

# Relationship between Short Term Traffic Flow Chaos Fractal Properties and the Necessary Data for Prediction

Gang Yin<sup>1</sup>

School of Measurement-Control and Mechanical-  
Electrical Engineering  
Inner Mongolia Vocational College of Chemical  
Engineering  
Hohhot, China  
e-mail: cheniyufeng1964@163.com

Yu Feng Chen<sup>2</sup>

School of Mechanical and Electrical Engineering  
Hohhot Vocational College  
Hohhot, China  
e-mail: cheniyufeng1964@163.com

**Abstract**—Analyze the traffic flow in multi-scale time window of a freeway by using the nonlinear analysis method such as Correlation dimension, recursive map and so on, we find that chaos and fractal still exist in wide observation scales. Traffic flow correlation dimension reduces when the length of time window increases, in the observation scale of minutes. However, traffic flow correlation dimension reduces when the length of time window reduces, in the observation scale of seconds, instead of fractal property disappearance as predicted before. We present that, from the view of prediction, the recording point which is 10 times of the correlation dimension is an essential length of the data to predict. The simple model we present, which includes speed difference between vehicles and observation scales of traffic flow, can explain some of the reasons of the traffic flow chaos.

**Keywords**—component; recursive map; correlation dimension; traffic flow

## I. INTRODUCTION

Road traffic system is a time-varying complex system with group involving in, which have a high degree of uncertainty. The real-time accurate prediction to the traffic is the key to intelligent traffic control and traffic guidance. Generally speaking, the two methods of prediction are “mechanistic model” and “identification model”. Mechanistic model is causal, and it comes from the input-output relationship of the various components of the system, and it reveals the variation law of the various components of the system, from the physical mechanism point of view; even for a deterministic chaotic behavior it also has a short-term prediction ability. But for complex systems, such as traffic flow, the possibility to find its deterministic mechanism model is small. As a specific method of empirical prediction, identification predictive becomes the primary means of prediction, such as time series, artificial neural networks, support vector machines, chaotic prediction.

Existing literatures suggest that the traffic flow collected from the highway has the nature of the chaos and fractal. The change of traffic fractal and chaotic nature with statistical time interval, the affection to traffic flow forecasting, and the reason for the formation of traffic fractal and chaotic nature are the main contents of the article.

## II. THE SIGNIFICANCE OF CHAOS AND FRACTAL NATURE TO TRAFFIC FLOW

### A. The General Principle of the Identification Model Prediction

Time series prediction, especially identification prediction, need to assume that the variation of things in the future is the same as that in the past. Prediction work is divided into three stages: (1) Using more historical data to determine an identification model; (2) Using the recent historical data to verify the availability of this identification model; (3) If the identification model can achieve the desired requirements, we can use the current data to forecasts, and then get a prediction point; if the identification model is not satisfactory, we need to optimize the parameters or replacement model. Identification model can be time series method AR, artificial neural networks, support vector machines, chaotic phase space reconstruction. In theoretical research, due to selection of historical data, we tend to omit the identification model validation phase.

To get a suitable identification model, we need a certain amount of historical data: first of all, it is determined by the nature of the data; secondly, it is related to the identification method. If historical data is too small, we cannot get a useful identification model. When using the identification model to predict the future, we also come to a question that the time length of effective prediction is restricted. The estimate of “the length of time data required for prediction” and “the length of the maximum prediction” is the basics of reliable prediction.

### B. Nonlinear Theory and Identification Predictive Data

Nonlinear theory can be used to predict “the length of time data required for prediction” and “the length of the maximum prediction”. For the chaotic behavior, the prediction of the necessary historical data length is related to phase space reconstruction embedding theorem, and the length of the maximum prediction can be estimated by maximum Lyapunov exponent.

#### 1) The record number of history data that prediction needs

Embedding theorem requires sneak into space dimension  $m \geq 2D_2 + 1$ . Here,  $D_2$  is the correlation dimension of the traffic

flow. This means that  $m$  is the necessary degree of freedom to describe the traffic stream. However, experience has shown that the identification of the prediction model can be credible only when the record number of points is above  $5m$ . Namely, the record number is 10 times higher than the correlation dimension.

