Proceedings in Manufacturing Systems

Proceedings in Manufacturing Systems, Volume 6, Issue 3, 2011 ISSN 2067-9238

TECHNOLOGICAL RELIABILITY ANALYSIS WITH THE MULTIVARIATE STATISTICS

Adrian Stere PARIS[1,*], Ioan TĂNASE[2], Constantin TÂRCOLEA[3]

[1)] PhD, Assoc. Prof., Department of Materials Technology and Welding, University Politehnica, Bucharest, Romania
[2)] PhD, Prof., Machines and Production Systems Department, University "Politehnica" of Bucharest, Romania
[3)] PhD, Prof., Department of Mathematics, University Politehnica of Bucharest, Romania

Abstract: *The lifetime evolution of the manufacturing accuracy of the machine-tools, defined as technological reliability, is an important research domain. The chosen practical case is connected with the accuracy of a set of universal milling machines. A pre-selection of field data using the Spearman technique was developed. Five characteristics of manufacturing accuracy were selected from the field experimental data for a detailed analysis. After a brief introduction in the multivariate analysis, the main side of the paper covers a dimensional reduction of the accuracy attributes. Based on the factor analysis method, using XLSTAT 7.5.2 software and determining the numbers of factors for a given level of significance, a given alpha risk is achieved.*

Key words: *technological reliability, formalization, factor analysis, XLSTAT software, case study*.

1. INTRODUCTION

The reliability and the safety analysis in the assessment of complex manufacturing systems are becoming a more difficult task. Productivity and accuracy of machine tools are important competition aspects. Rapidly changing operating conditions for machine tools, however, make it difficult to increase productivity and accuracy. In the manufacture of parts, increasingly small batch sizes have to be produced economically and yet accurately.

Each machining operation creates a feature which has certain geometric variations compared to its nominal geometry. Designers normally give design tolerance specifications on the nominal value, to specify how large these variations are allowed to be. One needs to estimate accuracy of various manufacturing processes in order to verify whether or not a given process plan will produce the desired design tolerances [5, 6]. In comparison with usual sense of reliability, the geometric errors of machine tools are much more difficult to observe and it needs a new term to describe this situation.

The technological reliability at the moment *t* can be quantitatively defined as the probability of a manufacturing equipment (namely a machine-tool) to maintain her working accuracy limits by the time *t*. This means to check the machine-tool accuracy at different time moments and establish the corresponding function of technological reliability [5, 6]. It follows a short description of the experimental researches of the authors in the field of technological reliability of a family of milling machines [5].

[*] Corresponding author: Splaiul Independenței 313, sector 6, 060042, Bucharest, Romania
Tel.: 0214029337; 0726342625; Fax:0213155956
E-mail addresses: *ıubomir_dimitrov@tu-sofia.bg (*A.S. Paris), *tanase.ioan@yahoo.com* (I. Tănase)

The results were processed using some applications of multivariate data analysis, especially correlation theory and factor analysis.

Multivariate analysis is used to denote the study of data which are multidimensional in the sense that each object bears the values of several characteristics of interest. In order to perform multivariate exploratory statistics, these data must be interpreted as an attributes/objects table [7]. Multivariate data analysis contains two classes of methods: analyzing data and advanced data analysis. In the first class are, among others, Factor Analysis (FA), Principal Component Analysis (PCA), Biplot, Discriminant analysis (DA), Correspondence Analysis (CA), Multiple Correspondence Analysis (MCA), Multidimensional Scaling (MDS), Agglomerative Hierarchical Clustering (AHC), k-means Clustering, Univariate Clustering. The second group contains: Canonical Correspondence Analysis(CCA and partial CCA), Generalized Procrustean Analysis(GPA), Multiple Factor Analysis (MFA), Redundancy analysis (RDA), Coordinate Analysis, useful for a variety of applications, ranging from ecology to marketing [2, 3]. An important branch of multivariate analysis is factor analysis.

An important branch of multivariate analysis is factor analysis. The kernel of Factor Analysis is to identify a number of underlying factors that explains the relationship between correlated variables and, in the same time, to have a smaller alpha risk [1].

Factor Analysis is deeply related to Principal Component Analysis, but while the Factor Analysis assumes that the correlation between variables is due to a set of latent variables that are being measures by the variables [6], Principal Component Analysis is a method for reducing the number of variables and is not based on the idea that there are underlying factors, that are being measured [10].

