Acta Mechanica Slovaca 15 (3): 52 - 61, 2011

* Corresponding author
*Phone +421 55 602 2214*
*E-mail address: miriam.andrejiova@tuke.sk*
(Miriam Andrejiova, RNDr., PhD.)

Article information
*Article history: AMS-Volume15-No.3-00118-11*
Received 23 March 2011
Accepted 26 May 2011

# Analysis of Water Pollution Indicators with the Use of Selected Statistical Methods

**Miriam Andrejiová*, Zuzana Kimáková, Dušan Knežo**

*Department of Industrial Engineering and Management, Faculty of Mechanical Engineering, Nemcovej 32, 042 00 Košice, Slovak Republic*

## BIOGRAPHICAL NOTES

RNDr. Miriam Andrejiová, PhD. graduated in 1997 at Faculty of Science at Pavol Jozef Šafárik University in Košice in field of mathematics and physics. She obtained her post-graduate degrees in „Theory of Physics Teaching" at Faculty of Science at University of Pavol Jozef Šafárik in Košice. Since 1998 she has worked as assistant and special assistant at Department of Applied Mathematics and Informatics at Faculty of Mechanical Engineering at Technical University in Košice.

RNDr. Zuzana Kimáková, PhD. graduated in 1985 at Faculty of Science at Pavol Jozef Šafárik University in Košice in field of mathematics and physics. She obtained her post-graduate degree in „Environmental Studies" at Faculty of Mechanical Engineering at Technical University in Košice. Since 1987 she has been working as an assistant and specialized assistant at Department of Applied Mathematics and Informatics at Faculty of Mechanical Engineering at Technical University in Košice.

prof. RNDr. Dušan Knežo, CSc. is head of the Department of Applied Mathematics. He graduated at the Faculty of Natural Sciences, University of P.J. Šafarik in Košice. He received the PhD.-degree at the Mathematic-Physics Faculty, University of J.A. Comenius in Bratislava, and he was habilitated at the Faculty of Mechanical Engineering, University of Žilina, in the scientific branch of applied mathematics. His scientific focus is oriented to the modelling of processes in the biomedical engineering, to the research of materials, as well as to the measurement and diagnostic equipment, prognostics and evaluation of medical proceed. He is a co-author of one monograph, one academic textbook and author of several textbooks. He has published more than 60 papers in the scientific journals and in the proceedings of the conferences and he has more than 80 quotations concerning his professional works. He was also incorporated in various grant research projects and industrial projects.

## KEY WORDS

Waste water, pollution indicators, ANOVA, Kruskal-Wallis test, methods of multiple comparisons

## ABSTRACT

The need to protect the environment from pollution is still one of the most important issues in modern world. It is generally accepted that the environment covers a broad scale of elements; including air, water, soil, flora and fauna, as well as human health and security, and these are to be protected which is a part of global goal to ensure permanent and sustainable development.

Water is the basic substance and condition of human life on Earth. Until recently it

has been possible to satisfy humans´ requirements for water without any greater difficulties and that is thanks to regenerative and self-cleaning ability of water. The development of industries, agriculture, increasing number of human population and constant increase of life standards caused the pollution and decrease of water resources.

The main goal is to point out possibilities of use of selected statistical methods for analysis and evaluation of waste water pollution indicators. While evaluating data of indicators from Sokoľansky Creek are to be used. The data are withdrawn from monthly reports covering the range of years from 2007 until 2011.

## 1. Introduction

Water is substance responsible for life on Earth and it is often referred to as „coal oil of 21st century". A man constantly puts pressure on nature and man´s water demands are increasing. On the other hand, water that has been for centuries used solely for drinking and watering is now polluted with wastes and is quickly becoming a place for disposal of various wastes.

For water to continue to fulfill its important role in nature and in human society, the protection of water has to be a primary goal of each country´s politics. For this reason, environmental protection, protection of water sources and human health introduce new and crucial requirements in fields of agriculture and in water handling [4].

Draining waste water and polluting from point sources has a decreasing tendency since 1990. This decrease relates to higher price of water, more strict limitations of amount of emissions wasted into recipients, technological changes in production, more strict requirements for installing water cleaners for wasted waters and increased demand for their technical securing and changes in cleaning technology. According to Law about waters No. 384/2009 *waste water is water used in dwelling, manufacturing, agricultural, sanitarian and other fabrics and facilities or in means of conveyance, if it changes its quality after being used (content or temperature including), as well as percolated water from storage piles and settling pits.*

## 2. Material and Methods

### *Waste water cleaner*

Waste water cleaner (WWC hereafter) of U. S. Steel Košice, s. r. o. in Sokoľany was built and started working in 1991 with the view of final cleaning of all waste water coming from this company area. Waste water which is produced by U.S. Steel Košice, s. r. o. company is a mixture of surface drains, sewage, industrial waste water and permeated water from under dry waste dump [5].