For example, we can assume that  $D_2$  of traffic flow in a minute of statistical time is 1.7, so we can get that  $m \geq 2D_2 + 1 = 4.4$ ; the record number of history data that prediction needs is above  $5m = 22$ . That is the history data of more than 22 minutes.

2) *The length of the maximum prediction*

In accordance with the theory of chaotic dynamics,  $T_m$  is the reciprocal of Lyapunov exponent  $\lambda_1$ , which can represent the time upper bound of deterministic prediction of chaotic systems which can also be called the time of the maximum prediction. Of course, others may use the reciprocals of the sum of the Lyapunov exponent. In the above formula, if during a minute statistical interval,  $\lambda_1 = 0.0318$ , then we can get  $T_m = 1/\lambda_1 = 31.4$ , which means the length of the maximum prediction is 31 min. The predictive value over 31 minutes is not reliable.

C. *Chaos and Fractal Nature of the Traffic Flow*

Highway traffic flow is complex time series which are both deterministic and stochastic. In different "time window" (minimum time interval of counting effective number of vehicles) observation, the proportion of uncertainty and randomness are different. Common situation is that when the time window is longer, the uncertainty of the traffic flow is more prominent; conversely, when the window of time is shorter, the randomness of the traffic flow is more prominent. Traffic flow in the longer window is visible in literature [1].

On the existence of fractal traffic flow, fractal is no recognized strict mathematical definition, so the action considering traffic flow which cannot be described by classic geometric concepts as fractal is not in dispute. We can use G-P method to estimate the correlation dimension  $D_2$  [2] of the traffic flow. A non-integer  $D_2$  can not only show that the traffic flow is fractal, but also give out the length of the history data which the traffic flow prediction needs.

On whether the traffic flow is chaotic, at present, the core of the definition of chaos that people are familiar with is "the initial value sensitivity existing in the deterministic system". The degree of sensitivity is generally represented by positive Lyapunov exponent. For low-dimensional chaotic, we commonly use the largest Lyapunov exponent  $\lambda_1 > 0$  as chaos judge sufficient conditions. The bounded regression behavior is generally satisfied with that "a Lyapunov exponent is 0, then the sum of all the Lyapunov exponents is less than 0". If  $\lambda_1$  is larger, it means that the initial sensitivity is stronger.

Since we are not sure if the traffic behavior comes from a strictly deterministic system, it is not appropriate to say "traffic flow is strictly chaos" now. It is appropriate to say that traffic flow has the nature of chaos. This formulation which needs to give up "from deterministic dynamical

systems" and to broaden the definition of chaos has been put forward in a related document [3].

III. ANALYSIS AND CALCULATION OF EXAMPLE

A. *Estimate of the Correlation Dimension and the Largest Lyapunov Exponent*

We generally use G-P algorithm [2] proposed by Grassberg and Procaccia to estimate  $D_2$ .

To estimate  $\lambda_1$ , we can use the method proposed by Kantz in 1994 except the method [4] proposed by Wolf in 1985. The method proposed by Kantz has good stability. This is its focus [5]:

For time-series  $\{x_t\}$ ,  $t=1,2, \dots$ , the largest Lyapunov exponent is defined as

$$\lambda_{max} = \lim_{t \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \frac{1}{t} \ln \left( \frac{x(t) - x_\epsilon(t)}{\epsilon} \right) \tag{1}$$

$$|x(0) - x_\epsilon(t)| = \epsilon. \tag{2}$$

Here,  $x(t)$  is the point of  $x_t$  in  $m$ -dimensional reconstructed space.

