URBAN TRAFFIC DETECTORS DATA MINING FOR DETERMINATION OF VARIATIONS IN TRAFFIC VOLUMES

LADISLAV BARTUSKA¹, JIRI HANZL², JAN LIZBETIN³

Abstract

This paper analyses road traffic volumes in the urban environment for the purpose of traffic planning and creation of traffic models. For modelling traffic in a certain area, the initial information about transport demand and distribution in given area is required. The demand for transport is further re-distributed to the transport network and measured against the current road traffic volumes / intensity of traffic. Traffic volumes over time are characterized by various periodic and non-periodic influences (variations). By studying these variations, the tools can be specified for making the final estimate of traffic volumes for a specific time period, a specific type of road or specific vehicle category, and for improving the traffic models for a specific area. In this paper, the authors study time variations in traffic volumes using the data obtained from vehicle detectors for monitoring traffic located on roads in the city of Ceske Budejovice, the Czech Republic.

Keywords: urban road traffic, traffic variations, road traffic data mining, classification of roads

1. Introduction

The empirical knowledge based on many years of experience recognizes the factors that form the resulting traffic volumes on the roads (the determination of traffic volumes on the basis of traffic surveys). These factors are known as traffic variations and if their weights (i.e. expansion factors) are determined, the results of statistical evaluation can predict and estimate traffic volumes on a particular road.

There are known variations of transport in the form of daily variations (weekdays, weekends), monthly variations or seasonal variations (winter season, holidays, etc.). Other factors have a more or less significant effect on the resulting traffic volumes on the roads. If there is a significant influence in the time scale, the given factor needs to be further studied to what extent the intensity of traffic is affected by this variation. Many authors have tried to identify and studied these variations. Seasonal variations in relation to transport demand in the area were studied, for example, by a team of authors led by Li [13],

¹ The Institute of Technology and Business in Ceske Budejovice, Department of Transport and Logistics, Okruzni 517/10, 37001, Ceske Budejovice, Czech Republic, e-mail: bartuska.vste@seznam.cz
² The Institute of Technology and Business in Ceske Budejovice, Department of Transport and Logistics, Okruzni 517/10, 37001, Ceske Budejovice, Czech Republic, e-mail: hanzl@mail.vstecb.cz
³ The Institute of Technology and Business in Ceske Budejovice, Department of Transport and Logistics, Okruzni 517/10, 37001, Ceske Budejovice, Czech Republic, e-mail: lizbetin@mail.vstecb.cz
Stathopoulos and Karlaftis [23] or Aunet [2]. In their studies, some have identified different variations of traffic in different working weeks relating to the activities of the population in the area [5, 10, 25]. Due to the fact that traffic volumes on the roads are formed under a great influence of the catchment area, other authors also studied the relationships between traffic variations and catchment areas. For example, these authors managed to find correlations between variations in long-term transport and a given type of area [4, 7, 27]. Other authors studied variations associated with weather and weather conditions on a given day where the weather effect on the resulting traffic volumes was found negligible, except on extreme weather conditions [12, 26].

The basis for data classification is to identify systematic variations of transport. These systematic variations are expected to be the most frequent and regularly recurring. However, non-periodic, non-systematic traffic variations may also be observed on the roads. Systematic variations can be summarized in two main groups:

- variations in time, and
- variations in space.

Systematic variations in time usually vary from minute to minute (but more often they are simplified as hourly intervals) in a day, day to day of the week, month to month of the year, and they even vary across years. In connection with time changes in traffic volumes, the literature mainly describes daily trends of traffic volumes and their changes in every day of the week and with the seasonal effect – the effect of seasons on the total daily traffic volumes. The volumes usually do not differ on a normal working day (except for a road with irregular traffic without a systematic pattern) – the peak values are usually observed in the morning and then in the afternoon congestion.

The trend of daily traffic volumes may differ on different days of the week. In the vast majority of busy roads, you can distinguish between normal working days, Friday traffic and Saturday and Sunday traffic - even the trend of daily traffic volumes between each of the weekend days is mostly different [10, 16]. However, it may be challenging to study these differences in urban areas where other types of daily traffic volumes can also be observed. Significant changes can also occur during public holidays, school holidays as well as on working days before holidays or on the days before school holidays [10].

