Journal of Computing and Information Technology - CIT 20, 2012, 3, 209–215
doi:10.2498/cit.1002099

Analysis of Heavy Metals Concentration in Wastewater along Highways in Croatia

Dijana Grd[1], Jasminka Dobsa[1], Vesna Simunic-Meznaric[2] and Teuta Tompic[2]

[1]Faculty of Organization and Informatics, University of Zagreb, Varazdin, Croatia
[2]Laboratory of Ecology, Bioinstitut d.o.o., Cakovec, Croatia

In this paper we have analysed concetrations of heavy metals (lead, copper, nickel, zink, mercury, cadmium, and chromium) in wastewater along highways in Croatia. We have used standard statistical methods: analysis of variance, Kruskal-Wallis test and principal analysis. Analysis of variance and Kruskal-Wallis test were used to detect factors that influence the concentration of lead, copper, nickel, and zink in wastewater. We have investigated the influence of the highway sampling location, the side of a highway, and the influence of the season of the year. Principal components were used to identify groups of elements with similar characteristics in wastewater.

Keywords: wastewater, heavy metals, highway, ANOVA, Kruskal-Wallis test, principal components analysis

1. Introduction

As the awareness of the importance of environmental protection increases, determining the concentration of some elements that can be found in wastewater receives increasing interest. Wastewater treatment is a very important communal activity, both for environmental protection and for human health. Wastewater testing is performed to evaluate its quality, detect possible changes, adjust the measures of protection if necessary, and to evaluate the functionality of the drainage systems. There are various factors that affect the quality of wastewater, including the following: location, wind, rainfall, human impact. Therefore, in this study the critical elements will be evaluated. Harmful elements like copper, zinc, lead and nickel are commonly found in wastewater along the highways. These metals accumulate along the roads due to consumption of automotive parts (tires, breaks), combustion of fuel, road infrastructure (eg. road surface-layer fraying, corrosion of galvanized road buffers). Thus, zinc comes from tires which contain ZnO and galvanized parts of the car, while nickel comes from the abrasion of metals in vehicles and gasoline usage.

This study will evaluate the factors influencing the increasing occurence of certain metals in wastewater along the highways in Croatia. Such a research has not yet been conducted in Croatia.

Contamination of soil by heavy metals was investigated in [5], [6] and [7]. Contamination along a motorway was investigated in [5], in an industrial zone in Iran in [7], while an investigation along a highway in Slovenia is shown in [6]. Usage of multivariate techniques for analysis of water quality is often used. The technique of principal components in [1] and [4] was used for the determination of main factors which are a linear combination of standard variables used for the assessment of water quality.

The rest of the paper is organized in the following way: the second section describes the used statistical methods (analysis of variance, Kruskal-Wallis test and method of principal components), the third section describes data set on which the research has been conducted, the fourth section gives the results. The last section presents the conclusion and disscusion.

2. Methods

Analysis of variance, often abbreviated to ANOVA, is a broad group of techniques for identifying and measuring different sources of variation within the data set. It consists of a set of procedures by which a variance of the random variable is broken down by certain sources of variation of its value. With the components of variance, depending on the sources, one can conclude if there is a significant difference between the values of dependent variable for different levels of the observed factor variables [9].

There are certain assumptions for conducting a test of analysis of variance on random samples, and they are: samples selected from the basic sets are independent, the distribution of dependent variables in basic sets is normal and distribution of basic sets has equal variance [8].

In this paper, we will use a one-way analysis of variance which is used to compare two or more groups having different levels of one variable (level).

If the above-mentioned assumptions for ANOVA are not met, the Kruskal-Wallis test is usually performed. Kruskal-Wallis test is a nonparametric test which is used for determining whether three or more independent samples originate from the same population. When this test leads to significant results, at least one sample differs from the others.

A principal component analysis is a standard tool in modern data analysis used in various fields of science. It is a simple, nonparametric method for extracting relevant information out of confusing data sets. Principal component analysis is concerned with the interpretation of the variance and covariance structure of the original set of variables through a small number of their linear combinations. The general objectives of principal component analysis are data reduction and interpretation [3]. In order to reduce the number of variables, it is necessary to find a criterion by which we will discard components that are not as informative and keep the ones that contain most of the information from the initial variables. In this paper we will use the Kaiser criterion which says that we can retain only factors with eigenvalues greater than 1.

