



# Statistical analysis of time of water outflow on the soil surface after the failure of a buried water pipe

M. Iwanek, P. Suchorab, A. Skrzypek & M. Budzioch  
*Faculty of Environmental Engineering,  
Lublin University of Technology, Poland*

## Abstract

Descriptive and inferential statistics are two main methodologies used in statistical data analysis. Descriptive statistics summarize data sets obtained as a result of an experiment, whereas inferential statistics enables generalization of observed data properties on a larger population. The present article is the outcome of the statistical analysis of results of time measurements of water outflow on the soil surface after the failure of a buried water pipe during laboratory investigations. During each of the 28 experiment variants, the pipe under internal water pressure was intentionally damaged causing the leakage. The time between the moment of the breakage and the moment of the water outflow on the soil surface was measured. A pressure head in the pipe and a leak area were two parameters varying during investigations. The first part of the statistical analysis included calculations of a mean, a mode, quartiles, a range, a standard deviation, a dispersion, a skewness and an excess kurtosis for data obtained in laboratory investigations. The normality of the results distribution was verified with Kolmogorov–Smirnov tests – original and modified by Lilliefors as well as the Shapiro–Wilk test. Relationships between varying analysed time and parameters during investigations were estimated on the basis of the regression analysis. All needed parameters were calculated with the Statistica 12 (StatSoft, Inc.) and MS Excel software. Results of analyses enabled to determine average values of time for investigated conditions, necessary in further studies. Laboratory investigations indicated that higher pressure head in a pipe results in tendency of time of water outflow to be lower (as per expectations), but no regularity between time and the leak area was proved during the analysis.

*Keywords: statistical analysis, water pipe failure, water outflow time.*



# 1 Introduction

The statistical analysis of experimental results is an important part of any investigations. The most basic statistical analysis is descriptive statistics that summarizes data set obtained as a result of an experiment, using measure of a central tendency and dispersion [1]. Theoretically, descriptive statistics should be used only for data set obtained in identical conditions. However, it is practiced that descriptive statistics is used to describe data sets also from different conditions [2–4]. Among many descriptive statistics techniques there are tabular description, graphic presentations and calculation of summary statistics (measure of location, of statistical dispersion, of asymmetry and concentration). Basing on information contained in a sample of measurements, the inferential statistics enables to make estimates, predictions or generalizations about a population of measurements. There are two basic types of statistical inference: the estimation and the hypothesis testing [5]. To calculate these statistics, tests of Kolmogorov–Smirnov (K–S), Kolmogorov–Smirnov with Lilliefors modification (K–S–L) and Shapiro–Wilk (S–H) are frequently used [6]. Investigations of a water pipe failures is a difficult task mainly because of its complex character. Statistical analysis of results of a water pipe breakage laboratory investigation should start with determination of average characteristic parameters and assessment of data sets normality in order to enable further, more advanced analysis.

Water supply pipes failures are inseparable and natural phenomenon [7], which occurs during the water supply network exploitation. The most often occurring water supply pipe failures are: seal material damage, longitudinal and cross cracks, welds raptures, fittings damages and pipe corrosion damages. Generally, among many different water supply pipes failures there are failures that generate and do not generate water leakages. The random character of failures and insufficient knowledge about them [8] is caused mainly by the variety of potential failure reasons, influenced both by static (material, diameter, age of a pipe and a kind of a soil) and dynamic (climate, changes of pressure zones) factors [9–11]. The variety of failure implicating factors causes that both place and time of a water pipe failure is impossible to precise defining [12]. However, the random character of a water supply pipe failure is typical mainly for pipes in a good technical condition. Together with the aging of a pipe, also a frequency of failure occurring is rising, due to the fact of material aging. There are also several dependences between pipe failure intensity and various water supply network parameters (pressure, diameter, function and material of a pipe) [13].

