Delta t_{max}=1/0.1$) according to metric entropy value. Considering the sampling time of the data, it can be said that the predictions up to 100 minutes will have high level of accuracy. In the literature, a similar application was made for the Beijing Highways (Shang, Li, and Kaman, 2005) and predictability was calculated via maximum Lyapunov exponent. We believe that these results are unreliable, due to fact that the research of Shang and his colleagues was based only on largest Lyapunov exponent. According to Pesin's Identity, Metric Entropy is equal to sum of the positive Lyapunov exponents, where K_2 is the lower limit for this summation. Predictions that are only based on the largest Lyapunov exponent could be misleading. In case of multiple positive Lyapunov exponent existence, the prediction times would be greater than actual values.

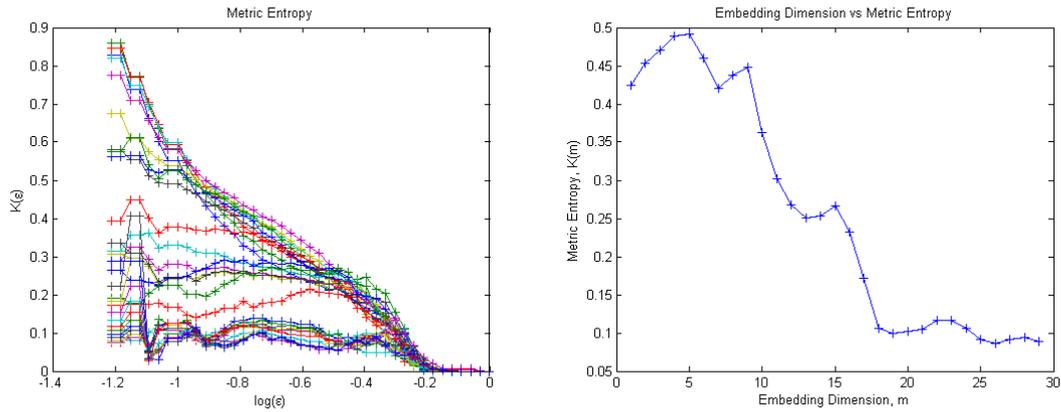


Figure 5: Metric Entropy and Metric Entropy variation plots for traffic flow data

4. CONCLUSION

In this study, on the traffic flow data collected from one of the major traffic artery of Istanbul, the direction of Ikitelli – Mahmutbey, a series of nonlinear analysis have been performed and exhibit of chaotic behavior is shown. Although not given here, the average mutual information analysis also has been made and the data was determined as almost periodic (23.33 hours). The reason of the result is just under 24 hours may depend on computational sensitivity. This almost periodicity led to a low K_2 value. The data is therefore highly deterministic, but it has a low-dimensional chaotic characteristic. In almost all chaos literature, it is recommended to approach low K_2 values with suspicion. The basis of this concern relies on the fact that real- K_2 values could be blocked under the noise. The authors think that it is important to place this warning here.

The effect of noise on traffic flow data as well as determining the type and magnitude of that noise is left as a future work. Obtained value of $K_2 = 0.1$ leads to a 100 min predictability, which provides a longer time frame for prediction and reaction by the authorities compared to highway sections around the world. If it is assumed that the results are accurate, it can be compared with the result of 8 min found by Shang and his colleagues as mentioned above. The predictability value of 8 min indicates that the analyzed dynamic of Beijing Highways segment traffic is more complex and less predictable, in contrast to the traffic in the direction Ikitelli-Mahmutbey. Relatively low predictability values are not beneficial for traffic planners. As a future study, the parameters, which affect the predictability, can be investigated and be adjusted in such way to increase the predictability time.

Acknowledgments

We would like to thank to Department of Transportation of Istanbul Greater Municipality for making the data available.

References

Journal Article

1. C. Fraizer and Kara M. Kockelman, Chaos theory and transportation systems: an instructive example, *Transportation Research Record*, 1897, 9-17 (2004).
2. L. W. Lan, Feng-Yu Lin, and A. Y. Kuo, Testing and prediction of traffic flow dynamics with chaos. *Journal of the Eastern Asia Society for Transportation Studies*, 5, (2003).
3. X. Li and P. Shang, Multifractal classification of road traffic flows. *Chaos, Solitons & Fractals*, 31, 1089-1094 (2007).
4. P. Shang, X. Li, and S. Kamae, Chaotic analysis of traffic time series. *Chaos, Solutions and Fractals*, 25, 121-128 (2005).
5. P. Shang, X. Li, and S. Kamae, Nonlinear analysis of traffic time series at different temporal scales, *Physics Letters A*, 357, 314-318 (2006).
6. P. Shang, Y. Lu, and S. Kama, The application of hölder exponent to traffic congestion warning. *Physica A*, 370, 769 – 776 (2006).
7. P. Shang, Y. Lu, and S. Kamae, Detecting long-range correlations of traffic time series with multifractal detrended fluctuation analysis, *Chaos, Solutions and Fractals*, 36, 82-90 (2008).
8. P. Shang, M. Wan, and S. Kama, Fractal nature of highway traffic data, *Computers and Mathematics with Applications*, 54, 107-116 (2007).
9. Na Xu, P. Shang, and S. Kamae, Minimizing the effect of exponential trends in detrended fluctuation analysis, *Chaos, Solutions and Fractals*, Article in Press (2008).

Book

10. H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 1997.

Conferences

11. D. Belomestny and H. Siegel, Stochastic and self-similar nature of highway traffic data, ISBN 3-88722-568-6 (2003).
12. H. E. Sevil and S. Ozdemir, Prediction of a Small Diameter Drill Bit Breakage Using Metric Entropy. Proceedings of Int. Colloquium on Math. in Eng. and Num. Physics, Bucharest, Romania, pp. 120-123, 2006.
13. Attoor Sanju Nair, Jyh-Charn Liu, Laurence Rilett and Saurabh Gupta, Non-Linear Analysis of Traffic Flow, IEEE Intelligent Transportation systems Conference Proceedings, Oakland, USA, (2001).

Thesis

14. H. E. Sevil, *On the Predictability of Time Series by Metric Entropy*. M.Sc. Thesis, Izmir Institute of Technology, Izmir, Turkey, 2006.