WWC Sokoľany cleans waste water with the help of physical-mechanical processes which are brought by single canalization system. Approximately 88 percent of cleaned waste water is released to Sokoľansky Creek and the rest of water returns to production process as industrial water after filtration. All operations, factories and other businesses are obliged to follow rules of U.S. Steel Košice, s.r.o regarding this canalization while emitting waste waters. These rules specify limited degree for emitting waste waters into single canalization system. Cleaned waste water is carried away by single tube to emitting apparatus. At the place of emission a concrete dry valley is situated through which water is emitted and on which there is a measuring device installed.

Control of quality is ensured by physical and chemical laboratory, which [5]

- *controls quality of flowing water during emission,*
- *controls water quality at its in pour into fining reactors,*
- *controls filtrated water quality which is returning to the system of company,*
- *controls quality of water emitted into recipients.*

Place and time of taking samples, multitude of samplings, manner and control of individual indicators together with limited degrees (Tab. 1) are manifested in the decision of the court No. 2007/00019 of Regional Environmental Office in Košice from day 18.01.2007 [9].

### *Statistical methods*

Statistics has an irreplaceable place in modern society. It is used while analyzing various phenomena - not only in science and research. Before contemplating actual analysis it is suitable to determine basic number characteristics of selected files through descriptive statistics. Research analysis which uses graphical methods mostly (histogram, box plot, Q-Q graph etc.) allows us to consider statistical peculiarities of measured data.

Dependency between individual indicators can be described through correlation coefficient.

Table 1: *Approved concentration values of selected indicators. Note: p – average, m – maximal approved concentration values.*

| Indicator | Concentration value (mg/l) | | Indicator | Concentration value (mg/l) | |
|---|---|---|---|---|---|
| | p | m | | p | m |
| **Chemical Oxygen Demand (COD)** | 1,5 | 1,5 | **Soluble Substances (105°C)** | 900 | 1000 |
| **Chlorides** | 250 | 300 | **Soluble Substances (550°C)** | 740 | 800 |
| **Insoluble Substances** | 35 | 40 | **Sulfates** | 200 | 250 |
| **$N\text{-}NH4^+$** | 2,0 | 3,5 | **Total Cyanides** | 0,1 | 0,2 |
| **pH** | 6,9-9,0 | 9,0 | **Total Iron** | 2,7 | 3,0 |

Absolute value of correlation coefficient which is higher than 0,7 is interpreted as very high correlation (dependency); value from 0,5 to 0,7 is interpreted as high correlation; value from 0,3 to 0,5 is mild; and value which is less than 0,3 is trivial correlation.

ANOVA (ANalysis Of VAriance) is a method which enables the comparison of average values of various independent basic files. Its objective is to detect whether differences of average values of independent files are statistically important or only accidental. According to number of researched factors we distinguish analysis of diffusion as one-way ANOVA in which a single factor´s impact is monitored, and multi-way ANOVA - in which impact of multiple factors on quantitative variable is monitored. [1,2]

According to the extent of selective files we distinguish between balanced model (extent of selective files is the same) and unbalanced model (extent of selective files differs). To basic assumptions of diffuse analysis use belong; independency (selections are independent), normality (selective files come from basic files with normal distribution), homoscedasticity (homogeneity of variances) [6].

While testing statistical hypothesis with help of statistical programs it is often decided about dismissal or admission of null hypothesis with the use of p-value (significance level) which represents the lowest possible degree of importance based on which the null hypothesis can be rejected. For decision about acceptance or rejection of the null hypothesis a simple rule is in effect: if $p<\alpha$, then the null the is rejected on degree of importance $\alpha$; and if $p\geq\alpha$, then the null hypothesis is not rejected. [3] Rejection of null hypothesis which validates concord of mediate values offers information that in group significant differences in mediate values exist. Usually, we are interested also in information between which two files statistically important differences exist. For that information methods of multiple comparisons or so-called post hoc tests are used. To most common methods belong Scheffe method, Tukey method, Modified LSD method, etc. [6]

If assumptions about normality of distribution are not fulfilled then ANOVA will not be laid out. For comparison of multiple independent basic files Kruskal-Wallis test can be used. This test is non-parametric version of one-way ANOVA which does not require fulfillment of assumptions about form of distribution of basic files.

Similarly as one-way ANOVA, Kruskal-Wallis test determines only whether compared files are significantly different or not. It cannot be determined which two files are different- for that methods of multiple comparisons are used. When selections have the same latitude we can use Nemenyi method. When latitudes differ we can use general method of multiple comparisons. [1]

Statistical elaboration and evaluation of indicators of pollution was held by R package and EXCEL. R package is a language and integrated environment for data analysis, mathematical calculations and graphical data elaboration and depiction in one. It is free distributed tool which comes from similar language-S. It belongs to programs with open source code - Open Source Software and its actual version can be downloaded directly from homepage http://www.r-project.org/.