Consequences of failures can be direct and indirect. Direct consequences are mainly the possibility of water quality worsening, water losses and the water supply delivery discontinuity. Indirect consequences are mainly a surface subsidence, foundations wash out and water entering into buildings. Therefore, methods of detecting water pipe failures are continuously developing [7, 14, 15]. Identification of a leakage problem is relatively the easiest when the water outflow is visible on the soil surface. However, the outflow does not always occur on the soil surface immediately after a pipe failure and sometimes it does not occur at all. Time of water outflow on the soil surface after a failure of a buried pipe depends

on many different parameters – both hydraulic water pipe characteristics as well as properties of soil [16]. The longer the time between the water pipe failure and the water outflow, the more serious possible consequences of a failure.

This paper presents descriptive and inferential statistics of a data set obtained from laboratory experiments of a physical simulation of a failure of a buried water pipe. During experiments the pipe under internal water pressure was intentionally damaged causing the leakage. Obtained and interpreted data represent time of water outflow on the soil surface after a failure.

## 2 Materials and methods

Time of water outflow on the soil surface area after a failure of a buried water pipe was investigated during laboratory simulation. Tests required construction of the setup (Figure 1), reflecting natural operation conditions in scale 1:10. The laboratory setup consisted of an intentionally damaged water pipe (2) buried 18 cm below the soil surface of medium sand filling a cuboid box, with dimensions of 1.5 by 1.5 by 0.5 m (1). The pipe was supplied by water from a container (4) located on the assumed height.

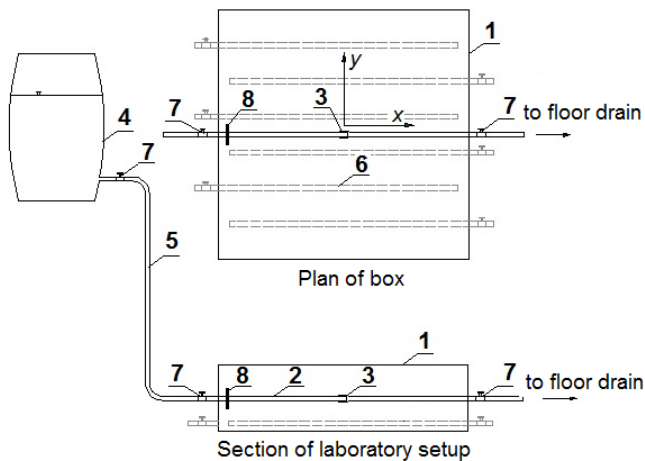


Figure 1: Scheme of laboratory setup for physical simulation of water supply failure [17, 18]: 1 – sand-filled box, 2 – water pipe, 3 – bell-and-spigot connection (place of leakage), 4 – container, 5 – hose, 6 – drainage system, 7 – valves, 8 – holder.

Tests were conducted for 4 different leak areas  $A$ : 4.71 cm<sup>2</sup>, 5.58 cm<sup>2</sup>, 9.42 cm<sup>2</sup> and 12.25 cm<sup>2</sup>. Internal water pressure head in the pipe varied from 3.0 to 6.0 H<sub>2</sub>O depending on the high of the container (4) and the water level in it. A pressure head in the pipe and a leak area were the only parameters varying during investigations. Overall, tests were conducted for 28 variants. The detailed description of laboratory setup, sand parameters and experiment methodology are given in the article of Iwanek *et al.* [18].

The next part of investigations was the statistical analysis of test results of measured time  $t$  between the moment of intentional water pipe failure and the moment of water outflow on the soil surface. The first part of statistical analysis included calculations of a mean, quartiles, a range, a standard deviation, a coefficient of variation, a skewness and a kurtosis for data  $t$  obtained in laboratory investigations. Data were selected according to two parameters varied in laboratory tests (both leak area and pressure head), for data selected according to only one parameter (leak area or pressure head) as well as for all data obtained during experiments, without any selection. The normality of the results distribution was verified with Kolmogorov–Smirnov tests – original and modified by Lilliefors as well as the Shapiro–Wilk test, at significance level  $\alpha = 0.05$ . Relationships between the analyzed time and parameters varying during the investigations were estimated on the basis of the regression analysis. All needed parameters were calculated with Statistica 12 (StatSoft, Inc.) and MS Excel software.