As far as transport variations in space are concerned, they can be associated with individual road sections, or directly road categories. In their research, Martolos et al. also studied transport variations in space and time. Based on long-term traffic surveys of various types of roads in the Czech Republic, they categorize roads with different traffic patterns as follows [16]:

- 1st class motorways;
- 2nd class motorways;
- 1st class roads with international traffic status (including municipal transit sections);
- 1st class roads (including municipal transit sections);
- 2nd and 3rd class roads (including municipal transit sections);
- Urban streets and roads or service roads.
The category "Urban streets and roads" in the Czech Republic is owned and managed by the municipality (city). It forms an extensive network of streets and roads in the area of cities, towns and villages supplemented by stretches of 1st, 2nd and 3rd class roads, or motorways.

For example, Schmidt further categorized roads in German cities according to their daily traffic volumes into three groups [19]: (1) streets in and near the city center with predominantly commercial and shopping traffic showing a flat daily traffic volumes, (2) radials towards the city center, access roads to the city and roads between the main attractions within the city show certain peaks in the morning and afternoon rush hours, (3) streets on the outskirts of the city, access roads to the outskirts of the city and roads between the attractions of these areas with a high proportion of commuters using motor vehicles show very high peaks in the morning and afternoon rush hours. For group 3 roads, he also observed differences in terms of longer morning peak due to occasional traffic or people commuting to work later.

Another possibility of road data classification is, for example, to categorize roads using the hierarchical method of data analysis described by Sharma et al. who classified groups and types of roads based on travel patterns of the population in the area (e.g. purpose, distance, etc.). The result of this division was four groups of roads that can be characterized as follows: "commuter", "regional commuter", "recreational" and "rural long-distance" [20].

2. Data Processing Methodology

The authors’ goal is to determine road traffic variations in urban areas. Specifically, in the city of Ceske Budejovice where detectors were installed in specific sections of busy roads across the city in 2018 to track the passage of vehicles with a long-term goal of traffic monitoring and streamlining the control of traffic lights at adjacent intersections (i.e. streamlining the road network node carry capacity) [21].

The data comes from automatic detection devices – magnetic sensors built into the road (lane). The sensors are designed to measure and compare the horizontal and vertical components of the earth’s magnetic field (magnetometers). Algorithms based on changes in the magnetic field above the road can evaluate the frequency (vehicle passage), the lane occupancy as well as the speed of vehicles, or classify vehicles into categories based on the length of the vehicle. At least two sensors must be placed in each direction at a sufficient distance from each other to identify other patterns of the traffic flow (e.g. speed) as well as classify any recorded vehicle in one of the reference categories. The more detectors installed one after the other in the lane axis, the more accurate the output of the algorithm. Magnetometers working as described above have been evaluated as accurate vehicle detection devices [4, 14, 18].

Magnetic sensors are located on the selected busy 1st, 2nd and 3rd class roads and urban streets / roads in the inner city. The number of sensors depends on the location – in some sections the sensors are installed only in one direction (mostly access roads to the city center), in others they are installed in both lanes (especially the backbone roads in the
Data collected from automatic detectors show hourly frequencies of vehicles (at hourly intervals) in each day of the entire year 2019 (from 1 January 2019 to 31 December 2019). When sorting data collected from detectors, serious deficiencies were identified in most of the datasets. These were often fatal errors in the form of invalid data. Invalid data was evaluated by the author as a result of measurement errors which was also confirmed by the City Town Hall of Ceske Budejovice personnel (for example, failure of electronic devices in the system or directly in the sensor, shutdown of the detector during road construction work or maintenance of technologies, or disturbing traffic environment).

These errors were identified based on clearly specified criteria. Due to the fact that these were common errors lasting weeks or even months, the datasets were incomplete. For the
subsequent analysis, it was therefore necessary to select datasets that showed a certain homogeneity [17, 18]. Such files can further be analysed for traffic variations in the short term, or possibly in the long term of the whole year in the case of a uniform dataset (complete data for the whole year).