## 3. Results and Discussion

We have average monthly values to our disposal (in mg/l) for period from January 2007 until June 2011 with indicators: Chemical Oxygen Demand (COD), Chlorides, Insoluble Substances (105°C), $N\text{-}NH_4^+$, pH, Soluble Substances (105°C), Soluble Substances (550°C), Sulfates, Total Cyanides and

Total Iron. Used data are set from 24hours alloy samples taken from Sokoľansky Creek recipient. Because since 2009 Biological Oxygen Demand (BOD) is not tracked, this indicator will not have an importance in this research. [7]

### 3.1 Descriptive statistics

Basic average numeral characteristics of individual indicators for tracked periods are shown in the following chart (Tab. 2). All measured values fulfill li-mited values of indicators which are specified in the Decision no. 2007/00019.

Graphical description of data with use of box-plots gives an image about distribution of values, about variance, symmetry and it enables us to identify eventual extreme values. Horizontal line inside of the rectangle matches median. (Fig. 1 - Fig. 10).

Median, mean value and variance of first tracked indicator of waste water pollution (COD) have decreasing tendency. In 2010 indicator Chlorides had the lowest values of median, mean and variance.

Table 2: *Descriptive statistics. COD =Chemical Oxygen Demand, Chlor. = Chlorides, IS= Insoluble Substances, SS105= Soluble Substances (105°C), SS550= Soluble Substances (550°C), Sulf. = Sulfates, TC = Total Cyanides, TI = Total Iron, SD – Standard deviation.*

| Year | | COD | Chlor. | IS | $N\text{-}NH_4^+$ | pH | SS105 | SS550 | Sulf. | TC | TI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2007 | Mean | 19,83 | 172,75 | 8,25 | 0,31 | 7,80 | 750,7 | 601,8 | 150,9 | 0,03 | 0,36 |
| | SD | 3,22 | 21,13 | 1,49 | 0,12 | 0,07 | 59,56 | 44,35 | 10,2 | 0,01 | 0,12 |
| 2008 | Mean | 18,25 | 186,33 | 9,58 | 0,15 | 7,71 | 743,7 | 573,8 | 166,1 | 0,03 | 0,34 |
| | SD | 2,01 | 25,88 | 2,07 | 0,07 | 0,14 | 76,77 | 71,5 | 16,5 | 0,01 | 0,17 |
| 2009 | Mean | 17,95 | 167,67 | 7,70 | 0,18 | 7,71 | 714,3 | 568,2 | 141,2 | 0,05 | 0,88 |
| | SD | 2,98 | 18,06 | 0,92 | 0,11 | 0,14 | 44,29 | 39,0 | 15,4 | 0,01 | 0,22 |
| 2010 | Mean | 15,82 | 157,42 | 9,81 | 0,23 | 7,90 | 754,9 | 620,0 | 171,0 | 0,05 | 0,78 |
| | SD | 1,40 | 19,91 | 1,61 | 0,09 | 0,11 | 49,02 | 59,0 | 16,7 | 0,02 | 0,15 |
| 2011 | Mean | 15,33 | 168,50 | 7,50 | 0,23 | 07,90 | 769,5 | 634,0 | 168,7 | 0,03 | 0,63 |
| | SD | 1,75 | 23,23 | 1,05 | 0,05 | 0,11 | 78,85 | 70,1 | 13,6 | 0,01 | 0,10 |

