# Research on Traffic Flow Prediction based on Chaotic Time Series

Xiaobo Yang and Lianggui Liu

*Abstract*—**In order to dramatically improve the precision of traffic flow estimates, this study proposes a prediction technique based on chaotic time series. Here is a breakdown of the exact research process. First, the chaos principle of traffic flow and two indexes—delay time and dimension selection—that affect system reconstruction are looked at. Then, in order to obtain the maximum amount of time that can be forecasted, the chaotic traffic flow is predicted using an improved local approach and the Lyapunov index. Finally, a comparison is made between the traffic flow predictions made using the traditional local technique and the modified local method. According to the outcomes of the predictions, chaotic time series can be utilized to forecast traffic flow, and the prediction error is lower than that of both the widely used neural network prediction method and the least squares support vector machine prediction method, proving the effectiveness of the method proposed in this study.**

*Index Terms*—**Chaos principle, Traffic flow prediction, Time series, Improved local area method**

## I. Introduction

THE traffic system is flexible, erratic, and dynamic. The routine operation of the traffic system is disrupted simultaneously by a number of unpredictable variables, leading to chaos. Stochastic techniques were once utilized in traffic flow forecasting. Relevant scholars suggested that the traffic system is a complicated system that is mostly made up of chaotic variables with nonlinear dynamic changes as the field's research progressed [1]. Chaos theory can be used to examine the nonlinear system's time series change law [2]. By directly predicting the traffic flow change law from the traffic flow sequence, this theory can help to eliminate the interference of human variables and boost forecast accuracy. Greenberg's model was used by Zhang Zhiyong et al. [3] to examine chaotic traffic flow, and system simulation was employed to identify the chaotic features of the flow. Attoor et al. [4] used nonlinear time series to examine traffic flow data and came to the conclusion that the data was chaotic. By fusing chaos and fractal theory, Tang Ming and colleagues [5] examined short-term traffic flow and made local predictions about it. By reconstructing the phase space of the data, XUE et al. [6] use chaos theory to enhance the polynomial and create a prediction model to estimate the traffic flow over a short period of time. Forecasts can only be made in a single step using the prediction model. These chaos theory-based prediction algorithms have some drawbacks, including restrictive assumptions, subpar real-time prediction, a slow algorithm convergence speed, etc. However, they can assist with some forecasting issues with traffic flow. In this paper, a method for forecasting traffic flow based on chaotic time series is suggested to overcome the inadequacies in existing methods, enhance prediction accuracy, and lower processing costs.

Manuscript received October 1, 2022; revised June 23, 2023.

This work was supported by the Natural Science Foundation of Zhejiang Province. Grant number: Y1110023

Xiaobo Yang is an associate professor of the Information Technology Department, Zhejiang Shuren University, Hangzhou, ZIP Code 310015 China (email: yxb71520@163.com).

Lianggui Liu is a professor of the Information Technology Department, Zhejiang Shuren University, Hangzhou, ZIP Code 310015 China (email: felix@zjsru.edu.cn).

## II. Chaos Theory and Traffic Flow Delay Reconstruction

In addition to the environment and objective conditions, a number of human factors affect how traffic travels through cities. The root reason of the traffic flow issue is these ambiguous variables. Travelers use a variety of forms of transportation, and changes in traffic flow adhere to established regulations. In other words, the rule will be followed early on in the traffic flow, but there will be significant chaos later. As a result, the internal operations of the traffic system are the main cause causing the chaos in the flow of traffic.

The theory underlying the forecasting of chaotic time series was established by Safonov et al.'s [7] concept that the chaotic state will progressively transform into a regular state by delayed reconstruction of variables. This theory attempted to create a regular moving track out of the chaotic phenomenon of the traffic system. It is feasible to rebuild the initial characteristics of the traffic system in a chaotic condition thanks to the relevant state point indicators, which have a direct impact on the system characteristics. Delay duration and dimension selection are the two factors that this study focuses on most since they affect system reconstruction.

The choice of the delay time affects the reconstruction of a chaotic system significantly since it affects both the complexity and volume of information. In order to select the ideal delay time, this paper improves the conventional auto-correlation approach [8].

