Article citation info:
PROCHOWSKI L. Evaluation of the process of mileage growth during the operation of motor trucks, in several categories of engine cubic capacity. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2018; 20 (3): 359–370, http://dx.doi.org/10.17531/ein.2018.3.3.

Leon PROCHOWSKI

EVALUATION OF THE PROCESS OF MILEAGE GROWTH DURING THE OPERATION OF MOTOR TRUCKS, IN SEVERAL CATEGORIES OF ENGINE CUBIC CAPACITY

OCENA PROCESU NARASTANIA PRZEBIEGU PODCZAS EKSPLOATACJI SAMOCHODÓW CIĘŻAROWYCH W KILKU KATEGORIACH POJEMNOŚCI SILNIKA*

The intensity of use of motor vehicles and the range of the transport jobs performed change during the many-year operation of such vehicles. The changes are introduced in result of ongoing analyses of actual current vehicle operation costs, reliability, and performance characteristics. The changes in the vehicle operation process, taking place with vehicle's age, were analysed on the grounds of the mileage of over 9 000 motor trucks. The analysis covered a 20-year period of vehicle operation. The analysis results were used to estimate the mathematical models that describe the basic characteristics of the mileage growth process, with changes in these characteristics as observed in the recent years along with an intensive development of the road transport in Poland. Models of vehicle mileage growth have been developed for seven categories of engine cubic capacity. The coefficients of model equations have been given and the accuracy of the mileage values calculated on these grounds has been comprehensively evaluated. The relative measures of the scatter in the mileage values obtained from these models do not exceed 12 % of the average values determined from experimental data. A procedure has been proposed that leads to evaluating the mileage growth process and is based on the experience having already been gained in this field. It has been shown that the mileage growth process is strongly related to the engine cubic capacity. The mileage growth process is an important source of information for the planning of vehicle operation, forecasting of costs, and estimating of exhaust emissions and energy consumption in the whole cycle of operation of motor vehicles by transport companies.

Keywords: *motor vehicle mileage, motor vehicle operation, motor trucks, estimation of motor vehicle mileage.*

W czasie wieloletniej eksploatacji samochodów zmianie ulega intensywność ich użytkowania oraz zakres realizowanych zadań transportowych. Zmiany te są rezultatem analizy kosztów eksploatacji, niezawodności oraz osiąganych parametrów trakcyjnych. Zmiany w procesie eksploatacji, które zachodzą wraz z wiekiem samochodów, analizowano na podstawie przebiegu prawie 9000 samochodów ciężarowych. Analizowany jest dwudziestoletni okres eksploatacji pojazdów. Rezultaty tej analizy wykorzystano podczas estymacji modeli matematycznych, które opisują zasadnicze cechy procesu narastania wartości przebiegu, pojawiające się w ostatnich latach wraz z intensywnym wzrostem transportu drogowego w Polsce. Opracowano modele narastania przebiegu w siedmiu kategoriach pojemności silnika. Podano współczynniki równań tych modeli oraz dokonano wszechstronnej oceny dokładności obliczonych na ich podstawie wartości przebiegu. Obliczone względne miary rozrzutu wartości przebiegu z tych modeli nie przekraczają 12% od wartości średnich, wyznaczonych z danych eksperymentalnych. Zaproponowano procedurę postępowania, która prowadzi do wyznaczenia oceny procesu narastania przebiegu i jest oparta na dotychczasowych doświadczeniach w tym zakresie. Pokazano, że proces narastania przebiegu ma silny związek z pojemnością silnika. Proces narastania przebiegu jest ważnym źródłem informacji podczas planowania eksploatacji pojazdów, prognozowania kosztów, a także oceny emisji spalin i zużycia energii w całym cyklu eksploatacji samochodów w firmach transportowych.

Słowa kluczowe: *przebieg samochodów, eksploatacja samochodów, samochody ciężarowe, ocena przebiegu samochodów.*

1. Introduction and analysis of the current state of knowledge in this field

Motor trucks are selected according to the transport tasks planned. There are many factors of considerable importance for the vehicle selection, such as load capacity, unladen mass, fuel consumption, or engine cubic capacity. They are interrelated, e.g. the engine capacity has an impact on the power output and fuel consumption. During the many-year operation of motor vehicles, changes take place in the intensity of use of such vehicles and in the range of the transport jobs performed. The intensity of vehicle use is chiefly measured by the rate of growth in vehicle mileage. The vehicle mileage values constitute a basis for vehicle operation planning and cost forecasting. The vehicle mileage is taken into account in various ways to estimate e.g. insurance risk [6, 15], exhaust emissions [1, 5, 12, 27], or costs of fuel, tyres, or spare parts [16, 23]. The publications concerning this subject matter are predominantly dedicated to the intensity of use of motor cars [5, 6, 9, 12, 23]. However, the average annual mileage of heavy goods vehicles (HGV) is many times as high as that of motor cars (MC). As an example, the HGV to MC mileage ratio reported for the UK in 2006 exceeded 3.5 [25]. Moreover, the intensity of MC operation has been recently decreasing [13, 20].

In Poland, a high rate of growth in the motor truck traffic is now observed [14]. The mileage of motor trucks with up to 3 500 kg (3.5 t) gross vehicle mass (GVM) is often compared with that of motor cars [23, 26]. The light motor trucks, popularly referred to as local transport vehicles (LTV) or light commercial vehicles (LCV), are chiefly used for distribution-type jobs. According to catalogues [7],

(*) Tekst artykułu w polskiej wersji językowej dostępny w elektronicznym wydaniu kwartalnika na stronie ***www.ein.org.pl*** the mileage of motor trucks with up to 3.5 t GVM, after 10 years in service, exceeds that of motor cars with comparable engine cubic capacity by 30 %. By contrast, the mileage values for HGVs (with GVM exceeding 3.5 t) are higher than those for motor cars are by more than 300 %. Increasingly often attention is paid to the negative effects of road transport, which include the external costs such as exhaust emission effects and high fatality of road accidents [4, 18]. In the UK, the number of fatal accidents with HGV drivers is 1.8 per 100 million km and it is twice as high as that with motor car drivers [2]. The high intensity of motor truck operation determines the rate of replacement of the fleet of such vehicles in Poland. In consequence, the motor trucks aged up to 5 years perform 49.5 % of cargo transport jobs (in ton-kilometres) and those aged up to 10 years carry out as much as 82.4 % of such jobs [22, 24]. Motor trucks having been operated for more than 20 years are rarely seen on roads and their share in the transport work is 0.5 % [19].

The objective of this work is to evaluate the process of growth in the mileage of motor trucks with successive years of the vehicles being in service. The evaluation will be based on indicators related to several measures of the deviations of model data from empirical data. The problem will be analysed in connection with the engine cubic capacity value. The models having been developed will facilitate the analysis of the mileage growth rate, recognition of the major tendencies, and identification of the development trends in the successive periods of vehicle operation. The models of the mileage growth process are often the main source of information in the prognostic calculations of the trends to develop and of the values of the indicators that help to manage the operation of a motor vehicle fleet. The research conducted in this field provides anticipating information about vehicles' mileage growth, rate of approaching the planned or target mileage, current advancement in reaching this mileage and the maximum time of cost-effective vehicle use, termination of the vehicle use, replacement of vehicle fleet, or modifications to the operation of the fleet [3, 10, 11, 21].

This work consists of several successive stages, which include analysis of data and estimation of the models and model accuracy. Detailed and current data about the intensity of operation of motor trucks (i.e. HGVs and LTVs), based on a large dataset, are not easily available. There is a lack of data about the LTV mileage growth process and about the relation between this process and the engine cubic capacity. There is also a lack of current information how the increasing number and diversity of motor trucks influenced the intensity of operation of such vehicles in Poland.

2. Characterization of the dataset

The information about motor vehicles and their mileage was collected on the grounds of:

- surveys carried out among motor truck drivers;
- information about vehicles' mileage and characteristics recorded during mandatory periodical inspections carried out at Vehicle Testing Stations (VTS);
- analysis of post-accident motor truck inspection records.