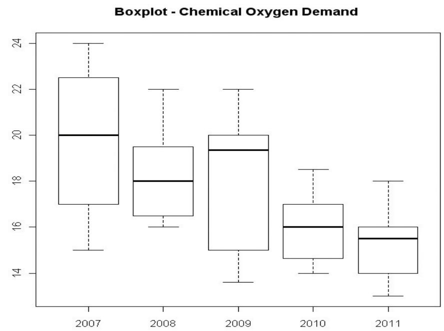


Fig. 1: *Boxplot – Insoluble COD.*

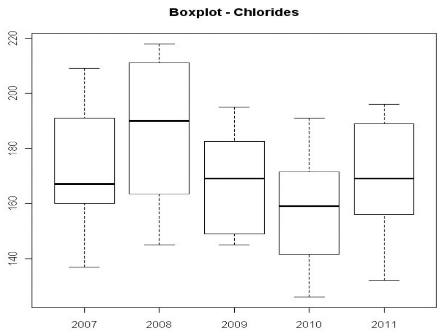


Fig. 2: *Boxplot - Chlorides.*

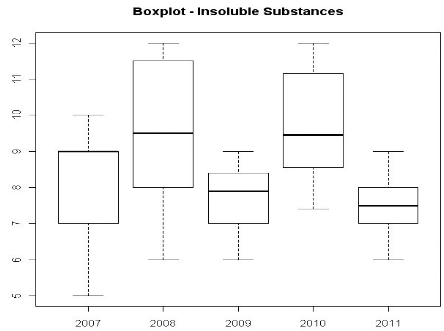


Fig. 3: *Boxplot – Insoluble Substances.*

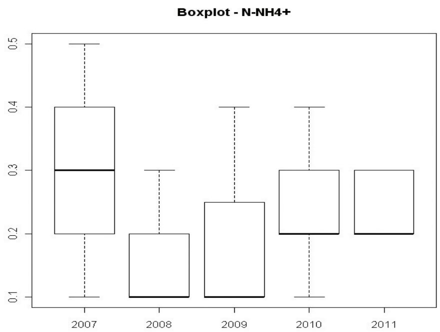


Fig. 4: *Boxplot - $N\text{-}NH_4^+$.*

Mean value of pH decreased in 2008 while for this year a largest dispersal of values is characteristic. Amount of Soluble Substances (550°C) is approximately constant during tracked period (Fig. 7). On the other side, content of disulphide's in 2009 decreased visibly and next year it has reached its largest mean value (Fig. 8).

From box-plot (Fig. 9) it follows that in 2009 content of Total Cyanides indicator in waste waters has increased. Its amount is lowly decreasing. The same applies to Total Iron indicator (Fig. 10).

Dependency between individual indicators in 2007 is depicted with the use of correlation matrix (Tab. 3). In 2007 a very high positive dependency existed between indicators SS550 and SS105 (r=0,939), SS550 a Chlorides (r=0,827), SS105 a Chlorides (r=0,855), SS550 a Sulfates (r=0,715), SS105 a Sulfates (r=0,756).

Dependency of indicators in following years is depicted in Tab. 4.

In a following analysis we will examine an influence of seasons on content of selected indicators in waste waters in Sokoľansky Creek recipient. At first, we will verify requirements for ANOVA use: homogeneity of variance and normality of basic files.

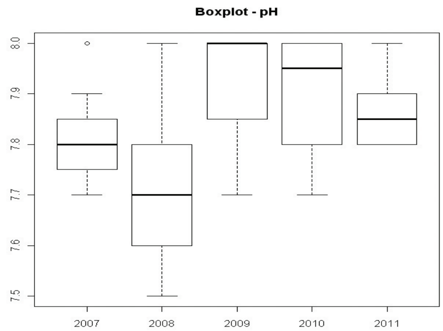


Fig. 5: *Boxplot – pH.*

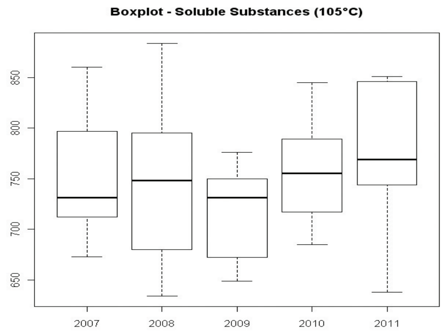


Fig. 6: *Boxplot - Soluble Substances (105°C).*

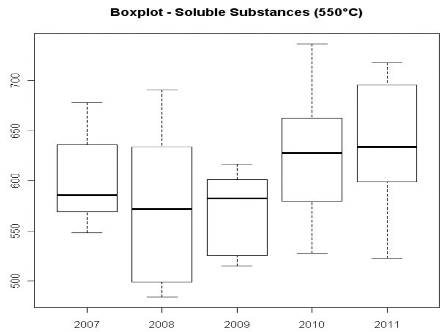


Fig. 7: *Boxplot – Soluble Substances (550°C).*

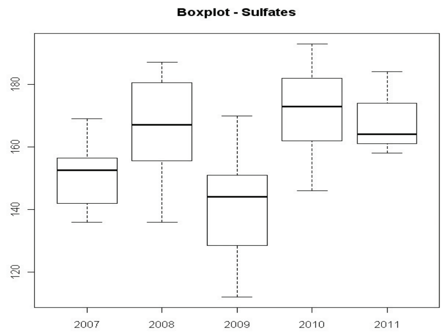


Fig. 8: *Boxplot - Sulfates.*

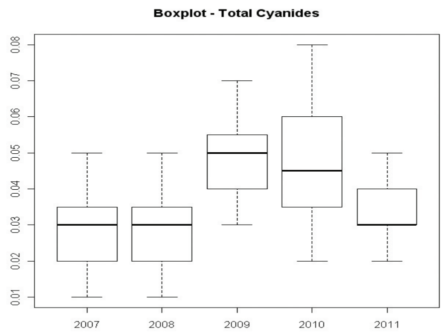


Fig. 9: *Boxplot – Total Cyanides.*

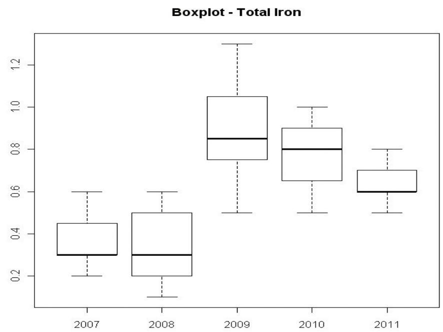


Fig. 10: *Boxplot – Total Iron.*

Table 3: *Correlation matrix – year 2007. TC =Total Cyanides, TI – Total Iron, IS = Insoluble Substances, SS = Soluble Substances.*

