Archives of
Transport System Telematics

Volume 11
Issue 1
February 2018

Trip Volume Seasonal Variations at Regional Level – Case Study of Małopolska GSM OD Matrices

R. KUCHARSKI, A. SZARATA, J. MIELCZAREK, A. DRABICKI
CRACOW UNIVERSITY OF TECHNOLOGY, Warszawska 24 31-155, Kraków, Poland
EMAIL: rkucharski@pk.edu.pl

ABSTRACT

In this paper we analyze the big-data set of GSM origin-destination matrices collected between 360 districts (Powiat) in Poland to indirectly observe the trip volume fluctuations: weekly and seasonal. We utilized the dataset obtained from BTS position for all clients registered to a single provider. The data was collected over several days, which allows for a valuable temporal analysis of trip volumes. We analyze internal and external (inbound and outbound) trips in Małopolska region, intra-zonal trips (within district), inter-zonal (between districts of Małopolska). We discuss the general variability of observed trip volumes. We present the weekly fluctuations (working day, Friday, Saturday, Sunday) and the seasonal ones (winter, holiday, etc.). We verify the results obtained from GSM data with the traffic counts and their seasonal variations provided by national road administration (GDDKiA). Main contribution of the paper is presenting the observed fluctuations from the GSM data and comparing them with the classically collected data, as we demonstrate the results are comparable.

KEYWORDS: demand model; floating car data; GSM traces; seasonal variations; GSM OD data

1. Introduction

In the age of growing development of various Intelligent Transportation Systems (ITS) there is a need to collect, analyze and synthesize high quality traffic information. Traditional and the most commonly used method for road traffic data collection are manual counts, which seem to be the most accurate, but often have inaccessible costs [1]. This costs are bigger when observation area is extended or time span is long, such as in regional scale. For this reason it is frequently observed, that traffic management authorities have to rely on an infrequent, incomplete, out-to-date surveys and/or small population samples [4]. Such dataset, in general, does not provide a realistic and credible insight on the traffic situation.

Therefore various alternative information sources, e.g. big data such as the real-time traffic counts [3] or GSM trip traces seems to has nowadays a huge potential to improve quality of traffic measurement and reduce its costs [13, 14, 15].

In the contemporary society almost every single person uses mobile phones during every-day movement. That is why the cellular trip observation has a potential to lead near real-time traffic information at a huge scale. Toole et al. [11] elaborate system to estimate multiple aspects of travel demand using GSM records data. Järv et al. [11] used mobile phone tracking to modeled traffic flow at Tallin - Tartu highway during several periods (daily, weekly and monthly) and compare them to traffic counts. In the result they achieved coefficient factor between modelled traffic and observation of 0,98. Becker et al. [2] conducted an investigation over using a cellular position records to estimate relative traffic volumes. His research showed high correlation between estimated traffic volumes and vehicle counts from loop detector (correlation coefficient of 0.77). In similar study Vaccari et al. [12] compared GSM data to the traffic flow data between Brooklyn and Manhattan and also achieved correlation coefficient of 0.77. Thiessenhusen et al. [10] also proved relationships between traffic volume peaks (morning and afternoon) comparable for GSM and traffic counts data.

Big data such a GSM *od* matrices are also useful in the traffic flow forecasting and prediction. Several deep-learning-based traffic flow prediction methods were elaborated using mobile phone trip traces [5, 8].

Large scale, strategic modelling of transport system, typically relies on average values. Frequently, appraisal is made upon the so-called average annual daily traffic (AADT) where the mean, representative daily traffic volume is obtained with established methodology. This allows to measure some set of specific days and use them as a sample to estimate the annual average. Those days are typically sampled throughout the seasons days of week, holidays and weekends. Obtained value is representative for strategic purpose and applicable in most of use cases.

Yet, since traffic volume is a random variable, apart from its mean, it is reasonable to understand its variability, i.e. how the actual traffic volumes deviate from AADT. This is crucial when for some days the traffic volumes are significantly higher than average, which may lead to oversaturation and delays. That is why it becomes popular to rely in the planning and control problems, not only on the AADT, but also on its variability.