Based on the analysis of datasets considering the measurement errors identified as explained above, the following time periods were selected for specific road sections (detector locations) to provide reliable sampling data:

- 1st class roads in the city – 2 annual two-way volume datasets (D56 and D35);
- 2nd class roads in the city – 1 annual two-way volume datasets (D25);
- Urban streets / roads – 3 annual two-way volume datasets (D18, D43 and D65).

The above datasets were further classified based on time series validated by the scientific community. Datasets contained all-day traffic volumes for each day of the week (Monday to Sunday) with regard to weekly variations – each working day and each weekend day and public holidays were evaluated separately taking into account the seasonal variations in traffic volumes. These time series can be defined as traffic intensity $I_{d} = (q_{1d}, ..., q_{nd})$ where $n$ is the number of intervals in a series for $d$ day. For these datasets, traffic intensity details were provided in hourly intervals ($n$=24).

### 2.1. Data Filtration and Statistical Methods

As mentioned above, data from the specified time intervals and locations were analysed due to measurement errors. In this preparatory phase, data were sorted according to the given criteria to obtain an almost homogeneous statistical dataset. Subsequently, the datasets were analysed to identify other anomalies in the datasets (data filtering). For each communication, the initial traffic volume data were broken down using a pivot table as follows:

- traffic volumes on working days in the school year;
- traffic volumes on working days during the summer holidays;
- traffic volumes on weekends in the school year;
- traffic volumes on weekends during the summer holidays.

The interquartile margin rule was applied to each category of data classified as above. This rule allows to identify values that do not fall into the overall structure of other data in the dataset [28, 29]. To apply the rule, the 1st quartile ($Q_1$) and the 3rd quartile ($Q_3$) of the statistical dataset must first be determined. Consequently, the interquartile range must be calculated from the following relation:

$$IQR = Q_3 - Q_1$$  \hspace{1cm} (1)

Consequently, the lower and upper limits must be determined to separate the data that do not fall into the overall data structure of the statistical dataset. The lower limit is calculated from the relation:

$$MIN = Q_1 - 1.5 \times IQR$$  \hspace{1cm} (2)
The values of the statistical dataset that are lower than the calculated MIN value are outliers.

By analogy, the upper limit is calculated from the relation:

\[ MAX = Q_3 + 1.5 \times IQR \]  

(3)

The values of the statistical dataset that are higher than the calculated MAX value are outliers.

By applying the above mentioned rule, outliers were deleted from the datasets of specific locations that might cause distortion of the calculated outputs. Mainly public holiday data have been removed. Almost all public holidays have been removed. Calendar holidays that have not been removed fall mainly on Sundays when the trends of traffic volumes largely correspond to the standard Sunday traffic volumes.

The data from which the outliers were removed were subsequently merged into one dataset within each communication. The arithmetic mean of traffic volumes in hourly intervals at D56 and D35 locations (1st class roads) and at D65, D43 locations and D18 (urban roads and streets) was determined for the purpose of further calculations. The arithmetic mean was solely calculated from data of the days on which the data were available for both roads. Data from 337 days were used for the calculation in case of 1st class roads and data from 275 days in case of urban roads and streets.

For the D25 location (2nd class road), it was a single set of data and significant fluctuations in daily traffic volumes were identified in a short period of time. Due to the expected erroneous measurement, all data with a daily traffic volumes of less than 318 vehicles were deleted, beyond the filtration technique in the form of an interquartile range. In this way, data for 107 days were deleted from the statistical dataset. A total of 253 days were used for calculation of the given 2nd class road after the outliers and other measurement errors were deleted.

The calculation of average daily traffic volumes and hourly traffic volumes for each day of the week was only based on common data of datasets from locations corresponding to the specific road categories. In these categories, traffic volumes from the locations in the set were compared using a variation coefficient with a positive result - the data showed similar trends in traffic volumes on given days of the year [11, 15]. The correlation coefficient of hourly traffic volumes from locations in the specific road categories was also high (more than 0.90) - therefore these are almost homogeneous datasets.

