THE PROBLEM OF IMPUTATION OF THE MISSING DATA FROM THE CONTINUOUS COUNTS OF ROAD TRAFFIC

M. SPŁAWIŃSKA[1]

Missing traffic data is an important issue for road administration. Although numerous ways can be found to impute them in foreign literature (inter alia, the most effective method, that is Box-Jenkins models), in Poland, still only proven and simplified methods are applied. The article presents the analyses including an assessment of the completeness of the existing traffic data and works related to the construction of SARIMA model. The study was conducted on the basis of hourly traffic volumes, derived from the continuous traffic counts stations located in the national road network in Poland (Golden River stations) from the years 2005 – 2010. As a result, the proposed model was used to impute the missing data in the form of SARIMA $(1.1,1)(0,1,1)_{168}$. The newly developed model can be used effectively to fill in the missing required days of measurement for estimating AADT by AASHTO method. In other cases, due to its accuracy and laboriousness of the process, it is not recommended.

KEYWORDS: roads, traffic data collection, imputation of the missing traffic data, model SARIMA

1. INTRODUCTION

Continuous automatic recording of road traffic provides valuable information necessary for the planning, design and operation of road facilities, including the selection of types of intersections and their design, determining the peak hours in the morning and afternoon, planning short-term measurements and transposing their results into the daily volume, economic analysis of feasibility studies or road surface analysis and determining the level of noise.

[1] Cracow University of Technology, Faculty of Civil Engineering, ul. Warszawska 24, 31-155 Krakow, Poland, e-mail: msplaw@pk.edu.pl

In Poland, since 1974, such measurements have been conducted using Fischer-Porter, RPP – 2, RPP – 5, Golden River, PAT and EasyCOUNT recorders (total of about 63 continuous counts stations, depending on the year) [1, 2]. The recorders allow for continuous automatic recording of traffic at one-hour intervals throughout the year. This makes it possible to directly determine the basic parameters describing the traffic such as Annual Average Daily Traffic (AADT) and Design Hourly Volume (DHV). These values are a measure of the burden of the analyzed elements of the communication system and thus provide a basis for the design and operational analysis of road buildings and their surroundings. Moreover, thanks to the continuous recording of traffic, it is possible to determine seasonal, weekly and daily traffic fluctuations enabling the performance of the simplified conversion of volumes from short-term measurements into AADT (due to the cost of collecting and processing, in most road sections, there are no continuous measurement stations; therefore, AADT must be determined based on short periods of random measurements). Unfortunately, not all stations work properly (failure of the measuring equipment, repairing the section of the road on which the station is located or memory overflow of the recorders) and only a few of them contain the complete data set of 365 or 366 days of measurement. Although every year new stations are created and improved recorders are gradually introduced in the existing stations, there are no actions to improve the quality of databases. For years, it has been known that the most commonly used recorders Marksman 660 are subject to failure and data derived from them are incomplete and erroneous. The problem of incomplete data is present not only in Poland but also all over the world, however, its scale is difficult to compare. Although it is estimated that in North America, more than 50% of the continuous counts station does not have a complete set of data (as is apparent from S. Sharma, P. Lingras, M. Zhong [3, 4] in some years even up to 70 – 90%), at such a large number of continuous measurement stations, it is not nearly as problematic as in Poland (for instance, 293 continuous counts stations work in the state of Florida and approximately 350 in the Alberta province). In addition, AASHTO guidelines [5, 6] do not allow for the replenishment of databases for statistical reports, including the definition of AADT. Lack of even one hour in the day excludes the whole day, and, in the absence of the required number of days in a year, the whole station. Nevertheless, it is hard to imagine a total resignation from the incomplete data, especially due to the fact that sometimes they are the only ones available (for example, because of technological reasons: the lack of one hour every day). Therefore, also in America, they are imputed with the so-called principle of Truth-in-date, that is with a clear indication of the place of interference. The identification of the so-called completeness

indicator C (formula 1), which refers to the number of available data in relation to the expected volume (no universal classification of the indicator) is also recommended.

$$C = (\frac{n_d}{n_o}) \times 100 \, [\%] \tag{1}$$

where:

n_d – the available number of records (hourly volumes) or days with the correct data,

n_o – the expected number of records (hourly volumes) or days.