In this paper we analyzed big data set of GSM *od* matrices, present seasonal and weekly fluctuations and compare them with the traffic counts and their seasonal variations provided by national road administration (GDDKiA). We propose a method to estimate traffic flow fluctuation based on big data set.

2. Method

Usually, quantitative analysis of transport system performance are based on typical (expected, mean) values. This refers both to the typical demand pattern and the typical road network. For most use cases it is acceptable to rely on expected values in the decision making process. However, for some specific cases, understanding an intrinsic variability of transport system is crucial. This may refer to variations:

1. within-day (to capture hours of highest demand),
2. weekly (to capture days of highest demand) and
3. seasonal (to capture seasons of highest demand).

In this paper we use empirically observed values to estimate weekly and seasonal variations with a following method. We star from the generic variable q denoting the trip volume (quantity), obtained from a generic demand model D. In the baseline approach, this is understood as a typical trip volumes $\overline{q}$ obtained from the typical demand model $\overline{D}$. However, due to some variation, the observed trip volumes $\hat{q}$ may be different. We express this with:

$$\hat{q} = \alpha \cdot \overline{q} \qquad \textbf{(1)}$$

where α is the factor allowing to cover seasonal, or weekly trip volume variation. While the root of the variation typically lays in the demand model (D), in this paper we assume the demand model remains typical and the variation is handled at the level of resulting trips volumes only. This is motivated by the nature of the data that we used to analysis, which did not allow a reasonable understanding of demand pattern (e.g. trip generation rates, trip purposes composition, mode choice, etc.) and its variations. Similarly, we assume linearity between trip volumes at different levels, i.e. seasonal variation α of 1.05 means equally the growth in total number of trips, number of trips at each *od* pair, number of trips and cars on each path and on each road segment by factor of 1.05.

We propose the method derive the trip volume variations indirectly, from the variations of the observed GSM *od* matrices. The GSM *od* matrices report the number (quantity) of trips between given origins o and destinations d. Trip is understood as the event of transferring a device registered to a given provider between GSM stations. Such definition significantly reduces accuracy of short trips (which may be conducted within range of a single GSM station), yet for longer, regional trips, obtained values are reasonable. We collected a set of such GSM *od* matrices recorded over several days, with each matrix containing trip volumes between origins and destinations observed during a single day. In the specific context of this paper, origins and destinations of matrices were counties (powiaty) of Poland, i.e. for each trips its' GSM station at the origin and at the destination is uniquely assigned to a county in Poland. Thus, the trips with origin and destination in the same count are reported on the diagonal of the *od* matrix. We focused on one selected region of ca. 20 counties out of over 300 in Poland. In the above setting, the generic trip volume q can be now specified as either:

- number of trips between a given origin o and destination d: q_{od}
- number of trips originating from a given county: $q_o = \sum_{d \in D} q_{od}$;
- number of trips finished in a given county: $q_d = \sum_{o \in O} q_{od}$;
- number of intra-zonal trips (inside the county) q_{oo};
- total number of trips recorded: $q = \sum_{o \in O, d \in D} q_{od}$.

For further analysis we categorize origin and destination counties into:

- External counties *E*□*Z* (outside of analyzed region)
- Internal counties *I*□*Z* (inside the analyzed region)
- Capital county (central agglomeration) *c*□*I*□*Z*.

With such categorization of origins and destinations we can categorize trips into:

- Inter-zonal: between internal counties $q_{od} : o, d \in I$
- Intra-zonal: within one internal county $q_{od} : o, d \in I, o = d$
- Outbound: from internal counties outside (to external counties) $q_{od} : o \in I, d \in E$
- Inbound: from external counties inside (to internal counties) $q_{od} : o \in E, d \in I$
- To/from central agglomeration $q_{od} : o = c \vee d = c$

Since we have observed *od* matrices over several days, we denote each of them with a day index i so that each of above variables is identified as a tuple: $\mathbf{q} = \{q_i : i \in I\}$. Observation period can be categorized into either seasons (winter, spring, summer, autumn) and days of the week (1 to 7), or day types (working days, Fridays, weekends, holidays).