### 3 Results and discussion

Results of data  $t$  descriptive statistics for cases of  $A = 4.71 \text{ cm}^2$  are presented in Figures 2–5. Values of an arithmetic mean and a median of data  $t$  occurred comparable in most presented cases (Figure 2). The highest relative percentage difference (comparing to the mean) between these parameters (20%) was observed for the case of data unselected according to pressure head in the pipe. In cases of selected data, a relative percentage difference was between 2.81% and 7.41%. The median occurred higher than the mean for the cases of  $H = 4.0 \text{ m H}_2\text{O}$  and  $H = 6.0 \text{ m H}_2\text{O}$ , which indicated left asymmetrical distribution of data set for these pressure values. In other cases, higher values of the mean indicated right asymmetrical distribution.

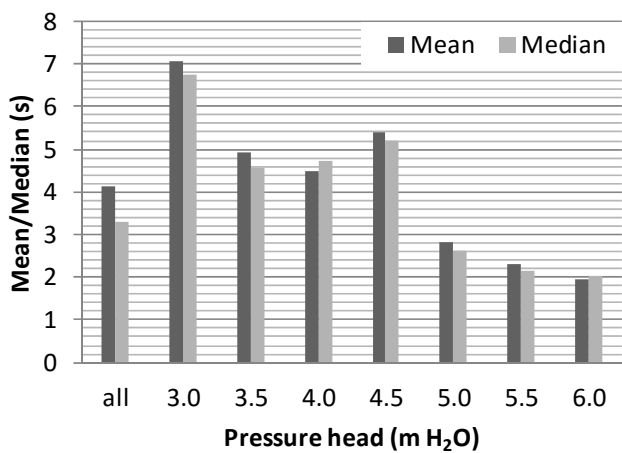


Figure 2: Arithmetic mean and median of data  $t$  obtained for cases of leak area of  $4.71 \text{ cm}^2$ .



The difference between maximal and minimal values of time (range) in presented cases tended to decrease with higher pressure head in the water pipe (Figure 3), from 16.71 s for  $H = 3.0$  m H<sub>2</sub>O to 0.12 s for  $H = 6.0$  m H<sub>2</sub>O. An exception in this tendency was the case of  $H = 4.5$  m H<sub>2</sub>O. Placement of quartiles with regard to external values suggested data distribution irregularity for  $H = 3.5$  m H<sub>2</sub>O,  $H = 4.0$  m H<sub>2</sub>O,  $H = 5.5$  m H<sub>2</sub>O and in the case of all pressure heads without selecting.

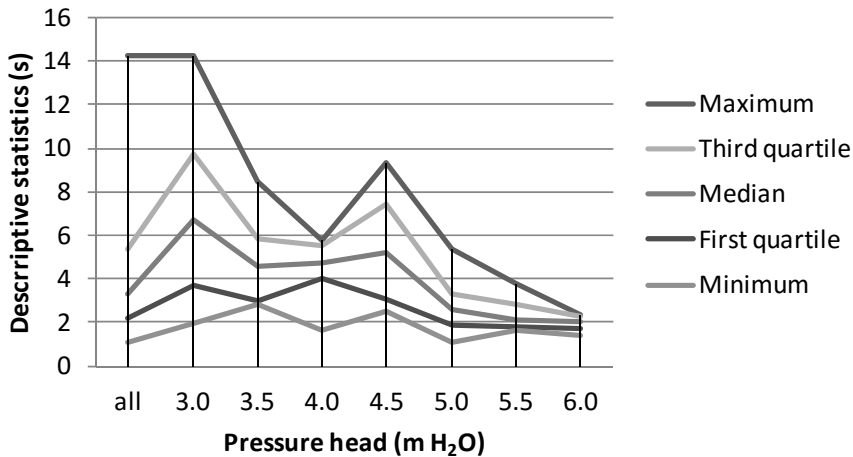


Figure 3: Maximum, minimum and quartiles of data  $t$  obtained for cases of leak area of 4.71 cm<sup>2</sup>.

The next calculated measure was a standard deviation, which enabled to quantify dispersion of a set of data  $t$ . Results shown in Figure 4 indicated that data points are spread out over a wide range. A relative standard deviation (a ratio of a standard deviation to a mean) exceeded 50% in the case of data unselected according to pressure head and for  $H = 3.0$  m H<sub>2</sub>O. The lowest value (17.74%) was reached for  $H = 6.0$  m H<sub>2</sub>O.