2. WAYS TO IMPUTE DATA

It is possible to find different ways to impute the data in the foreign literature: from the simplest consisting of the manual replacement of missing data with historical data, or using simple computer programs (solutions analogous to the indicator method of estimating AADT based on short term measurements) described by S. Datla, S. Sharma [7], M. Zhong, P. Lingras, S. Sharma [4, 8, 9], for example:

- in Alabama, if fewer than six hours are missing, they are replaced with hours from the same period of the previous year or just another day of the same month of the current year. When there is more than 6 hours missing, the day is discarded,
- in Delaware Department of Transportation, DOT, the missing data are filled in on the principle of the simple average of the same hours and days of the previous and next month,
- in Dakota, DOT, they are replaced by the average of the previous three years of the period from which the data are missing,
- in Saskatchewan Highways and Transportation, they are replaced with the data from the same day of the week and the month of the previous year. The maximum duration of which they can be imputed in such a way is 21 days,
- in France, the MELODIE computer system that imputes the data using the same hour and day of the week from the previous month is used,
- in the UK, two approaches are used, namely: By-eye, consisting of manual data imputing, and automatic, mainly based on simple moving average, weighted average, exponential smoothing or combination thereof. The missing data are imputed on the basis of the previous week,
- in the Netherlands, the INTENS computer system fills in missing data based on linear interpolation between the available and missing data.

- – to more sophisticated techniques, using, inter alia:
- – Box-Jenkins models (ARIMA, SARIMA) in which it is assumed that the projected volume depends on the volatility of its past values (proposed by B. Ghosh, B. Basu, M. O'Mahony [10], M. Sabry, H. Abd-El-Latif, S. Yousef, N. Badra [11], S. Sharma, P. Lingras, M. Zhong [3, 4, 8, 9, 13], Y. Zhang, L. Yuncai [12]),
- – artificial neural networks (with time delay, TDNN) that use both past and future data (proposed by P. Lingras, S. Sharma , M. Zhong [3, 4, 8, 14, 15, 16]),
- – genetic algorithms (Gas) proposed by M. Zhong, S. Sharma, P. Lingras [4, 15, 16],
- – locally weighted regression (LWR) proposed by S. Sharma, P. Lingras, M. Zhong [3, 4, 8, 15, 16],
- – support-vector networks (SVR, LS-SVMs) proposed by Y. Zhang, Yuncai [12],
- – and hybrid models such as genetic algorithms supporting neural networks (GA – TDNN) proposed by P. Lingras, S. Sharma , M. Zhong [3, 4, 8, 13, 14, 15, 16].

However, tests determining the effectiveness of these methods are not conducted. In the literature, it is hard to find statistical studies on their accuracy and the impact on the volumes determined on their basis (among other things, AADT, DHV). A few studies (S. Sharma, P. Lingras, M. Zhong [3, 13]) suggest that, although hybrid models give the best accuracy, there is no need to use such a complicated method as comparable accuracy is obtained by using Box-Jenkins models (ARIMA, SARIMA). Unfortunately, at present, only simplified ways of imputing missing data are used in the country. According to foreign literature (S. Sharma, P. Lingras, M. Zhong [3]), the impact of the data imputed by these methods on the estimated values of AADT and DHV is very big (estimation error, respectively: 13,5% and 18,5%). Therefore, the aim of this article is to develop, in Polish conditions, the algorithm of imputing the missing data from continuous recording, using SARIMA model.

3. DATABASE

For the analysis, the data from the Golden River station from the years 2005 – 2010 was used. Throughout the country, there are about 34 stations (depending on the year [1, 2]). For the total length of the national roads, amounting to 17247 km in 2010, on average, there is one station for every 507.3 km length of the road. Unfortunately, very often the data is incomplete and, therefore, before the construction of the SARIMA model, the number of missing days of the week (in subsequent years, in relation to the total number of vehicles in the section of the road),

the number of missing days in each month, and the number of missing hours were determined for each of the stations. According to the US guidelines [5, 6], the lack of even one hour means the elimination of the whole day from the analysis. Table 1 presents summary statistics.

In the years 2005 – 2010 (Table 1), the percentage of incomplete data is very high, ranging from 75,8 to 93,5%. In case of missing days and hours, it is as follows: a) in relation to the entire data set: from 7,7% to 14,9% and from 7,1% to 14,4%; b) in relation to a set of incomplete data: from 9,4% to 18,4% and from 8,7% to 17,7%. It should be noted that the above analysis applies to all vehicles in the section of the road. Unfortunately, there are errors in relation to the direction of traffic and the type structure. It is difficult to assess whether the failures were related to the loop (lack of data) or classification (poorly defined category of the vehicle). However, it can be noted that the missing data distribution is roughly uniform in the case of days of the week but not for the months of the year. In addition, their scope and duration of occurrence vary from year to year. Therefore, the SARIMA model used must have a wide range of applications, including imputing the data within all days throughout the year. Further and detailed analyses were conducted on data from station number 20 (national road no. 16, section: Barczewo – Biskupiec). This point was chosen because of the lowest number of missing hours in the period of three years (1 hour missing). The analyses were carried out for the road section in relation to all vehicles.