Thanks to the above tuple representation we can understand and describe the variability. Specifically, we understand the mean of a tuple as the typical value. Usually, this refers to the so called

AADT, here we avoid strongly stating that this value represents an annual average and barely state that it is a mean of observed values. From the tuple and its mean we can reproduce the α_i as the ratio between volume observed at day *i* and the mean value:

$$\alpha_i = q_i / \overline{q} \qquad \textbf{(2)}$$

This tuple can be further analyzed statistically, specifically we are interested in coefficient of variation *v* observed on a set of daily matrices:

$$v_I = \sigma(\mathbf{q}_I) / \overline{q}_I \qquad \textbf{(3)}$$

where σ is the standard deviation obtained from the sample consisting of daily matrices observed over *I* days.

In the following section we use the proposed method to derive the traffic variations in Małopolska region in Poland. We will focus on seasonal factors α for various weekdays, seasons in various relations (inbound, inter-zonal, etc.) and variability *v* of trip volumes observed during those days.

3. Results

3.1. Data

We illustrate the above method with the data obtained from one of GSM providers in Poland. He shared with us the GSM *od* matrices where the trips between counties in Poland were reported. We used the slice of dataset, where trips with either origin or destination in Małopolska (one of the regions in southern Poland, of ca. 3M inhabitants). The diagonal of the matrix contained trips made within a county, but between the GSM receiving stations. We collected data from 42 observation days covering all days of the week, seasons and day types. We observed 7 winter, 14 spring, 14 summer, 7 autumn days. 8 Sundays and holidays, 5 Mondays, 6 Fridays, 6 Saturdays and 17 Other working days. In total, we observed over 20M trips, 12M of which were inter-zonal, 8M intra-zonal, 1.7M outbound and 1.7M inbound. Around one fourth of the trips were related to the central agglomeration (Table 1).

Table 1. Observed trip volumes [own study]

OD relation	Trip volume	Trip volume to/from Central Agglomeration (Kraków)
Intra – zonal	11 982 072	1 457 252
Inter – zonal	8 304 584	1 166 261
Outbound	1 746 493	258 154
Inbound	1 751 117	265 673

In the first part we present the variability of observed trips volumes in general and per day types through their variability *v* (eq. 2). While in the second part we report the weekly and seasonal variations through their factor α (eq. 1).

3.2. Trip volume variability

We present variability on a table with generic statistics and with box-whiskers plots below. We divided 42 observation days into 17 working days, 14 weekend days and 14 autumn days. The variability of the total dataset is around 11%, yet it drops to 6% on working days and 9% on weekends (Table 2). Average number of trips is 142 000 per day, which grows to 154 000 on working days and drops to 123 000 during weekends. Box and whiskers plots (Fig.1) show that general variability is high, yet working days are somehow similar, with inter-quantile range way smaller than for all 42 observed days. The variability though remain relatively high for weekends, where many atypical days were still observed.

Table 2. Trip volume variability [own study]

Description	Complete dataset	Working day	Weekend	Autumn
Count	42	17	14	14
Average	142 644	154 141	123 886	136 878
Standard deviation	16 666,4	8 782,04	10 996,9	15 039,7
Coeff. of variation v	11,68%	5,69%	8,8%	10,98%
Minimum	11 0076	14 6904	11 0076	11 0076
Maximum	17 3279	17 3279	14 9611	15 2072
Range	63 203	26 375	39535	41996

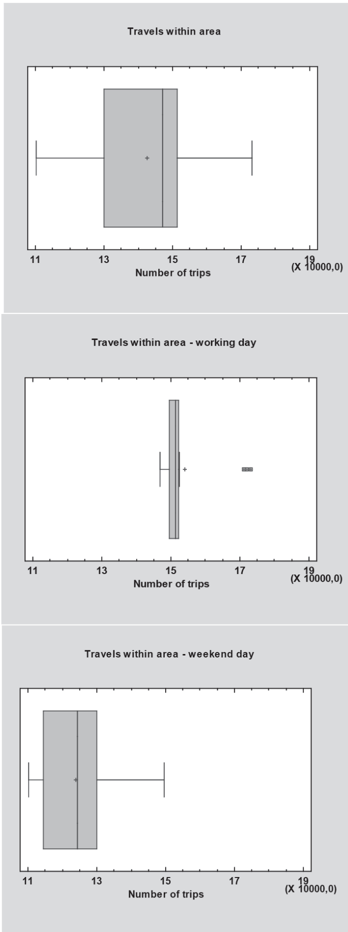


Fig. 1. Variability of trip volumes observed during various day types and seasons [own study]