| 2007 | COD | Chlorid. | IS | $N\text{-}NH_4^+$ | pH | SS105 | SS550 | Sulfates | TC | TI |
|---|---|---|---|---|---|---|---|---|---|---|
| **COD** | 1 | | | | | | | | | |
| **Chlorid.** | -0,200 | 1 | | | | | | | | |
| **IS** | 0,581 | 0,063 | 1 | | | | | | | |
| **$N\text{-}NH_4^+$** | 0,141 | 0,070 | 0,481 | 1 | | | | | | |
| **pH** | 0,162 | 0,049 | 0,323 | 0,482 | 1 | | | | | |
| **SS105** | -0,176 | **0,855** | 0,046 | -0,243 | -0,164 | 1 | | | | |
| **SS550** | 0,074 | **0,827** | 0,224 | -0,124 | -0,152 | **0,939** | 1 | | | |
| **Sulfates** | -0,503 | 0,671 | -0,257 | -0,165 | -0,415 | **0,756** | **0,715** | 1 | | |
| **TC** | 0,296 | -0,299 | 0,467 | -0,055 | 0,196 | -0,290 | -0,196 | -0,489 | 1 | |
| **TI** | -0,133 | 0,266 | -0,284 | -0,389 | 0,115 | 0,224 | 0,075 | -0,017 | -0,055 | 1 |

Table 4: *Very high and high dependency of indicators in period from 2008 until 2011. IS = Insoluble Substances, SS105= Soluble Substances (105°C), SS550= Soluble Substances (550°C).*

| Year | Very high dependency | High dependency |
|---|---|---|
| **2008** | SS550 and $N\text{-}NH_4^+$ (r=0,828),<br>SS105 and $N\text{-}NH_4^+$ (r=0,736),<br>SS105 and Chlorides (r=0,723) | Sulfates and Chlorides (r=0,6025),<br>Total Iron and $N\text{-}NH_4^+$ (r=0,585),<br>Total Cyanides and Chlorides (r=0,582),<br>pH and Chlorides (r=-0,519),<br>pH and Sulfates (r=-0,504) |
| **2009** | Chlorides and COD (r=0,825),<br>SS105 and SS (550°C) (r=0,985),<br>SS105 and Chlorides (r=0,907),<br>SS550 and Chlorides (r=0,881) | IS and COD (r=-0,573),<br>$N\text{-}NH_4^+$ and Total Cyanides (r=0,580),<br>pH and Chlorides (r=-0,56),<br>pH and IS (r=0,536),<br>SS105 and Sulfates (r=0,535),<br>SS550 and Sulfates (r=0,57) |
| **2010** | SS105 and Total Cyanides (r=0,763),<br>Total Cyanides and Chlorides (r=0,743),<br>SS105 and SS550 (r=0,979),<br>SS105 and Chlorides (r=0,777) | $N\text{-}NH_4^+$ and Total Iron (r=-0,697),<br>$N\text{-}NH_4^+$ and IS (r=-0,556),<br>SS550 and Chlorides (r=0,696),<br>SS550 and Sulfates (r=0,609) |
| **2011** | IS and Total Iron (r=0,739),<br>IS and Chlorides (r=0,702),<br>IS and $N\text{-}NH_4^+$ (r=0,739),<br>Sulfates and pH (r=-0,735),<br>Sulfates and Total Cyanides (r=-0,735),<br>SS105 and SS550 (r=0,99),<br>SS105 and Chlorides (r=0,991),<br>SS550 and Chlorides (r=0,999) | Chlorides and Total Iron (r=0,6),<br>Sulfates and Total Iron (r=-0,679),<br>SS105 and Total Iron (r=0,621),<br>SS550 and Total Iron (r=0,597),<br>SS105 and IS (r=0,671). |

### 3.2 Verification of homogeneity of variance

Homogeneity of variance is verified by Bartlett test. We are testing the null hypothesis $H_0$: $\sigma^2_{2007} = \sigma^2_{2008} = \sigma^2_{2009} = \sigma^2_{2010} = \sigma^2_{2011}$ against $H_1$ non $H_0$. P-values for individual parameters are shown in Tab. 5. From results follows that the null hypothesis about dispersal equality is not dismissed for any tracked parameter.