2. MATHEMATICAL FORMALIZATION OF THE PROBLEM

At the beginning the data are given, as a n*p-matrix, objects/attributes, a table $y = (y_{ij})$, $i = 1, 2, ..., n$; $j = 1, 2, …, m$. Each row of the matrix represents an object (individual) with his attributes, and each column is an attribute (property, variable). The number of observable attributes gives the dimension of the initial representation space of the objects. With other words it is considered an m-dimensional coordinate system, each coordinate being an attribute. Instead of realer attributes the FA uses new factors, but only a few, which are artificial ones [8].

The problem can be mathematical formulate; it is supposed that $\mathbf{y}^t = (y_1, y_2, …, y_m)$ is a random vector with the center of dispersion m and the covariance matrix $\sum$. The FA procedures try to identify new uncorrelated variables $z_1, z_2, …, z_m$, whose variance decreases when the index increases from 1 to m [12]. The first FC explains the maximum variance in the data; the second FC explains the maximum variance that has not been accounted by the first FC, and so one. The FA solves the problem of finding the directions of the greatest variance of the linear combination of the old coordinates. In other words it seeks the a set of the coefficient vectors $a_1, a_2, …, a_k$, each new variable is a linear combination of the initial variables. The first principal component:

$$z_1 = a_{11} y_1 + ... + a_{m1} y_m \quad (1)$$

is chosen, so that:

$$Var(z_1) = Var(a_1^{tr} y) = a_1^{tr} \sum \cdot a_1 \quad (2)$$

is maximal, under the restriction:

$$a_1^{tr} a_1 - 1 = 0 . \quad (3)$$

To find the conditional extreme of a function, given a relationship it is used a so-called Lagrange function:

$$L(a_1;\lambda) = a_1^{tr} \sum \cdot a_1 - \lambda(a_1^{tr} a_1 - 1), \quad (4)$$

where λ is an undetermined multiplier. The necessary conditions for the extreme are:

$$\begin{cases} 2\sum \cdot a_1 - 2\lambda a_1 = 0 \\ a_1^{tr} a_1 - 1 = 0 \ . \end{cases} \quad (5)$$

The directions of the new coordinate axes, called principal components, or factors, have been chosen, in such a way, that the deformations of the original cloud implied by this representation are minimal [9]. The coordinates of the objects (samples) in the new system are called scores. The corresponding relationships between the original variables and the new principal components are called loadings.

3. MATERIALS AND METHOD CASE STUDY

The case study is referring to technological reliability $R_{th}(t)$ of a universal milling machine with a high level of geometric accuracy parameters.

The criterion of the time variation of the geometric accuracy was selected for the establishing of technological reliability indicators. It was monitored a set of universal milling machines in the factories working conditions for a time span for the reliability evaluations. Especially a set of measurements was made for the accuracy of geometry.

In the first step was applied Spearman’s test for the field data processing. In statistics, Spearman's rank correlation coefficient often denoted by the Greek character ρ or as r_s, is a non-parametric measure of statistical dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

In applications where ties are known to be absent, a simpler procedure can be used to calculate ρ [4]. Differences $d_i = x_i − y_i$ between the ranks of each observation on the two variables are calculated, and ρ is given by:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} . \quad (6)$$

As a result of the Spearman test [4] it were selected the following accuracy checks:

- parallelism between the table surface and and the direction of her longitudinal movement (B_3 norm test);
- perpendicularity between of the milling head rotation axis and the table surface in both directions longitudinal and transversal (C_{3a} and C_{3b} norm test);
- parallelism between upper surface and base surface when machining with the vertical milling head (K_{3b} norm test);
- perpendicularity between side surfaces and base surface when machining with the horizontal milling head (K_{2b} norm test).

The data are presented in Table 1.

As another statistical preliminary test was applied the ANOVA method, to verify if the attributes are statistical identical; the numerical results shows that the null hypothesis it is not rejected, based on the F-test and p-value (Table 2).

