Луценко Владислав Вячеславович
[Lutsenko Vladislav Viacheslavovich],
Кучеров Николай Николаевич
[Kucherov Nikolay Nikolaevich],
Гладков Андрей Владимирович
[Gladkov Andrei Vladimirovich]

УДК: 519.246.8
DOI: 10.37493/2307-910X.2023.1.4

ПРОГНОЗИРОВАНИЕ ДОРОЖНЫХ ЗАТОРОВ НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

PREDICTING TRAFFIC CONGESTION BASED ON TIME SERIES ANALYSIS

Северо-Кавказский Федеральный Университет, факультет математики и компьютерных наук имени профессора Н.И. Червякова, Ставрополь, РФ, E-mail: vvlutcenko@ncfu.ru; nkucherov@ncfu.ru; agladkov@ncfu.ru. / North-Caucasus Federal University, Faculty of mathematics and computer science named after Professor N. I. Chervyakov, Stavropol, Russian Federation, E-mail: vvlutcenko@ncfu.ru; nkucherov@ncfu.ru; agladkov@ncfu.ru

Аннотация

Дорожные заторы являются серьезной проблемой во многих городах, приводя к потере времени, увеличению загрязнения воздуха и снижению качества жизни. В последние несколько лет модели временных рядов широко используются для прогнозирования транспортных потоков и пробок на основе статистических данных. В данном исследовании анализируются данные о дорожном движении, собранные за несколько лет, и разрабатывается прогнозная модель на основе методов анализа временных рядов. Модель учитывает различные факторы, способствующие возникновению заторов, такие как время суток, день недели, перекресток. Результаты показывают, что модель эффективно прогнозирует загруженность дорог с высокой степенью точности, что может быть использовано для принятия рациональных решений и снижения загруженности дорог в городах.

Ключевые слова: прогнозирование загруженности дорог, метод Хольта-Винтера, модель ARIMA, интеллектуальная транспортная система, прогнозирование временных рядов.

Финансирование: Исследование было поддержано грантом Российского научного фонда № 22-71-10046, https://rscf.ru/en/project/22-71-10046/

Abstract

Traffic congestion is a serious problem in many cities, resulting in lost time, increased air pollution, and reduced quality of life. In the past few years, time series models have been widely used to predict traffic flows and congestion. This study analyzes traffic data collected over several years and develops a predictive model based on time series analysis techniques. The model takes into account various factors that contribute to congestion, such as time of day, day of the week, and junction. The results show that the model effectively predicts traffic congestion with a high degree of accuracy, which can be used to make rational decisions and reduce urban traffic congestion.

Key words: traffic forecasting, Holt-Winter method, ARIMA model, intelligent transportation system, time series forecasting.

Funding: The research was supported by the Russian Science Foundation Grant No. 22-71-10046, https://rscf.ru/en/project/22-71-10046/

Introduction

The development and widespread active introduction of modern electronic communication systems, global navigation systems, computer vision systems, active and passive sensors of various types and purposes has led to the possibility of solving extremely complex problems, the very statement of which two decades ago seemed impossible. Such problems undoubtedly include the problems of creating smart cities [1] and intelligent transportation systems (ITS) [2, 4]. In this article, we will explore the use of time series analysis for predicting traffic congestion, and examine some of the techniques and tools that can be used to generate accurate and actionable insights. In large cities, this task is one of the urgent ones that have to be solved in order to solve problems efficiently and completely. Due to the processes of urbanization and economic growth, there are more and more cars on the streets of cities. Traffic congestion is one of the most common problems we encounter on a daily basis. The current system of junctions with traffic lights, cannot adapt to the changes in traffic on the roads.

Time series analysis is a powerful tool for understanding and predicting patterns in data that change over time. In the context of traffic congestion, time series analysis can be used to analyze data from traffic sensors, GPS-enabled vehicles, and other sources to identify patterns in traffic volume, speed, and congestion levels over time. By analyzing this data, researchers and transportation professionals can identify key factors that contribute to traffic congestion, such as rush hour traffic, road construction, and special events.