### 3.3 Normality verification

To verify normality we use e.g. Shapiro-Wilk test for normality which is most commonly used in cases of small or average extent of values to value of 2000. At the level of importance $\alpha$ = 0,05 we are testing the null hypothesis accidental selection comes from basic file with normal distribution, against the hypothesis accidental selection comes from basic file with distribution different from normal. Tab. 6 shows calculated p-values for individual indicators in every year.

From these results follow that files of almost every indicator, except for indicators pH and $N\text{-}NH_4^+$(pH: year 2007, 2009, 2010, $N\text{-}NH_4^+$: 2008, 2009, 2011), fulfill normality requirements. In those two cases we cannot use analysis of variance, therefore non-parametrical Kruskal-Wallis test is used instead which does not require fulfillment of normality of basic files.

### 3.4 ANOVA

8 of 10 indicators fulfill requirements for analysis of variance. By these indicators we will examine dependency of seasons on concentration of selected parameter in waste water with use of one-way ANOVA. Examined period is represented by factor A which influence on content of concentration of pollution indicators we examine and which covers years 2007, 2008, 2009, 2010, 2011 hat make up for 5 levels of factor A. At the level of importance $\alpha = 0{,}05$ we are testing the null hypothesis $H_0$: $m_{2007} = m_{2008} = m_{2009} = m_{2010} = m_{2011}$ against the alternative hypothesis $H_1$: at least one of equalities is not fulfilled. Final ANOVA table for Chlorides is in Tab. 7.

Table 5: *Results of Bartlett test ($\alpha$=0,05).*

| Indicator | p-value | Conclusion for null hypothesis |
|---|---|---|
| COD | 0,0578>$\alpha$ | Not rejected |
| Chlorides | 0,8133>$\alpha$ | Not rejected |
| Insoluble Substances | 0,1106>$\alpha$ | Not rejected |
| $N\text{-}NH_4^+$ | 0,1542>$\alpha$ | Not rejected |
| pH | 0,2338>$\alpha$ | Not rejected |
| Soluble Substances (105°C) | 0,3189>$\alpha$ | Not rejected |
| Soluble Substances (550°C) | 0,2604>$\alpha$ | Not rejected |
| Sulfates | 0,5886>$\alpha$ | Not rejected |
| Total Cyanides | 0,3130>$\alpha$ | Not rejected |
| Total iron | 0,2839>$\alpha$ | Not rejected |

Table 6: *Shapiro - Wilk test for normality ($\alpha$=0,05).*

| Indicator | p-value | | | | | Conclusion for null hypothesis |
|---|---|---|---|---|---|---|
| | 2007 | 2008 | 2009 | 2010 | 2011 | |
| COD | 0,276 | 0,305 | 0,086 | 0,430 | 0,918 | Not rejected for any period |
| Chlorides | 0,499 | 0,256 | 0,270 | 0,824 | 0,884 | Not rejected for any period |
| Insoluble Substances | 0,083 | 0,304 | 0,312 | 0,297 | 0,820 | Not rejected for any period |
| **$N\text{-}NH_4^+$** | 0,421 | **0,002** | **0,003** | 0,099 | **0,001** | **Rejected (2008, 2009,2011)** |
| **pH** | **0,020** | 0,370 | **0,003** | **0,006** | 0,091 | **Rejected (2007, 2009,2010)** |
| Soluble Subst. (105°C) | 0,533 | 0,802 | 0,175 | 0,851 | 0,479 | Not rejected for any period |
| Soluble Subst. (550°C) | 0,138 | 300 | 0,062 | 0,849 | 0,860 | Not rejected or any period |
| Sulfates | 0,740 | 0,461 | 0,922 | 0,515 | 0,424 | Not rejected for any period |
| Total Cyanides | 0,440 | 0,470 | 0,440 | 0,487 | 0,473 | Not rejected for any period |
| Total iron | 0,172 | 0,143 | 0,935 | 0,381 | 0,473 | Not rejected for any period |

Table 7: *ANOVA Summary Table for Chlorides.*

| Source of Variability | SS | Degree of freedom | MS | F | p-value | F krit |
|---|---|---|---|---|---|---|
| Between groups ($SS_A$) | $SS_A$=5239,87 | $df_A$ =4 | $MS_A$=1309,9 | 2,80 | 0,0359 | 2,5611 |
| Residual ($SS_R$) | $SS_R$=22924 | $df_R$ =49 | $MS_R$=467,84 | | | |
| Total ($SS_T$) | $SS_T$=28163,87 | $df_T$ =53 | | | | |