3.3. Seasonal variations

In this section we present results of analyzing GSM *od* matrices for Małopolska region and compare them with the traffic counts and their seasonal variations provided by national road administration (GDDKiA).

Basic source of information about traffic flows on the motorways, national and province roads in Poland are General Traffic Counts (GPR) performs by central administration authority for issues related to the national road system (General Direction for National Roads and Motorways - GDDKiA). This observation are being done regularly – every five years within strictly defined days during the year. It provides complex information about traffic volumes, structure and variability during different periods of time.

In our investigation we focused on a traffic seasonal and weekly fluctuations in reference to Average Annual Daily Traffic (AADT). Seasonal factors published by GDDKiA are presented in the Fig. 2

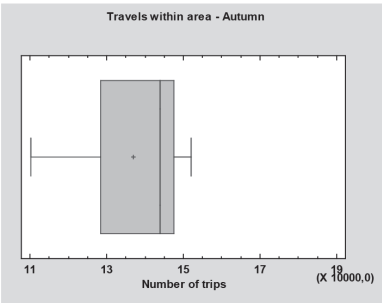


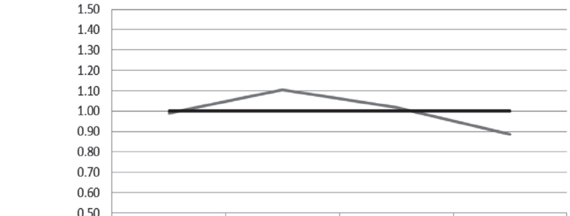
Fig. 2 Seasonal factors based on GPR traffic counts [16-18]

Factors presented in the Fig. 2 refer to the general traffic i.e. traffic observed during working day for whole national road network in Poland. We can observe increased volume during the summer months (over AADT) and decreased in winter (below AADT). In other seasons it remain stable ca. AADT level.

In the next step we compared factors published by GDDKiA to the same factors based on GSM *od* matrices. We stratified big data set into samples depending on origin and destination position in reference to Małopolska Regional Demand Model. Our investigations shows several differences between data sets. As we can observe in Fig. 3 the highest value of seasonal factor based on GSM *od* matrices occur in autumn. Summer traffic is comparable to AADT. For winter and spring observed traffic variation factor is similar to same factor based on big data.

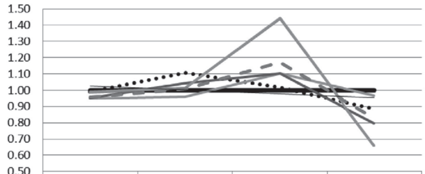
This differences may result from the sample size for each seasons in GSM *od* matrices (as we mentioned in point 3.1, we analyzed 14 summer and only 7 autumn days) or to be caused by dissimilarity of observation days for both data sets.

We also analyzed traffic fluctuation in trips to central agglomeration in the Region (Kraków) based on GSM *od* matrices (Fig. 4). They differ significantly from observations, but in some points are quite similar to other variations received from GSM *od* matrices. The highest traffic in this dataset is observed in autumn and the lowest in summer months. It can be caused by touristic trips influence and specification and attractiveness of the city. For example traffic volume drop in summer months may result from smaller number of trips in the relation with schools and university.