3. Data Set

In this paper the data about wastewater along the highway on the sections between Varaždin and Split were analyzed. Figure 1 shows these highway sections. They are: A4 highway between Zagreb and Varaždin, A3 highway between Ivanja Reka and Lučko and the A1 highway to Split.

Our data set consists of 1418 samples, which are described by 13 variables relevant to our study.

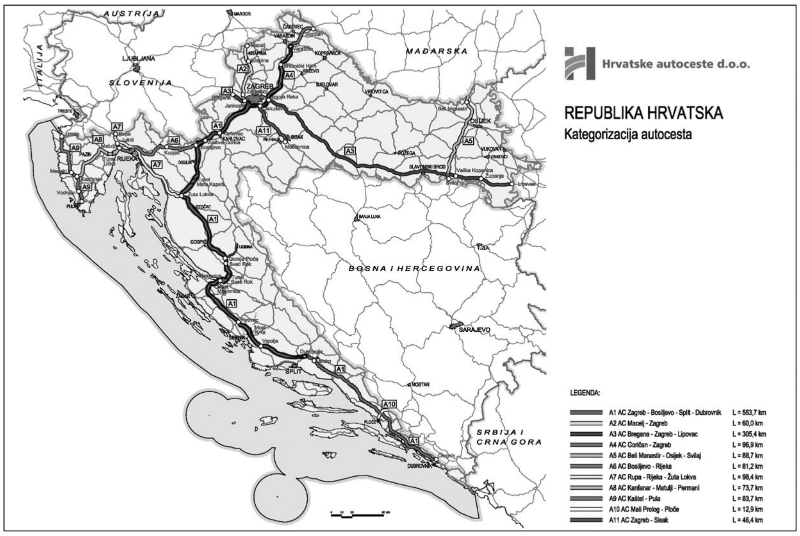

Figure 1. Analyzed highway sections [2].

For each sample we have variables that contain information about the time and the location of sampling.

More precisely, location information include the section of the highway where the sample was taken, and also the side of the road or the bridge, tunnel, viaduct or portal where it was taken. Variables containing information about time are: cycle when the sample was collected (there were five cycles) and the exact date of sampling. Samples were collected in the period between 2008 and 2010. Also, there are variables describing physical and chemical characteristics of water. For this research we have analysed variables containing information about the concentration values of copper, zinc, lead, and nickel in water samples. In cases where the values of observed variables are less than the minimum measurable value, the value is approximated to zero.

4. Results

In this paper we have analysed whether there are significant differences in the average element amount considering sampling location, side and the season in which the data was collected. This was conducted by the analysis of variance, while the relationship between different variables was explored by the method of principal components.

Experiments were conducted using the Statistica 10.0 software.

4.1. Analysis of Variance Considering the Sampling Location

One-way ANOVA was conducted to resolve the question whether there is a significant difference in the average amount of lead, nickel, zinc and copper, depending on the sampling location, i.e. whether there are differences depending on the specific part of the highway on which the sample was taken. We have 12 basic sets – sampling locations (Benkovac, Brinje, Ivanja Reka, Lučko, Maslenica, Ogulin, Perušić, Split, Sveti Rok, Šibenik, Varaždin and Zagvozd). Two variables were defined – the amount of element (dependent variable) and sampling location (grouping variable). The null hypothesis stated that there was no significant difference in the average amount of elements considering the sampling location. Accordingly, an alternative hypothesis argued that these differences exist. The defined level of significance is 0,05.

Firstly, the requirements for the application of ANOVA, normality of distribution and equality of variances, were tested. The assumption about distribution normality was satisfied, but Levene's test for equality of variances showed that this assumption was not justified. Therefore, we needed to further analyze the results obtained by ANOVA using a nonparametric method to ensure that the level of statistical significance (p) was not increased because of unfulfilled assumptions.

The results for lead, copper and nickel showed that there were significant differences in the average amount of these elements, depending on the sampling location, i.e. null hypothesis was rejected. Thus, Figure 2 shows the arithmetic mean of lead depending on the section.

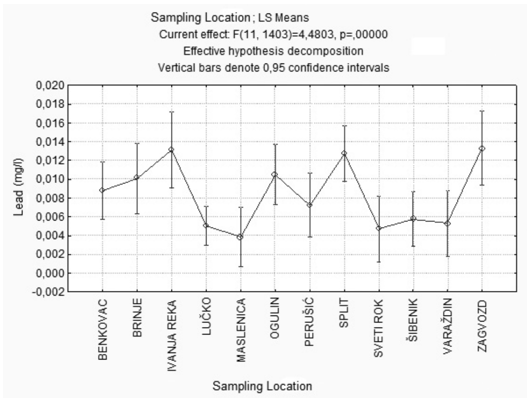

Figure 2. Lead mean value for monitored sampling locations.