Results of a skewness and an excess kurtosis calculations for cases of leak area of 4.71 cm<sup>2</sup> are shown in Figure 5. Negative values of a skewness observed in two cases ( $H = 4.0$  m H<sub>2</sub>O and  $H = 6.0$  m H<sub>2</sub>O) suggested left-asymmetrical data  $t$  distribution. In rest cases, positive values of a skewness indicated right-asymmetrical shape of data distribution. In the case of  $H = 4.5$  m H<sub>2</sub>O, the absolute value of a skewness (0.35) was nearest to zero – characteristic for symmetrical distributions. In other cases the parameter ranged from 0.63 to 1.65.

The second measure presented in Figure 5 is an excess kurtosis, equal to 0 for normal distribution. The nearest to this value result (-0.39) was reached for  $H = 3.0$  m H<sub>2</sub>O. The highest absolute values exceeded 2 in two cases: in the case of data unselected according to pressure head and for  $H = 4.0$  m H<sub>2</sub>O. Negative values of an excess kurtosis observed in half of presented cases mean that the distribution is less peaked and has less frequent extreme values than normal

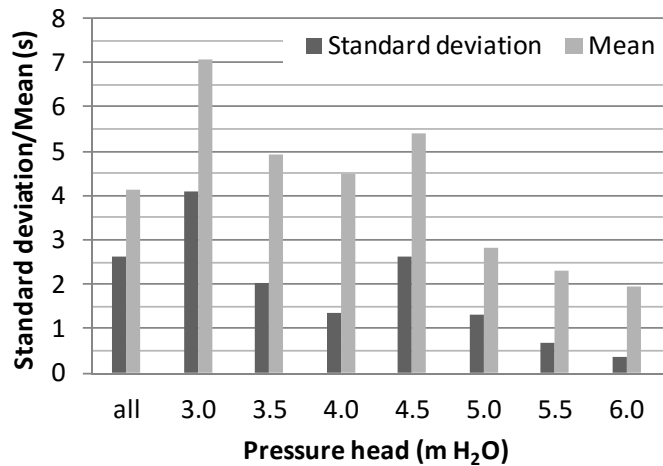


Figure 4: Maximum, minimum and quartiles of data  $t$  obtained for cases of leak area of 4.71 cm<sup>2</sup>.

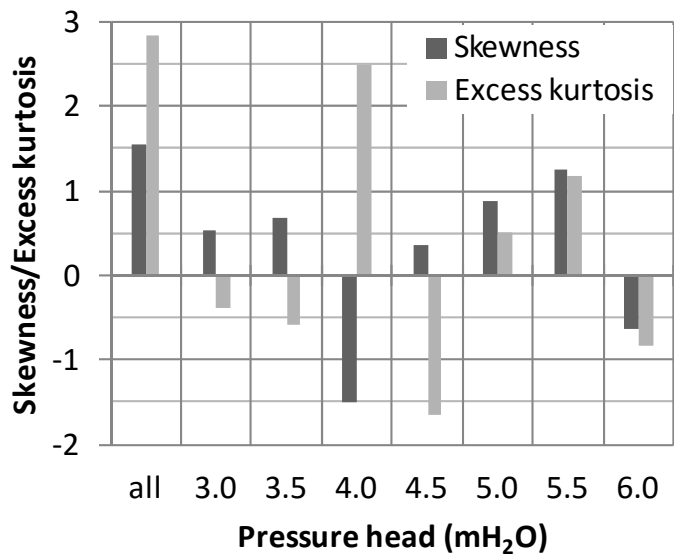


Figure 5: Skewness and excess kurtosis of data  $t$  obtained for cases of leak area of 4.71 cm<sup>2</sup>.

distribution (platykurtic distribution). The second half of cases with positive excess kurtosis was characterized by leptokurtic distribution (observations close to a mean were more frequent than in normal distribution).