Table 1. Annual statistics for 2005-2010

YEAR	2010	2009	2008	2007	2006	2005
number of days in a year	365	365	366	365	365	365
number of hours in a year	8760	8760	8784	8760	8760	8760
number of traffic recorder stations	33	33	33	32	31	29
number of traffic recorder stations with complete data	6	8	7	6	2	2
number of traffic recorder stations with incomplete data	27	25	26	26	29	27
total number of missing days	927	982	1024	1744	1540	1136
total number of missing hours	20493	22284	23055	40410	35497	25036
% of incomplete data in total [%]	81,8	75,8	78,8	81,3	93,5	93,1
% of missing days in total [%]	7,7	8,2	8,5	14,9	13,6	10,7
% of missing days in the incomplete data [%]	9,4	10,8	10,8	18,4	14,5	11,5
% of the missing hour in total [%]	7,1	7,7	8,0	14,4	13,1	9,9
% of the missing hour in incomplete data [%]	8,7	10,2	10,1	17,7	14,0	10,6

4. METHODOLOGY OF TIME SERIES

On the basis of the conclusions of the literature review, it was decided to impute the missing data using seasonal SARIMA model (generalization and development of the ARIMA model, in which seasonal regularities are repeated in time). It is characterized by the functional relationship between the values of the variable predicted in the period t and the same variable values from previous periods $t - 1$, $t - 2$, ..., $t - p$ with accuracy of the random component. The SARIMA (Seasonal Autoregressive Integrated Moving Average) model construction is based on the autoregressive method and moving average. Autoregressive process (described in A. Sokołowski [17] and A. Zeliaś, B. Pawełek, S. Wanat [18]) of p, as indicated in the abbreviation AR(p), has the following form:

$$\check{z}_t = \varphi_1 \times \check{z}_{t-1} + \varphi_2 \times \check{z}_{t-2} + ... + \varphi_p \times \check{z}_{t-p} + \varepsilon_t \tag{2}$$

where:

φ_1, φ_2, ..., φ_p – process parameters,

ε_t – random disturbance (random variable with normal distribution $N(0, \sigma_\varepsilon^2)$,

p – the amount of delay.

In this process, the current values are the sum of a finite linear combination of previous values of the series and random disturbance (random variable $\check{z}_t$ is explained by regression due to its past value). The process of moving average of q, as indicated in the abbreviation MA(q) is as follows:

$$\check{z}_t = \varepsilon_t - \theta_1 \times \varepsilon_{t-1} - \theta_2 \times \varepsilon_{t-2} - ... - \theta_q \times \varepsilon_{t-q} \tag{3}$$

where:

ε_t, ε_{t-1}, ..., ε_{t-q} – random disturbance in periods t,

$t - 1$, ..., $t - q$ (random variables with normal distribution $N(0, \sigma_\varepsilon^2)$,

θ_1, θ_2, ..., θ_q – model parameters,

q – amount of delay.

Every observation consists of a random component and the linear combination of the random components of the past. SARIMA models can be used for modeling and forecasting stationary series (constant average in time, variance and autocorrelation) or non-stationary series which can be changed into the stationary ones by the operation of differentiation (d-fold calculation of the difference of neighboring words of the series) and logarithmic transformation (to stabilize

the variance). All of these processes can be saved using a universal notation SARIMA(p, d, q)(P, D, Q)$_s$ where small letters denote not seasonal parameters but large seasonal ones, respectively: autoregression (p, P), the operator's difference (the multiple of differentiation – d, D), the amount of delay of the moving average (q, Q) for a given delay (s). In the process of constructing an appropriate model, the following steps are distinguished (described in A. Sokołowski [17] and A. Zeliaś, B. Pawełek, S. Wanat [18]):

– identification – a model is pre-selected by determining p, d, q, P, D, Q, s based on data analysis and experience. The basic tool used in the identification phase is series graphs, the AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF),

– estimation – the value of the parameter estimates of the pre-selected model is calculated. In general, in the phase of parameter estimation, the function minimization algorithm is used (minimization of the sum of squares of residuals) in order to maximize the reliability of receiving the observed series with the given values of the parameters,

– verification – on the basis of appropriate statistical tests and the evaluation of the quality of forecasts, it is checked whether the model meets the formulated requirements. The correctness check consists of: deleting residues, looking for systematic trends and studying the autocorrelogram of the residues (there should not be any autocorrelation). If the model successfully passes the test, it is ready to be used (for forecasting), and if it does not, the procedure should be repeated.

