Prediction of traffic flows by applying the statistical method

O A Lebedeva¹, J.O Poltavskaya²
¹,² Angarsk State Technical University, Str. Tchaikovsky, 60, Angarsk, Russian Federation

E-mail: kravhome@mail.ru

Abstract. Traffic flows in the streets network undergo changes not only during the day, but also on the weekdays. Existing models take these changes into account as unpredictable fluctuations. Flows may vary systematically due to differences in the schedules of industrial enterprises, warehouses, and terminals. By analyzing the statistics, weekdays or seasons are determined when the flows predictably differ. To process the data, it is necessary to develop models or amend the initial data of the model (in the case of everyday dynamic models) to verify the reliability of the proposed methods. In order to investigate daily variability, as well as take into account the dynamics during the day, the distribution of flows is analyzed using functional linear models. The result of the study is the ability to predict traffic flow by day of the week or season. Testing is carried out on real data. The daily circulation was found to vary significantly for each day of the week, including differences in time and peak flux duration, as well as the relationship between peak and peak peaks. This technique aims to increase the efficiency and reliability of the transport sector as a whole.

1. Introduction

There are many studies aimed at analyzing and modeling changes in traffic flows during the day due to differences in time, throughput, and spatial distribution of transport infrastructure [1–9]. One of the current trends in this area is the elimination of differences in traffic between days, known as daily variability. This emphasis on intraday and everyday variations is a topic of research. Existing models include daily changes and represent the probability distribution for each randomly changing indicator. The indicator «type of day» refers to a combined characteristic – a day of the week or a season. This characteristic can be divided into a complex combination of days of the week and months. The stochastic model of traffic flows includes various factors for forecasting with a changing set of data, an estimation of the functioning coefficients. When modeling traffic flows, the following features were revealed: cyclic schemes based on days of the week or seasons were noted in the user’s behavior; the use of mobile network data allows to get monthly changes in the activities of the population. But such schemes disappear after data aggregation; in some cases, external factors can cause systematic patterns in behavior. Standard changes include changes during the peak and off peak periods, a decrease in demand in the winter months, or changes in daily flows due to differences in the work hours on certain days [10-12].

This study proposes a forecasting method for traffic flows that differ from each other by day of the week using available data. Models are evaluated both in size and in the form of daily flows. Variation of the flow at the same place can occur depending on many factors, including demand, route selection, departure time, and traffic conditions throughout the network. The purpose of the study is forecasting the traffic flows on the days of the week under conditions of systematic changes, regardless of the cause of them. This methodology is relevant for modeling throughput (for example, because of road works or
changes in parking rules). Such studies can be used to examine the transport mobility of the population, optimize the road network in conditions of changing traffic and traffic congestion, which depend on the day of the week or season, travel patterns and inform carriers and the public for timely changes in the route network, and adjust the working hours of enterprises or trade centers; providing objects of gravity with public transport according to a flexible schedule, depending on the day of the week and seasonal needs of the population.

2. Theoretical basis
The methodology for processing model data and applying statistical significance tests to traffic data to obtain a representative sample is reduced to the following steps.

Stage 1:
- determination of the necessary data to solve the problem;
- collection of data on traffic flow and analysis of the statistics obtained;
- calculation of the total daily flow;
- conducting checks to eliminate erroneous data;
- seasonal updating;
- analysis of variance;
- verification of model conditions;
- the condition is not fulfilled — nonparametric test;
- the condition is met — a sample of days with a statistically significant effect on the total daily flow has been obtained;
- combining a common daily flow for each day.

Stage 2:
- calculation of the daily flow for each time interval;
- assessment of the smoothness of the daily flow using B-splines and Tikhonov's regularization;
- error checking and determining additional explanatory variables;
- selection of an appropriate functional linear model;
- model comparison using a parametric test;
- verification of residuals;
- the condition is fulfilled — types of days with a statistically significant effect on the flow are determined;
- the condition is not fulfilled — nonparametric test.

Total daily fluxes can be analyzed using standard methods, subject to certain conditions [13-16]. The total daily flow should be normally distributed, and the dispersion of the population in each group should be the same. Observations must be independent. Data should be checked for normality and uniformity of deviations prior to modeling. Where these assumptions are not fulfilled, alternatives exist, such as data conversion or nonparametric tests. The assumption of independence is required not only for a functional linear model, but also for nonparametric tests. To study the total diurnal fluxes, an adjustment may be necessary to eliminate seasonal variations suggesting a multiplicative relationship. This should provide a more stable basis for testing. The process of creating a flow-based model is not a standard method used in transport studies.