The same analysis as presented above for data  $t$  obtained during laboratory tests for cases of leak area of 4.71 cm<sup>2</sup> was conducted for the rest leak areas (9.42 cm<sup>2</sup>,

15.07 cm<sup>2</sup> and 18.84 cm<sup>2</sup>) and similar results were attained. The relative percentage difference between an arithmetic mean and a median did not exceed 10% for 21 out of 32 cases. Values of dispersion measures indicated a great scatter in the data for most cases. A relative standard deviation exceeded 40% for half of considered data sets and its lowest value equalled 16.95%. An absolute value of skewness exceeded 1 for 10 out of 32 cases, what suggests that most cases were characterized by a distribution close to symmetrical. An excess kurtosis exceeded 2 for 5 out of 32 cases, so it might be presumed that data distribution for these 5 cases would not occur normal.

The next step of the statistical analysis was assessment of data  $t$  sets normality on the basis of Kolmogorov–Smirnov tests – original and modified by Lilliefors as well as the Shapiro–Wilk test. A null hypothesis that a considered data distribution was normal, was rejected, if  $p < 0.05$  ( $p$  – probability level,  $p$ -value) in each test. The obtained results are shown in Table 1.

Table 1: Kind of data  $t$  set distribution.

| $\begin{matrix} A \\ (\text{cm}^2) \\ \backslash \\ H \\ (\text{m H}_2\text{O}) \end{matrix}$ | all | 4.71 | 9.42 | 15.07 | 18.84 |
|---|-----|------|------|-------|-------|
| all   | IR  | IR   | N    | IR    | IR    |
| 3.0   | IR  | N    | N    | N     | N     |
| 3.5   | N   | N    | N    | IR    | IR    |
| 4.0   | N   | N    | N    | IR    | N     |
| 4.5   | IR  | N    | N    | N     | N     |
| 5.0   | N   | N    | IR   | N     | N     |
| 5.5   | N   | N    | N    | N     | N     |
| 6.0   | N   | N    | N    | N     | N     |
| N – normal distribution, IR – irregular distribution  |     |      |      |       |       |

For sets of data  $t$  selected according to both pressure head in a pipe and leak area, 24 out of 28 were characterized by normal distribution. For data  $t$  selected according to no more than one parameter (pressure head or leak area), only half of sets had normal distribution. Thus, the next step of investigations – regression analysis, was conducted for sets of data selected according to two varied in laboratory tests parameters, excluding cases with irregular data distribution ( $H = 5.0$  m H<sub>2</sub>O and  $A = 9.42$  cm<sup>2</sup>,  $H = 3.5$  m H<sub>2</sub>O and  $A = 15.07$  cm<sup>2</sup>,  $H = 4.0$  m H<sub>2</sub>O and  $A = 15.07$  cm<sup>2</sup>,  $H = 3.5$  m H<sub>2</sub>O and  $A = 18.84$  cm<sup>2</sup>).

To determine an influence of hydraulic pressure head in a pipe on time  $t$  of water outflow on soil surface after a water pipe breakage, linear, exponential, logarithmic, and power regression models were used. Mean values of  $t$  tended to decrease with increasing pressure head for all analysed cases of leak areas and for all considered regression models (Figure 6, Table 2). For all but one ( $A = 15.07$  cm<sup>2</sup>) cases of leak areas, using the exponential model resulted in the highest value of a determination coefficient  $R^2$ , whereas the lowest  $R^2$  in these cases was

observed for the power model. For the case of  $A = 15.07 \text{ cm}^2$ , the best fitting was obtained by the logarithmic model and the purest by the exponential one. The coefficient of determination reached the lowest value for  $A = 9.42 \text{ cm}^2$  in every considered model, whereas values of  $R^2$  for  $A = 4.71 \text{ cm}^2$  and  $A = 18.84 \text{ cm}^2$  were comparable.

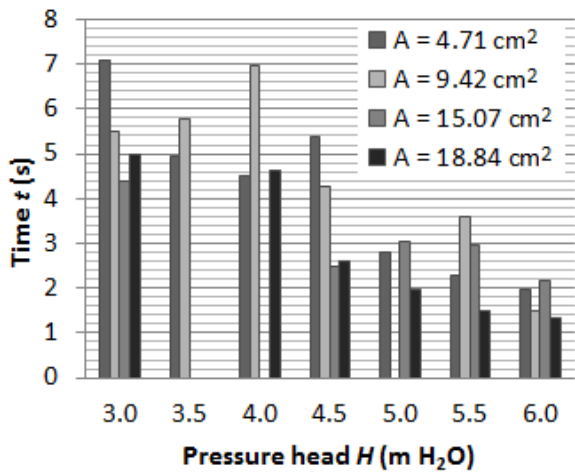


Figure 6: Arithmetic mean value of time  $t$  for different leak areas  $A$  and pressure head  $H$  in a water pipe.