Considering that ANOVA gives statistically significant results, Bonferroni's post-hoc test is conducted. In Table 1 we can see groups organized by arithmetic means, from the smallest to the largest. Arithmetic means of those that are not significantly different from each other have four stars in the same column. For example, group Maslenica significantly differs in amount of lead from Split, Ivanja Reka and Zagvozd.

Cell No.	Bonferroni test; variable Lead (mg/l) Homogenous Groups, alpha = ,05000 Error: Between MS = ,00030, df = 1403,0 Sampling Location	Lead (mg/l) Mean	1	2	3
5	MASLENICA	0,003812	****		
9	SVETI ROK	0,004726	****	****	
4	LUČKO	0,005034	****		
11	VARAŽDIN	0,005295	****	****	****
10	ŠIBENIK	0,005773	****	****	****
7	PERUŠIĆ	0,007245	****	****	****
1	BENKOVAC	0,008823	****	****	****
2	BRINJE	0,010108	****	****	****
6	OGULIN	0,010504	****	****	****
8	SPLIT	0,012733			****
3	IVANJA REKA	0,013129		****	****
12	ZAGVOZD	0,013307		****	****

Table 1. Bonferroni's test for lead.

The results for zinc, under the same conditions as in previous analyses, showed that there is no significant difference in the average amount of zinc depending on a sampling location.

Given that our assumptions were not met, we needed to conduct a nonparametric Kruskal-Wallis test. The results for lead are shown in Table 2, where we can see that p value is 0,00 which confirms that groups are significantly different.

Depend.: Lead (mg/l)	Kruskal-Wallis ANOVA by Ranks; Lead (mg/l) Independent (grouping) variable: Sampling Location Kruskal-Wallis test: H (11, N= 1415) =40,05366 p =,0000 Code	Valid N	Sum of Ranks	Mean Rank
BENKOVAC	101	124	87142,5	702,7621
BRINJE	102	83	59054,5	711,5000
IVANJA REKA	103	72	53897,0	748,5694
LUČKO	104	267	198896,5	744,9307
MASLENICA	105	117	74809,5	639,3974
OGULIN	106	113	74347,5	657,9425
PERUŠIĆ	107	102	65314,5	640,3382
SPLIT	108	131	109559,0	836,3282
SVETI ROK	109	95	54195,5	570,4789
ŠIBENIK	110	141	99461,0	705,3972
VARAŽDIN	111	95	66783,5	702,9842
ZAGVOZD	112	75	58359,0	778,1200

Table 2. Kruskal-Wallis test for the sampling location.

4.2. Analysis of Variance Considering Side of the Highway

Here we analyzed whether there is a significant difference in the average amount of lead, nickel, copper and zinc, depending on the side of the highway where a sample was taken (left or right). With this test we wanted to find out if there is a difference between northern and southern side of the highway, which would mean that wind and rainfall can affect the amount of elements in wastewater.

Again we had two hypotheses: null hypothesis stated that there is no significant difference in the average quantities of these elements depending on the side of the highway, and the alternative hypothesis stated the opposite. The test was performed for a significance level $\alpha = 0,05$.

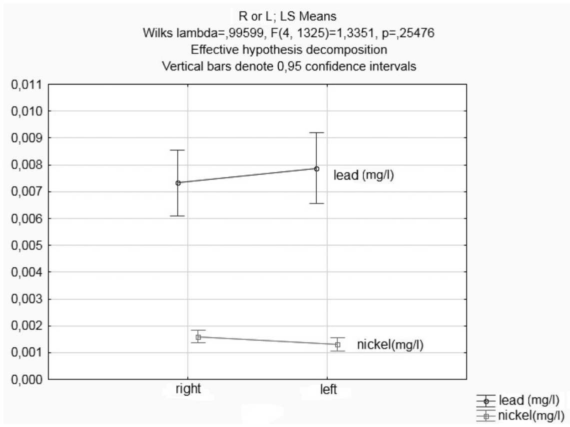

Figure 3. Lead and nickel mean values on each side of the highway.