The seasonal index was calculated for the average traffic volumes in the specific road categories. It is the ratio between the average weekly volumes and the annual average. Based on the seasonal index, the change in the average weekly volumes over the whole year was plotted as a graph. It made it possible to estimate the seasonal variation in road traffic in urban environments. Furthermore, the average daily volumes for all days of the week and the average daily volumes by specific day of the week were calculated for each group of roads [24, 25]. A graph was used to compare the daily traffic volumes by day of the week with the average daily volumes by week.
3. Seasonal variations in traffic volumes

From a global perspective, road traffic variations may vary from country to country and region to region due to many factors – especially cultural, social and economic differences. The given road locations also play a role here. The authors try to identify time variations of the road traffic in the urban environment. Seasonal variations and daily variations of traffic are studied to calculate the estimate of annual averages of daily volumes only based on the knowledge of a limited sample of hourly volumes from a short-term traffic survey [16].

The authors evaluated the seasonal variations in traffic volumes for three categories of roads in Ceske Budejovice using the seasonal index which is the ratio between the average weekly volume (average from locations on the same category roads) and the annual average of all weekly volumes in 2019. For illustration purposes, the average number of vehicles on urban roads was less than 5 million recorded vehicles for the whole year. The seasonal index was evaluated separately for three time intervals in a day: 6 am to 10 am (morning peak), 10 am to 2 pm (off-peak hours) and 2 pm to 6 pm (afternoon peak).

1st class roads – The data evaluated for this category of roads were collected from two road profiles marked I/3, namely the profile of Strakonická Street and Na Dlouhé louce Street. The traffic flow on the reference 1st class road is affected by both transit transport (being the backbone transit road for the transit flow in the city due to absence of a bypass) and the city public transport. Strakonická Street is a significant road connection of the city districts with the average traffic volumes of app. 45,000 vehicles in both directions on working days (according to data collected from detectors). Road I/3 is also part of the international E-road network marked E55.

The seasonal transportation services index calculated on the data collected from detectors on the sections of road I/3 is shown in Figure 2. Significantly above-average traffic flow volumes are evident during the morning and afternoon peaks (the annual average is corresponding to 1 on the y-axis) throughout the year 2019, except for the last weeks of December, the weeks of January and the beginning of February. The dotted curve shows the moving average of the weekly traffic volumes.

For the morning peak, a significant increase in traffic is evident in the period of May and June (weeks 20-26) where the seasonal index reaches its maximum. During the summer holidays, the average weekly volumes in the morning rush hour are close to the values in the off-peak hours which are probably a consequence of the increase in recreational traffic in the summer season with a view to the existence of popular tourist destinations in South Bohemia.
Fig. 2 Seasonal index of traffic volumes in the 1st class road category, divided by morning peak (top), off-peak hour (middle) and afternoon peak (bottom)

Fluctuations are mainly observed in the last weeks of December (due to the behaviour of the population before, during and after the Christmas holidays). There is heavy traffic in the afternoon rush hour especially in the period before the holidays (weeks 50 and 51) and the morning traffic shows a maximum in week 52. On the contrary, the morning rush hour has significantly below-average values - the morning rush hour clearly occurs much later in this period.

Given the facts above, these are not abnormal traffic patterns that would not be known to the professional public. Traffic on these roads corresponds to normal traffic on major roads across the Czech Republic [8, 16].

2nd class roads – For the category of 2nd class roads, the set of data was only obtained from one road II/157. It can be characterized as a through road connecting the North and
South of the city outside the center. The seasonal traffic index, especially during the morning rush hour, shows high fluctuations in terms of changes in weekly traffic volumes. This is probably due to the nature of the traffic which may be irregular on this road. However, incorrect detector function is also possible.

The graphs in Figure 3 show heavy road traffic in the morning and the afternoon rush hours when the weekly traffic volumes are above the annual average, especially in the spring (congestion in the morning prevails during the summer holidays). In the autumn, traffic volumes in the morning and afternoon congestions are more in line with the annual average while above-average volumes are outside the rush hour.

Fig. 3 Seasonal index of traffic volumes in the 2nd class road category, divided by morning peak (top), off-peak hour (middle) and afternoon peak (bottom)

**Urban streets and roads** – The traffic volume data were obtained from urban roads with specific traffic patterns – carrying traffic from the higher-category roads in the surroundings directly to the city center while connecting densely populated areas on the outskirts of the city with the natural administrative and cultural centre. Specifically, these are the profiles of Na Sadech, Husova and Pražská třída streets.
The evaluation of the seasonal index of traffic volumes separately by time intervals in a day revealed slight fluctuations in volumes (graphs in Figure 4). However, traffic volumes are very similar in these parts of the day for most of the year – the beginning of the year and spring and autumn show a similar trend of weekly traffic volumes. There is a notable change during the summer holidays (weeks 27 to 33) when the morning rush hour occurs a little later and traffic is heavier in the morning and afternoon.