Table 2: Equations of regression lines  $t(H)$ .

| Model type             | $A \text{ (cm}^2\text{)}$ | Regression equation         | $R^2$  |
|------------------------|---------------------------|-----------------------------|--------|
| Linear regression      | 4.71                      | $t = -0.7983H + 7.3309$     | 0.8522 |
|                        | 9.42                      | $t = -0.6979H + 7.2818$     | 0.7086 |
|                        | 15.07                     | $t = -0.3155H + 4.4541$     | 0.7374 |
|                        | 18.84                     | $t = -0.6978H + 5.8589$     | 0.8991 |
| Exponential regression | 4.71                      | $t = 8.6943e^{-0.209H}$     | 0.8814 |
|                        | 9.42                      | $t = 8.9281e^{-0.2H}$       | 0.7097 |
|                        | 15.07                     | $t = 4.5465e^{-0.097H}$     | 0.6989 |
|                        | 18.84                     | $t = 7.3256e^{-0.249H}$     | 0.9258 |
| Logarithmic regression | 4.71                      | $t = -2.49\ln(H) + 7.1698$  | 0.8330 |
|                        | 9.42                      | $t = -1.855\ln(H) + 6.7443$ | 0.4915 |
|                        | 15.07                     | $t = -0.974\ln(H) + 4.3149$ | 0.8098 |
|                        | 18.84                     | $t = -2.035\ln(H) + 5.492$  | 0.8172 |
| Power regression       | 4.71                      | $t = 8.019H^{-0.621}$       | 0.7799 |
|                        | 9.42                      | $t = 7.6476H^{-0.529}$      | 0.4904 |
|                        | 15.07                     | $t = 4.3245H^{-0.293}$      | 0.7392 |
|                        | 18.84                     | $t = 6.2132H^{-0.702}$      | 0.7830 |





The influence of the area of leak in a pipe on time  $t$  was analysed for  $H = 3.0$  m H<sub>2</sub>O,  $H = 4.5$  m H<sub>2</sub>O,  $H = 5.5$  m H<sub>2</sub>O and  $H = 6.0$  m H<sub>2</sub>O only, because for these pressure heads, all results of time measurements had been taken into consideration after laboratory results descriptive statistics and normality analysis (Figure 6). The obtained results suggested that bigger area of leak caused faster appearance of water on the soil surface after a failure. However, it is a suggestion only, because four cases of leak area are too small to evaluate goodness of fit of regression lines to  $t$  measurements results. Thus, on the present stage of the investigations there is no basis to claim that a regularity between time and the leak area occurred.

## 4 Summary and conclusions

Results of descriptive statistics of time of water outflow on the soil surface after a breakage of a buried water pipe suggested some derogations from the normality of data  $t$  set distribution. Nevertheless, the K-S, K-S-L and S-W tests validated the distribution normality of majority of data  $t$  set selected simultaneously according to two parameters varied during laboratory investigations – pressure head in a pipe and leak area. Thus, both parameters should be taken into account in further studies as significant and influential. According to common rules, an arithmetic mean of  $t$  measurements was taken as an average value, necessary in further studies, for these variants of laboratory investigation, which were characterized by the normal distribution.

Moreover, investigations outcomes indicated that higher pressure head in a pipe results in tendency of time of water outflow on the soil surface to be lower (as per expectations). However, although some regularity between time and the leak area in a damaged pipe was noticed during the analysis, there is not enough argument to prove the regularity occurrence. Thus, laboratory investigation should be continued for higher number of cases of leak area.

Results of the presented analysis of time between moment of a buried pipe failure and water outflow on the soil surface will be used in our further studies to verify effects of computer modelling of leakage from a damaged water pipe.

## Acknowledgement

This paper was financed by statutory activity of the Faculty of Environmental Engineering, Lublin University of Technology.