The method used should allow the modeling of traffic flows for use as input to dynamic models. When comparing daily flows, it is necessary to take into account differences in shape and height. Differences in form may be associated with restrictions on the working hours of industrial facilities, and in the magnitude of the flows are likely to be dictated by demand. However, the total size and forms of the flow are inextricably linked, and these two aspects should be considered together in order to evaluate the observed patterns. Thus, the main requirements for the method are as follows: provide data on the causes of differences in the size and shape of the stream; time of day - continuous variable; accounting for the correlation between the hourly flows of the day; identification of key characteristics; checking the statistical significance of day type variables.
The presented statistical approach is suitable for identifying systematic changes in the daily traffic flow based on known factors (day of the week / season). To assess the dynamics of traffic flow, we propose the study of the relationship between the day of the week and the daily flow using a functional linear model [13]:

$$\gamma(t) = z\beta(t) + e(t). \quad (1)$$

Or in an alternative entry:

$$
\begin{pmatrix}
y_1(t) \\
y_2(t) \\
y_3(t) \\
\vdots \\
y_n(t)
\end{pmatrix} = 
\begin{pmatrix}
Z_{1,1} & Z_{1,m-1} & Z_{1,m} \\
Z_{2,1} & \ldots & Z_{2,m-1} & Z_{2,m} \\
Z_{3,1} & \ldots & Z_{3,m-1} & Z_{3,m} \\
\vdots & \vdots & \vdots & \vdots \\
Z_{n-1,2} & Z_{n-1,m-1} & Z_{n-1,m} \\
Z_{n,1} & \ldots & Z_{n,m-1} & Z_{n,m}
\end{pmatrix}
\begin{pmatrix}
\beta_1(t) \\
\beta_2(t) \\
\beta_3(t) \\
\vdots \\
\beta_{m-1}(t) \\
\beta_m(t)
\end{pmatrix} + 
\begin{pmatrix}
e_1(t) \\
e_2(t) \\
e_3(t) \\
\vdots \\
e_{n-1}(t) \\
e_n(t)
\end{pmatrix}, \quad (2)
$$

where \(y(t)\) is the vector of functional responses with respect to continuous time \(t\), which in this case will be a daily flow; \(Z\) is the matrix of traffic flows, consisting of records \(Z_{k,l}\), which are equal to 1 if day \(k\) is of type \(l\), and 0 otherwise; \(\beta(t)\) is the vector of functional coefficients; \(e(t)\) is the vector of functional residues that represent inexplicable changes after variables of the day type are taken into account.

For the model to work properly, it is necessary to determine the variables that must be included in the model, and then evaluate the corresponding coefficients. For multicomponent problems, it is proposed to use direct stepwise regression. To estimate the coefficient of the day type \(\beta(t)\), the least squares method will be used. Since the existing range between the observed and predicted daily flux needs to be minimized, it can be expressed as the sum of the squared residuals:

$$\text{Sum} = \sum_{i=1}^{n} \sum_{j=0}^{m} \int [y_i(t) - Z_{i,j}\hat{\beta}_j(t)]^2.$$

To determine the most suitable model, \(F\)-type tests are used. This is equivalent to testing the null hypothesis that a «reduced» model, including fewer predictor variables, is preferable to a «full» model, which includes one or more additional predictor variables. As an alternative to point statistics, tests have been proposed to measure the overall significance of a functional model. The test involves calculating one value for each model comparison as follows:

$$F_{nest} = \frac{RSS_{red} - RSS_{full}}{RSS_{full}} \times \frac{df_{full}}{df_{red} - df_{full}}, \quad (4)$$

where \(RSS\) – is the residual sum of squares, \(df\) is the degree of freedom of the «complete» model. This \(F\) – type test assumes that residuals are Gaussian random processes.

The research was carried out on the basis of traffic flow data in the city of Angarsk. For analysis, all days of the week were identified as different streams. The results of the analysis are presented in figures 1 and 2.
Figure 1. Average flow on Saturdays and Sundays.