One of the key benefits of using time series analysis for predicting traffic congestion is that it allows for real-time monitoring and adjustments. By constantly monitoring traffic data, transportation professionals can adjust traffic signals, reroute traffic, or provide real-time alerts to drivers in order to alleviate congestion and improve traffic flow. Additionally, time series analysis can help transportation planners identify long-term trends in traffic patterns, allowing them to make more informed decisions about where to invest in infrastructure improvements, public transportation, and other strategies to reduce congestion.

In recent years, there have been significant advances in the use of machine learning and other data science techniques for analyzing time series data [3]. These techniques allow for more accurate and precise predictions, and can help transportation professionals identify patterns and trends that may be difficult to detect with traditional statistical methods. Additionally, machine learning algorithms can be trained to adapt to changing traffic conditions, improving their accuracy and reliability over time.

A great number of works are devoted to the task of forecasting traffic flows in the world press. Detailed reviews and detailed classifications can be found in review publications [5-7]. Based on these works, the following main approaches to solving the problem can be distinguished:

– regression models [8, 9];
– time series models [10-12];
– neural network models [13, 14];
– method of reference vectors [15].

The purpose of this paper is to develop a model of a traffic congestion prediction system based on time series analysis. The paper has the following structure. Section 1 discusses the proposed methodology in detail. Section 2 analyzes the performance of the system when using different time series analysis algorithms. In the conclusion, the results are summarized.

1. Methodology

The presented system of traffic forecasting is shown in Figure 1. The stages of model development are described below.

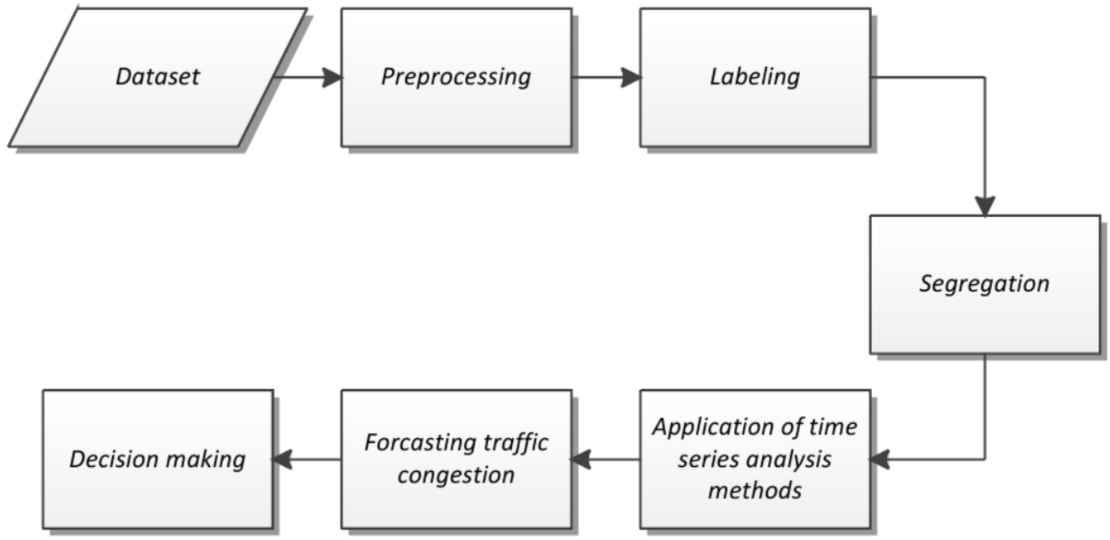

Figure 1 – Traffic congestion prediction model

Step 1: Data collection. The data for our model was the traffic dataset downloaded from the URL: https://www.kaggle.com/fedesoriano/traffic-prediction-dataset. This dataset is stored in .csv format. The dataset contains attributes such as time to the nearest second, junction number, number of vehicles, and a unique identifier. A program was developed to view and feed this data set into the proposed system. After feeding the data into the system using the pandas library in python, the proposed model reads the data set.

Step 2: Preprocessing and labeling. This step determines the size of the dataset, finding out that it consists of 48120 rows and 4 attribute columns. Then each column is checked for missing values, and the Senor ID column is discarded because it does not contribute to traffic prediction. The Datetime column, which contains the date as a string, is converted to the standard date format. Datetime is combined with junction and vehicle numbers. Finally, the resulting list is used to label these attributes for use in the next step, which is data partitioning.