In order to evaluate the model used, a comparison of theoretical and empirical values is conducted. For this purpose, the indicators of the assessment of the model are calculated (measurements of the accuracy of the fitting of the smoothed series to the original one). The primary indicator of the assessment of the model is the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\left(\frac{y_t - \hat{y}_t}{y_t}\right)\times 100\%\right| \tag{4}$$

5. CONSTRUCTION OF THE SARIMA MODEL

Due to the lack of publications concerning the imputing of traffic data with respect to the volume of traffic in the national literature, the analysis diagram was taken from the work on short-term forecasting of electricity demand proposed by A. Lichota [19]. In the first stage, individual components of the time series were separated, that is the cycle, trend and seasonality. To this end,

a visual analysis of the series constructed on the basis of the hourly traffic volumes and daily average volumes of the month was performed. Sample data (station 20) is shown in Figures 1 and 2.
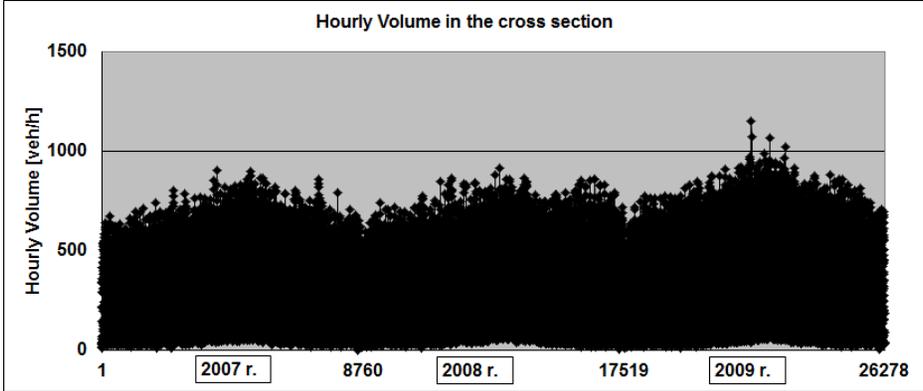

Fig. 1. The hourly volumes of the total number of vehicles in a section of the national road no. 16 from the period of three years (st. No. 20)
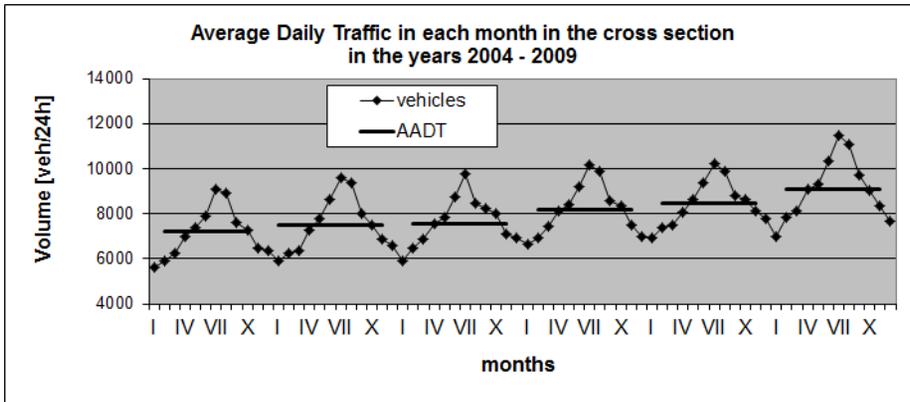

Fig. 2. The variability of the annual total number of vehicles in a section of the national road no. 16 (st. No. 20)

The source data in Figure 2 for the period of 6 years (a longer period provides a better observation of the cycle) are incomplete, where AADT in each month was calculated without the missing days. It can be seen that the volume of traffic varies to a small extent in the years 2007 – 2009 (the AADT value in subsequent years, respectively: 8186 veh/24h, 8439 veh/24h, 9085 veh/24h). Since

the purpose of the model are immediate and short-term forecasts, annual cycle can be omitted. Nevertheless, annual cyclical fluctuations associated with the seasons are noticeable.