In this situation follows the second step. The chosen features are detailed using multivariate statistics, Factor Analysis, with XLSTAT 7.5.2 software, a Microsoft

Table 1

Geometric accuracy experiments results

No.	C3lg	C3tr	B3	K_{3b}	K_{2b}
1	0.01	0.01	0.008	0.014	0.025
2	0.03	0.02	0.012	0.025	0.008
3	0.1	0.19	0.12	0.12	0.1
4	0.055	0.15	0.06	0.02	0.02
5	0.48	0.14	0.04	0.1	0.15
6	0.14	0.19	0.1	0.2	0.04
7	0.025	0.07	0.027	0.03	0.07
8	0.042	0.015	0.03	0.08	0.03
9	0.028	0.017	0.01	0.15	0.13

Table 2

Results of ANOVA test

Source of variation	*SS*	*df*	*MS*	*F*	*P* value	F_{crit}
Between Groups	0.017	4	0.0043	0.6	0.67	2.61
Within Groups	0.293	40	0.0073			
Total	0.31	44				

Table 3

Mean and standard deviation of the columns

Variables	Mean	Standard deviation
1	0.101	0.148
2	0.089	0.078
3	0.045	0.041
4	0.082	0.066
5	0.064	0.052

Table 4

Correlation matrix

	Var1	Var2	Var3	Var4	Var5
Var1	1	0.471	0.209	0.302	0.619
Var2	0.471	1	**0.908**	0.448	0.232
Var3	0.209	**0.908**	1	0.529	0.100
Var4	0.302	0.448	0.529	1	0.473
Var5	0.619	0.232	0.100	0.473	1

Table 5

Reproduced correlation matrix

	Var1	Var2	Var3	Var4	Var5
Var1	**0.762**	0.471	0.208	0.303	0.618
Var2	0.471	**0.981**	**0.908**	0.448	0.232
Var3	0.208	**0.908**	**0.992**	0.529	0.101
Var4	0.303	0.448	0.529	0.642	0.472
Var5	0.618	0.232	0.101	0.472	**0.722**

Excel add-in [11] and the obtained results are presented bellow. Data processing needed 25 iterations for a 0.001 convergence. Table 3 gives a statistical overview of the geometric accuracy checks, called variables in the following calculi.

It results (Table 3) that the variation coefficient, a normalized measure of dispersion of a probability distribution, also known as unitized risk**,** are comparable for all selected attributes of this milling machines set.

In bold are the significant values (except diagonals) at the level of significance α = 0.05 (two tailed test).

When the method converges with a sufficient precision, the values of the main diagonal are equal to specific variances.

In bold there are significant values (except diagonals) at the level of significance α = 0.05 (two tailed test).

If this model is correct, it is not possible that the factors will extract all variance from the items; rather, only that proportion that is due to the common factors and shared by several items. In the language of factor analysis, the proportion of variance of a particular item that is due to common factors (shared with other items) is called *communality*. Therefore, an additional task is to estimate the communalities for each variable, that is, the proportion of variance that each item has in common with other items (Fig. 1 and Table 6).

Number of removed trivial eigenvalues: 2.

The rectangles (Fig. 1) show the fraction of the total variance of the primary data for each factor.

The three largest eigenvalues are 2.6, 1.1 and 0.4 (Fig. 2 and Table 7). This suggests that the corresponding PC's (F1, F2, F3) are enough for the selection.

The representation of the data in a limited number of dimensions (three dimensions in this case) facilitates to a great extent this analysis.

The factor loadings, also called component loadings in FA, are the correlation coefficients between the variables (rows) and factors (columns). Similarly to Pearson's *r* coefficient the squared factor loading is the percent of variance in that indicator variable explained

Table 6

Maximum change in SQRT (communality)

Iteration	SQRT (communality)
1	0.281
2	0.053
3	0.030
4	0.018
5	0.011
6	0.007
7	0.005
8	0.003
9	0.002
10	0.001
11	0.001

Table 7

Eigenvalues

	F1	*F2*	*F3*
Eigenvalue	2.599	1.098	0.402
Total%variance	51.74	21.969	8.046
Cumulative%	51.974	73.944	81.990
Common %variance	63.391	26.795	9.813
Cumulative%	63.391	90.187	100.00

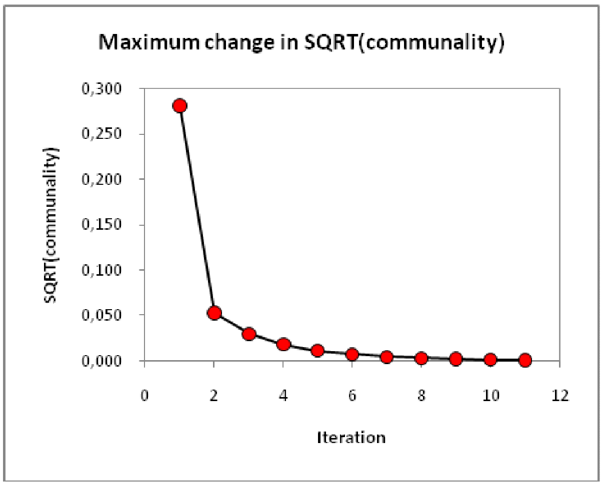

Fig. 1. SQRT change with iteration.

by the factor. To get the percent of variance in all the variables accounted for by each factor, add the sum of the squared factor loadings for that factor (column) and divide by the number of variables (Table 10).