When the auto-correlation curve reaches its initial value of (e-1)/e, the optimum delay time can be calculated for the known time series $\{x_i, i = 1,2,\cdots,n\}$. The corresponding auto-correlation function is as follows.

$$R(t) = \frac{1}{n}\sum_{i=1}^{n} x_i x_{i+t} \tag{1}$$

In Formula (1), $x_i$ denotes the value of the series at the ith instant, and R(t) is the auto-correlation function. High-dimensional values cannot be recovered; in general, this function can only deduce a linear connection between sequences. To solve this problem, the initial auto-correlation function can be improved, and the improved auto-correlation function is as follows.

$$C(t)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m-1}(x_i-\bar{x})(x_j-\bar{x})=R(t)-(m-1)(\bar{x})^2 \tag{2}$$

C(t) stands for the auto-correlation function of occurrence delay time, and the improved auto-correlation function in Formula (2) may extract high-dimensional values of time series. $\bar{x}$ is the time series' average value.

In a perfect world, the dimension selection just requires the minimal value that is greater than the correlation dimension. When selecting a dimension, it is important to adhere to the minimum dimension concept of system dynamics behavior because of the noise and data sequence length limits in the actual traffic environment. The vector field approach [13], Cao's method [11], Singular value decomposition method [12], Cao's method [11], Pseudo-nearest neighbor method [9], and correlation index method [10] are a few methods for determining the ideal dimension. This work extends Cao's technique, which has low sample size requirements and is more appropriate for nonlinear systems, to obtain the optimal dimension of traffic flow. Figure 1 shows how the algorithm works.

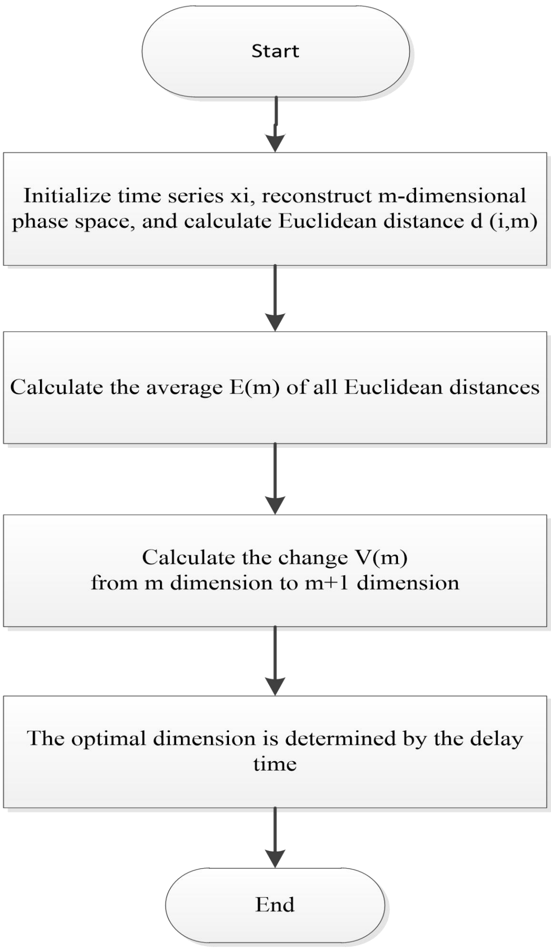


Fig. 1. Flow chart of the optimal dimensional algorithm

Figure 1 demonstrates how to compute the Euclidean distance of traffic samples, the average value of all Euclidean distances, the dimension change and delay time, and finally the ideal dimension value to rebuild phase space.

During the traffic flow measurement procedure, the time series data will be mixed with some noise due to the limitations of the external environment, measuring methodologies, and other considerations. If these noisy data can't be removed right away, it will have an impact on the measurement results itself. Therefore, it is essential to distinguish between accessible signals and noisy signals. There are already a number of qualitative and quantitative methods for recognizing chaos, including the Poincare section approach [14] and the auto-correlation function method [8]. The Lyapunov index [16] and alternative data technique [15] are techniques for quantitatively diagnosing chaos. Since the quantitative approach can more accurately describe the change process of traffic flow, this work will improve the Lyapunov index method to quantitatively detect chaotic traffic flow. Figure 2 shows the flow of the algorithm.