## References

- [1] Rossi F., Mirtchev V. Statistics for Food Scientists. Making Sense of the Numbers (Chapter 2). *Descriptive Statistics and Graphical Analysis*, Elsevier Inc. pp. 3-11, 2016.
- [2] Coultier V., Lefebvre R., Therrien R., Savard M.M. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical



- evolution of groundwater in a sedimentary rock aquifer system. *Journal of Hydrology*, 353(3), pp. 294-313, 2008.
- [3] Chłopek Z., Pawlicki M., Sypowicz R. Analiza statystyczna natężeń emisji zanieczyszczeń z silnika spalinowego w warunkach symulujących użytkowanie dynamiczne. *Archiwum Motoryzacji*, 1, pp. 77-87, 2005 (in Polish, with English abstract).
  - [4] Piotrkowski P. Analiza statystyczna danych do prognozowania ultrakrótkoterminowego produkcji energii elektrycznej w systemach fotowoltaicznych. *Przegląd Elektrotechniczny*, 90(4), pp. 1-4, 2014 (in Polish, with English abstract).
  - [5] Scott D.M. Statistics, Inferential. *International Encyclopedia of Human Geography*, Elsevier Inc., pp. 429-435, 2009.
  - [6] de Smith M.J. Statistical Analysis Handbook – a web-based statistics resource, The Winchelsea Press, Winchelsea, UK, pp. 12.10-12.17, 2015.
  - [7] Puust R., Kapelan Z., Savic D.A., Koppel T. A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7(1), pp. 25-45, 2010.
  - [8] Li W., Ling W., Liu S., Zhao J., Liu R., Chen Q., Qiang Z., Qu J. Development of systems for detection, early warning and control of pipeline leakage in drinking water distribution: A case study. *Journal of Environmental Sciences*, 23(11), pp. 1816-1822, 2011.
  - [9] Lahlou Z.M. Leak Detection and Water Loss Control. *Tech Brief – A national drinking water clearinghouse fact sheet*, pp. 1-4, 2001.
  - [10] Islam M.S., Sadiq R., Rodriguez M.J., Francisque A., Najjaran H., Naser B., Hoorfar M. Evaluating leakage potential in water distribution systems: a fuzzy-based methodology. *Journal of Water Supply: Research and Technology-AQUA*, 61(4), pp. 240-252, 2012.
  - [11] Kutylowska M., Hotłoś H. Failure analysis of water supply system in the Polish city of Głogów. *Engineering Failure Analysis*, 41, pp. 23-29, 2014.
  - [12] Piechurski F.G. Przyczyny i skutki awarii uszkodzeń w sieci wodociągowej. *INSTAL*, 4(317), pp. 42-47, 2011 (in Polish, with English abstract).
  - [13] Berardi L., Glustolisi O., Kapelan Z., Savic D.A. Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics*, 10(2), pp. 113-126, 2008.
  - [14] Okeya I., Hutton C., Kapelan Z. Locating pipe bursts in a district metered area via online hydraulic modelling. *Procedia Engineering*, 119, pp.101-110, 2015.
  - [15] Pérez R., Cugueró M.A., Cugueró J., Sanz G. Accuracy assessment of leak localisation method depending on available measurements. *Procedia Engineering*, 70, pp. 1304-1313, 2014.
  - [16] Iwanek M., Kowalski D., Kwietniewski M. Model studies of a water outflow from an underground pipeline upon its failure. *Ochrona Środowiska*, 37(4), pp.13-17, 2015 (in Polish, with English abstract).
  - [17] Iwanek M., Kowalski D., Kowalska B., Hawryluk E., Kondraciuk K. Experimental investigation of zones of leakage from damaged water network pipes. *Urban Water II. WIT Transactions on the Built Environment*,

- (139), eds. C. A. Brebbia, S. Mambretti, WIT Press, Boston, pp. 257-268, 2014.
- [18] Iwanek M., Kowalska B., Hawryluk E., Kondraciuk K. Distance and time of water effluence on soil surface after failure of buried water pipe. Laboratory investigations and statistical analysis. *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, 18(2), pp. 278-284, 2016.