For one-dimensional time series,

$$\lambda_r(t) = \lim_{\epsilon \rightarrow \infty} \frac{1}{\tau} \ln \left( \frac{|x(t+\tau) - x_\epsilon(t+\tau)|}{\epsilon} \right) \tag{3}$$

$$dist(x_t, x_i, \tau) = |x_{t+\tau} - x_{i+\tau}|. \tag{4}$$

Calculate

$$S(\tau) = \frac{1}{T} \sum_{t=1}^T \ln \left( \frac{1}{|U_t|} \sum_{i \in U_t} dist(x_t, x_i, \tau) \right). \tag{5}$$

The maximum Lyapunov exponent  $\lambda_1$  we want is the changing slope of the  $S$  along with the changes of  $\tau$ .

B. *Estimates of the Correlation Dimension and the Largest Lyapunov Exponent*

Using G-P method to calculate traffic flow data which was collected from the measured traffic flow data within one month of the one highway of Tianjin suburbs, we can get  $D_2$ , and we can get  $\lambda_1$  using the method of Kantz, which can be seen in TABLE I. What particularly need to note is that the existing estimation methods to  $\lambda_1$  are only applicable to low noise time series, for the reason that the current chaotic definition requires "deterministic dynamic system". For the traffic flow containing strong noise, the estimation to  $\lambda_1$  is erratic. Therefore the estimate provided is only for reference.

To this traffic flow, there are fractal and chaos phenomenon in intervals over 1s. A noteworthy phenomenon is that the maximum value comes when the interval is 15s.

TABLE I. THE NON-LINEAR NATURE OF THE TRAFFIC FLOW COUNTED AT DIFFERENT TIME INTERVALS

Statistics Time Window of Number of Vehicles (traffic flow)/s	Correlation Dimension D <sub>2</sub>	Largest Lyapunov Exponent λ <sub>1</sub>	The Record Number of History Data Prediction Needs/point	The Length of the Maximum Prediction /point
1	2.2362	0.018	18	55
5	2.4917	0.045	20	22
10	2.7147	0.036	19	27
15	3.1813	0.033	16	30
30	2.8286	0.032	15	31
45	2.3265	0.028	14	35
60	2.1976	0.028	14	35
2*60	1.9538	0.027	14	37
3*60	1.9054	0.026	13	38
4*60	1.8016	0.022	13	46
5*60	1.6835	0.018	12	54
6*60	1.6027	0.018	18	55
10*60	1.4359	0.045	20	22

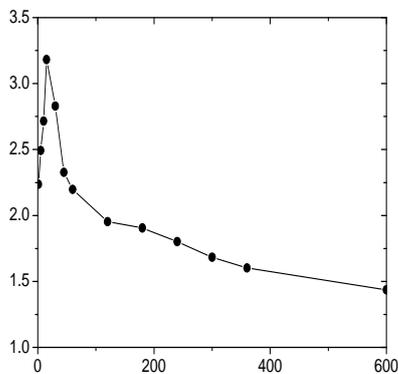


Figure 1. Changes of correlation dimension with the time window.

When the time window is over 15s, the larger the window is, smaller D<sub>2</sub> will be, which means certainty will be significantly enhanced. This phenomenon indicates that the traffic flow has the nature of inherent certainty, and randomness is due to the careless observation to certainty.

When the window is less than 15s, the shorter the window is, the smaller D<sub>2</sub> will be, which means certainty will be significantly enhanced. The reason for this phenomenon is that cars normally running on the highway need to keep a safe distance at least. When the window is very small, safe distance, as a relatively determining factor will have a significant performance; Though D<sub>2</sub> will decline if the interval is less than 15s, it will not directly lead to the disappearance of the fractal nature.

We found that chaos and fractals are present in a wide window of observation time, and this is a new discovery of this article.

C. Nonlinear Nature and Traffic Flow Forecasting

We can see from the TABLE I, that the length of the maximum prediction will generally not limit the short-term traffic flow forecast. What may limit the forecast is:

Traffic flow forecasting based on historical data (identification model), require enough historical data to forecast. The point given in TABLE I is required by nonlinear properties of traffic flow sequence, and it does not include the special case that special identification method requires more historical data. When we only see TABLE I, forecast needs the number of history data is 12 to 20 points. Specifically, traffic flow of 1, 3, 5 minutes needs 15, 14, 13 points to forecast, which the corresponding time with is 15, 42, 65 minutes. However, the nature of traffic flow will make certain changes in such a long time. This is the reason why it is so difficult to forecast traffic flow.