Apart from other information, the dataset includes basic technical data of the vehicle, date of first registration, and mileage covered in Poland. The data concerning special motor vehicles and vehicles with load capacity below 500 kg as well as vehicles with monthly mileage values below 100 km/m and above 25 000 km/m were excluded from the dataset; the upper limit was adopted because in practice the very high monthly mileage values are hardly achievable in the conditions of road traffic in Poland. The vehicle models that predominated in the dataset have been specified in Tables 1a and 1b.

Note: The GVM and engine cubic capacity values have been confirmed in catalogues [7].

In the dataset, the same vehicle makes can be seen that predominate in numerical summaries of the first registrations in Poland. For example: Citroen, Fiat, Ford, Mercedes-Benz, Peugeot, Renault, and Volkswagen predominate in the set of local transport vehicles. The same vehicle makes have been indicated as predominating among the first registrations of LTVs in the Automotive Industry Reports of 2008, 2014, and 2016 [22]. In the group of vehicles with GVM exceeding 3.5 tons (metric), DAF, Iveco, MAN, Mercedes-Benz, Renault Trucks, Scania, and Volvo Trucks are the makes that predominate in the number of HGV sales in the Polish market and in the number of first registrations of such vehicles [22]. The same makes predominate in Table 1b.

3. Procedure of evaluation of the mileage growth process

Very wide scatter can be observed in the motor truck mileage values after vehicle operation periods of the same duration. This scatter results from very diverse transport jobs performed and from variations in the intensity of vehicle operation. This characteristic feature of the motor truck mileage is confirmed by multiannual statistics. For illustration: the average annual mileage of such vehicles in the UK has been determined as 79 000 km with a standard deviation of 63 000 km, according to [17].

To find a relatively representative image of the process of growth in the motor truck mileage in Poland with taking into account the

Table 1a. Motor trucks belonging to the group of GVM ≤ 3.5 t

Engine cubic capacity category [cm^3]	Vehicle makes and models	GVM [tons (metric)]	Example engine capacity values [cm^3]
Up to 1 499	Citroen Berlingo 1.4; Fiat Doblo Cargo; Opel Combo CDTI; Peugeot Partner 1.4; Renault Kangoo Express III dCi; Volkswagen Caddy 1.4, Caddy Maxi	1.6; 1.7; 1.9; 2.2; 2.3; 2.4	1 197; 1 248; 1 360; 1 390; 1 461
1 500-1 999	Citroen Berlingo 1.9D, Berlingo II 1.6 HDI; Fiat Doblo Cargo JTD; Ford FT280; Opel Combo Diesel, Combo 1.7D, Vivaro CDTI; Peugeot Partner II HDI; Renault Kangoo D, Trafic DCI; Volkswagen Caddy 1.9SDI, Transporter T5 TDI	1.6; 1.8; 1.9; 2.0; 2.2; 2.7; 3.0	1 560; 1 686; 1 868; 1 870; 1 896; 1 910; 1 968; 1 995; 1 999
2 000-2 499	Citroen Jumper 35HDI; Fiat Ducato 30MJ; Ford FT 350 TDCi, Transit; Mercedes Sprinter 208,..., 313, Vito 110, 115; Opel Vivaro CDTI; Peugeot Boxer 335 HDI; Renault Master dCi, TD,Trafic dCi; Volkswagen Transporter T4 SD	2.6; 2.7; 2.8; 2.9; 3.0; 3.1; 3.5	2 148; 2 151; 2 163; 2 198; 2 299; 2 370; 2 402; 2 464;
2 500-2 999	Fiat Ducato 10, Maxi 35 MJ; Iveco Daily 35 C, 35; Mercedes Sprinter 210, Vito CDI; Peugeot Boxer 435 HDI; Renault Master Maxi	2.8; 2.9; 3.5	2 798; 2 800; 2 874; 2 953; 2 987; 2 998

Table 1b. Motor trucks belonging to the group of GVM > 3.5 t

Engine cubic capacity category [cm^3]	Vehicle makes and models	GVM [tons (metric)]	Example engine capacity values [cm^3]
2 500-9 999	Iveco Eurocargo, Stralis ML, ML180; MAN M2000, TGL; Mercedes-Benz 508D, 515 CDI , Sprinter, 814, 1114...1320, Atego; Volvo FL, FE, FM	4.6; 5.0; 6.0; 7.5; 8.0; 8.6; 10.0; 12.0; 13.0; 16; 18.0; 20.0	2 874; 3 908, 3 920; 4 249; 4 250; 4 580; 5 880; 5 958; 6 374; 6 871; 7 146; 9 364; 9 603
10 000-11 999	MAN TGA, TGX, TGS; Mercedes Benz Actros, Axor; Renault Premium, Premium Route; Scania 114, 124, R380, 124/420, 124/470, R420; R480;	18.0; 18.6; 19.0; 25.0; 26.0	10 300; 10 308; 10 318; 10 518; 10 520; 10 600; 10 635; 10 640; 10 837; 11 100; 11 116; 11 700; 11 705; 11 946; 11967
12 000+	DAF XF95, XF105; Iveco Stralis; MAN F2000, TG, TGA; Renault Magnum; Scania R440, R480, G440; Volvo FH12, FH16, FH400, FH440, FH480, FH500	18.0; 18.1; 18.6; 19.0; 26.0	12 100; 12 130; 12 771; 12 777; 12 780; 12 580; 12 740; 12 777; 12 780; 12 816; 12 880; 12 882; 12 895; 12 902

above as well as the results of the calculations carried out previously [19], a procedure consisting of the following stages was prepared:

- data collection;
- analysis of the data, with dividing the dataset into subsets;
- estimation of the average mileage values with determining the coefficient of variation;
- removal of outliers;
- approximation of the relation between the mileage and the vehicle operation time;
- estimation of the mileage growth models;
- evaluation of the accuracy of the estimators determined for the mileage growth models.

In the calculations, the following notation was adopted for the basic quantities:

L_i – mileage covered by the i^{th} vehicle during time t_i;
L_{10}, L_{20} – total vehicle mileage after 10 and 20 years of vehicle operation, respectively;
X, M, Q – arithmetic mean, median, and quantile, respectively;
S – standard deviation;
W – relative coefficient of variation;
R, φ – coefficient of determination and fraction of variance unexplained (FVU), respectively.

The scope of the work done at individual stages of the procedure, denoted by A1, A2, etc., has been described below.

A1. The data collection stage pertains to the motor trucks that are currently participating in the road cargo transport. In result of this stage, a representative dataset should be obtained that would include at least the technical specifications, mileage, and period of operation of individual vehicles. At this stage, it is important that various information sources should be used for the dataset formed to be representative.

A2. The analysis of the data collected should make it possible to clear the dataset, in a reasonable way, of the data that might adversely affect the calculation results, e.g. to remove the data of motor cars and car-based vehicles, which may be registered in Poland as accepted for cargo transport (or, more precisely, as passenger/cargo vehicles). At this stage, certain subsets are separated, which consist of uniform-category vehicles (e.g. vehicles for local, interurban, or international transport). The vehicles of individual categories are used for different transport jobs, which have an impact on the vehicle mileage. The division of a dataset depends on the dataset size and on the exact objective of the calculations.

The data analysis was based on an assumption made that it would cover a 20-year period of vehicle operation and the distance travelled by the end of such a period was referred to as "target mileage" (expressed in kilometres). The vehicles that were operated within this period for less than 3 months or that were more than 242 months old were not taken into account. The vehicles operated for only the first two months of the period under analysis were ignored because of very wide scatter in the values of the mileage covered by them during that time.

At this stage, the motor trucks of each category were divided into n subgroups. The vehicles whose operation time t_i was within the interval:

$$T_k < t_i \le T_{k+1} \tag{1}$$

were counted in the k^{th} subgroup. The limits of disjoint time intervals were defined with a monthly step, i.e.:

$$T_k = 12(k-1) \quad \text{for k = 2, 3, …, n} \tag{2}$$

and, additionally, T_1 = 2 months and T_{21} = 242 months. (3)

The separation of subgroups in which the numbers of vehicles would be adequate for research purposes (i.e. would be not less than a required minimum denoted by m_0) makes a basis for further calculations. In many cases, vehicles are divided into subgroups that cover successive vehicle operation periods whose length is T_0 = 12 months.