Table 8: *ANOVA results for examined indicators .*

| Indicator | p-value | Conclusion for null hypothesis | Eta factor |
|---|---|---|---|
| COD | 0.0008$<\alpha$ | Rejected | 31,59% |
| Chlorides | 0.0359$<\alpha$ | Rejected | 18,60% |
| Insoluble Substances | 0.0017$<\alpha$ | Rejected | 29,20% |
| **Soluble Substances (105°C)** | **0.3668>a** | **Not rejected** | **8,25%** |
| **Soluble Substances (550°C)** | **0.0604>a** | **Not rejected** | **16,53%** |
| Sulfates | 0.0001$<\alpha$ | Rejected | 41,43% |
| Total Cyanides | 0,00015$<\alpha$ | Rejected | 36,40% |
| Total Iron | $1.26.10^{-11}<\alpha$ | Rejected | 68,07% |

Table 9: *Kruskal-Wallis test (p –value).*

| Indicator | p-value | Conclusion for null hypothesis |
|---|---|---|
| pH | 0,2485$>\alpha$ | **Not rejected** |
| $N\text{-}NH_4^+$ | 0,0337$<\alpha$ | Rejected |

$SS_R$ represents variability inside groups (residual), $SS_A$ characterizes variability between groups (group) and $SS_T$ is total variability. Because p-value =0,0359 < $\alpha$, the null hypothesis is rejected. Similarly analysis of variance for other left indicators is held. Consequential p-values are shown in Tab. 8.

Because p-value for Soluble Substances (105°C) and Soluble Substances (550°C) is higher than the level of importance $\alpha$=0,05, the null hypothesis about equality of mean values of basic files is not rejected. From ANOVA follows that these two indicators do not have statistically important differences within individual years. Proportion of variability (factor effect) described by time is defined by coefficient (eta-squared)

$$\eta^2 = \frac{SS_A}{SS_T}.$$

It means that COD value in waste water depends from 31,59% on in which season it has been measured and from 68,41% it depends on other agents.

### 3.5 Kruskal- Wallis test

Indicators of pH and $N\text{-}NH_4^+$ do not fulfill normality requirement every year. Therefore we use non-parametric version of one-way ANOVA: Kruskal-Wallis test. In both cases we test the null hypothesis $H_0$: $m_{2007} = m_{2008} = m_{2009} = m_{2010} = m_{2011}$ against the alternative hypothesis $H_1$ at least one of equalities is not fulfilled. Final p-value results are shown in Tab. 9.

In case of pH indicator we can assume that mean values are equal. On the other hand, for $N\text{-}NH_4^+$ indicator the null hypothesis is rejected. There are at least two years when average values of $N\text{-}NH_4^+$ indicator are statistically of significant importance.

### 3.6 Post hoc tests

The null hypothesis about equality of mean values was rejected with use of one-way ANOVA for 6 indicators: COD, Chlorides, Insoluble Substances, Sulfates, Total Cyanides and Total Iron. In next step it needs to be decided which year pairs are significantly different from each other. Post hoc tests are used in this case. In our case we used Scheffe method, Tukey – Cramer and Modified LSD method.

In each case the null hypothesis is tested in which we consider differences of mean between year pairs as insignificant against the hypothesis that statistically significant differences between two given average values exist. The null hypothesis is tested: $H_0$: $m_i = m_j$ against $H_1$: $m_i \neq m_j$ for each pair $i, j=1, 2, \dots, k, i\neq j$, ($\alpha$=0,05). By each indicator then hypothesis are tested gradually.

According to Scheffe and Tukey-Cramer method are differences between pairs [2007, 2010] and [2007, 2011] statistically significant for indicators COD. Modificated LSD method is more sensitive and it considers pairs [2008, 2010], [2008, 2011], [2009, 2010], [2009, 2011] as also statistically significant in terms of difference. Results of post hoc tests for each indicator are shown in Tab. 10.

It is similar in case of $N\text{-}NH_4^+$ pollution indicator. With Kruskal-Wallis test we found out whether compared files are significantly different or not. With the use of general methods of multiple comparisons we found out that years [2007, 2008] are statistically different.

### 3.7 Simple statistical methods of comparisons

Our next goal was arrangement of individual months from January 2007 until June 2011 according to degree of waste water pollution in Sokoľansky Creek recipient. Degree of pollution has been specified with the help of two methods of multiple comparisons - method of standard variable and method of fictive object [8]. Each indicator has been consi-dered as de-stimulating variable (positive growth is its decrease). Final arrangement of 5 months with highest and 5 months with lowest degree of waste water pollution is shown in Tab. 11. Results acquired from use of these methods are comparable. Both methods show August 2008 and September 2008 as months with highest degree of water pollution. Similarly, these methods are in accord when indicating months with lowest degree of pollution (March 2007, February 2010).