In simple terms, for the specific prediction method, more history data will improve the forecast accuracy, if we assume that the nature of the traffic flow is unchanged; however, more data means more time, and more time will lead to more changes of traffic flow nature, then it will cause the decline of accuracy.

Good prediction method should consider the function of the two aspects. Particularly, we must select the appropriate length of the history data, and it is one important way to improve the accuracy of forecast.

IV. DISCUSSION ON THE CAUSE OF CHAOS IN TRAFFIC FLOW AND FRACTAL

Traffic flow is a complex time series, and its nature of chaos and fractal has many reasons. For the causes of chaotic characteristics, the author tries to analyze like the following:

A. One of the Reasons—Speed Difference

A fleet of cars run on highway, and observing its driving from A to B, assuming that we have known the traffic flow of observation point A. Only considering the different speed  $\Delta V = V_{max} - V_{min}$  of difference vehicles, when we observe the simple model of two factors of the time window T, we can explain the variation in minute scale D<sub>2</sub> and λ<sub>1</sub> which happened in traffic flow of point B.

Vehicles through point A will gradually arrive in point B after some times.

In particular window of time T, if speed difference ΔV is larger, the time span of vehicles from A to B will be larger

too, and this will strengthen overlapping of traffic flow of different T in point B, which will strengthen the uncertainty of traffic flow in point B, and make  $D_2$  and  $\lambda_1$  larger.

When the speed difference  $\Delta V$  is fixed, if the time window T is larger, it will make it more possible that the vehicles which through point A at certain T are counted in point B at a certain T. In other words, the certainty of traffic flow in point B will be larger, and it will make  $D_2$  and  $\lambda_1$  declined.

Though the model is simple, it basically has an ability to explain the phenomenon observed in the traffic flow,  $D_2$  of traffic flow in urban highway is generally larger than that in normal highway, and this is for the reason that the vehicle speed of urban highway changes a lot. No matter what the traffic flow is,  $D_2$  will decline with the increase of the time window. For the different time of one day in urban highway, when the number of vehicles is larger, the vehicle speed will generally be larger, and  $D_2$  will be larger too and vice versa[1].

For the complex system as the major role with random factor, when we add deterministic role to it, it will show the nature of chaos. This is similar to "Random behavior with memory" which also shows this nature [6]. The requirement to reduce the requirements defined of "From deterministic system" [3] comes from research to the practical problems.

#### B. The Second Cause—Difference of the Distribution of Various Types of Vehicles

When observed, if the time window T is larger, the overall number of vehicles under statistics will be larger, and type of vehicles (high-speed vehicles and low-speed vehicles) in adjacent time window will have more stable ratio of the number, and it also can reduce the affection to flow caused by the uneven distribution of different kinds of vehicles in highway observed. Therefore, the stability of whole traffic flow will increase, and the predictability will be enhanced. The number of vehicles can be observed will decline with the decreases of the time window and the difference between the numbers of vehicles distribution causes the growing instability of changes in traffic flow, and if the regularity is less obvious, the randomness will be greater and short-term forecasts will be more difficult.

#### C. The Third Cause—Different of Driving Habits

For the limit of automotive performance and difference of driver's preferences, personal vehicles usually keep its customary traveling speed, if the road condition permits. When traffic density on the road is small, the spacing of car head is large, and the vehicles flow approximately become free flow, and its orderly movement includes stabilizing factor of predictability. However, when the traffic flow density increase, high-speed vehicles always want to transcend low-speed vehicles, and uncertain state of disorder will come when they are adjusting speed.