In order to detect the periodicity of the time series, graphs were drawn for a shorter period, namely for two weeks of October (Figure 3). On their basis, the repetition of the weekly sequence (168 hours) and the daily sequence (24 hours) can be observed. In order to confirm the findings from the analysis of visual graphs, the correlogram (ACF) and partial autocorrelation (PACF) were performed. They showed that most of the residues of the autocorrelation process are: k = 1, or k = 24 hours (k is the delay of the analyzed time series). In view of this, the differentiation of the tested time series using these values was carried out (Figure 4). The maximum number of delays (k) was chosen in such a way as to demonstrate clear seasonal variations. These analyses confirm the two main periods of seasonality, that is, k = 24 and k = 168 hours, which are so strong that they cannot be removed completely by differentiation.
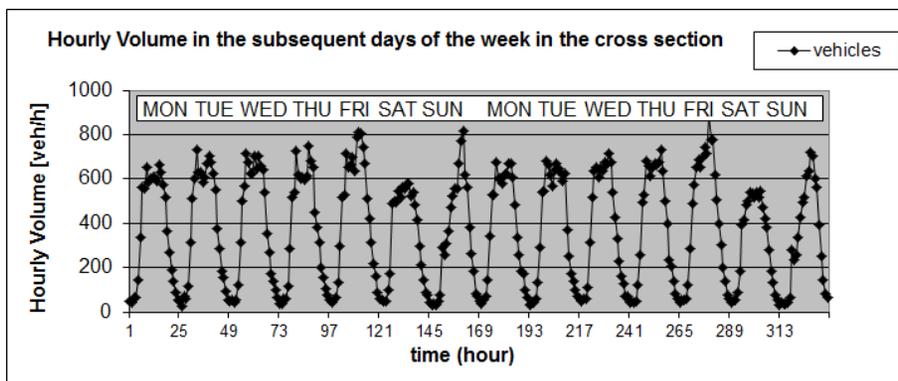

Fig. 3. The variability of the weekly total number of vehicles in a section of the national road no. 16 (st. No. 20)

In order to identify the trend, it was decided to use the analytical method, namely the determination of a mathematical equation expressing the form of the function of the trend. During the stepwise elimination procedure (removal of irrelevant time variables), the trend equation of statistically significant structural parameters was obtained in the following form:

$$\hat{y}_t = 354,4 + 0,0055 \times t \tag{4}$$

5	−,091 ,0107		283,0 0,000
19	+,011 ,0107		669,3 0,000
33	+,054 ,0107		2389, 0,000
47	+,004 ,0107		2456, 0,000
61	+,035 ,0107		2526, 0,000
75	−,004 ,0107		2578, 0,000
89	+,010 ,0106		2598, 0,000
103	−,005 ,0106		2651, 0,000
117	+,026 ,0106		2712, 0,000
131	−,028 ,0106		2777, 0,000
145	−,050 ,0106		3013, 0,000
159	−,074 ,0106		3315, 0,000
173	−,060 ,0106		4640, 0,000
187	+,016 ,0106		4990, 0,000

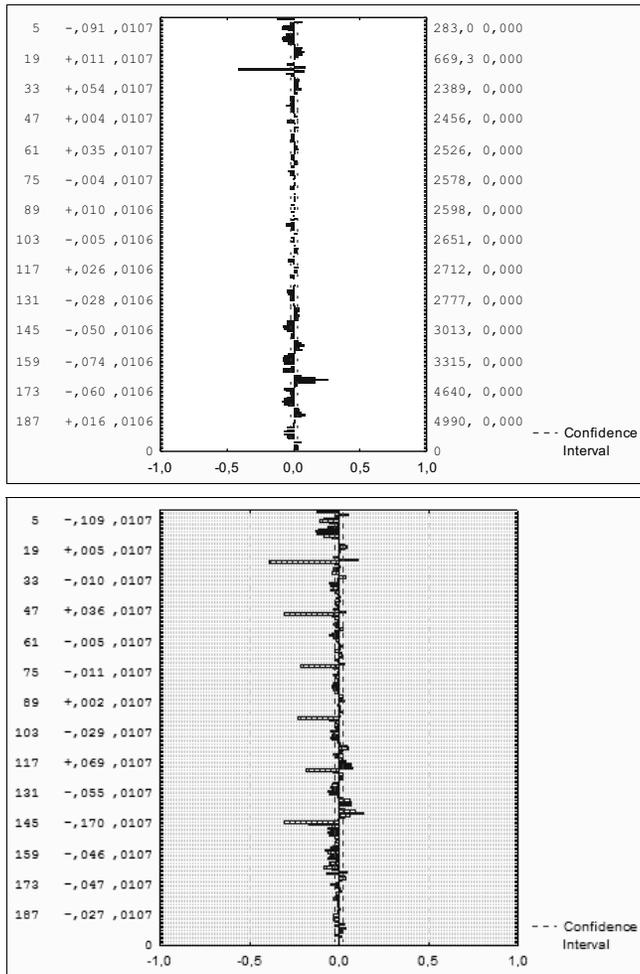|	
5	−,109 ,0107
19	+,005 ,0107
33	−,010 ,0107
47	+,036 ,0107
61	−,005 ,0107
75	−,011 ,0107
89	+,002 ,0107
103	−,029 ,0107
117	+,069 ,0107
131	−,055 ,0107
145	−,170 ,0107
159	−,046 ,0107
173	−,047 ,0107
187	−,027 ,0107

Fig. 4. Autocorrelation and partial autocorrelation plots (the number of delays k = 200, D - 1, D - 24), the data for the year 2009

The next step in the analysis was to determine the optimal parameters of the SARIMA model. For this purpose, a trial and error method was used based on checking (for the same data) alternative models.