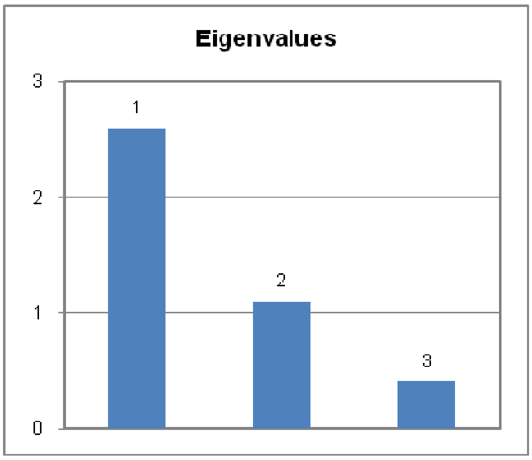

Fig. 2. Eigenvalues.

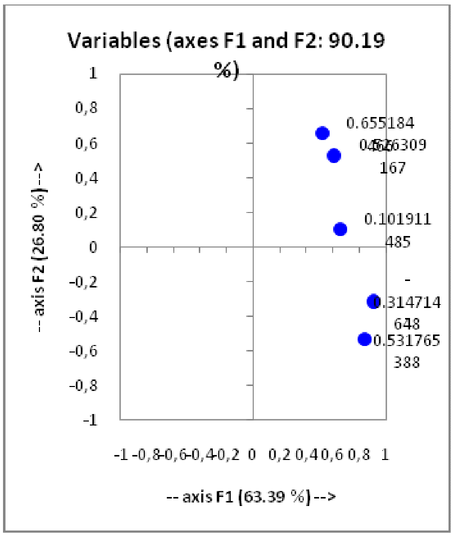

Fig. 3. Variables plot with F1 and F2 axis.

Table 8

Eigenvectors

	F1	*F2*	*F3*
Var1	0.379	0.502	–0.527
Var2	0.562	–0.300	–0.94
Var3	0.520	–0.507	0.118
Var4	0.406	0.097	0.711
Var5	0.324	0.625	0.219

Table 9

Factor loadings

Var	*F1*	*F2*	*F3*	**Initial com-munal-ity**	**Final com mu-nality**	**Sp. var.**
V1	0.611	0.526	–0.334	0.657	0.762	0.238
V2	0.905	–0.315	–0.250	0.918	0.981	0.019
V3	0.839	–0.532	0.075	0.914	0.992	0.008
V4	0.655	0.102	0.451	0.522	0.642	0.358
V5	0.523	0.655	0.139	0.517	0722	0.278

The observations (numerical results of the geometric accuracy tests) investigation (Table 11 and Figs. 5 and 6) (herein below having the axes F1 and F2 – Fig. 5, F1 and

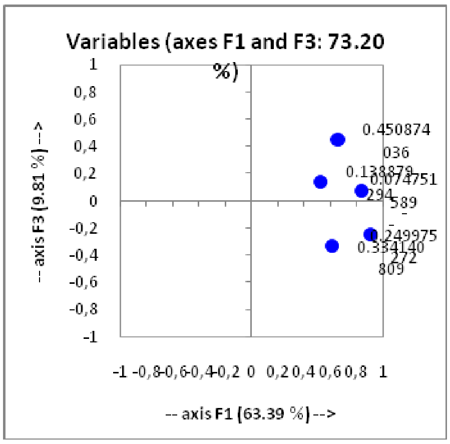

Fig. 4. Variables plot with F1 and F3 axis.

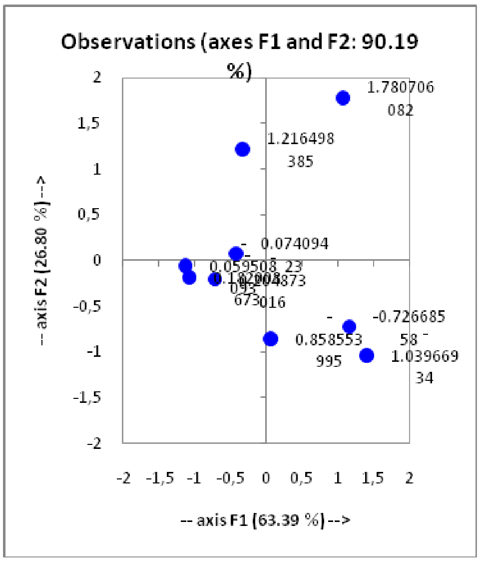

Fig. 5. Observations plot with F1 and F2 axis.