Step 3: Data Segregation. At this stage, the data is divided into two lists X and Y. X contains such attributes as time and junction number, and Y as an attribute contains the number of vehicles. Then, for each junction, this list is split into four lists X_train, X_test, Y_train and Y_test.

Step 4: Application of time series analysis methods. We will use the Holt's linear smoothing, the Holt-Winter method, and the integrated autoregressive moving average (ARIMA) model as traffic prediction methods.

1) Holt's linear smoothing is particularly useful when the data exhibits a trend, meaning that the values tend to increase or decrease over time [16]. It works by estimating the level and the trend of the time series separately and combining them to make a forecast.

Forecast equation:

$$\hat{y}_{t+h|t} = l_t + hb_t$$

The level is estimated by taking a weighted average of the observed values, where the weights decrease exponentially as the distance between the observation and the current time period increases. The trend is estimated by taking a weighted average of the differences between consecutive observations, with the same exponential weights.

Level equation:

$$l_t = \alpha y_t + (1-\alpha) l_{t-1}$$

Trend equation:

$$b_t = \beta (l_t - l_{t-1}) + (1-\beta) b_{t-1}$$

where b_t is the forecasted trend component, b_{t-1} is the previous forecasted trend and β is the trend smoothing factor that can take on values $0 \le \beta \le 1$.

The Holt's linear smoothing method is recursive, meaning that the estimates for the current time period are used to estimate the values for the next time period, and so on. The method can also be extended to include seasonality by using a seasonal factor.

2) The Holt-Winters method, also known as triple exponential smoothing, is a time series forecasting technique that uses a combination of smoothing techniques to make forecasts. The Holt-Winters method is useful for forecasting time series with trends and seasonal patterns. The method involves three types of smoothing: level smoothing, trend smoothing, and seasonal smoothing [17]. The level smoothing is used to estimate the overall level of the time series, the trend smoothing is used to estimate the trend in the time series, and the seasonal smoothing is used to estimate the seasonal variations in the time series.

Level equation:

$$L_t = \alpha\left(y_t - S_{t-s}\right) + \left(1-\alpha\right)\left(L_{t-1} + b_{t-1}\right)$$

Trend equation:

$$b_t = \beta\left(L_t - L_{t-1}\right) + \left(1-\beta\right)b_{t-1}$$

Seasonality equation:

$$S_t = \gamma\left(y_t - L_t\right) + \left(1-\gamma\right)S_{t-s}$$

Forecast equation:

$$F_{t+k} = L_t + kb_t + S_{t+k-s}$$

The Holt-Winters method produces forecasts by combining the estimated level, trend, and seasonal components of the time series. The forecasts are updated at each time step using the actual values of the time series, which makes the method adaptive to changes in the underlying patterns of the time series.

3) ARIMA (Autoregressive Integrated Moving Average) is a combination of three parts: the autoregressive (AR) part, the integrated (I) part, and the moving average (MA) part.

The AR part models the dependence of the current value of the time series on its past values. It assumes that the current value is a linear combination of the past *p* values of the series, where *p* is the order of the AR component [18].

The I part deals with the non-stationarity of the time series, which means that the statistical properties of the series change over time. It models the differences between the current and the previous values of the time series, and it is used to make the series stationary.

The MA part models the dependence of the current value of the time series on the past errors or residuals of the series. It assumes that the current value is a linear combination of the past *q* errors of the series, where *q* is the order of the MA component.

ARIMA models are generally denoted as ARIMA(*p, d, q*), where *p* is the order of the AR component, *d* is the degree of differencing required to make the series stationary, and *q* is the order of the MA component.

Step 5: Decision making for traffic congestion prediction. At this point, using the proposed methods, we obtain forecasts. Transportation agencies can use the predicted traffic levels to make informed traffic management decisions. For example, they can reroute traffic, adjust traffic signal timing, or use additional resources to manage traffic. By using predicted traffic levels to inform their decisions, transportation agencies can reduce congestion, improve safety, and reduce travel times.