Following the accuracy of the forecasts generated based on some data and equated to the real values and keeping in mind that the model must have the ability of generalization (regularity search), the selected model SARIMA $(1,1,1)(0,1,1)_{168}$ is as follows:

$$(1-B)\times(1-B^{168})\times(1-p\times B)\times\widetilde{z}_t=(1-q\times B)\times(1-Q_s\times B^{168})\times\varepsilon_t \tag{5}$$

where:

p – autoregressive model parameter,

q – moving average model parameter,

Q_s – moving average seasonal model parameter.

In order to check accuracy of the model used, its validation was performed on data derived from the station number 1 (national road no. 8, section Jeżewo – Choroszcz). Table 2 shows the results of exemplary one week forecast obtained based on data from the previous nine weeks.

The subsequent columns are as follows:

1. The number of week for which the forecast was performed,
2. The extent of weeks used to calibrate the model,
3. The period of time (date) for which the forecast was performed,
4. The period of time (date) for the weeks used to calibrate the model,
5. The mean absolute percentage error (*MAPE*) of the forecast for the station number 20,
6. The standard deviation of the mean absolute percentage error of the forecast for the station number 20,
7. The mean absolute percentage error (*MAPE*) of the forecast for the station number 1,
8. The standard deviation of the mean percentage error of the forecast for the station number 1,
9. The mean absolute percentage error (*MAPE*) of the forecast for the station 20 when the volumes of traffic on days of public holidays were replaced with data from the same but non-holiday days of the week in the same month,
10. The standard deviation of the mean absolute percentage error of the forecast for station number 20, with the values changed in accordance with the point above.

In addition, the table shows the average value of the average relative percentage error for all the weeks of the year and so-called improved average for data without weeks in which there were days of public holidays (marked in gray), or which were projected based on a few of such days (gray + bold).

An exemplary analysis of the distribution of residues comprising the autocorrelation and partial autocorrelation plots and fitting to a normal distribution for the period 1-9 was shown in Figure 5. Table 3 shows the statistical analysis and model parameters, while Figure 6 presents a graphical illustration of matching of the forecast for the tenth week to the original series.

Table 2. The results for the SARIMA $(1,1,1)(0,1,1)_{168}$ model for 2009

forecast [week]	calibration [week]	date of forecast	date of calibration	MAPE st.20 [%]	σ st.20 [%]	MAPE st.1 [%]	Σ st.1 [%]	MAPE st.20 corrected [%]	σcor. st. 20 [%]
1	2	3	4	5	6	7	8	9	10
10	1-9	09-15.03	05.01-08.03	10,3	8,3	6,4	5,0		
11	2-10	16-22.03	12.01-15.03	8,8	11,4	8,3	8,7		
12	3-11	23-29.03	19.01-22.03	9,5	10,9	8,6	6,6		
13	4-12	30.03-05.04	26.01-29.03	21,7	20,0	13,3	14,0		
14	5-13	06-12.04	02.02-05.04	26,6	34,0	23,6	35,6		
15	6-14	13-19.04	09.02-12.04	40,6	44,5	33,5	44,8		
16	7-15	20-26.04	16.02-19.04	**22,5**	**15,1**	**8,4**	**7,5**	12,1	10,0
17	8-16	27.04-03.05	23.02-26.04	25,8	26,2	22,9	28,5		
18	9-17	04-10.05	02.03-03.05	**16,5**	**13,9**	**40,7**	**13,6**		
19	10-18	11-17.05	09.03-10.05	9,9	10,5	14,8	10,2		
20	11-19	18-24.05	16.03-17.05	8,4	7,0	7,0	5,7		
average total				15,3	15,2	15,2	13,5	16,3	10,8
average corrected				11,3	11,2	9,2	9,1		