A3. At the stage of estimation of the average mileage values in individual subgroups, the scope of calculations was defined with taking into account, *inter alia*, the following premises:

- the arithmetic mean is an estimator rather insusceptible to an error in the evaluation of the properties of a dataset but it is susceptible to outliers (extreme values).

– the median is rather insusceptible to outliers, but it is unsuitable for statistical calculations.
– the average monthly mileage value calculated by dividing the total mileage value by the number of months of the vehicle operation period is burdened with an error arising from changes in the mileage growth rate with vehicle operation time.

In each subgroup, the vehicle mileage L was treated as a random variable and L_i was the i^{th} value of this variable. In the k^{th} subgroup that consisted of $m \geq m_0$ vehicles, the following operations were carried out:

B1. Forming of a series of mileage values $L_{0k}\{L_1,...,L_i,...,L_m\}$ consisting of m members arranged in ascending order.
B2. Calculation of the estimators: arithmetic mean X_{0k} (τ_k), median M_{0k} (τ_k), quantiles Q_{1k} (τ_k) and Q_{3k} (τ_k), and relative coefficient of variation W_{0k} (τ) as the following quotient:

$$W_{0k} = \frac{S_{0k}}{X_{0k}} \quad (4)$$

where S_{0k} is standard deviation, τ_k is the central value of the vehicle operation period in the k^{th} subgroup, and Q_{1k} (τ_k) and Q_{3k} (τ_k) are the lower and upper quantile, respectively.

In result of these calculations, the following series of discrete values were obtained for every vehicle category:

$$X_0\{X_{0k}(\tau_k), k=1,2,...,n\},\ M_0\{M_{0k}(\tau_k), k=1,2,...,n\},\ W_0\{W_{0k}(\tau_k), k=1,2,...,n\} \quad (5)$$

A4. Results of the calculations carried out at stage A3 make it possible to estimate the coefficient of mileage variation, the value of which shows whether the removal of outliers is necessary. High W_{0k} values indicate excessive scatter in the mileage values in the k^{th} subgroup. If the scatter is wide, the average mileage value in a specific subgroup not always adequately characterizes the mean mileage level. In such a case, efforts should be made to reduce the impact of the values that most differ from the average in the subgroup on the results of further calculations. The steps of this kind taken in the k^{th} subgroup may be divided into two sub-stages. At the first one, a series denoted by L_{Ek} was formed by removing 10 % of the first and last members of L_{0k}, i.e. the lowest and highest mileage values. Thus, the L_{Ek} series consisted of $m - 2m_E$ members, i.e.

$$L_{Ek}\ \{L_{m_E+1}, L_{m_E+2}, \ldots, L_{m-m_E}\},\ m_E = E(0.1m) \quad (6)$$

where operator E has the meaning of rounding off to the nearest integer.

At the second sub-stage, a series denoted by L_{Qk} was formed by clearing L_{0k} of the members where the L_i values were covered by the criterion:

$$L_i \leq Q_{1k} \text{ or } L_i \geq Q_{3k} \quad (7)$$

Following this operation, the L_{Qk} series consisted of members situated centrally in L_{0k} and falling within the range $Q_{1k} < L_i < Q_{3k}$.
After the removal of the outlying mileage values, the numbers of vehicles in individual subgroups decreases. In consequence, the numbers of vehicles in some subgroups may become $m < m_0$. If the k^{th} subgroup consists of less than m_0 vehicles, it should be merged with subgroup k+1. The subgroup thus combined will consist of vehicles whose operation time would fall within the interval:

$$T_k < t_i \leq T_{k+2}, \quad (8)$$

which is twice as wide as those considered previously.

Therefore, an operation was carried out to reduce the width of such "double" intervals to T_0 months with maintaining the natural scatter in the vehicle mileage values. With this objective in view, the mean mileage value $X_{(k)}$ and the central value of the vehicle operation time $\tau_{(k)}$ for the combined subgroup was calculated. Now, the $X_{(k)}$ and $\tau_{(k)}$ values were used to determine the deviation of the monthly mileage (PM_i) for the i^{th} vehicle of the subgroups combined together:

$$PM_i = \frac{L_i - X_{(k)}}{t_i - \tau_{(k)}} \quad (9)$$

For the vehicles where $|t_i - \tau_{(k)}| \geq 0.5T_0$ (i.e. whose operation time was longer or shorter than $\tau_{(k)}$ by more than $0.5T_0$ months), a reduced mileage value $L_{(i)}$ and a reduced vehicle operation time $t_{(i)}$ were calculated as follows:

$$L_{(i)} = L_i + 0.5PM_i\,(t_i - \tau_{(k)}) \quad (10)$$

$$t_{(i)} = t_i + 0.5(t_i - \tau_{(k)}) \quad (11)$$

In result of the above calculations, new parameters denoted by $L_{(i)}$ and $t_{(i)}$ were obtained for the i^{th} vehicle. They were substituted for the parameters L_i and t_i used previously and they already fell within the time interval:

$$\tau_{(k)} - 0.5T_0 < t_{(i)} \leq \tau_{(k)} + 0.5T_0 \quad (12)$$

The new subgroup, formed by merger of subgroups k and k+1, has characteristics similar to those of the subgroups that consisted of at least m_0 vehicles when they were initially defined. After these calculations, the diversity of the vehicle mileage and operation time values remained close to that recorded originally, although the said values were reduced to fall within the interval as defined by (12). The diversity of the mileage values L_i must be maintained in the subgroup merging process for the same procedure, with two sub-stages of removing the mileage outliers, to remain applicable when forming the L_{Ek} and L_{Qk} series.
The values of members of the L_{Ek} and L_{Qk} series were used to determine series of the following discrete values for each vehicle category:

$$X_{E,Q}\{X_{Ek,Qk}(\tau_k), k = 1, 2, ..., n_0\} \quad (13)$$

where: $X_{E,Q}$ represents series X_E and X_Q, respectively, and $n_0 < n$ in result of the merger of some subgroups.

A5. At this stage, a transition takes place from the series of discrete averaged values M_0, X_E, and X_Q in vehicle subgroups to a continuous function. This is done by approximating the dependence of the averaged vehicle mileage values on the vehicle operation time. In this work, the approximation was based on regression lines. In result of the approximation process, continuous functions in the form as follows were obtained, describing the vehicle mileage growth with vehicle operation time:

$$\hat{y}_a = f_a(\tau) \quad (14)$$

In consideration of the properties of the process of approximation based on regression lines, including strong dependence of the course of such lines on the degree of the approximating polynomial and on the number of the data available and on their nature, the approximating functions were determined in two ways:

- by using models with a 3rd degree polynomial based on three series, i.e. M_0, X_E, and X_Q, the following was calculated:

$$\hat{y}_{E,Q,M} = f_{E,Q,M}(\tau); \tag{15}$$

- by selecting models with 2nd and 4th degree polynomials based on the X_Q series of values, the following was obtained:

$$\hat{y}_{W2,W4} = f_{W2,W4}(\tau) \tag{16}$$

where subscripts $_{W2}$ and $_{W4}$ correspond to the degree of the polynomial.

In total, five approximating functions were calculated for each vehicle category.

A6. The estimation of the vehicle mileage growth models so that they were burdened with the smallest possible error was based on the combined use of approximating functions (15) and (16) for determining averaged functions on these grounds. The method of moving average was used to calculate the values:

$$Y_A(\tau),\ k = 1, 2, \ldots, n_0$$

that provided a basis for the estimation of the averaged approximating function:

$$\hat{y}_A = f_A(\tau) \tag{17}$$

The averaged approximating function (17) was treated as an estimator of the mileage growth model prepared for motor trucks.