2. Results and discussions

The traffic congestion prediction model that has been suggested is put into operation on a Windows machine that is equipped with a Core i5 processor and has 8 GB of RAM. Python programming language is utilized, along with the Spyder Integrated Development Environment (IDE), to carry out the implementation of the model.

The proposed model is evaluated for the estimation of the RMSE (Root Mean Square Error). This is a continuous evolution of the error rates between the actual values of traffic congestion and predicted values of the traffic congestion. The evaluation is carried out using the following equation:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{\left(z_{f_i} - z_{o_i}\right)^2}{N}}$$

where, $\left(z_{f_i} - z_{o_i}\right)^2$ – differences squared for the expected and predicted traffic congestion data, N - number of predicted data.

The results are shown in two graphs in Figures 2, 3, 4 and 5.

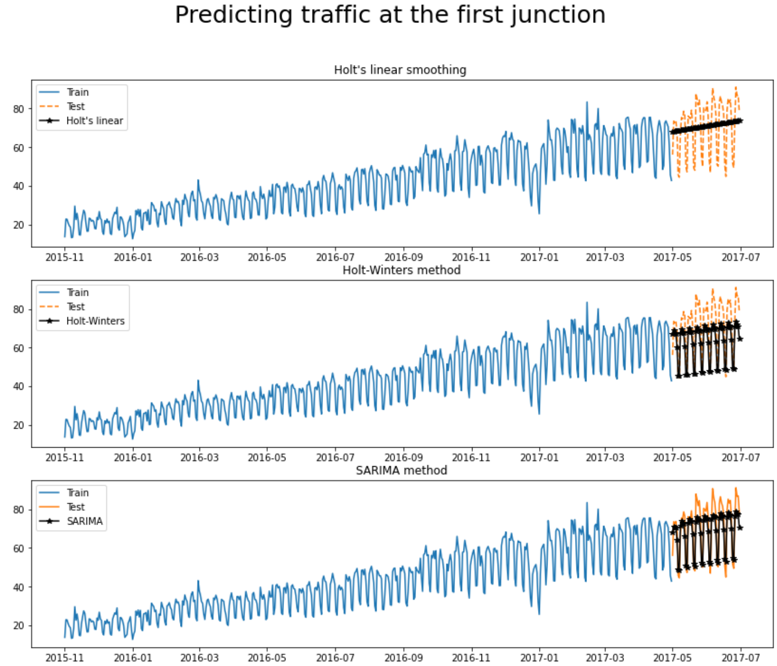

Figure 2 – RMSE estimation for predicting traffic congestion at the first junction