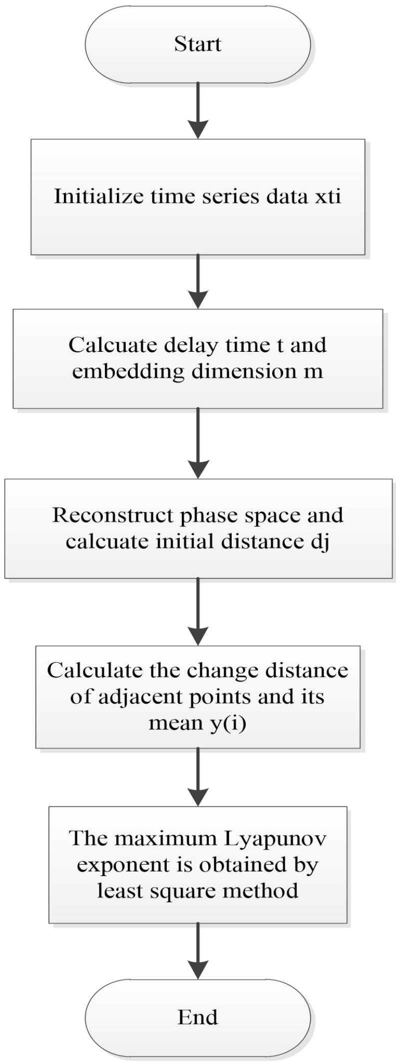


Fig. 2. Flow chart of quantitative identification of chaotic traffic

As seen in Figure 2, initialization of the time series data comes first, then calculations of the delay time and best-embedded dimension, reconstruction of the phase space to ascertain the change distance between adjacent nodes, and the least square method to ascertain the maximum Lyapunov exponent for quantitative chaos identification.

## III. CHAOTIC TIME SERIES PREDICTION ALGORITHM

The use of chaos theory to predict traffic flow is a nonlinear analysis technique. It first examines the function relation of the traffic flow time series through the trajectory of state points, then forecasts the future trend of state points to forecast the expected value of traffic flow at the trajectory points.

To decrease the effects of noise interference, the local approach is frequently employed to forecast chaotic traffic flow [17]. The classic local method has the advantage of being simple to use, but it also has limitations like high standards for data quality, inadequate noise reduction, and imprecise prediction. In this study, the conventional local method for predicting short-term traffic flow is enhanced to solve its shortcomings.

Assume that the center point's neighbor is $Y_{mi}(i=1,2,\ldots,n)$ and that $d_i$ represents $Y_m$'s Euclidean distance. The weight of the neighboring point $Y_{mi}$, presuming that dmin is the smallest value, is as follows.

$$w_i = \frac{exp(d_i - d_{min})}{\sum_{i=1}^{n} exp(d_i - d_{min})} \tag{3}$$

Then the linear fitting of the local method is as follows.

$$Y_{m(i+1)} = ae + bY_{mi},\ i = 1,2,\cdots,n \tag{4}$$

Where a and b are the fitting coefficients, e is the dimension vector, and $Y_{m(i+1)}$ is the predicted value as a result of $Y_{mi}$ development.

The selection of appropriate nearby places directly affects the forecast's accuracy when a local approach is used to predict traffic flow. The Euclidean distance between the state point and the prediction center point is frequently first calculated after the threshold value has been set at a low value. When the estimated Euclidean distance is below the threshold value, the state point is considered to be close by.

This strategy is improved in this work since adding fake neighbors, which lowers prediction accuracy, is straightforward. This means that while determining a point's weight $Y_{mi}$, both the correlation with the projected center point and the Euclidean distance between a point and that center point should be taken into account. The correlation coefficient can be calculated using this formula.

$$S_{mi} = \frac{Cov(Y_m, Y_{mi})}{\sqrt{D(Y_m)D(Y_{mi})}} \tag{5}$$

Where $Cov(Y_m, Y_{mi})$ represents the covariance coefficient between the adjacent point and the predicted center point, and and represent the respective variances between the adjacent point and the predicted center point. is the correlation coefficient between the adjacent point and the predicted center point, and the weight Ymi of the adjacent point may be adjusted as necessary.

$$w_i = a * b * S_{mi} \tag{6}$$

The estimated weight $W_i$ of the nearby point $Y_{mi}$ is then linearly fitted using the local approach of formula (4) to produce the predicted value.