### V. CONCLUSION

On the basis of the analysis to a highway traffic flow, in this paper, three aspects of work are: finding that the traffic flow correlation dimension is the one of the main reasons

that limits the enhancement of the prediction accuracy; we find that chaos and fractal nature of the traffic flow on a wide observation window between seconds to minutes are existing. Generally speaking, in the minute scale, if the time window is wider, correlation dimension will be dropped and if the time window is small, correlation dimension will be dropped, too. Previous studies suggest that it would cause the disappearance of fractal nature; however, this study does not support this opinion; we propose a simplified model which can qualitatively explain some of the causes of the chaos in traffic flow characteristics, and the causes are that the speed difference is the main factor to affect the correlation dimension in minute scale observation window. In certain time window, if the speed difference of different vehicles is larger, the correlation dimension of the traffic flow will increase. When the speed difference is fixed, the correlation dimension of the traffic flow will decline if the time window becomes larger.

The above work is the basic work to improve traffic flow forecasting accuracy. The optimal quantity of historical data traffic flow prediction accuracy needed and changes of the predictive accuracy with the changes of traffic flow time interval are the main points of future research.

### REFERENCES

- [1] Wang Zhengwu, Huang Zhongxiang, and Kuang Aiwu, "Short term traffic flow chaos identification and forecasting precision analysis," Changsha Institute of transport, 2004, pp. 73–76.
- [2] Feng Weidong, Chen Jian, He Guoguang, Liu bao, "Traffic flow in fractal research," High technology letters, 2003, pp. 59–65.
- [3] DENNIS B, Desharnais RA, Cushing JM, Henson SM, Costantino RF, "Can noise induce chaos," OIKOS, 2003, pp. 329–339.
- [4] WOLF, A., Swift, J., Swiney, H., Vastano, J, "Determining Lyapunov exponents from a time series," Physica D, 1985, pp. 285–317.
- [5] H. KANTZ, "A robust method to estimate the maximal Lyapunov exponent," Physics Letters A, 1994, pp. 77–87.
- [6] Yang Zhengling, Lin Kongyuan, "Power system load records discussion on causes of chaos," Automation of electric power systems, 2002, pp. 18–23.
- [7] Gao Wei, Lu Baichuan, and Yun Tianli, "Short-term traffic flow forecasting based on spatiotemporal characteristics and RBF neural network," Journal of transport information and safety, 2011, pp. 16–19.
- [8] Wu C H, Ho J M, and Lee D T, "Travel-time prediction with support vector regression," IEEE Transactions on Intelligent Transportation Systems, 2004, pp. 276–281.
- [9] Shan Huang, and Adel W Sadek, "A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting," Transportation Research Part C, 2009, pp. 510–525
- [10] Bidisha Ghosh, and Biswajit Basu, "Member, IEEE, and Margaret O' Mahony. Multivariate short-term traffic flow forecasting using time-series analysis," IEEE Transactions on Intelligent Transportation Systems, 2009, pp. 246–254.
- [11] Ding Shicheng, "Multivariate analysis method and its application," Changchun: Jilin people's publishing house, 1981.
- [12] Yao Zhisheng, and Shao Chunfu, "Road traffic state prediction research," Journal of Harbin Institute of Technology, 2009, pp. 247–249.
- [13] Wang Jin, and Shi Qixin, "Short-term traffic flow prediction model in China," public safety, 2005, pp. 92–98.

- [14] Shi Zongxin, and Zheng Weizhong, "The road network traffic flow forecasting methods," *Journal of traffic and transportation engineering*, 2004, pp. 68–71.
- [15] Wen Dong, "Influence Modeling of Complex Stochastic Processes," Cambridge, MA, USA: Massachusetts Institute of Technology, 2006.
- [16] Dong Honghui, Sun Xiaoliang, and Jia Limin, "The multi-modal traffic flow prediction model," *Journal of Jilin University: Science edition*, 2011, pp. 645–649.
- [17] Yang Chao, and Wang Zhiwei, "The optimal GA WNN application in traffic flow forecasting," *Computer engineering*, 2011, pp. 149–151.