	Spring	Summer	Autumn	Winter
General traffic	0.99	1.11	1.02	0.89
AADT	1.00	1.00	1.00	1.00

Fig. 3. Seasonal factor based on GSM od matrices [own study]

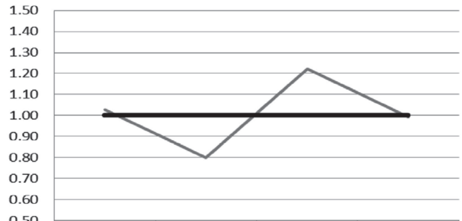


	Spring	Summer	Autumn	Winter
AADT	1.00	1.00	1.00	1.00
General Trafic Measurement	0.99	1.11	1.02	0.89
intra-zonal traffic	0.95	0.96	1.10	0.97
Inter-zonal trafic	0.95	1.04	1.10	0.80
inbound traffic	0.99	1.01	1.44	0.66
outbound traffic	0.99	1.01	1.44	0.66
Summarized traffic from Big Data set	0.96	1.01	1.17	0.83

Fig. 4. Seasonal factor based on GSM *od* matrices in trips to central agglomeration of the region [own study]

3.4. Weekly variations

Traffic volume fluctuations for weekly period in reference to AADT based on General Traffic Counts are shown on *Fig. 5*. Numbers on x-axis (1-7) describe following days-of-week (from Monday to Sunday).

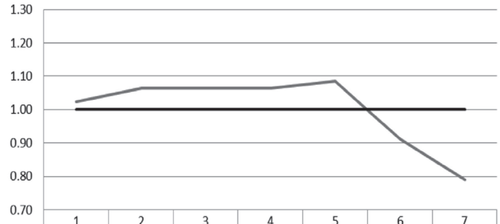


	Spring	Summer	Autumn	Winter
Seasonal traffic	1.03	0.80	1.22	0.99
AADT	1.00	1.00	1.00	1.00

Fig. 5. Weekly factors based on GPR traffic counts [16, 17, 18]

Traffic variation factors based on this data set remain stable at the level ca. 6% higher than AADT from Monday to Thursday, Then we can observe peak in Friday traffic (14% higher traffic than AADT) and fall in weekend days (ca. 10% below AADT).

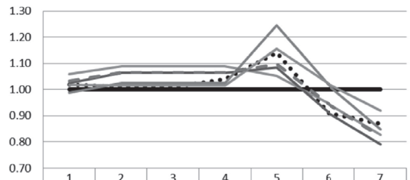
As in the previous example we compared weekly factors from GPR counts with GSM *od* matrices. This factors refers to average number of trips for each day-of-week throughout a year in each *od* demand strata described earlier. In the result we can observe (in *Fig. 6*) that the variability of factors received on the basis of two different data sets are comparable in reference to weekly traffic fluctuations.



	1	2	3	4	5	6	7
Weekly traffic	1.02	1.06	1.06	1.06	1.08	0.91	0.79
AADT	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Fig. 6. Weekly factors based on GSM *od* matrices [own study]

Fig. 7 presents the same factor for trips to central agglomeration. In this case traffic variations differ from the observed values by all odds. From Monday to Thursday it remains stable significantly above the average level, then fall in Friday (departures from the city before weekend) and in Saturday and Sunday drop under AADT.



	1	2	3	4	5	6	7
AADT	1.00	1.00	1.00	1.00	1.00	1.00	1.00
General Trafic Measurement	1.02	1.01	1.01	1.04	1.14	0.91	0.87
intra-zonal traffic	1.06	1.09	1.09	1.09	1.05	0.95	0.83
Inter-zonal trafic	1.02	1.06	1.06	1.06	1.08	0.91	0.79
inbound traffic	0.99	1.02	1.02	1.02	1.25	1.02	0.85
outbound traffic	1.02	1.02	1.02	1.02	1.16	1.02	0.92
Summarized traffic from Big Data set	1.03	1.06	1.06	1.06	1.10	0.95	0.82

Fig. 7. Weekly factor based on GSM *od* matrices in trips to central agglomeration of the region [own study]

4. Conclusion

In this paper we presented a method where trip volume variability can be indirectly analyzed through analysis of GSM *od* matrices. We utilized a big dataset of over 20M trips conducted between counties in Poland. We stratified the data into several meaningful subsets to understand its variability and fluctuations. The total variability of observed data was around 11%, yet it significantly dropped for weekdays (working days). Our analysis revealed that the seasonal variations are significant and each season has different expected trip volumes. Similarly, trip volumes within a week varied significantly, with weekdays being roughly stable, expect Friday when significantly higher trip volumes were observed. On weekends lower traffic flows were observed with minimum on Sunday. The seasonal variation was more pronounced for inbound and outbound traffic than for internal traffic. Interestingly, the intrazonal traffic on Friday was lower than for other working days.