Chaos theory typically relies a lot on the starting point. The trajectory index will diverge rapidly with small changes in the initial condition of the traffic flow, which will reduce prediction precision. As a result, it is impossible to separate the trajectory evolution inside the traffic flow system from the precision of the traffic flow prediction.

When analyzing the trajectory index's divergence statistically, the Lyapunov exponent can be utilized to express the trajectory's divergence or aggregation ratio. If the value is more than zero, the local trajectory is unstable and the nearby state points are gradually separating. If the value is less than zero, the neighboring state points are still accumulating. Because the maximum exponent $\lambda_{max}$ of the Lyapunov equation can quantitatively represent the divergence degree of the adjacent trajectory and the divergence degree is related to the maximum predictable time, the maximum exponent max of the traffic flow can be used to calculate the maximum predictable time.

The Lyapunov maximum index can be used to quantitatively characterize the divergence between close state points. Since chaos obeys deterministic law, traffic flow may be predicted within a critical time, and the critical time $t_{max}$, denoted by the symbol, is the longest period that can be predicted. The link between $t_{max}$ and the greatest Lyapunov exponent $\lambda_{max}$, is expressed as follows.

$$t_{max} \approx \frac{1}{\lambda_{max}} \tag{7}$$

Equation (7) shows that the maximum predictable time $t_{max}$ and the maximum exponent $\lambda_{max}$ have an inverse relationship. As max increases, the maximum anticipated time $t_{max}$ falls, which lowers prediction accuracy.

## IV. COMPARATIVE EXPERIMENTS

To assess the method proposed in this research, traffic flow data on the north-south bridge of Hangzhou is collected by the detecting coil once every 10 minutes. To forecast the traffic flow for October 2021, the daily traffic flow data from September 2021 are used as the known data.

Using the known time series traffic flow, the optimal system reconstruction dimension and delay time were calculated. The results of utilizing the auto-correlation technique to calculate the delay time are shown in Figure 3.

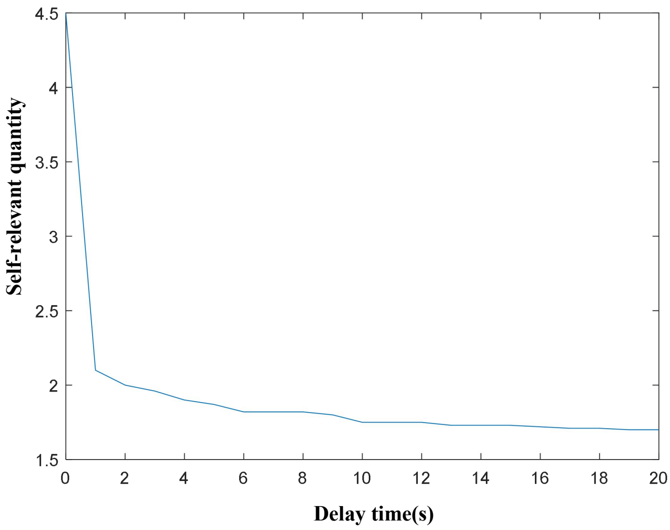


Fig. 3. Use autocorrelation to calculate the delay time

Figure 3 shows that it takes 6s following the self-correlation function curve's first minimum point to reconstruct the traffic flow system in a chaotic situation.

Figure 4 shows the outcomes of applying Cao's method to identify the ideal traffic flow dimension.

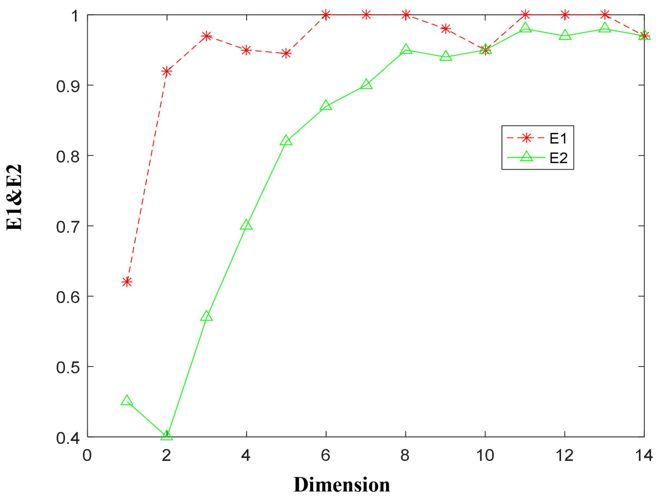


Fig. 4. Using the Cao method to calculate the optimal dimension of traffic flow

Figure 4 shows how, as size grows, the E1 curve progressively approaches saturation. As E1 changes the least when dimension m is 9, it may be concluded that 9 is the best dimension for traffic flow. The E2 curve is used to distinguish between chaotic and random signals. The simulation results show that the E2 curve is chaotic because it fluctuates around 1 as opposed to being a fixed constant.