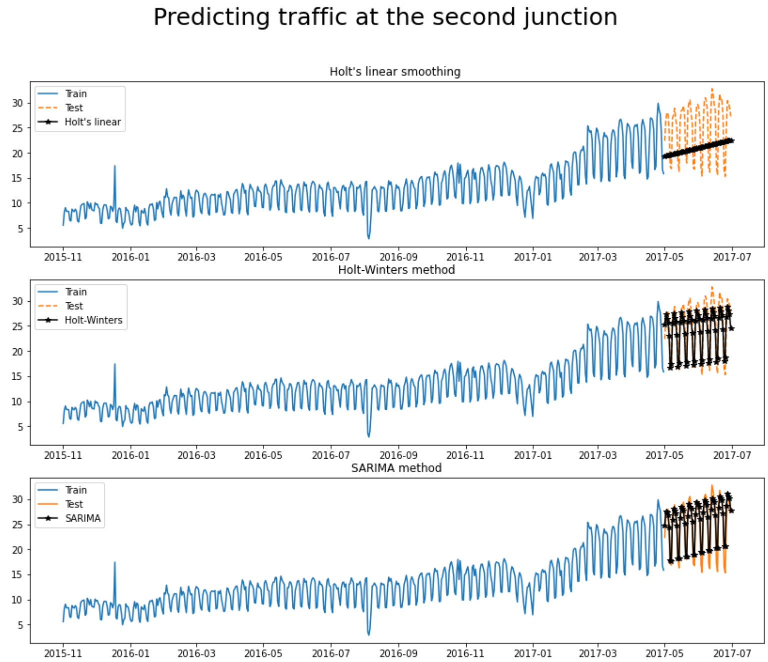

Figure 3 – RMSE estimation to predict traffic congestion at the second junction

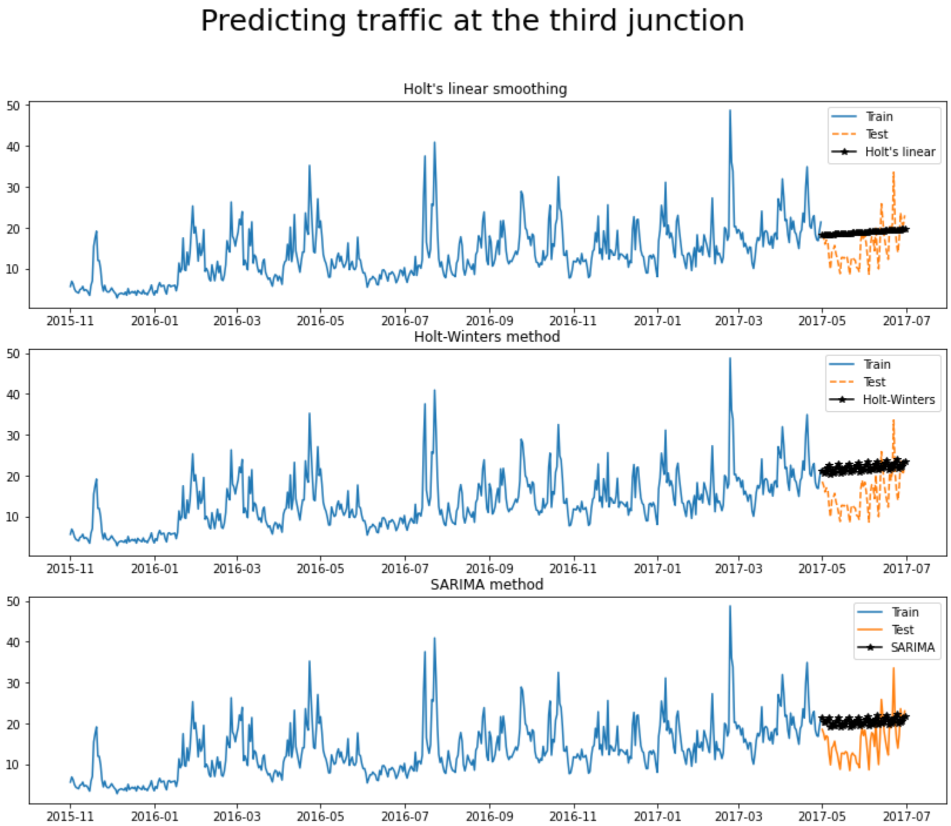

Figure 4 – RMSE estimation to predict traffic congestion at the third junction

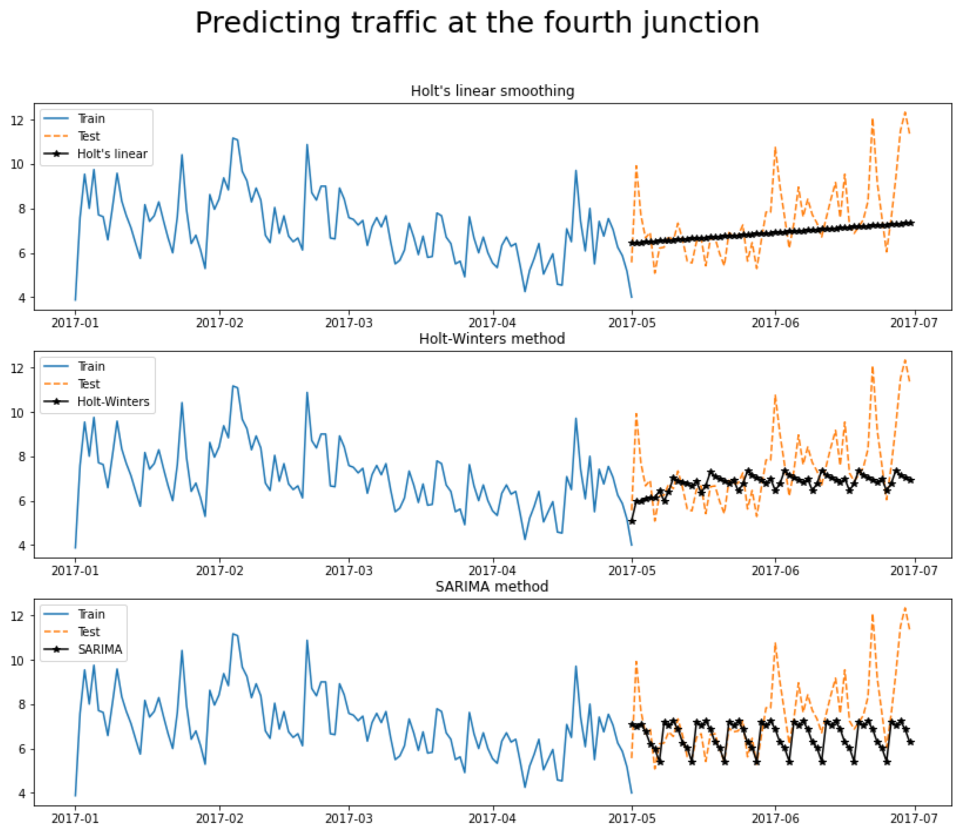

Figure 5 – RMSE estimation to predict traffic congestion at the fourth junction

The X-axis graphs the dates, and the Y-axis graphs the number of cars passing through that junction. For each of the junctions for the test dataset, which includes readings of the number of cars over two months, a prediction is created. The estimated RMSE for each of the predictions is shown in Table 1.

Table 1 – Comparative analysis of traffic forecasting methods for each junction

	RMSE		
	Holt's linear smoothing	The Holt-Winters method	ARIMA
Junction 1	14.043	10.452	6.443
Junction 2	6.898	2.33	1.849
Junction 3	5.523	7.081	5.0755
Junction 4	1.63	2.634	1.804

The results show that the forecasting of time series using the integrated autoregressive model shows an advantage compared to other selected methods, which is associated with the seasonality of the data under study. However, it should be noted that the ARIMA model requires more resources and time.

Conclusion

In this paper, we developed a traffic prediction model for predicting traffic congestion at junctions.

The proposed model has the following properties:

– It allows you to generate a prediction with pre-filtered data.
– The forecast is generated for individual junctions, which saves computing power.