A7. The evaluation of the accuracy of this estimator was based on absolute and relative indicators. The following absolute indicators were used (based on [8, 28]):

- standard error of estimation:

$$S_{E,Q,M} = \sqrt{\frac{1}{n_0 - z - 1}\sum_{k=1}^{n_0} e_{Ek,Qk,Mk}^2} \tag{18}$$

where: $e_{Ek,Qk,Mk} = y_{Ek,Qk,Mk} - \hat{y}_{Ak}$; $y_{Ek,Qk,Mk}$ – empirical values, i.e. the k^{th} members of series X_E, X_Q, and M_0, respectively; $\hat{y}_{Ak}$ – mileage value obtained from model (17); z – number of explanatory variables in the model;

- mean absolute error:

$$\overline{X}_{E,Q,M} = \frac{1}{n_0}\sum_{k=1}^{n_0}\left|e_{Ek,Qk,Mk}\right| \tag{19}$$

The relative indicators used included:

- coefficient of determination related to the values of X_E, X_Q, and M_0, i.e.

$$R_{E,Q,M}^2 = \frac{\sum_{k=1}^{n_0}(\hat{y}_{Ak} - \overline{y}_{E,Q,M})^2}{\sum_{k=1}^{n_0}(y_{Ek,Qk.Mk} - \overline{y}_{E,Q,M})^2} \tag{20}$$

- coefficient of variation in the error of estimation:

$$V_{E,Q,M} = \frac{S_{E,Q,M}}{\overline{y}_{E,Q,M}}100\% \tag{21}$$

where: $\overline{y}_{E,Q,M}$ – average value of series X_E, X_Q, and M_0, respectively;

- fraction of variance unexplained (FVU):

$$\phi_{E,Q,M}^2 = \frac{\sum_{k=1}^{n_0}(y_{Ek,Qk,Mk} - \hat{y}_{Ak})^2}{\sum_{k=1}^{n_0}(y_{Ek,Qk,Mk} - \overline{y}_{E,Q,M})^2} \tag{22}$$

The symbol $S_{E,Q,M}$ covers three quantities, i.e. S_E, S_Q, and S_M, which are calculated from X_E, X_Q, or M_0, as appropriate. The same subscripts have been used in equations (18)-(22).

The determination of the estimator evaluation indicators is the final step of the procedure, which was followed in practice at the evaluation of the process of growth in the motor truck mileage during the 20-year vehicle operation period.

4. Example implementation of the procedure developed

At the initial data collection stage, the vehicles were divided into two groups, based on the GVM value, i.e. up to and above 3.5 tons (metric). Thus, two separate datasets were formed, which covered 5 084 vehicles of the first group and 3 855 vehicles of the other one. The distribution of the number of vehicles, by age, in both groups has shown in Fig. 1. Both in the dataset and in the real road cargo transport, the vehicles having been operated for up to 10 years predominate.

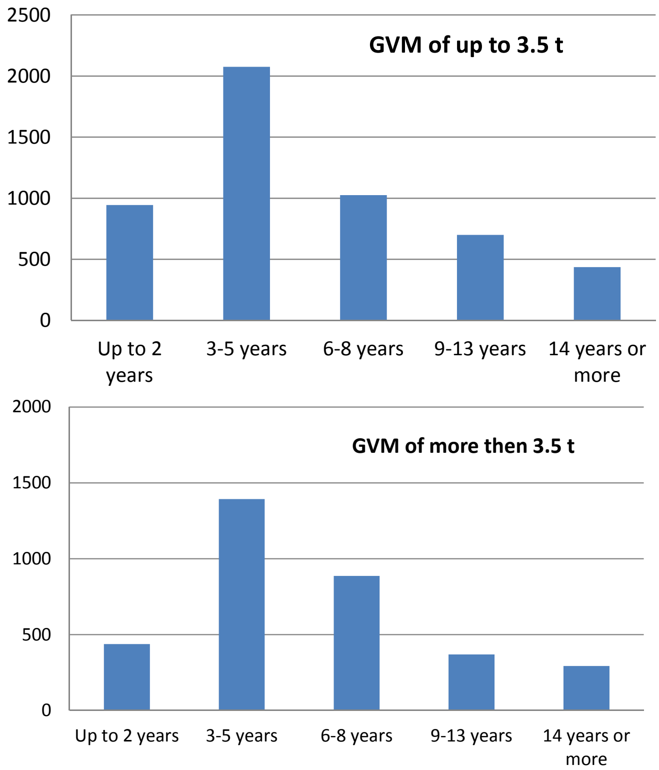

Fig. 1. Age (years of operation) of motor trucks in groups of up to and above 3.5 t GVM

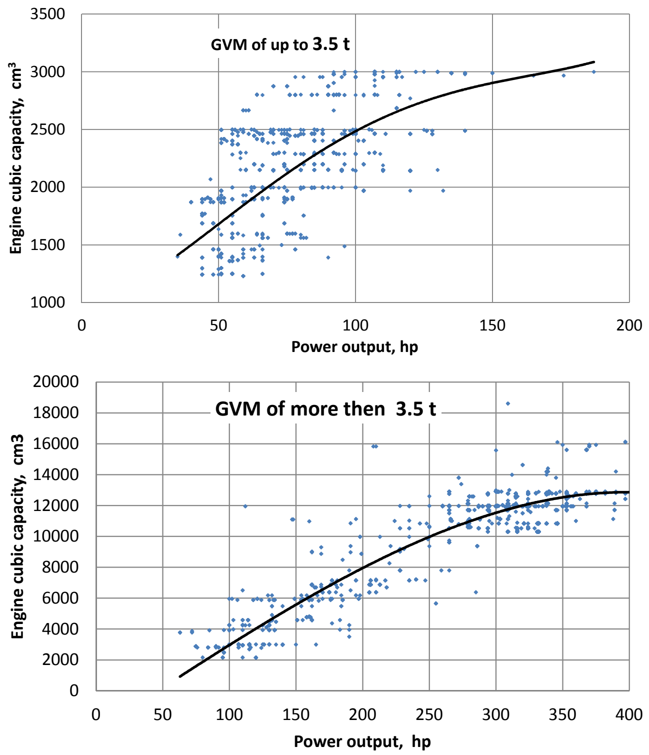

Fig. 2. Engine power output vs cubic capacity curves characterizing the datasets

Fig. 2 supplements the information about the vehicles under analysis that has been provided in Table 1. It shows that the engine capacity values gather in a few major ranges: 1 850-1 999, 2 400-2 499, 2 800-2 999, 5 900-6 900, 10 000-11 999, and 12 000-13 000 cm^3; in this study, therefore, this finding has been adopted as a basis for vehicle categorization.

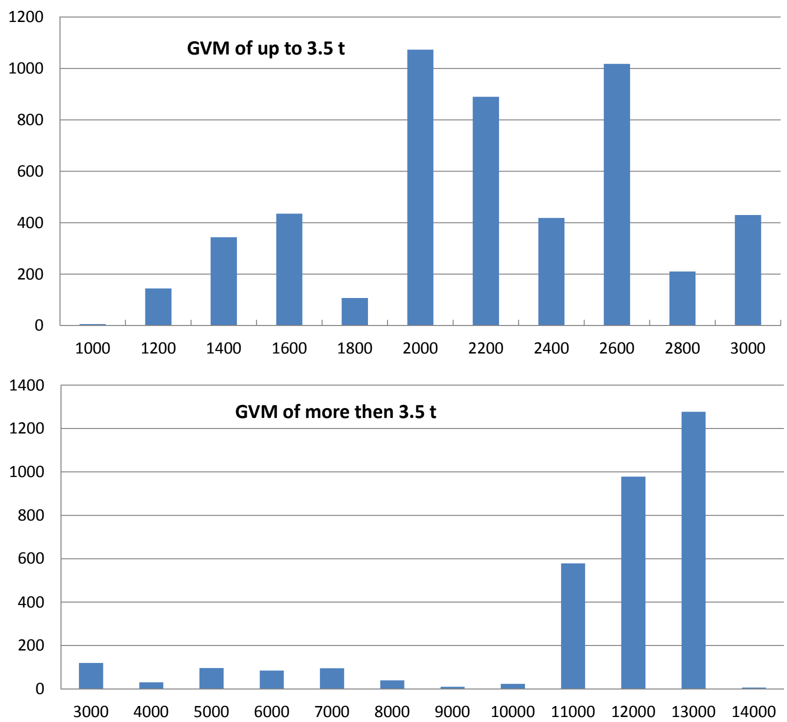

Fig. 3. Numbers of vehicles within successive ranges of engine cubic capacity

In Fig. 3, the engine cubic capacity values have been grouped in ranges of up to 1 000 cm^3, 1 001-1 200 cm^3, and so on. In consideration of the above, the distribution of engine capacity values as shown in Fig. 2, and classification adopted in [22, 24], the vehicles under analysis have been divided into engine capacity categories as follows:

- four categories in the group of local transport (delivery) vehicles (LTV);
- three categories in the group of heavy goods vehicles (HGV, with more than 3.5 metric tons GVM).