The ideal dimension and delay time can be used to calculate the maximum Lyapunov index of traffic flow, as shown in Figure 5.

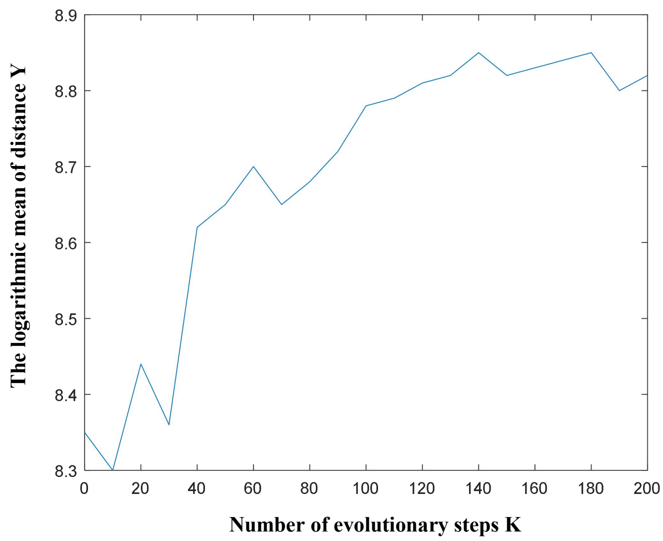


Fig. 5. Lyapunov exponential curve of maximum traffic flow

Figure 5 shows this relationship: where $K$ is the total number of time evolution steps, and $Y$ is the average logarithmic distance between all state points. Figure 5 depicts the approximate shape of the Lyapunov exponential curve as a straight line for k values in the range [0,55], the intersection of the straight line and curve line as the ideal value of k, and the least squares method computation of the slope of the straight line. Because the slope is 0.0078, the maximum Lyapunov exponential $\lambda_{max}$ of the traffic flow may be determined. The forecast error will rise until it exceeds the maximum predicted time of two hours. After evolution, formula (4) can be used to calculate the expected value $Y_{m(i+1)}$ of the state point, and after multiple iterations, the traffic flow prediction outcomes for each subsequent period can be found. The results of the new local method and the previous local approach, both employed to anticipate traffic flow, are compared in Figure 6.

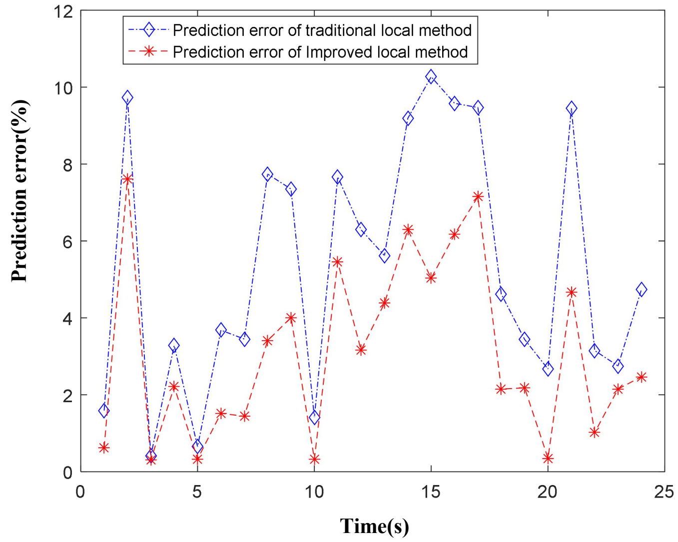


Fig. 6. Comparison of traffic flow prediction errors between traditional local method and improved local method

According to the results of the predictions in Figure 6, the old method's average absolute error for estimating traffic flow is 1.59%; there are five state points with absolute percentage errors that are greater than 9% and twelve with absolute percentage errors that are less than 5%. The enhanced local technique results in an average absolute error of 0.63 percent when predicting traffic flow; there are no absolute percentage errors larger than 9% and 18 absolute percentage mistakes under 5%. It proves that the improved local technique put forth in this paper has higher prediction accuracy.