Analysis of the results showed that the day of the week has a statistically significant effect on the total daily flow (flows on Tuesdays and Wednesdays and flows on Thursdays and Fridays do not differ significantly). To eliminate the effect of the magnitude of the flow, they were converted to the percentage of flows during the day relative to the total flow for the day.

The simulated schedules for the weekend and weekdays have different trends, so their comparison is carried out separately. The simulation results on weekends showed that the maximum coefficient of 0.06 is observed from 09.00 to 10.00, and the minimum coefficient is observed in the morning from 06.00 to 07.00 and after 22.00 (Figure 1).

Figure 2. Average flow on weekdays.

The simulated schedules for the weekend and weekdays have different trends, so their comparison is carried out separately. The simulation results on weekends showed that the maximum coefficient of 0.06 is observed from 09.00 to 10.00, and the minimum coefficient is observed in the morning from 06.00 to 07.00 and after 22.00 (Figure 2).
3. Conclusion
The study aims to develop a methodology for assessing the daily traffic flow and its changes by day of the week or season. The methodology includes a functional data analysis that is practically not used in transport planning. This approach has advantages, since it allows you to save the complexity of the flow dynamics, while having conceptual ease of observation. The statistical method for determining the wording of the model allows taking into account indicators that have a statistically significant effect on the flow. The analysis of real traffic flow data showed that all seven days of the week have distinctive differences in the form of daily traffic. These differences include peak periods, nighttime differences. This methodology can be used to estimate the volume of traffic flows and does not require the purchase of any special software. The methodology allows you to apply options for adjusting traffic flows with the following goals: minimizing traffic congestion depending on the time of day; influence the mode of work and rest depending on the day of the week or season; providing public transport with a more flexible schedule depending on the day of the week and the seasonal needs of passengers.

References
[1] Poltavskaya Yu O 2018 Nadezhnost’ kak pokazatel’ effektivnogo funktsionirovaniya transportnoy sistemy V sbornike: Transport Rossi: problemy i perspektivy Materiały mezhdunarodnoy-nauchno-prakticheskoy konferentsii pp 206–9
[2] Lebedeva O 2018 Modeling of public transport waiting time indicator for the transport network of a large city MATEC Web of Conferences p 04018
[3] Lebedeva O 2018 Increasing effectiveness of the transportation network by using the automation of a Voronoi diagram Transportation Research Procedia 36 pp 427–433
[4] Popova O 2018 Principles of modern route systems planning for urban passenger transport Transportation Research Procedia 36 pp 603–9
[5] Levashev A 2018 Special generators in tasks of transportation demand assessment Transportation Research Procedia 36 pp 434–9
[6] Xu D 2020 Real–time road traffic state prediction based on kernel KNN Transportmetrica A: Transport Science 16(1) pp 104–118
[7] Nemchinov D 2019 Assessment of the required street and road network development in urban agglomerations WIT Transactions on the Built Environment 186 pp 1–10
[8] Shesterov E 2017 Elaboration of a coordinated transport system in course of territorial planning of urban areas development Transportation Research Procedia 20 pp 608–612
[9] Yakimov M 2018 Methods for spatial analysis of city structure distribution to estimate city agglomeration boundaries Transportation Research Procedia 36 pp 794–800
[10] Kopylova T 2018 A Level-of-Service concept regarding intermodal hubs of urban public passenger transport Transportation Research Procedia 36 pp 303–7
[11] Nemchinov D 2017 Traffic flows merging cooperation analysis Transport Infrastructure and Systems - Proceedings of the AIIT International Congress on Transport Infrastructure and Systems, TIS pp 975–8
[12] Sechi L 2020 Planning tourist infrastructures to regenerate marginalised territories: the study case of North Sardinia, Italy City, Territory and Architecture 7(1) p 5
[13] Ramsay J 1997 Functional Data Analysis (New York Springer)
[14] Watling D 2013 Model representation & decision-making in an everchanging world: the role of stochastic process models of transportation systems Networks and Spatial Economics
[15] Watling D 2013 Modelling sources of variation in transportation systems: theoretical foundations of day-to-day dynamic models Transportmetrica B Transport Dynamics 1 pp 3–32
[16] Rasay H 2020 Modeling and analyzing incremental quantity discounts in transportation costs for a joint economic lot sizing problem Iranian journal of management studies 13 (1) pp 23–49