The limits of individual engine capacity categories and percentage of vehicles of individual categories in both groups have been presented in Table 2.

In each category, the vehicles were divided into $n = 20$ subgroups, based on the vehicle operation time, which was determined with 1 month accuracy, according to (1)-(3). The subgroups varied in size, with 63 vehicles per subgroup, on average, in the group of vehicles with GVM of up to 3.5 t and 64 vehicles per subgroup, on average, in the group of vehicles with GVM of above 3.5 t. The first subgroup consisted of vehicles whose operation time ranged from 3 months to 12 months; in the other one, there were vehicles having been operated for 13-24 months. When forming the subgroups, an assumption was made that a subgroup should consist of not less than $m_0 = 20$ vehicles. However, there were more than 10 subgroups where the number of vehicles remained below this minimum, in spite of a long data collection time (3 years). For these subgroups, the calculations to be done at stage A3 were not carried out. Only when these subgroups were merged with the next ones (at stage A4), the data about the vehicles of these subgroups were taken for further computations.

In the subgroups, the estimators of arithmetic mean X_{0k} (τ_k), median M_{0k} (τ_k), quantiles Q_{1k} (τ_k) and Q_{3k} (τ_k), and relative coefficient of variation W_{0k} (τ_k) were calculated. Example results of calculations carried out for the subgroups of vehicles belonging to category S02 have been presented in Fig. 4. Gradual growth in X_{0k} and M_{0k} with increasing τ_k as well as high values of the coefficient of variation W_{0k} can be seen in the graphs.

High values of the relative coefficient of variation ($W > 0.4$) can be seen in Fig. 4 and for many vehicle subgroups, especially for $k \leq 4$. This shows that the average value not always adequately characterizes the mean mileage level. Therefore, 10 % of the first and last members were removed at stage A4 from L_{0k} in the subgroups and thus the L_{Ek} series were formed. Then, on the grounds of results of calculation of quantiles Q_{1k} (τ_k) and Q_{3k} (τ_k), the L_{0k} series were cleared of the vehicles whose mileage was covered by the following criterion:

$$L_i \leq Q_{1k} \text{ or } L_i \geq Q_{3k}$$

In result of the above, the series L_{Ek} and L_{Qk} of mileage values were formed, whose members were situated centrally in L_{0k}.

When these operations were completed, all the $20 \times 7 = 140$ subgroups under analysis included 37 ones with $m < 20$ (i.e. which consisted of less than 20 vehicles). Therefore, the undersized subgroups were merged with the next ones. The new series of mileage values L_{Ek} and L_{Qk} were used for determining discrete estimator values, based on (4) and (13), i.e.:

Table 2. Limits of individual engine capacity categories [cm³] and percentage of vehicles of individual categories in both groups

Engine cubic capacity [cm³], GVM ≤ 3.5 t	Vehicle category symbol	Percentage [%] in the group of GVM ≤ 3.5 t	Engine cubic capacity [cm³], GVM > 3.5 t	Vehicle category symbol	Percentage [%] in the group of GVM > 3.5 t
Up to 1 499	S01	13.0	2 500-9 999	S2	14.1
1 500-1 999	S02	28.6	10 000-11 999	S10	46.3
2 000-2 499	S03	45.4	12 000 and more	S12	39.6
2 500-2 999	S04	13.0	–	–	–

Table 3. Values of the mileage and coefficient of variation for a few values of τ_k

Vehicle category: S02						
Years of vehicle operation, τ_k	X_{0k} [km]	X_{Ek} [km]	X_{Qk} [km]	W_{0k}	W_{Ek}	W_{Qk}
2.5	75 120	68 320	68 320	0.627	0.403	0.242
8.5	212 850	207 140	202 930	0.422	0.261	0.123
16.0	265 190	256 870	256 780	0.401	0.207	0.121
Target mileage, L_{20}	298 000	285 000	284 000	**0.454***	**0.294***	**0.169***
Vehicle category: S12						
Years of vehicle operation, τ_k	X_{0k} [km]	X_{Ek} [km]	X_{Qk} [km]	W_{0k}	W_{Ek}	W_{Qk}
2.5	202 540	205 360	209 380	0.435	0.289	0.153
8.5	870 360	882 180	883 440	0.243	0.152	0.075
15.0	1 009 120	1 002 240	1 007 810	0.205	0.056	0.036
Target mileage, L_{20}	1 390 000	1 310 000	1 255 000	**0.351***	**0.198***	**0.118***

** Average values of coefficient W, calculated for 20 years of vehicle operation; the L_{20} values were calculated from the approximating functions $\hat{y}$ according to (15).*

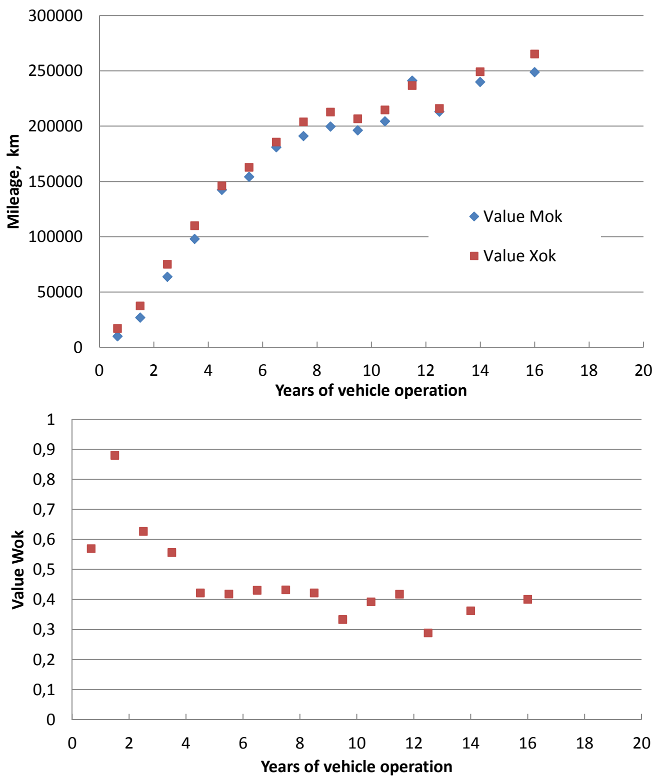

Fig. 4. Results of calculations of the X_{0k}, M_{0k}, and W_{0k} values in the subgroups of vehicles belonging to category S02

$$X_{E,Q}\{X_{Ek,Qk}(\tau_k),\ k=1,2,\ldots,n_0\},\ W_{E,Q}\{W_{Ek,Qk}(\tau_k),\ k=1,2,\ldots,n_0\}$$

where, as an example: $$W_{Ek} = \frac{S_{Ek}}{X_{Ek}}\ ,\ \ W_{Qk} = \frac{S_{Qk}}{X_{Qk}}$$

S_{Ek} and S_{Qk} are standard deviations and W_E and W_Q are coefficients of variation in mileage values in series L_{Ek} and L_{Qk}.

Examples of the calculation results have been presented in Table 3.

These calculation results were used to evaluate the effectiveness of the outliers-removal operation. Attention was paid to the impact of the outliers-removal operation used on the average mileage values in individual subgroups and on the values of the coefficient of variation. The X_{0k}, X_{Ek}, and X_{Qk} values in Table 3 do not differ very much from each other. They insignificantly changed with the removal of outliers in individual subgroups, in particular:

– $X_{0k} > X_{Ek} > X_{Qk}$, when the average values were chiefly affected by very high mileage values L_i;
– $X_{0k} < X_{Ek} < X_{Qk}$, when a predominating impact was exerted by very low values L_i.

An analysis of values $W\{W_0, W_E, W_Q\}$, based on calculation results obtained for all the vehicle categories, revealed beneficial effects of the removal of outliers from L_E and L_Q. This is confirmed by the W_E and W_Q values being definitely lower than W_0 (see Table 3). Desirable effects of the outliers-removal operation has also been illustrated in Fig. 5 by the relations between the extreme L_i values in series L_0, L_E, and L_Q and the X_{0k}, X_{Ek}, and X_{Qk} values. The calculation results have been presented in percentage terms, where the average values in individual subsets were assumed as 100 %.