Assumptions check gave the same results as in the previous analysis. The results indicated that the null hypothesis was accepted, which means that the premise of there being a difference in the average amount of elements depending on the side on which a sample was taken cannot be accepted. For example, in Figure 3, lead and nickel mean values for left and right side are shown. We can see that there is no big difference in those values.

Depend.: Lead (mg/l)	Kruskal-Wallis ANOVA by Ranks; Lead (mg/l) Independent (grouping) variable: R or L Kruskal-Wallis test: H (1, N= 1402) =,0901797 p =,7639 Code	Valid N	Sum of Ranks	Mean Rank
Right	101	755	527461,5	698,6245
Left	102	647	456041,5	704,8555

Table 3. Kruskal-Wallis test for the side of the road.

Again, because our assumptions were not met, we needed to conduct Kruskal-Wallis test. The results for lead are shown in Table 3, where we can see that p value is 0,7639, which confirms that groups are not significantly different.

4.3. Analysis of Variance Considering Season

This one-way ANOVA test resolves the question whether there is a significant difference in the average amount of lead, nickel, copper and zinc depending on the season in which samples were collected. Samples were collected in different months, so new season variable with four modes – winter, spring, summer and autumn, was defined. The null hypothesis stated that there is no significant difference in the average amount of elements considering the season, while the alternative hypothesis stated that significant differences exist. Again, a test was performed for a significance level $\alpha = 0,05$.

Firstly, the assumptions check gave the same results as in the previous analysis. After conducting the ANOVA test, the results showed a significant difference in the average amount of lead, copper and nickel, depending on the season in which samples were taken (null hypothesis rejected). Again, the test for zinc gave different results and showed no significant difference in the average amount of zinc, depending on the season in which samples were taken (null hypothesis accepted).

For example, mean values of lead per season are shown in Figure 4.

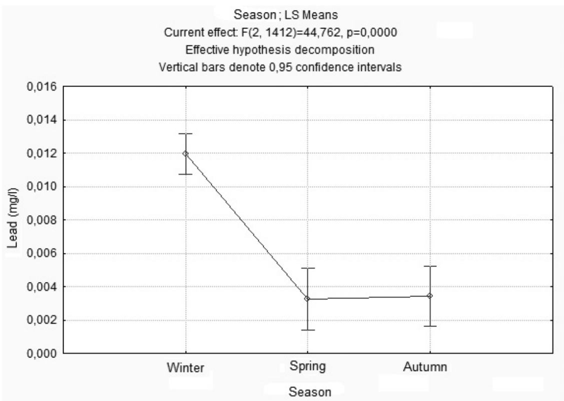

Figure 4. Mean values of lead per season.

We can see that in winter mean value of lead is higher than in spring or autumn, which also shows Bonferroni's post-hoc test shown in Table 4.

Cell No.	Bonferroni test; variable Lead (mg/l) Homogenous Groups, alpha = ,05000 Error: Between MS = ,00029, df = 1412,0 Season	Lead (mg/l) Mean	1	2
2	spring	0,003233	****	
3	autumn	0,003425	****	
1	winter	0,011943		****

Table 4. Bonferroni's post hoc test.

As in previous examples, we conducted a Kruskal-Wallis test. The results for lead are shown in Table 5, where we can see that p value is 0,00, which confirms that groups are significantly different. The highest amount of lead in winter time is ascribed to increased maintenance of the highway during that season. Lead is often an impurity in the technical sodium chloride, which is extensively used for sprinkling over slippery roads.

Depend.: Lead (mg/l)	Kruskal-Wallis ANOVA by Ranks; Lead (mg/l) Independent (grouping) variable: Season Kruskal-Wallis test: H (2, N= 1415) =63,44795 p =,0000 Code	Valid N	Sum of Ranks	Mean Rank
Winter	1	733	564496,0	770,1173
Spring	2	331	237748,5	718,2734
Autumn	3	351	199575,5	568,5912

Table 5. Kruskal-Wallis test for seasons.

4.4. Principal Components Analysis

The goal of principal components analysis is to reduce the number of variables and detect the structure in the relationship between them. Analysis for lead, nickel, copper, zinc, mercury, cadmium and chromium values was conducted.

Eigenvalues and associated indicators are shown in Table 6. Eigenvalues are the variances of principal components and their sum is 7, which shows us the cumulative number of these values in the third column. The second column shows the percentage of the total variance that each principal component explains, while in the last column we have a series of cumulative percentages of the second column.

