1. Introduction

The Faculty of Civil Engineering at the University of Zilina has been involved with the project, the goal of which is the monitoring of particulate matter creation resulting from the road traffic as well as its contamination in a direct relation to the surface of roads and the traffic load [1, 2, 3].

Heavy metals belong among the most basic groups of contaminants which are monitored in the various parts of the environment. The concerned group of contaminants is quite a large one, including the contaminants having a variable source of their origination and in many cases, the contaminants concerned share a different platform of professional approach from the scientific circles related to their impact and influence as related to the health of population. The subject of monitoring, pursuant to the general law [4] are the following elements: As, Cd, Hg, Pb and Ni. These are generally considered as the most harmful to people and animals. Their limits are governed by the Public Notice [5]. Even some of the other elements may be of dangerous nature, which is to be found in soils where they are necessary in small amounts; however, when accumulated in large quantities, they may have an even more toxic impact than the aforementioned elements. In this way, the following elements are to be considered: Cr, Co, Sn, Sb, Cu, Ni, Ag, Au, Zn, Mo, V, Mn, Fe and other [6]. These elements are bound with fine particles of aerosols. Heavy metals do enter the environment by virtue of natural and anthropogenic processes. Natural sources include in situ weathering processes and atmospheric deposition of metals, oceanic processes and volcanic eruptions. Anthropogenic sources include burning fossil fuels in order to generate electricity, raw materials excavation and ore processing, industrial processes, agricultural activities and a continuously increased usage of motor vehicles.

2. Particulate matter amount determination and chemical analysis of metals

Readings of particulate matter were performed in the close vicinity of the urban connecting route, road in regular intervals, namely 4 times per year. The goal was a long-term monitoring of a proportional representation of particulate matter (PM) and 17 heavy metals in the PM fraction. The aim was to identify sources of particulate matter by means of statistical methods. The problem solving assumes the knowledge of multivariate statistical data analysis methods as, for instance, principal components analysis (PCA), factor analysis (FA) and multivariate regression and vector algebra. For the application of methodology suitable software may prove appropriate.

The STATISTICA software, a computer program and the method of PCA and FA were used to quantify the contributions of individual sources of air pollution to solid particles. Data matrix with rows corresponding to individual measurements (samples) and columns corresponding to variables (measured pollutants) served as an input for calculations. Pollutants characterizing and defining the sources of pollution were selected as variables.

Keywords: Particulate matter, road traffic, air pollution, PM, heavy metals, statistical methods.
standards of STN EN 12341 [7] and STN EN 14907 [8] was used. The measurement readings were performed by means of low volume flow samplers of LECKEL LVS3 (Single Filter Gravimetric Sampler) amounting to the total number of 3 pieces. Concurrently, the three fractions of particulate matter were monitored, which include the following: PM$_{10}$, PM$_{2.5}$ and PM$_{1.0}$. They were trapped in nitrocellulose filters of the 47 mm diameter and assessed in a gravimetric manner. The monitoring of atmosphere along the concerned road, route was performed in weekly cycles. The particulate matter is bound with various elements and compounds. To test all the chemical components of the particulate matter concerned would be ineffective, quite demanding and financially unsustainable. In the first phase did we concentrate on the monitoring of the selected heavy metals found in the fraction of PM$_{10}$ [9]. Each of these metals may come from a specific source (Table 1). Based on a sufficiently comprehensive database of data it seems possible by the utilization of multidimensional statistical methods (for instance factor analysis), to more closely specify the possible source of these particles.

In order to identify or to determine the present chemical form of an observed element in the sample of particulate matter the spectroscopic methods are utilized. The analyses of filters and the determination of metals present in the fraction of PM$_{10}$ were performed pursuant to the standard of STN EN 14902 [16].

Prior to specifying inorganic pollutants, the filters were reacted by the mixture of acids (HNO$_3$ and HF) and oxidizing agents with a resultant specification by means of the mass spectrometry method with inductively coupled plasma mass spectrometer ICP MS (Perkin – Elmer ELAN 6000 – USA) in cooperation with the Institute of Geology at the Faculty of Natural Sciences of Comenius University in Bratislava [17].

3. Used statistical methods

Multivariate statistical analyses of PCA – Principal component analysis and FA – Factor analysis were used for statistical assessment.