Fig. 6 shows example curves representing the approximating functions determined according to (15) and (16). Figs 6a and 6c shows values X_{Ek}, X_{Qk}, and M_{0k} (points in the graphs) and the functions that approximate the mileage growth process for vehicles of categories S03 and S12, based on models with a 3rd degree polynomial and marked

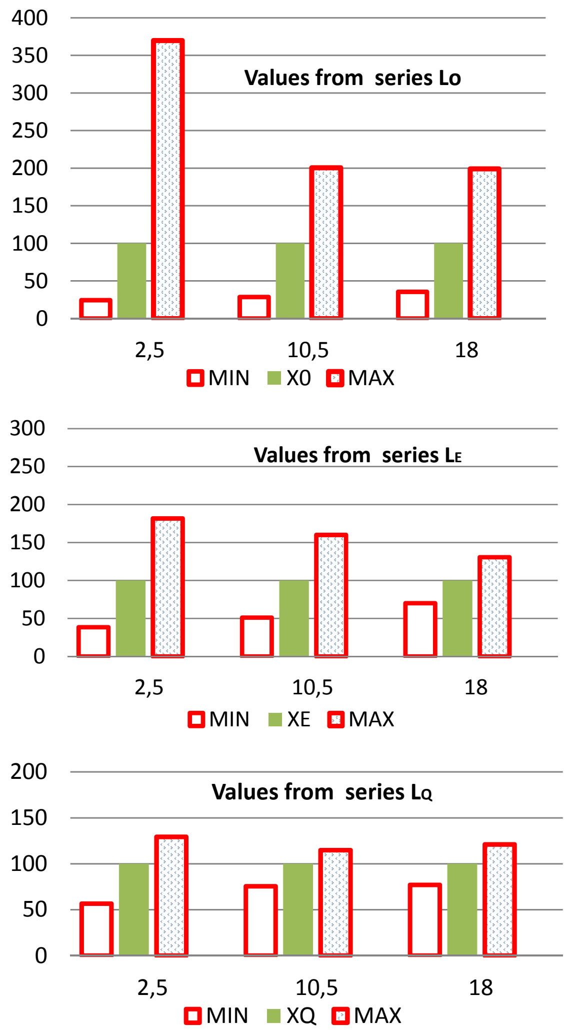

Fig. 5. Extreme (MIN and MAX) and mean mileage values in series L_0, L_E, and L_Q, calculated for the S02 category vehicles (in percentage terms) and compared with each other for τ_k = 2.5, 10.5, and 18.0 years of vehicle operation

Polynomial XE, *Polynomial XQ*, and *Polynomial M*0, respectively. In Figs 6b and 6d, the X_{Qk} values and the approximating functions based on models with 2nd and 4th degree polynomials and marked as *Polynomial XQ2* and *Polynomial XQ4*, respectively, have been presented.

In result of an analysis of the approximating functions, the following findings have been formulated:

- The comparison of the approximating functions calculated according to (15) showed small differences between them within the range of 6-14 years of vehicle operation (e.g. in Fig. 6a).
- In several cases, the use of a 3rd degree polynomial resulted in the showing of an excessive growth in the mileage value in the 18th, 19th, and 20th year of vehicle operation (Fig. 6c).
- The use of 2nd degree polynomials eliminated the deviations mentioned above but the value of the coefficient of determination became considerably lower than that for polynomials of higher degrees.
- The use of a 4th degree polynomial resulted in a formal improvement in the model quality, manifested in a growth in the coefficient of determination; however, the approximating function showed excessively dynamic changes in the final part of the vehicle operation period (e.g. fluctuations in the function values, see Figs 6b and 6d).

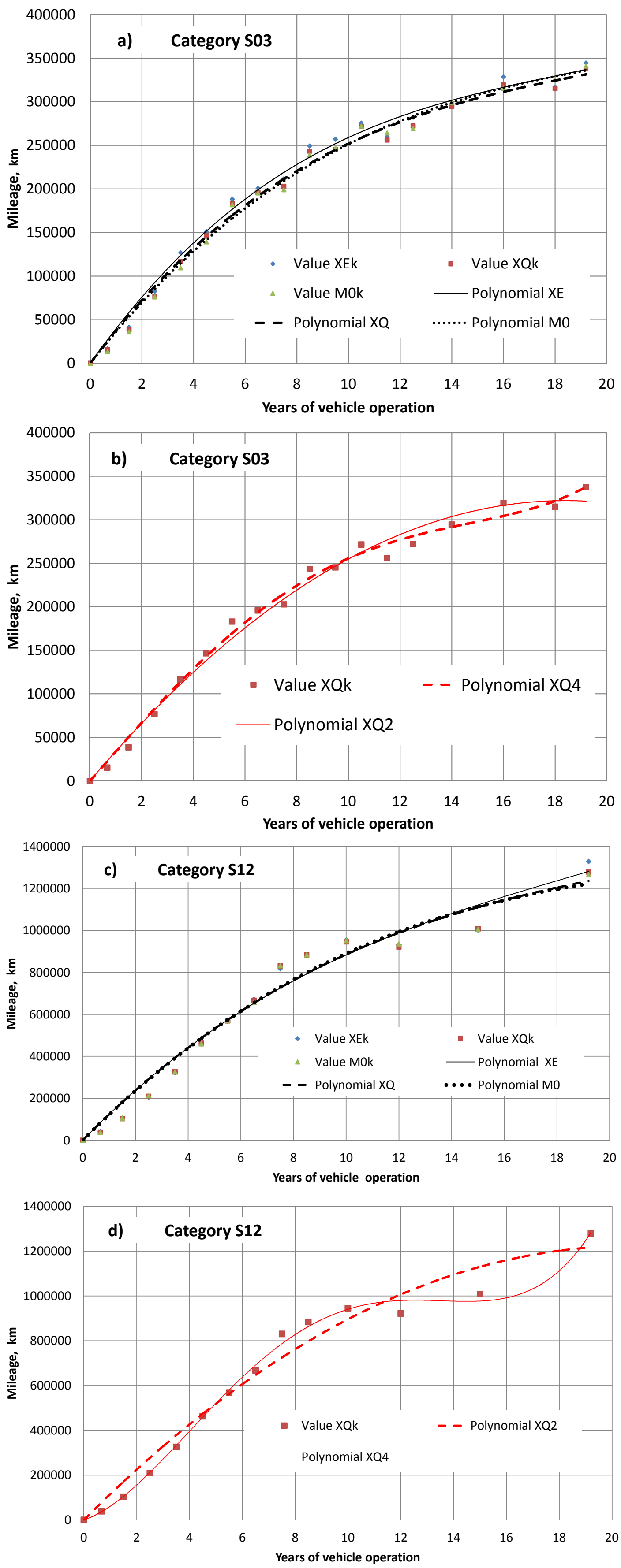

Fig. 6. Results of the calculations of X_{Ek}, X_{Qk}, and M_{0k} (points) and approximating functions $\hat{y}$ according to (15) and (16): a and b – for vehicles of category S03; c and d – for vehicles of category S12