![Fig. 4. Seasonal index of traffic volumes in the category of urban streets and roads, divided by morning peak (top), off-peak hour (middle) and afternoon peak (bottom)](image)

The above graphs show significant traffic volumes in the last weeks of August and in the first weeks of September (weeks 34-39), and heavy traffic in the afternoon rush hour in the autumn. Above-average traffic volumes during the off-peak period in the last weeks of December is also evident while the morning and afternoon peaks are much below the average.
4. Day-to-day variations in traffic volumes

Day-to-day variations in traffic volumes on each of the reference roads are usually expressed by the daily trend of volumes in a given day – these are hourly volumes around the clock. The authors used a different approach to identify day-to-day traffic variations. Variations in traffic volumes in the following graphs are expressed as changes between the average daily volumes in a particular day of the week and the daily traffic volumes averaged for all days of the week. Those are the average hourly traffic volumes per working day of the week and weekend days (Saturday and Sunday altogether) and the hourly frequency deviations from the weekly average. The summer holidays are removed from these datasets. Figure 5 illustrates the change in daily variations in traffic volumes at detector locations on 1st class roads. The graphs clearly show similar changes in traffic variations on working days – the daily trend of traffic volumes is almost identical but for smaller differences when, for example, there is heavier traffic in the off-peak period and in the evening on Thursdays.

Fig. 5. Day-to-day variations in 1st class road traffic volumes
As expected, Friday's traffic congestion on 1st class roads is experienced in the morning and afternoon as well as in the evening. At the same time, traffic demand is significantly lower on weekends, especially in the morning when traffic only increases in the late morning unlike working days. In overall, the daily traffic volumes on Saturday and Sunday are very much below the average.

The interesting fact about the 2nd class road (D25 location) is that a roughly similar trend of daily traffic volumes was observed in all days of the week, including weekends, with significant morning and afternoon rush hours which is unusual compared to other categories of roads [11, 22]. By comparing the deviations of hourly traffic volumes in a day to the average values during the week, changes in traffic variations on different days are identified as can be seen from the graphs in Figure 6.

![Fig. 6. Day-to-day variations in 2nd class road traffic volumes](image)

Monday's traffic is comparable to weekend traffic when the morning rush hour is on average lower than on other working days. The Monday afternoon congestion is also below the
average values during the week. It should be noted the deviation from the average lies in the absolute numbers of vehicles, being in the order of tens to hundreds of vehicles. A quite similar trend of daily vehicle volumes is experienced on Wednesday, Thursday and Friday as evidenced by the deviation of values from the weekly average. The observed daily traffic volumes and deviations of values from the weekly average indicate the specific traffic patterns that can be attributed to the irregular and specific use of the road.

In the case of urban streets and roads, above-average traffic volumes are observed on weekdays for most of the time interval from early morning to late afternoon. Friday's traffic is characterized by the fact that the afternoon congestion is not as pronounced as in the case of 1st class road results, with heavier evening traffic (Figure 7).

In overall, the daily trend of traffic volumes corresponds to daily trends on other types of roads – this paper can only compare traffic volumes on 1st class and 2nd class roads and urban streets and roads. Based on previous empirical evidence, similar systematic variations in traffic volumes are experienced on such other types of frequented roads that are
influenced by primary factors such as the size of traffic demand between different areas, and socio-economic factors, division of transport labour, or other aspects of traffic behaviour [3, 24].

5. Discussion and Conclusions

In this paper, the authors discussed the time factors that affect road traffic volumes in the urban environment with regard to the importance of the road. The so-called seasonal index was calculated for the average traffic volumes of different categories of roads, indicating the ratio between the average weekly traffic volume and the annual average volume. Based on the seasonal index, the change in the average weekly volumes over the whole year was expressed by a graph. In addition, the average daily volumes for all days of the week and the average daily volumes for each specific day of the week were calculated for each group of roads. The daily traffic volume trend by specific day of the week was compared to the average daily volume trend by week and the results were expressed by a graph. Only two-way traffic volumes were subject to evaluation without consideration given to the direction of vehicles and vehicle categories.