Table 10

Standardized factor score coefficients

	F1	*F2*	*F3*
Var1	0.178	0.327	–0.069
Var2	0.372	0.141	–1.876
Var3	0.373	–0.882	1.588
Var4	0.129	0.214	0.328
Var5	0.228	0.406	0.304

Table 11

Estimated factor scores

	F1	*F2*	*F3*
Obs1	–1.133	–0.060	–0.081
Obs2	–1.078	–0.182	–0.219
Obs3	1.402	–1.040	0.909
Obs4	0.057	–0.859	–1.429
Obs5	1.066	1.781	–1.008
Obs6	1.158	–0.727	0.152
Obs7	–0.425	0.074	–0.442
Obs8	–0.717	–0.205	1.004
Obs9	–0.331	1.216	1.115

F3 – Fig. 6) shows a projection of the initial variables in the factors space. In Fig. 5 the observations are in majority close to the centre and variables 1, 2, 3, 5, 6 are significantly correlated with F1, and variables 3, 5, 9 are correlated with F2 . This can be confirmed either by looking at the correlation matrix.

Next, it is applied the varimax rotation, that has changed the way each factor explains part of the variance. The varimax rotation makes the interpretation easier by maximizing the variance of the squared factors loadings by column. For a given factor, high loadings become higher, low loadings become lower, and intermediate loadings become either lower or higher.

Once the results have been obtained, they may be transformed in order to make them easier to interpret, for example by trying to arrange that the coordinates of the variables against the factors are either high (in absolute value), or close to zero.

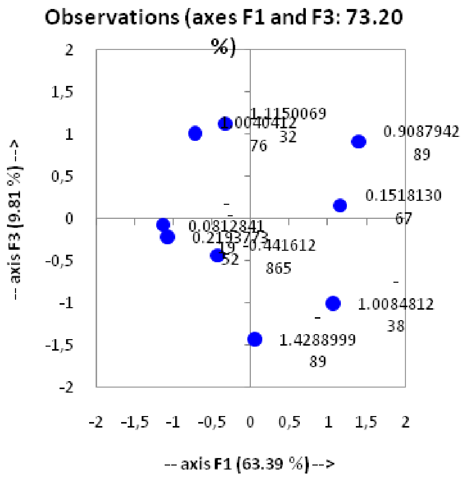

Fig. 6. Observations plot with F1 and F3 axis.

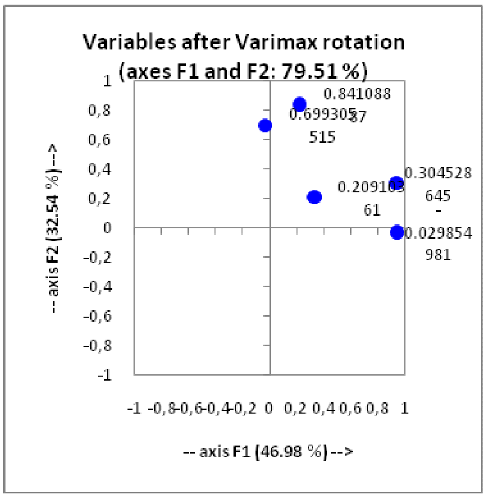

Fig. 7. Variables plot with F1 and F2 axis after the Varimax rotation.

Table 12

Rotated matrix

	F1	*F2*	*F3*
F 1	0.759	0.488	0.431
F 2	–0.609	0.767	0.204

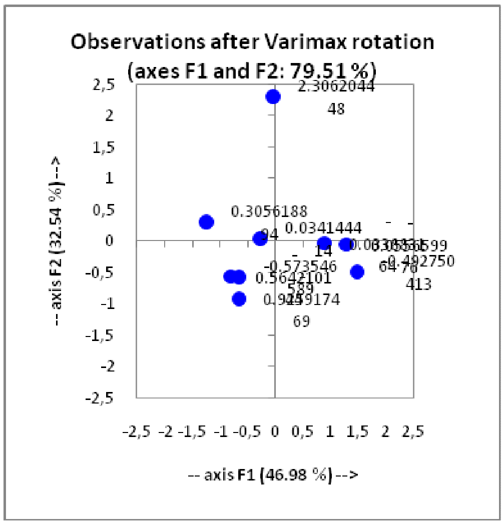

Fig. 8. Observations plot with F1 and F2 axis after the Varimax rotation.