To further support the value of the improved local approach, the optimal dimension threshold can be selected, with the threshold range being [10,15]. Both the simple Euclidean distance approach and the improved Euclidean distance method are used to anticipate the traffic flow. The MAPE (Mean Absolute Percentage Error) results are shown in Figure 7.

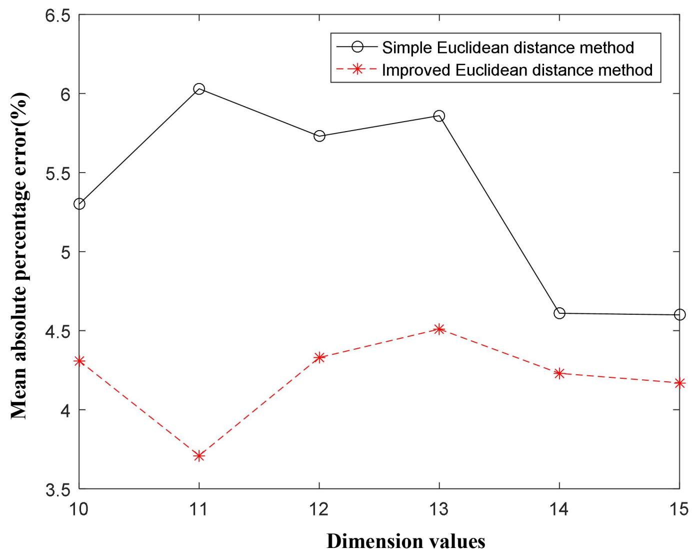


Fig. 7. Comparison of traffic flow prediction errors of different dimension values

Figure 7 demonstrates the effectiveness of the improved local method by showing that whereas the average absolute percentage error of the simple Euclidean distance technique is 5.35%, that of the improved Euclidean distance method is 4.21% and the overall error is decreased by 1.14%.

The results are shown in Figure 8 after three different methodologies are used to estimate and compare the traffic flow.

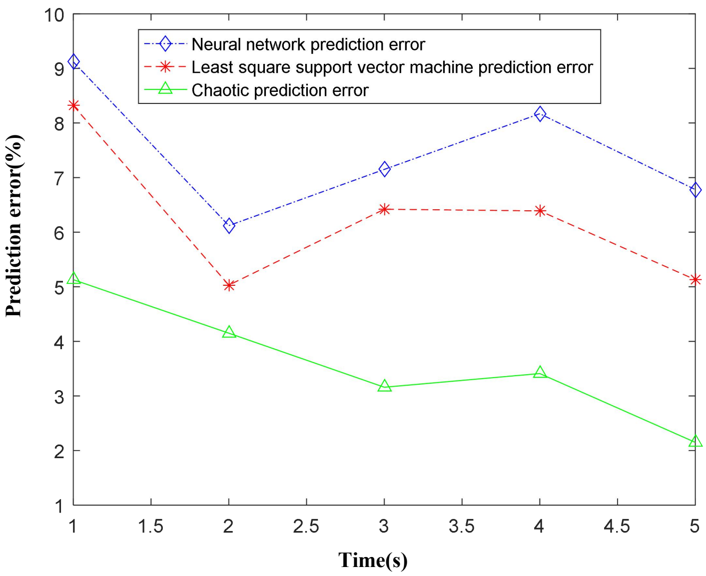


Fig. 8. Comparison of three different methods for traffic flow prediction

Figure 8 demonstrates how the fact that the method proposed in this paper has a smaller prediction error than both the commonly used neural network prediction method and the least squares support vector machine prediction method further supports its utility.

## V. Conclusion

In this work, which suggests a prediction method based on chaotic time series to forecast traffic flow, the following results are attained by developing a chaotic model to foresee traffic circumstances.