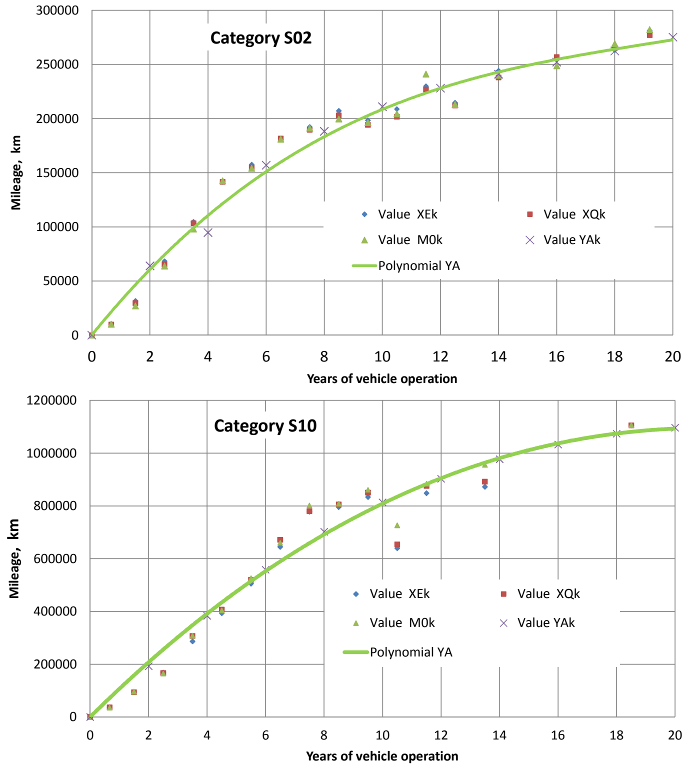

Fig. 7. Curves representing the averaged approximating function $\hat{y}_A(\tau)$ as well as positions of points X_{Ek}, X_{Qk}, M_{0k}, and average values Y_{Ak}; the example pertains to vehicles of categories S02 and S10

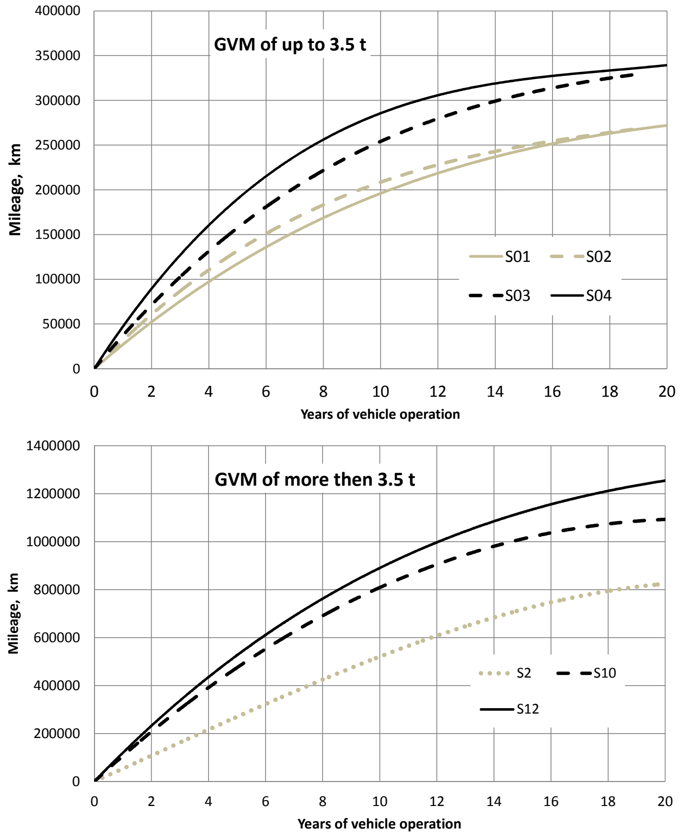

Fig. 8. Models of the process of vehicle mileage growth with vehicle operation time

Table 4. Values of coefficients a_r of the equations that model the vehicle mileage growth process for seven categories of engine cubic capacity

Vehicle category	a_1	a_2	a_3
S01	27 924	−948.56	11.623
S02	33 147	−1484.20	25.437
S03	38 692	−1554.90	22.677
S04	49 684	−2589.40	47.680
S2	53 590	314.20	−46.636
S10	109 769	−3022.80	13.339
S12	123 752	−3891.50	42.025

These findings give reasons for a need of going through the next stage of the procedure presented, i.e. the averaging of the approximating functions calculated as described above.

5. Estimation of the mileage growth models

An example of using the approximating functions (15) and (16) to determine their average values Y_{Ak} and to define an averaged approximating function has been presented in Fig. 7. In the graphs, the average mileage values directly calculated from the initial data (points X_{Ek}, X_{Qk}, and M_{0k}) have been compared with the final result shown in the form of a mileage growth model $\hat{y}_A(\tau)$ or two vehicle categories S02 and S10. It can be seen in Fig. 7 that the curves obtained from the mileage growth model are situated close to the points that represent the average values directly calculated from the experimental data.

Fig. 8 shows a comparison between the curves representing the averaged approximating functions (17) determined at stage A6 of this procedure. The mileage growth models have been based here on polynomials:

$$L = a_3\tau^3 + a_2\tau^2 + a_1\tau \tag{23}$$

where: L – vehicle mileage [km] and τ – vehicle operation time [years].

The values of coefficients $a_r(a_1, a_2, a_3)$ of equation (23) have been given in Table 4.

6. Evaluation of the conformity of models with empirical (initial) data

The mileage growth process described by the estimation of the averaged approximating function (17) was evaluated in several steps. Within this evaluation, the mileage values calculated from the models were compared with average values X_E, X_Q, and median M_0, directly based on experimental data. Thus, the evaluation of the accuracy of the estimators determined for the mileage growth models was based on indicators of conformity of model values with empirical data. Results of the calculation of values of the indicators of conformity have been given in Table 5.

The values of standard errors of estimation S_E, S_Q, and S_M are the basic measure of the scatter in mileage values around the values determined from the model equations; the V_E, V_Q, and V_M values define this scatter in percentage terms. The scatter in values X_E, X_Q, and M_0 around the mileage values determined from the model does not exceed 12 %. The values of the coefficient of determination R^2 of regression equations (23) for the coefficients specified in Table 4 are 0.99, i.e. these equations represent 99 % of the information contained in the sets of values Y_{Ak} (cf. Fig. 7). Figures of special importance for

Table 5. Results of the calculation of values of the indicators of conformity of model values with experimental data

Calculations based on data from series X_E					
Vehicle category	S_E [km]	$\bar{X}_E$ [km]	V_E [%]	R_E^2 .	φ_E^2
S01	14 900	10 400	7.87	0.99	0.023
S02	12 800	9 940	6.49	0.98	0.022
S03	11 620	9 370	4.82	0.97	0.012
S04	28 690	22 040	10.91	0.88	0.055
S2	41 640	31 660	7.94	0.94	0.017
S10	87 100	66 080	11.26	0.87	0.046
S12	76 410	61 440	8.86	0.87	0.026
Calculations based on data from series X_Q					
Vehicle category	S_Q [km]	$\bar{X}_Q$ [km]	V_Q [%]	R_Q^2 .	φ_Q^2
S01	15 860	11 070	8.38	0.98	0.026
S02	12 940	10 280	6.55	0.97	0.022
S03	10 320	8 200	4.28	0.95	0.009
S04	27 700	22 350	10.54	0.86	0.050
S2	42 850	30 460	8.17	0.91	0.017
S10	84 690	63 100	10.95	0.88	0.044
S12	72 730	57 570	8.43	0.89	0.024
Calculations based on data from series M_0					
Vehicle category	S_M [km]	$\bar{X}_M$ [km]	V_M [%]	R_M^2 .	φ_M^2
S01	16 630	11 050	8.78	0.97	0.028
S02	13 800	11 300	6.99	0.94	0.025
S03	10 130	7 800	4.20	0.91	0.008
S04	31 400	27 010	11.94	0.83	0.062
S2	49 750	33 740	9.49	0.92	0.023
S10	74 490	55 380	9.63	0.88	0.034
S12	71 510	55 610	8.29	0.90	0.023

Table 6. Values of the mileage covered for 10 years of vehicle operation, with estimation of their relative accuracy

Vehicle category	S01	S02	S03	S04	S2	S10	S12
Mileage L_{10} [km]	196 010	208 490	254 110	285 590	520 680	808 750	890 390
$(S_E/L_{10})100$ [%]	7.60	6.14	4.57	10.04	8.00	10.77	8.58
$(S_Q/L_{10})100$ [%]	3.82	3.64	2.81	6.49	3.84	6.05	4.90
$(\bar{X}_E/L_{10})100$ [%]	8.09	6.21	4.06	9.70	8.23	10.47	8.17
$(\bar{X}_Q/L_{10})100$ [%]	4.07	3.77	2.46	6.58	3.69	5.77	4.59

the evaluation of mileage growth are the R_E^2 R_Q^2 and R_M^2 coefficient values, which indicate how much, in percentage terms (83-99 %, see Table 5), of the information contained in X_E, X_Q, and M_0 of the experimental data has been taken into account in the models calculated.