The primary goal of PCA is the transformation of the original characters of $x_j$, $j = 1, ..., m$, into a smaller amount of latent variables $y_i$. These latent variables possess more appropriate and comprehensive properties: their presence is less significant, they capture and represent almost the entire variability of the original characteristics, properties and they are mutually uncorrelated. Latent variables are known as the main components and they represent linear combinations of former variables: the first principal component (PC) $y_1$ describes the greatest part of variability, hence the dissipation, spread of the original data, the second principal component $y_2$ on the other hand the greatest part of variability, spread not-contained within $y_1$, etc.. Mathematically speaking, the first PC is viewed as a linear combination of input characters which covers the greatest variability, spread among the other linear combinations. The disparity between the coordinates of object in the former characters and the principal components, thence the loss of information by the projection into the smaller number of dimensions is known as the fitness rate of PCA model interpretation or it is also called the PCA model error. For the first PC the following relation prevails:

$$y_1 = \sum_{j=1}^{m} v_{1j} x_j$$  \hspace{1cm} (1)

where:

- $x_j$, former character, input variable, $j = 1, ..., m$,
- $v_{1j}$, coefficients of eigenvectors.

Within the process of factor analysis (FA) so called factor loadings are estimated for particular variables (pollutants) for a generated factor. Factor loadings are the expression of correlations between the particular variables and acquired factors. Based on the values of factor loadings it is possible to specify a group of variables for each factor, those ones which correlate with it in the closest-possible manner. And vice versa, by means of factor loadings, the identified factor is appended with an extent of impact on each of particular variables. The variables with the highest factor loadings for a generated factor are considered as decisive even when interpreting such a factor. A data matrix serves the purpose of input for calculations, whose lines correlate with particular measurements (objects) and bars of variables, i.e. measured pollutant (character).

Sources of metals contained in the particulate matter – in general [10, 11, 12, 13, 14,15] Table 1

<table>
<thead>
<tr>
<th>Source</th>
<th>Selection of produced metals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transportation</td>
<td>road surface abrasion Al, Si, Ca, Mg, C, Na, K</td>
</tr>
<tr>
<td>car-body components abrasion</td>
<td>Cu, Sn, Cr, Pb, Cd, As, Sb, Fe, Al</td>
</tr>
<tr>
<td>brake callipers, pads and rotors abrasion</td>
<td>Cu, Sb, Ba</td>
</tr>
<tr>
<td>tyre wear</td>
<td>Zn, Cd, Pb, Cu</td>
</tr>
<tr>
<td>fuel combustion</td>
<td>Mo, V</td>
</tr>
<tr>
<td>catalytic converter</td>
<td>Pt, Pa, Rh (Platinum metals)</td>
</tr>
<tr>
<td>Burning coal and wastes</td>
<td>Zn, Sb, Cu, Cd, Hg, Se, As, Cr, Co, Al</td>
</tr>
<tr>
<td>Industry</td>
<td>Sh, Ag, V, Ni, As, In, Cu, Mn, Cr, Co, Cr, Pb</td>
</tr>
<tr>
<td>Biomass burning</td>
<td>K</td>
</tr>
<tr>
<td>Soil and city-generated dust</td>
<td>Ti, Al, Fe, Mn, Cr, Cu, Th</td>
</tr>
</tbody>
</table>
The variables to be used are those pollutants which are able to specify anticipated sources of pollution.

The basic principle of factor analysis lies in the fact that each of monitored values \( X_j \ (j = 1, \ldots, p) \) may be expressed as a sum of a linear combination of a lesser amount \( m \) non-observed (hypothetical) random values \( F_1, \ldots, F_m \) - so called common factors and the further source of variability \( E_j \ (j = 1, \ldots, p) \) - so called specific (residual) elements. Let us suppose that the following model prevails:

\[
X_j = \sum_{i=1}^{m} \alpha_{ij} F_i + E_j, \quad j = 1, \ldots, p, \quad (2)
\]

where:
- \( \alpha_{ij} \): factor loadings of the \( k \) common factor relevant to the \( j \) variable and of \( k \) factor, \( k = 1, \ldots, m, \)
- \( F_k \): the \( k \)th common factor,
- \( E_j \): a random deviation of the exact model, relevant to the \( k \)th variable, \( j = 1, \ldots, p. \)

Values \( X_1, \ldots, X_p \) are standardized, i.e. they have a zero main value and unit variance. In the matrix form we get the FA model of the following string:

\[
X = A \times F + E \quad (3)
\]

The \( F \) matrix is called factor matrix. Furthermore, let us suppose that the random vectors \( F \) and \( E \) are non-correlated. In the FA method it is recommended to have at least 5 samples, while the optimum number of samples could reach 20 per each variable. The appropriateness of factor analysis used is proven by the test of KMO (Kaiser-Meyer-Olkin) and MSA - the extent of internal correlation [18], [19].