It was established that the daily road traffic volume trends of the reference categories in the urban environment do not noticeably differ in relation to the expression of the absolute frequency of vehicles. There are differences concerning the morning and afternoon congestion when similar traffic volumes were recorded on urban streets and roads in the morning and afternoon congestion on weekdays while traffic volumes were significantly higher in the afternoon rush hour on 1st class roads (Friday's morning traffic congestion seems even weaker compared to the next period of the day on the 1st class road and urban streets and roads).

On the 2nd class road, significant differences were identified between the two peaks and the period between them (10 am – 2 pm) on weekdays – both peaks reached significant extremes compared to off-peak traffic. Similarly, different traffic patterns were observed on this type of road on weekends (compared to other categories of roads), getting close to the traffic volumes on weekdays. It was a set of data from only one road registered as a 2nd class road, but with clearly specific traffic. This may be evidenced by the fact that road traffic can vary significantly in urban environments on different road categories.

The application of the seasonal index identified relative deviations of weekly averages from the average values of weekly traffic volumes for the whole of 2019 to be further broken down by time intervals in a day. The authors consider the fact the afternoon congestion on urban roads is above average in the autumn period while it showed below-average values throughout the spring period to be an interesting finding. Significantly above-average values in both peaks and off-peak periods were observed in the first weeks of September when traffic volumes reached the maximum values. Differences were identified by comparing the traffic volumes on other reference categories of roads as related to traffic patterns of said road category.

For the analysis of traffic volumes on urban roads, heavy traffic was observed in the morning and afternoon peaks similar to the road traffic volumes on most other roads with similar
importance. In the urban environment, however, it is possible to identify other roads with specific traffic patterns that may not correspond to traffic on such frequented roads [9, 30]. For the purposes of research into traffic variations in this paper, however, it was not possible to identify this traffic on other urban streets and roads due to the absence of data. In particular, access roads in commercial zones, service roads in residential buildings, and service roads in administrative centers, or urban roads connecting industrial zones have different traffic patterns with specific trends of daily traffic volumes.

The hierarchy of roads within the road network indicates to some extent the magnitude of traffic volumes and the trend of daily traffic variations. However, the classification of roads based on categories according to their importance or affiliation to a given road administrator (motorways, 1st and 2nd and 3rd class roads, urban streets and roads, etc.) is a secondary factor that is taken into account when estimating the volume of roads. From this point of view, the road category itself is a kind of precursor indicating the resulting traffic volumes. In addition to the primary factors generating traffic volumes, there is also the effect of so-called traffic induction where higher capacity roads induce additional traffic volumes as a result of better traffic conditions.

It is very costly to collect long-term data concerning traffic volumes from the largest possible sample of different road categories if we consider the deployment of vehicle detection sensors (automatic traffic counters). In an urban environment, the existing stationary detectors and CCTVs (video surveillance) at intersections can be used to obtain this data and use it for cluster data analysis (e.g., using neural networks) without the initial data classification by road categories. The cluster analysis of traffic volume datasets from all detectors would allow to identify datasets that show similarities [6]. If so-called "clusters" (similar datasets) were consequently assigned to the road network, the types of roads with similar traffic patterns could be identified.

The traffic analysis can be used to improve the existing tools for estimating traffic volumes and for research of road traffic flows, or other traffic engineering applications with regard to the year-round type of analysis. The authors are also aiming to create a classification of roads within the urban road infrastructure to reflect the specific patterns of traffic on given categories of roads. However, this is a very challenging task due to the complexity of traffic in the urban environment and traffic variations in the cities of different regions although it would help to improve the transport models needed for transport and land-use planning.

**Nomenclature**

CCTV  Closed-circuit Television

**Acknowledgement**

This manuscript was supported within solving the research project entitled "Autonomous mobility in the context of regional development LTC19009" of the INTER-EXCELLENCE program, the VES 19 INTER-COST subprogram.
References