The indicators calculated may also constitute a measure averaged for the vehicle operation period. Example values of this measure, expressed in percentage terms as the quotient of the values of absolute indicators (18 and 19) to mileage value L_{10}, have been given in Table 6. Thus, the average values of the indicators have been considered in relation to the value of the mileage achieved in the middle of the vehicle operation period. These calculations were carried out with using the mileage values obtained from the models having been developed.

The relative error values specified above, calculated for the middle of the target mileage, do not exceed 11 % of the mileage covered for 10 years of vehicle operation.

7. Recapitulation and conclusions

A several-stage process of the estimation of coefficients of the mileage growth models prepared for motor trucks was carried out and then the accuracy of these calculations was evaluated. The values of the evaluation measures, determined in this process, characterize the accuracy of calculations of the mileage growth during a 20-year period of vehicle operation. The calculations covered seven categories of engine cubic capacity and two vehicle groups, with GVM of up to and above 3.5 tons (metric). The mileage growth process was evaluated on the grounds of the values of five different indicators defined by equations (18)-(22), based on a comparison of mileage values directly calculated from experimental data with values obtained from the models. Both absolute indicators (standard error of estimation, mean absolute error) and relative indicators (coefficients of determination, coefficient of variation in the error of estimation, and fraction of variance unexplained – FVU) were used.

The conclusions that can be drawn from the evaluation of the process of growth in the motor truck mileage may be formulated as follows:

– Due to the wide scatter in mileage values (coefficient of variation W > 0.5), which resulted from diversity of transport jobs, a specially prepared estimation procedure had to be used in order to obtain adequate reliability of the description of the mileage growth process.
– The outliers-removal procedures applied had a favourable impact on concentration of the mileage values around the mean level and, at the same time, did not considerably affect the said mean value (see Table 3).
– The values of the coefficients of determination $R^2_{E,Q,M}$ show that a very big part of the initial information (0.83-0.99, see Table 5) contained in the experimental data (X_E, X_Q, and M_0) was retained in the mileage growth models having been created.
– The high quality of the procedure adopted to evaluate the mileage growth process was confirmed by the coefficients of variation in the error of estimation V_E, V_Q, and V_M; as a relative measure of the scatter in the mileage values, they did not exceed 12 % of the average values of the experimental data.

The calculations carried out made it possible to obtain original values of the coefficients of equations of the models representing the current processes of growth in the motor truck mileage values in Poland. The calculations covered practically all the categories of the motor vehicles used for road cargo transport. The values of the indicators of evaluation of the mileage growth models having been developed within this work confirmed high quality of the calculations carried out, the results of which will find application in many areas, e.g. planning, forecasting, and managing of the motor truck operation or life cycle assessment (LCA) of motor trucks in terms of exhaust emissions and energy consumption.

References

1. Chłopek Z, Bebkiewicz K. Model of the structure of motor vehicles for the criterion of the technical level on account of pollutant emission. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(4): 501-507, https://doi.org/10.17531/ein.2017.4.2.
2. Clarke D D, Ward P, Bartle C, Truman W. An in-depth study of work-related road traffic accidents. Road Safety Research Report No. 58. London: Department for Transport, 2005.
3. Droździel P, Krzywonos L. A model of the economic effectiveness of the truck transportation services. Transport Problems 2010; 5(4): 49-56.
4. External costs of transport - Update study. Final Report. Zurich/Karlsruhe: INFRAS, 2004.
5. Highway Statistics 2008. Washington: Department of Transport, 2008.
6. Hultkrantz L, Nilsson J-E, Arvidsson S. Voluntary internalization of speeding externalities with vehicle insurance. Transportation Research Part A 2012; 46: 926-937, https://doi.org/10.1016/j.tra.2012.02.011.
7. Info-Ekspert Computer system. Catalogs. www.info-ekspert.pl.
8. Krysicki W, Bartos J, Dyczka W, Królikowska K, Wasilewski M. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach (Probability theory and mathematical statistics in problems). Part I. Rachunek prawdopodobieństwa (Probability Theory); Part II. Statystyka matematyczna (Mathematical statistics). Warszawa: PWN, 2002.
9. Lapparent M (de), Cernicchiaro G. How long to own and how much to use a car? A dynamic discrete choice model to explain holding duration and driven mileage. Economic Modelling 2012; 29: 1737-1744, https://doi.org/10.1016/j.econmod.2012.05.018.
10. Macián V, Tormos B, Riechi J. Time replacement optimization model: comparative analysis of urban transport fleets using Monte Carlo simulation. Eksploatacja i Niezawodnosc - Maintenance and Reliability 2017; 19(2): 151-156, https://doi.org/10.17531/ein.2017.2.1.
11. Mendyk E. Ekonomika transportu (Economics of transport). Poznań: Publishing House of the Poznan School of Logistics, 2009.
12. Meurs H, Haaijer R, Geurs K T. Modeling the effects of environmentally differentiated distance-based car-use charges in the Netherlands. Transportation Research Part D 2013; 22: 1-9, https://doi.org/10.1016/j.trd.2013.02.013.
13. National Travel Survey. Statistical Release. Department of Transport. UK 2013. national.travelsurvey@dft.gsi.gov.uk.
14. Opoczyński K. Synteza wyników GPR 2015 na zamiejskiej sieci dróg krajowych (Synthesis of results of the General Measurement of Road Traffic on the extra-urban network of national roads in 2015). Warszawa: Transprojekt-Warszawa, 2016.
15. Paefgen J, Staake T, Fleisch E. Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data. Transportation Research Part A 2014; 61: 27-40, https://doi.org/10.1016/j.tra.2013.11.010.
16. Piernikarski D. Reduction of fuel consumption in commercial vehicles as a factor influencing transport efficiency. Combustion Engines 2013; 154(3): 663-669.
17. Poulter D R, Chapman P, Bibby P A, Clarke D D, Crundall D. An application of the theory of planned behaviour to truck driving behaviour and compliance with regulations. Accident Analysis and Prevention 2008; 40: 2058-2064, https://doi.org/10.1016/j.aap.2008.09.002.
18. Prochowski L. Analysis of the Relations between the Road Category and the Fatality of Road Transport Accidents. Logistics and Transport 2012: 14(1): 5-16.
19. Prochowski L. Characteristic features of the relation between motor truck mileage and engine cubic capacity. Journal of KONES Powertrain and Transport 2016; 23(4): 403-412, https://doi.org/10.5604/12314005.1217257.
20. Puentes R, Tomer A. The Road Less Travelled: An Analysis of Vehicle Miles Travelled Trends in the U.S. Metropolitan Infrastructure Initiative Series. Washington: Brooking Institution, 2009.
21. Redmer A, Kiciński M, Rybak R. Freight Fleet Management - the methods. Gospodarka materiałowa i logistyka 2014; 4: 11-18.
22. Automotive Industry Reports. Year Books 2008, 2012, 2014, 2016. Warszawa: Polish Automotive Industry Association, 2008- 2016.

23. Small K A. Energy policies for passenger motor vehicles. Transportation Research Part A, 2012; 46: 874-889, https://doi.org/10.1016/j.tra.2012.02.017.
24. Transport Activity Results in 2014. Warszawa: Central Statistical Office, 2015.
25. Transport Statistics Great Britain, The Stationery Office. London: Department for Transport, 2006.
26. Vehicle Survivability and Travel Mileage Schedules. Report NHTSA's. Washington: National Center for Statistics and Analysis, 2006.
27. Vierth I, Schleussner H. Impacts of different environmentally differentiated truck charges on mileage, fleet composition and emissions in Germany and Sweden. Stockholm: CTS Working Paper, 2012.
28. Zeliaś A, Pawełek B, Wanat S. Prognozowanie ekonomiczne. Teoria, przykłady, zadania (Economic forecasting. Theory, examples, problems). Warszawa: PWN, 2003.

Leon PROCHOWSKI
Military University of Technology
ul. Gen. Urbanowicza 2, 00-908 Warsaw, Poland

E-mail: leon.prochowski@wat.edu.pl