Value number	Eigenvalues of correlation matrix, and related statistic			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	4,962017	70,88595	4,962017	70,8860
2	1,000183	14,28833	5,962200	85,1743
3	0,604912	8,64159	6,567111	93,8159
4	0,290696	4,15279	6,857807	97,9687
5	0,073844	1,05491	6,931651	99,0236
6	0,049245	0,70349	6,980895	99,7271
7	0,019105	0,27293	7,000000	100,0000

Table 6. Eigenvalues and associated indicators.

These results show us that the first principal component explains 70,89% of the total variance, and the second explains 14,29% of total variance. We can see how important the first principal component is in relation to the others. Using the Kaiser criterion, we take the first two principal components which explain 85,17% of total variance. In Figure 5 we can see that the first axis is mostly correlated with the variables mercury, lead, cadmium, chromium, nickel and copper, while the second axis is mostly correlated with the variable zinc.

This circuit shows how well each of the variables is represented by the current set of principal components. Given that all variables are close to the circle, representation of all variables is satisfactory.

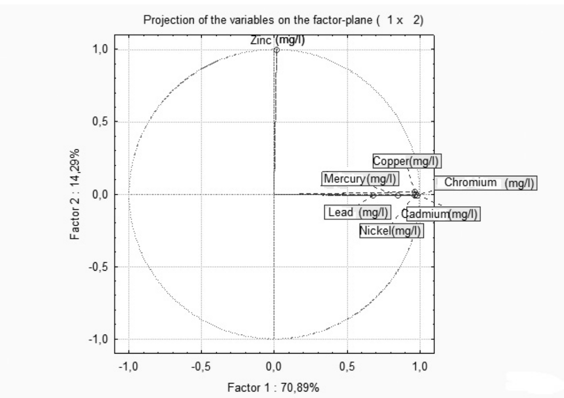

Figure 5. Variables in the components determined plane.

This test shows us which groups of elements have similar characteristics in wastewater. As we can see in Figure 5, zinc has different characteristics from other elements, which explains the results of the previous analysis.

5. Conclusion and Discussion

In this paper we have analyzed which factors affect amount values of the elements contained in wastewater along highways in Croatia. As we have seen, amounts of metals depend on the sampling location. For instance there is more lead near Zagreb and Split, which is expected because these are two big cities where traffic is heavier than in other locations. This is not true for zinc, which is present in all samples in larger quantities than the other metals. Zinc is present in all types of water, therefore it is not a true indicator of pollution. We also proved that there is no difference between left and right side of the highway, which means that there are no differences between northern and southern sides of the road. Thus, there is no influence of wind and rainfall. The results also show the dependence of metal amount on the season, which is a really interesting subject and can mean that there are differences due to traffic density in certain seasons. We are missing the data for the summer owing to the absence of rainfall.

References

[1] R. A. HEIKKA, Multivariate monitoring of water quality: a case study of Lake Simpele, Finland. *Journele of Chemometrics*, 22 (2008), 747-751.

[2] HIGHWAYS IN CROATIA (image online) `http://www.indalmazia.com/id/wp-content/uploads/Autostrade_croazia.jpg` [02/01/2012]

[3] R. A. JOHNSON, D. W. WICHERN, *Applied multivariate statistical analysis*. Prentice Hall, USA, 2007.

[4] H. JUAHIR, S. M. ZAIN, M. K. YUSOFF, T. I. HANIDZA, A. S. ARMI, M. E. TORIMAN, M. MOKHTAR, Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment*, 173 (2011), 625-641.

[5] P.-K. LEE, J.C. TOURAY, Characteristics of a Polluted Artificial Soil Located Along a Motorway and Effects of Acidification on the Leaching Behavior of Heavy Metals. *Water Research*, 32 (1998), 3425-3435.

[6] A. PLESNIČAR, N. ZUPANČIĆ, Heavy Metal Contamination of Roadside Soil Along Ljubljana – Obrežje Highway. *RMZ – Materials and Geoenvironment*, 52 (2005), 403-418.

[7] S. SHAKERI, F. MOORE, S. MODABBERI, Heavy Metal Contamination and Distribution in the Shiraz Industrial Complex Zone Soil, South Shiraz, Iran. *World Applied Sciences Journal*, 6 (2009), 413-425.

[8] I. ŠOŠIĆ, *Primijenjena statistika*. Školska knjiga, Zagreb, 2004.

[9] I. ŠOŠIĆ, V. SERDAR, *Uvod u statistiku*. Školska knjiga, Zagreb, 1997.