– Uses statistical data from individual vehicles (GPS/GLONASS data) rather than information from traffic sensors as data sources.
– Analyzes the efficiency of individual prediction algorithms.

Further areas of work include:

– Research related to generating forecasts simultaneously for the entire street and road network of the city.
– Research related to the use of neural networks.
– Research based on the traffic data of urban passenger transport in the city of Stavropol.

ЛИТЕРАТУРА / REFERENCES

1. Batty M. Smart cities of the future // The European Physical Journal Special Topics, 2012. Vol. 214. P. 481-518.
2. Arena F., Pau G., Severino A. A review on IEEE 802.11 p for intelligent transportation systems // Journal of Sensor and Actuator Networks, 2020. Vol. 9, No. 2. P. 22.
3. Allahar H. What are the Challenges of Building a Smart City? // Technology Innovation Management Review, 2020. Vol. 10, No. 9.
4. Qu W. Short-term intersection traffic flow forecasting // Sustainability, 2020. Vol. 12, No. 19.
5. Vlahogianni E.I., Karlaftis M.G., Golias J.C. Short-term traffic forecasting: Where we are and where we’re going // Transportation Research Part C: Emerging Technologies, 2014. Vol. 43, part 1. P. 3-19.
6. Pavlyuk D. Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review // European Transport Research Review, 2019. Vol. 11, No. 1. P. 6.
7. Kashyap A.A. Traffic flow prediction models. A review of deep learning techniques // Cogent Engineering, 2022. Vol. 9, No. 1.

8. Cai L. A sample-rebalanced outlier-rejected k-nearest neighbor regression model for short-term traffic flow forecasting // IEEE access, 2020. Vol. 8.