a) KMO (Kaiser-Meyer-Olkin) criterion is to be considered as an index for the comparison of magnitudes in experimental correlation coefficients when compared with magnitudes of partial correlation coefficients. Should the sum of squares in partial coefficients among all the pairs of characters remain low in comparison with the sum of squares of pair-forming correlation coefficients, the KMO measure of sampling adequacy shall be close to 1. Low values in KMO on the other hand signify that FA of former characters shall not be interpreted as a good approach taken, as the correlation among all the character pairs may not be substantiated by mediation of other characters. Kaiser defined the measure of KMO, in case the correlation measure is to be assessed, in the following way:

\[
KMO = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} r_{ij}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2} \quad (4)
\]

Where \( r_{ij} \) is a pair correlation coefficient between the \( i \)th sign and \( j \)th sign and \( a_{ij} \) is to be considered a partial correlation coefficient between \( i \) sign and \( j \) sign.

a) MSA represents the measure of internal correlation degree among the former signs and by the very same means the extent of appropriateness for adopting the factor analysis. This index gains values within the interval between 0 up to 1:

- \( MSA_i = 1 \), je \( i \)th sign predicated in the most perfect way by means of other signs, \( MSA_i = 0.8 \), predicated still well within range, \( MSA_i = 0.7 \), predicated on an average base, basis, \( MSA_i = 0.6 \), predicated to a medium extent, \( MSA_i = 0.5 \), predicated to a weak extent and if the MSA shall remain lower than 0.5 the predication is unacceptable as well as inappropriate.

In order to carry out the FA successfully, a high value in the MSA measure is required. By means of MSA it is also equally possible to monitor the addition and subtraction of former signs, operators and characteristics, as by means of the MSA method, every character, sign or operator may be enumerated and expressed numerically.

\[
MSA_i = \frac{\sum_{j=1}^{n} r_{ij}^2}{\sum_{j=1}^{n} r_{ij}^2 + \sum_{j=1}^{n} a_{ij}^2} \quad (5)
\]

In order to approximate the contribution of specific sources of PM the method of Absolute Principal Component Scores is adopted, abbreviated as – APCS in conjunction with Multiple Regression Analysis - MRA. First of all the values of factor scores \( F \) are estimated. In order to acquire to factor scores with the physics-oriented meaning, the APCS matrix is calculated, in the way that the reference value of factor scores \( F_{i0} \) is calculated, which is subtracted from the calculated \( F \) matrix following the relation

\[
APCS = F - F_{0} \quad (6)
\]

By means of utilization of Multiple Regression Analysis, where PM is a dependent variable and APCS are independent variables, one can ascertain the contributions of separately-identified sources. These are added up, while being summarized, as of the determined regression coefficients by means of their multiplication using the mean value of APCS for each common factor as an absolute contribution, which is out of the total concentration of PM re-calculated and transferred into the percentile contribution of given sources [20].

4. **Statistical analysis of measured results**

Multilayer statistical analyses were applied on the data file consisting of measurements of concentrations in analysed metals (Na, Mg, Al, Ca, Cu, Nb, Co, Pb, Cd, Cr, As, Mo, V, Mn, Fe, Ni, Zn) as well as PM\(_{10}\) concentrations. The overall matrix contained 18 columns (variables) and 108 rows (objects, operators - measurements). This very matrix contains the concentration of metals and PM\(_{10}\) for daily and nightly sampling.
In order to establish the appropriateness of the factor analysis adoption the criteria of KMO and MSA were calculated. In accordance with them the adoption of factor/factorial analysis is substantiated. The results of statistical tests for the appropriateness of factor analysis adoption:

- MSA criterion: Na - 0.86, Mg - 0.73, Al - 0.83, Ca - 0.81, Cu - 0.91, Sb - 0.95, Ba - 0.93, Pb - 0.87, Cd - 0.9, Cr - 0.89, As - 0.91, Mo - 0.88, V - 0.86, Mn - 0.88, Fe - 0.92, Ni - 0.88, Zn - 0.91, PM10 - 0.83 → particular characters are predicated with the other ones very precisely.
- KMO criterion: KMO = 0.88 → a very good correlation of characters.
- Bartlett Sphericity Test: Criterion = 2501.7, Degrees of variance = 153, P = 0.000.

The calculated significance level of P is significantly lower than the given level of α = 0.05, that is why the PCA may be applied on this set of data, readings [21].

Whereas, at the factor analysis it is deemed to be necessary to specify the number of factors and then the calculation may be run, during the first step the analysis of principal components was performed. As a result of which we could conclude the possible number of principal components which, to a sufficient measure, specify the variability, variety of characters. As observed in the Fig. 1 we are able to select the number of principal components 3. The three principal components define 79.66 % of the total variability of the original characters [21].