Table 10: *Results of post hoc tests for each indicator of pollution .*

| Indicator | Method | | |
|---|---|---|---|
| | Scheffe | Tukey-Cramer | Modificated LSD |
| COD | [2007, 2010], [ 2007, 2011] | [2007, 2010], [ 2007, 2011] | [2007, 2010], [ 2007, 2011], [2008, 2010], [2008, 2011], [2009, 2010], [2009, 2011] |
| Chlorides | [2008, 2010] | [2008, 2010] | [2008, 2009], [2008, 2010] |
| IS | [2009, 2010] | [2008, 2009], [2008, 2011], [2009, 2010], [2010, 2011] | [2007, 2008], [2007, 2010], [2008, 2009], [2008, 2011], [2009, 2010], [2010, 2011] |
| Sulfates | [2007, 2010] | [2007, 2010], [2008, 2009], [2009, 2010] | [2007, 2008], [2007, 2010], [ 2008, 2009], [2009, 2010], [ 2009, 2011] |
| Total Cyanides | [2007, 2009], [2008, 2009], [2007, 2010], [2008, 2010] | [2007, 2009], [2008, 2009], [2007, 2010], [2008, 2010] | [2007, 2009], [2007, 2010], [2008, 2009], [2008, 2010], [2009, 2011], [2010, 2011] |
| Total Iron | [2007, 2009], [2007, 2010], [2007, 2011], [2008, 2009], [2008, 2010], [2008, 2011], [2009, 2011] | [2007, 2009], [2007, 2010], [2007, 2011], [2008, 2009], [2008, 2010], [2008, 2011], [2009, 2011] | [2007, 2009], [2007, 2010], [2007, 2011], [2008, 2009], [2008, 2010], [2008, 2011], [2009, 2011] |

Table 11: *Arrangement according to degree of pollution.*

| Number | 48 | 47 | 46 | 45 | 44 |
|---|---|---|---|---|---|
| Method of standard variable | VIII.08 | IX.08 | VII.09 | VIII.07 | VIII.09 |
| Method of fictive object | VIII.08 | IX.08 | VIII.07 | VII.09 | VIII.09 |
| **Number** | **5** | **4** | **3** | **2** | **1** |
| Method of standard variable | I.10 | X.09 | IV.10 | III.07 | II.10 |
| Method of fictive object | III.11 | V.08 | I.10 | III.07 | II.10 |

## 4. Conclusion

In this article we treated evaluation of indicators of waste water pollution with help of selected statistical methods while focusing on period from year 2007 until year 2011. With the use of one-way ANOVA, Kruskal-Wallis test and following post-hoc tests we have shown that statistically significant differences between average values of selected indicators during tracked period exist (COD, Chlorides, Insoluble Substances, N-$NH_4^+$, Sulfates, Total Cyanides, Total Iron). Differences in average values during tracked period of indicators Soluble Substances (105°C), Soluble Substances (550°C) and pH are statistically not significant. With use of methods of multiple comparisons we discovered that August 2008, September 2008, July 2009, August 2007 and August 2009 belong to months with highest degree of waste water pollution. During years from 2007 until 2011 many indicators had shown decrease which relates mostly to limited values and more strict requirements for waste water cleaning, technological securing and cleaning technology.

## 5. Acknowledge

This paper was prepared with the support of grant project VEGA 1/0543/10.

## 6. References

[1] Anděl, J. (1998). Statistical methods (in Slovak). MATFYZ-PRESS, Praha

[2] Berthouex, P.M., Brown, L. C. (2002). Statistics for Environmental Engineers. Lewis Publishers, USA

[3] Hanke, J. E., Reitsch, A. G. (1991). Understanding business statistics. Irwin, Boston

[4] Králiková, R., Badida, M., Halász, J. Technology protection of environment (in Slovak). Elfa, Košice

[5] The handling procedure for WWC Sokoľany. Internal Documentation U.S.Steel Košice, s.r.o.

[6] Meloun, M., Militký, J. (2004). Statistical analysis of experimental data (in Czech). Academia, Praha

[7] Monthly report – Indexes of waste water contamination. Documentation U.S.Steel Košice, s.r.o. [online] [cit 2011-07-11], http: www.usske.sk/corpinfo/eco/

[8] Pažitná, M., Labudová, V. (2007). Methods of statistical comparisons (in Slovak). EKONÓM, Bratislava

[9] Decision of the Regional Environmental Office in Kosice on the authorization of discharges from WWC Sokoľany in Sokoľanského Creek (in Slovak)

[10] Patkó, I. - Juvancz, Z. - Túri, M. Szerletics - Lumnitzer, E. - Badodová, D. - Hricová, B.: Emission of dioxins in biomass using furnaces. In: Acta Mechanica Slovaca. Roč. 14, č. 1 (2010), s. 94-99. - ISSN 1335-2393

[11] LUMNITZER, E.: Research and development trends in the area of environmental protection at the Faculty of Mechanical Engineering . In: Acta Mechanica Slovaca. Roč. 14, č. 1 (2010), s. 4-5. - ISSN 1335-2393