Acknowledgement

This paper was developed within R&D project RID2A "Trip modelling and forecasting using different modes of transport" financed by The National Centre for Research and Development and National Roads and Motorway Administration in Poland.

Bibliography

[1] ANTONIOU C., BALAKRISHNA R., KOUTSOPOULOS H.N.: A synthesis of emerging data collection technologies and their impact on traffic management applications. European Transport Research Review, 3(3), 2011, p.139-148

[2] BECKER R.A., et al.: Route classification using cellular handoff patterns. In: Proceedings of the 13th International Conference on Ubiquitous Computing. ACM, Beijing, China, 2011

[3] CALABRESE F., et al.: Real-time urban monitoring using cellphones: a case study in Rome. IEEE Transactions on Intelligent Transportation Systems 12 (1), 2011, p.141–151

[4] DEMISSIE M.G., et al.: Intelligent road traffic status detection system through cellular networks handover information: An exploratory study. Transportation research part C: emerging technologies, 32, 2013, p.76-88

[5] HOU Z., LI X.: Repeatability and similarity of freeway traffic flow and long-term prediction under big data. IEEE Transactions on Intelligent Transportation Systems, 17(6), 2016, p.1786-1796

[6] JÄRV O., et al.: Mobile phones in a traffic flow: a geographical perspective to evening rush hour traffic analysis using call detail records. PloS one, 7(11), e49171, 2012.

[7] LEDUC G., et al.: Road traffic data: collection methods and applications. In: Working Papers on Energy, Transport and Climate Change N.1. EUROPEAN Comission-Joint Research Centre-Institute for Prospective Technologica, 2008

[8] LV Y., et al.: Traffic flow prediction with big data: a deep learning approach. IEEE Transactions on Intelligent Transportation Systems, 16(2), 2015, p.865-873

[9] SPŁAWIŃSKA M.: Charakterystyki zmienności natężeń ruchu i ich wpływ na eksploatację wybranych obiektów drogowych: praca doktorska, Politechnika Krakowka, Kraków, 2013

[10] THIESSENHUSEN K.U., SCHÄFER R.P., LANG T.: Traffic Data from Cellphones: A Comparison with Loops and Probe Vehicle Data, 2003

[11] TOOLE J.L., et al.: The path most traveled: Travel demand estimation using big data resources. Transportation Research Part C: Emerging Technologies, 58, 2015, p.162-177

[12] VACCARI A., et al.: A holistic framework for the study of urban traces and the profiling of urban processes and dynamics. In: 12th International IEEE Conference on Intelligent Transportation Systems. IEEE, St. Louis, MO, 2009

[13] VALERIO D.: Road Traffic Information from Cellular Network Signaling. Forschungszentrum Telekommunikation Wien, Vienna, Austria, 2009

[14] VALERIO D., et al.: Exploiting cellular networks for road traffic estimation: a survey and a research roadmap. In: IEEE 69th Vehicular Technology Conference, Barcelona, Spain, 2009a

[15] VALERIO D., et al.: Road traffic estimation from cellular network monitoring: A hands-on investigation. In: Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on, Tokyo, 13–16 September 2009b, pp. 3035–3039
[16] https://www.gddkia.gov.pl/userfiles/articles/s/stacje-ciaglych-pomiarow-ruchu-d_26174/R02_05_2016.pdf [date of access: 31.01.2018]
[17] https://www.gddkia.gov.pl/userfiles/articles/s/stacje-ciaglych-pomiarow-ruchu-d_26174/R02_04_2016.pdf [date of access: 31.01.2018]
[18] https://www.gddkia.gov.pl/userfiles/articles/g/godzina-miarodajna-i-wahania-ruc_24176/Metoda_szacowania_SDRR.pdf [date of access: 31.01.2018]