1) To restore the original characteristics of the traffic system in a chaotic state, delay time and dimension selection — two essential indices — are explored. The increased Lyapunov exponent approach is a quantitative technique for detecting chaotic traffic flow.
2) To evaluate the effectiveness of the chaotic time series prediction system reported in this research, monthly traffic flow data were collected by detecting coil. To predict the traffic flow, the modified local approach and the traditional local technique were both used, and comparisons were done. The prediction results show that the average absolute error and absolute percentage error of the improved local technique are lower than those of the conventional local method, proving the value of the proposed methodology.

## Acknowledgment

Xiaobo Yang thanks to the scientific research team of Zhejiang Shuren University.

## References

[1] M. Cetin, G. Comer. "Short-term traffic flow prediction with regime-switching models," Transportation Research Record, 1965, 2006,pp. 23-31.
[2] J. M. HU, et al. "An applicable short-term traffic flow forecasting method based on chaotic theory," Proc. Of IEEE 6th International Conference on Intelligent Transportation Systems, 2003, pp.608-613.
[3] Z. Y. Zhang, J. Rong, F. T. Ren. "Study on chaos phenomena in following motorcade," Journal of Civil Engineering and Transportation Engineering,2001,1(1), pp.58-59.
[4] S. N. Attoor, J. C. Liu, et al. "Non-linear analysis of traffic flow," IEEE Conference on Intelligent Transportation Systems. Oakland, USA: IEEE, 2001, pp. 681-685.
[5] M. Tang, B. X. Chen, et al. "Short-term traffic fractal research based on phase space reconstruction," Journal of Shandong Jiaotong University, 2004,12(1), pp. 50-55.
[6] J. N. XUE, Z. K. SHI. "Short-Time Traffic Flow Prediction Using Chaos Time Series Theory," Journal of Transportation Systems Engineering and Information Technology, 2008,8(5), pp. 68-72.
[7] L.A. Safonov, E. Tomer, et al. "Delay-induced chaos with multifractal attractor in a traffic flow model," Europhysis Letters, 2002,57(2), pp.151-157.
[8] M. S. Desti. "auto-correlation Method for Cyclic Prefix OFDM Estimation.," International Journal of Information Technology and Electrical Engineering, 2019,3(3), pp.91-94.
[9] B. K. Matthew, B. Michael. "Estimating good discrete partitions from observed data symbolic false nearest neighbors, " Physical Review Letters, 2003, 91(8), pp.84-102.
[10] G. G. Richard, F. A. J. Edward. "Correlations redux: asymptotic confidence limits for partial and squared multiple correlations," Applied Psychological Measurement, 1999, 23, pp.116-119.
[11] T. A. Xiong, B.B. Liu, et al. "Method for Prediction of Landslide by Phase Space Reconstruction," Geospatial Information, 2011,3(54), pp.162-164.
[12] W. Xu, Y. Chen, et al. "Application of SVD Method in Ill-posed Problem," Bulletin of Surveying and Mapping, 2016,0(1), pp.62-63.
[13] V. M. Goncalves, C. A. Pimental, et al. "Vector fields for robot navigation along time varying curves in n dimensions," IEEE Transactions on Robotics, 2010,26(4), pp.647-659.
[14] S. Y. Liu, S. J. Zhu, et al. "Study on the application of poincare section algorithm to chaotic vibration identification," Applied Physics, 2011,1, pp.108-115.
[15] H. Y.Sun, F. Wang F. "Effect of the data processing for surrogate data tests," Mathematics in Practice and Theory, 2006,36(1), pp.160-164.
[16] S. Y. Chen, W. Wang. "Chaos forecasting for traffic flow based on lyapunov exponents," China Civil Engineering Journal, 2004,37(9), pp.96-99.
[17] T. Ding, H. C. Zhou. "Local prediction research of chaotic time series," Systems Engineering and Electronics, 2004, 26(3), pp. 338-367.

**Xiaobo Yang:** Chinese nationality, Ph.D. in Computer Application, Donghua University, Shanghai, China, was born on May 20, 1971. In May 2002, he obtained his doctorate. Data mining, for example, and intelligent transportation are among the key research areas.

He is employed as a college professor in Hangzhou, Zhejiang province, China. He has more than 30 research articles to his credit. Current research areas include intelligent transportation systems and computer applications.

Dr. Xiaobo Yang is third-level training person for the New Century 151 Personnel Project and an IEEE member.