There were 3 factors selected for the factor analysis. The used model of factors rotation was Varimax. In the Table 2 the factor loadings of characters are quoted in relation to factors. They may be explained as the correlation between the factors and characters. They represent the most important unit of information the interpretation of factors is based on.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na</td>
<td>0.731243</td>
<td>0.239047</td>
<td>0.174231</td>
</tr>
<tr>
<td>Mg</td>
<td>0.131402</td>
<td>0.104391</td>
<td>0.948492</td>
</tr>
<tr>
<td>Al</td>
<td>0.079466</td>
<td>0.681083</td>
<td>0.564493</td>
</tr>
<tr>
<td>Ca</td>
<td>0.211769</td>
<td>0.390023</td>
<td>0.857842</td>
</tr>
<tr>
<td>Cu</td>
<td>0.523754</td>
<td>0.744955</td>
<td>0.227409</td>
</tr>
<tr>
<td>Sb</td>
<td>0.427920</td>
<td>0.741372</td>
<td>0.256502</td>
</tr>
<tr>
<td>Ba</td>
<td>0.273923</td>
<td>0.783796</td>
<td>0.486153</td>
</tr>
<tr>
<td>Pb</td>
<td>0.856858</td>
<td>0.239686</td>
<td>0.118103</td>
</tr>
<tr>
<td>Cd</td>
<td>0.898693</td>
<td>0.237152</td>
<td>0.047900</td>
</tr>
<tr>
<td>Cr</td>
<td>0.699834</td>
<td>0.339355</td>
<td>0.121727</td>
</tr>
<tr>
<td>As</td>
<td>0.395564</td>
<td>0.477598</td>
<td>0.199278</td>
</tr>
<tr>
<td>Mo</td>
<td>0.210323</td>
<td>0.906938</td>
<td>0.213498</td>
</tr>
<tr>
<td>V</td>
<td>0.244208</td>
<td>0.483371</td>
<td>0.619649</td>
</tr>
<tr>
<td>Mn</td>
<td>0.619768</td>
<td>0.670952</td>
<td>0.260199</td>
</tr>
<tr>
<td>Fe</td>
<td>0.314326</td>
<td>0.892339</td>
<td>0.092341</td>
</tr>
<tr>
<td>Ni</td>
<td>0.680881</td>
<td>0.39069</td>
<td>0.202344</td>
</tr>
<tr>
<td>Zn</td>
<td>0.785544</td>
<td>0.361700</td>
<td>0.215592</td>
</tr>
</tbody>
</table>

Fig. 1 Eigenvalues of Correlation Matrix Chart [21]

Each factor is contributed by several elements (characters). As the most decisive factor of loads the values close to or greater than 0.7 were selected. Based on the representation and presence of elements in particular, selected factors, the following factors may be named.

**Factor 1.** Na, Pb, Cd, Zn, PM10  
**Factor 2.** Cu, Sb, Ba, Mo, Fe  
**Factor 3.** Mg, Ca
5. Discussion on the achieved results

In order to identify the origin of the particulate matter at the aforementioned monitoring station, the usage of factor analysis proved as the appropriate one. As resulted from the factor analysis, the original dataset may be explained using several factors. The four factors explain 79.66% of the total variability of the original characters. Naming these factors is enabled based on the characters, which specify unique factors. This is, at the same time, conditioned by a supposition of the fact where separate elements – heavy metals – may originate from. The survey on the origin may be drawn from the former research surveys performed by various authors. In the monitored station, the primary cause of particulate matter originates as a consequence of the vehicular traffic. We follow the supposition that the observed metals are traced to specific automotive components and the road surface itself (Table 1). Based on the results of factor analysis and the supposed sources, the factors of 1 and 3 were named, designated in the following manner: Factor 1 – Tyre wear and tear and gritting materials. Factor 2 – Brake pads lining tear and wear, brake pads, car components and combustion of propulsive, propulsion fuels. Factor 3 – Road pavement wear – aggregate, road tar binders and road dust.

Using the APCS method while supported by adoption of the input data the following may be implicated: the contribution of the tyre tear and wear and gritting materials to the PM$_{10}$ formation by 52.45%, brake pads lining tear and wear, brake pads, car components and combustion or propulsive, propulsion fuels by 16.77%, road pavement wear – aggregate, road tar binders and road dust by 14.16% and unresolved part by 16.62%. A detailed specification of contribution towards particular sources of PM$_{10}$ formation may be narrowly specified by a number of chemical elements and compounds, based on which statistical analyses are performed. In our case, the analyses performed incorporated the selection of 17 heavy metals, which are typical for vehicular traffic and road transportation [21].

Acknowledgements
The paper originated thanks to being supported by means of a grant for the scientific research task VEGA 1/0508/2011 entitled Overall characteristics and chemical composition of particulates as created by the vehicular traffic.

References


[16] STN EN 14902: 2005 Ambient air quality. Standard method for the measurement of Pb, Cd, As and Ni in the PM$_{10}$ fraction of suspended particulate matter.