9. Sun H., Liu H., Xiao H., He R., Ran B. Short term traffic forecasting using the local linear regression model // Journal of Transportation Research Board, 2003. Vol. 1836. P. 143-150.

10. Mai T., Ghosh B., Wilson S. Short-term traffic flow forecasting using dynamic linear models // Irish Transport Research Network, 2011.

11. Ma T., Antoniou C., Toledo T. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast // Transportation Research Part C: Emerging Technologies, 2020. Vol. 111. P. 352-372.

12. Ghosh B., Basu B., O'Mahony M. Bayesian time-series model for short-term traffic flow forecasting // Journal of transportation engineering, 2007. Vol. 133, No. 3. P. 180-189.

13. Çetiner B. G., Sari M., Borat O. A neural network based traffic-flow prediction model // Mathematical and Computational Applications, 2010. Vol. 15, No. 2. P. 269-278.

14. Yi H., Jung H. J., Bae S. Deep neural networks for traffic flow prediction // 2017 IEEE international conference on big data and smart computing (BigComp). IEEE, 2017. P. 328-331.

15. Guorong G., Yanping L. Traffic Flow Forecasting based on PCA and Wavelet Neural Network // Information Science and Management Engineering (ISME), 2010. Vol. 1. P. 158-161.

16. Gardner Jr E. S. Exponential smoothing: The state of the art // Journal of forecasting, 1985. Vol. 4, No. 1. P. 1-28.

17. Tratar L. F., Strmčnik E. The comparison of Holt–Winters method and Multiple regression method: A case study // Energy, 2016. Vol. 109. P. 266-276.

18. Lin X., Huang Y. Short-term high-speed traffic flow prediction based on ARIMA-GARCH-M model // Wireless Personal Communications, 2021. Vol. 117, No. 4. P. 3421-3430.

ОБ АВТОРАХ / ABOUT THE AUTHORS

Луценко Владислав Вячеславович, аспирант, кафедры вычислительной математики и кибернетики факультета математики и компьютерных наук имени профессора Н.И. Червякова ФГАОУ ВО «Северо-Кавказский федеральный университет», 355000, Россия, г. Ставрополь, тел. 8-960-424-33-09, E-mail: vvlutcenko@ncfu.ru

Vladislav V. Lutsenko, Postgraduate Student, Department of computational mathematics and cybernetics, Faculty of mathematics and computer science named after Professor N. I. Chervyakov, North Caucasus Federal University, 355000, Russia, Stavropol, 355000, tel. 8-960-424-33-09, E-mail: vvlutcenko@ncfu.ru

Кучеров Николай Николаевич, старший научный сотрудник учебно-научного центра «Вычислительной математики и параллельного программирования на суперЭВМ» факультета математики и компьютерных наук имени профессора Н.И. Червякова ФГАОУ ВО «Северо-Кавказский федеральный университет», 355000, Россия, г. Ставрополь, тел. 8-865-295-68-00, E-mail: nkucherov@ncfu.ru

Nikolay N. Kucherov, Senior Researcher, Educational and Scientific Center "Computational Mathematics and Parallel Programming on Supercomputers", Faculty of mathematics and computer science named after Professor N. I. Chervyakov, North Caucasus

Federal University, 355000, Russia, Stavropol, 355000, tel. 8-865-295-68-00, E-mail: nkucherov@ncfu.ru

Гладков Андрей Владимирович, младший научный сотрудник учебно-научного центра «Вычислительной математики и параллельного программирования на суперЭВМ» факультета математики и компьютерных наук имени профессора Н.И. Червякова ФГАОУ ВО «Северо-Кавказский федеральный университет», 355000, Россия, г. Ставрополь, тел. 8-865-233-06-60, E-mail: agladkov@ncfu.ru

Andrei V. Gladkov, Junior Researcher, Educational and Scientific Center "Computational Mathematics and Parallel Programming on Supercomputers", Faculty of mathematics and computer science named after Professor N. I. Chervyakov, North Caucasus Federal University, 355000, Russia, Stavropol, 355000, tel. 8-865-233-06-60, E-mail: agladkov@ncfu.ru

Дата поступления в редакцию: 12.12.2022
После рецензирования: 13.01.2023
Дата принятия к публикации: 07.02.2023