Received: June, 2012
Accepted: August, 2012

Contact addresses:

Dijana Grd
Faculty of organization and informatics
University of Zagreb
Pavlinska 2
42 000 Varazdin
Croatia
e-mail: dijana.grd@foi.hr

Jasminka Dobsa
Faculty of organization and informatics
University of Zagreb
Pavlinska 2
42 000 Varazdin
Croatia
e-mail: jasminka.dobsa@foi.hr

Vesna Simunic-Meznaric
Laboratory of Ecology
Bioinstitut d.o.o.
R. Steinera 7
40 000 Cakovec
Croatia
e-mail: vesna.meznaric@bioinstitut.hr

Teuta Tompic
Laboratory of Ecology
Bioinstitut d.o.o.
R. Steinera 7
40 000 Cakovec
Croatia
e-mail: teuta.tompic@bioinstitut.hr

DIJANA GRD is a graduate student of information and programming engineering at the Faculty of Organization and Informatics Varazdin, University of Zagreb, Croatia. She received her B.Sc. degree with honorary mention in information technology from the Faculty of Organization and Informatics in 2011. Throughout her education she has worked as a student assistant in various courses and volunteered in student organizations. She received an Award for Excellence in the category baccalaureate graduates and graduates from Rotary Club Varaždin for a research which included data analysis in the field of environmental protection.

JASMINKA DOBSA obtained her BSc and MSc degrees from the University of Zagreb, Department of Mathematics in 1995 and 1999, respectively. She obtained her PhD degree from the University of Zagreb, Faculty of Electrical Engineering and Computing in 2006. From 1995-1997 she worked as an assistant at the University of Zagreb, Faculty of Science and Mathematics, Department of Mathematics. From 1997 till 2007 she was working at the Faculty of Organization and Informatics as an assistant, and since 2007 she works as assistant professor at the same institution. She teaches courses related to statistics and data analysis at undergraduate and graduate levels. Her research interest includes statistics, data mining and text mining. In 2003 she spent 3 months at the Institute Jožef Stefan in Ljubljana, Department for Intelligent Systems, working on a research in the field of text mining. From 2005 till 2006 she led bilateral Croatian-Slovenian project *System for automatic classification of Croatian and Slovenian Web pages*. She published 9 scientific papers in journals of which 3 are indexed in CC or SCIE, 14 scientific papers published in conference proceedings and is coauthor of a textbook "Statistika (deskriptivna i inferencijalna) i vjerojatnost" published in 2008.

VESNA SIMUNIC-MEZNARIC obtained her B. Sc., M. Sc. and Ph. D. degrees in chemistry from the University of Zagreb, Faculty of Science and Mathematics, in 1996, 2000 and 2003, respectively. From 1996-2004 she was an assistant at the University of Zagreb, Faculty of Science and Mathematics, Department of Chemistry. Since 2004 she has been working at the Laboratory of Ecology, Bioinstitute, as quality manager and analyst in the field of water, waste and soil quality and pollution. Her professional interest includes accreditation and quality system in laboratories and good laboratory practice (GLP). She has participated in more than twenty international and Croatian conferences and she has eight published research and professional papers in journals and conference proceedings.

TEUTA TOMPIC, born Jurasic, was born on 8th of January 1975 in Ogulin, Republic of Croatia, where she completed elementary and high school (mathematics and physics department). In 1994 she enrolled in the study at Faculty of Food Technology and Biotechnology, Department of Food Engineering, from which she graduated in 1999, with excellent grade. Average grade of her study in general was 4,25. In 2000 she got employed in Podravka, in central physics and chemistry laboratory where she worked on gas chromatography instrumental technique. In 2004, she joined Bioinstitut Ltd. She works in the laboratory on gas chromatography, high performance liquid chromatography and atomic apsorption spectrophotometry. She gets professional training in the domain of food managing in analytical laboratory. In this company she also works as consultant for food safety. In 2006 she completed her education for EOQ Food Safety System Manager and got certificate from the European organisation for quality (EOQ). In 2011, she completed her PhD study at the Faculty of Food Technology and Biotechnology in Zagreb and earlier this year got promoted to the position of Head of the Laboratory of Ecology in Bioinstitut Ltd.

Works indexed according to SCI/CC:1

Works reviewed and published in Proceedings of international scientific and professional conferences (works presented like posters): 6

Professional and popularization works: 1