Table 3. The result of the statistical analysis for the SARIMA $(1,1,1)(0,1,1)_{168}$ model

| PARAMETER	DATA: Hourly Volume for vehicles in the cross section STATION NUMBER 20 TRANSFORMATION: ln(x),D(1),D(168) MODEL: (1,1,1)(0,1,1) DELAY SEASONAL: 168 RESIDUAL MS=,01506					
	Parameter	Asympt. std error	Asympt. t(7844)	p	lower limit 95% CI	upper limit 95% CI
p(1)	0,141003	0,028659	4,9200	0,000001	0,084781	0,197225
q(1)	0,973686	0,007207	135,0936	0,000000	0,959547	0,987825
$Q_s(1)$	0,805157	0,040698	19,7835	0,000000	0,725318	0,884996
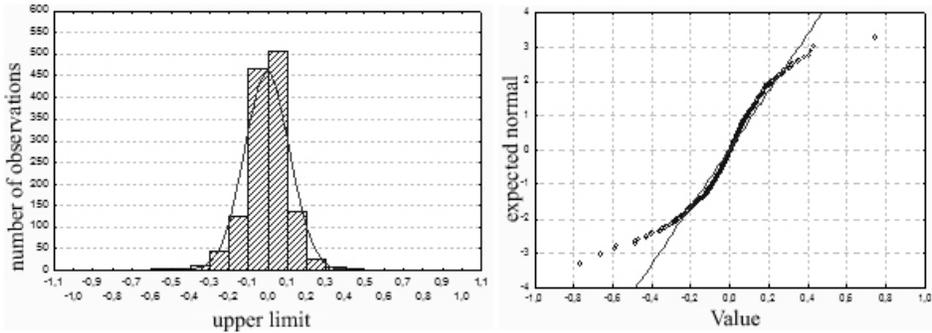
Fig. 5. Graph of the analysis of the residuals for the SARIMA$(1,1,1)(0,1,1)_{168}$ model
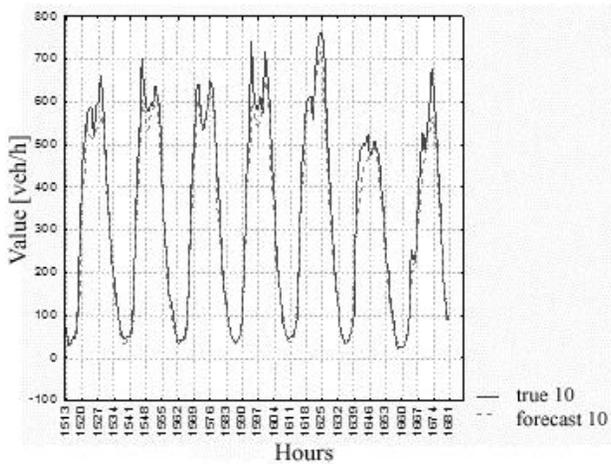

Fig. 6. Graph of matching the forecast to the real series for the tenth week

Despite the fact that the analysis of residues showed the correctness of the adopted model, and the result of validation study was positive, the accuracy of the forecast was inadequate. The biggest error (MAPE) was obtained in the weeks in which there are statutory holidays, used for calibration. The highest accuracy is observed in March, during the period from July to October, in the second half of November and the first half of December. In order to increase the accuracy in the weeks for which the calibration of the model was based on data from a few days of public holidays (MAPE > 20%), the values of traffic volume on these days were replaced with data from the same

but "non-holiday" day of the week in the same month. Thus, a reduction in error for the station number 20 was achieved on average of more than 16%.

6. CONCLUDING REMARKS

The conducted analyses showed that only a few of the stations of continuous automatic counts of traffic function properly and contain a complete set of data. Therefore, these data must be imputation. The article presents, as indicated by the literature, the most effective method of filling them in (the best ratio of the financial input to the resulting accuracy), that is the Box-Jenkins method. Although the analyzes referred years 2005 – 2010, because of established by M. Spławińska [20] stability traffic fluctuations in longer period of time, it can be concluded that the specified depending will also occur in other years.

The proposed SARIMA model due to the large number of days of public holidays is not suitable for the period from Easter to the end of June, and from the Christmas Eve to the end of the year. Although the imputed data from the remaining period should not be used to aggregate statistics, they can be used for general analyses (for instance to determine the AADT). Moreover, it is not recommended to use the proposed model to impute missing data for determining the design volume (because of the low accuracy – MAPE for the station number 20 from 7,4% to 40,6%) and to estimate the AADT in absence of single, randomly selected days (because of the laboriousness of the process). As is apparent from T. Wright, P. S. Hu, J. A. Young, Lu [21], in the absence of up to 50% of the days, a very high accuracy (error of less than 1,0%) is obtained.

The above model, despite its low efficiency, can be used to impute the missing required measurement days for estimating AADT according to AASHTO method (so-called average of the average described in [5, 6] and by A. Olma [22]). Can also be used by road administration for databases updates. It should be remembered to respect the principle of Truth-in-date and set the completeness index C every time. This allows for assessing the reliability and statistical usefulness of values determined based on incomplete data in an easy way. In addition, the threshold of "completeness" should be determined, below which the imputed data would not be used for statistical analyses. This threshold should be determined on the basis of experience and empirical studies of traffic.

REFERENCES

1. Zbieranie, archiwizacja i analizy danych ze stacji ciągłych pomiarów ruchu, Etap III, Archiwizacja wyników pomiarów ze stacji GR i PAT oraz analiza wyników pomiarów ze stacji F-P, RPP-2 i RPP-5 za okres od 01.11 do 31.12.2004, Transprojekt – Warszawa Sp. z o.o., Warszawa 2005
2. Zbieranie, archiwizacja i analizy danych ze stacji ciągłych pomiarów ruchu w roku 2008, Etap III, Analiza roczna i edycja wyników pomiarów prowadzonych w stacjach GR i PAT w roku 2009, Transprojekt – Warszawa Sp. z o.o., Warszawa 2010
3. S. Sharma, P. Lingras, M. Zhong "Effect of Missing Values Estimations on Traffic Parameters", Transporatation Planning and Technology, Vol.27, No.2, 119-144, 04.2004
4. M. Zhong, P. Lingras, S. Sharma "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques", Transport Research Part C: Emerging Technologies Volume 12, Issue 2, 139-166, 04.2004
5. AASHTO Guidelines for Traffic Data Programs, American Association of State Highway and Transportation Officials, 1992
6. Federal Highway Administration (FHWA), Traffic Monitoring Guide, 2001
7. S. Datla, S. Sharma "Consideration of Weather Conditions to Estimate Missing Traffic Data", Transportation Research Record: Journal of the Transportation Research Board 2049, 71-80, Washington DC 2008
8. M. Zhong, P. Lingras "Genetically Designed Models for Accurate Imputations of Missing Traffic Counts", Transport Research Record 1879, 71-79, Washington DC 2004
9. M. Zhong, S. Sharma, P. Lingras "Matching Patterns for Updating Missing Values of Traffic Counts. Transportation Planning and Technology", Vol. 29, No.2, 141 ÷ 156, 04.2006
10. B. Ghosh, B. Basu, M. O'Mahony "A Bayesian Time-Series model for Short-Term Traffic Flow Forecasting", Journal of Transportation Engineering, Volume 133, Issue 3, 180-189, 03.2007
11. M. Sabry, H. Abd-El-Latif, S. Yousef, N. Badra "Use of Box and Jenkins Time Series Technique in Traffic Volume Forecasting", Research Journal of Social Sciences, 83-90, 2007
12. Y. Zhang, L. Yuncai "Comparison of Parametric and Nonparametric Techniques for Non-peak Traffic Forecasting", World Academy of Science, Engineering and Technology 51, 236-242, 2009
13. M. Zhong, S. Sharma, P. Lingras "Genetically-Designed Time Delay Neural Networks for Multiple-interval Urban Freeway Traffic Flow Forecasting", Neural Information Processing – Letters and Reviews, Vol.10, Nos. 8-9, 201-209, 08.-09.2006
14. P. Lingras, S. Sharma, M. Zhong "Prediction of Recreational Travel using Genetically Designed Regression and Time Delay Neural Network Models", Transport Research Record 1805, 16-24, Washington DC 2002
15. M. Zhong, S. Sharma, P. Lingras "Short-Term Traffic Prediction on Different Types of Roads with Genetically Designed Regression and Time Deley Neural Network Models", Journal of Computing in Civil Engineering, 94-103, 01.2005
16. M. Zhong, S. Sharma, P. Lingras "Refining Genetically designed Models for Improved Traffic Prediction on Rural Roads", Transporatation Planning and Technology, Vol.28, No.3, 213-236, 06.2005
17. A. Sokołowski, Prognozowanie i analiza szeregów czasowych. Kraków, 25 –26.06.2009
18. A. Zeliaś, B. Pawełek, S. Wanat "Prognozowanie ekonomiczne teoria, przykłady, zadania", PWN, Warszawa 2003
19. A. Lichota "Prognozowanie krótkoterminowe na lokalnym rynku energii elektrycznej", Rozprawa doktorska, Akademia Górniczo – Hutnicza, Kraków 2006
20. M. Spławińska „Analiza stabilności charakterystyk zmienności natężeń ruchu w dłuższym okresie", 9902-9911, Logistyka 6/2014
21. T. Wright, P. S. Hu, J. Young, A. Lu "Variability in Traffic Monitoring Data", Final Summary Report, Oak Ridge National Liboratory, US Department of Energy, 08.1997
22. A. Olma „Określenie współczynników przeliczeniowych do szacowania natężeń ruchu drogowego w obszarach miejskich", Praca doktorska, Politechnika Śląska, Gliwice 2005