Table 13

Total % variance after Varimax rotation

	F1	*F2*	*F3*
Total%variance	46.979	32.536	20.485
Cumulative%	46.979	79.515	100.00

Table 14

Factor loadings after Varimax rotation

Var	*F1*	*F2*	*F3*
V1	0.221	0.841	0.077
V2	0.936	0.305	0.106
V3	0.943	–0.030	0.319
V4	0.331	0.209	0.699
V5	–0.034	0.699	0.481

Table 15

Standardized rotated factor score coefficients

Var	*F1*	*F2*	*F3*
V1	–0.048	0.367	0.083
V2	0.629	1.073	–1.460
V3	0.454	–1.158	1.376
V4	–0.108	0.090	0.388
V5	–0.145	0.296	0.449

Table 16

Estimated factor scores after Varimax rotation

	F1	*F2*	*F3*
Obs1	–0.805	–0.564	–0.572
Obs2	–0.657	–0.574	–0.695
Obs3	1.487	–0.493	1.191
Obs4	0.896	–0.034	–1.407
Obs5	–0.042	2.306	–0.063
Obs6	1.287	–0.056	0.484
Obs7	–0.265	0.034	–0.556
Obs8	–0.651	–0.926	0.531
Obs9	–1.249	0.306	1.086

4. CONCLUSIONS

The paper presents a kind of useful procedure in the experimental researches case, offering a simplification of tests and of consequent effort. Nevertheless at the first sight seems to present a major difficulty for engineering practice, due to the complex mathematical formulation, but in reality, how it was illustrated in the article, processing the data with a specialized software offers rapid solutions. The presented example, on the testing of geometrical accuracy of machine tools, targeting the following calculus of the technological reliability, was a stable research domain for the authors, starting from 70s [4 and 5].

As model development it is important to compare the outputs of the Factor Analysis to those of the Principal Component Analysis (PCA) [2]. In work [7] it is applied the technique of PCA for reducing the number of variables by finding artificial variables, using Pearson and Jöreskog [9] procedures.

Depending of the chosen dimension and technique is resulted different risks, as proportion of neglected dimensions in the general variance. It can say that the procedure can be choosing for each practical application.

Usually in the first stage, for each check, there are taken into consideration as many accuracy tests as possible. In the second stage, based on FA, there were chosen two or three, given by the principal components. An artificial subspace with three (two) dimensions [10] with XLSTAT 2011 software is developed in the present research. The initial attributes for each tool should be expressed with a precision of 80% as function of two artificial axes. The application of this model simplifies the technological reliability evaluation. The presented method enables many other possible extensions in the exploratory field analysis of reliability.

REFERENCES

[1] A. Afifi, A. Clark, V. May, *Computer-aided multivariate analysis*, Chapman and Hall/CRC Press, London, 2004.

[2] Gh. Amza, A.S. Paris, C. Târcolea, *Risk and Factor Analysis*, Proceedings of the 3td WSEAS International Conference on RISK MANAGEMENT, ASSESSSMENT AND MITIGATION RIMA'10, Univ. Politehnica, Bucharest, April 2010, in vol. Recent Advances in Finite Differences-Finite Elements-Finite Volumes-Boundaries Elements, WSEAS Press, pp. 146–151.

[3] C. Croux, A. Ruiz-Gazen, *High Breakdowns Estimators for Principal Components The Projection-Pursuit Approach Revisited 2000*, Journal of Multivariate Analysis, Vol. 95, Iss. 1, July 2005, Academic Press, Orlando, FL, USA.

[4] D. Drimer, A. S. Paris, C. Târcolea, D. Dimitrescu, G. M. Dumitru, *Metoda operativa de apreciere a corelaţiei unor parametri de fiabilitate tehnologică* (Operative metod of assessment of the correlation of technological reliability parameters), Sesiunea de comunicari I.U.G. Brăila, 1976, pp. 411–414.

[5] A. Oprean, A. Dorin, D. Drimer, A.S. Paris, A. Ionescu, *Fiabilitatea maşinilor–unelte* (Reliability of machine tools), Edit. Tehnică, Bucharest, 1979.

[6] A.S. Paris, C. Târcolea, *Regression models applied to manufacturing systems*,Proceedings in Manufacturing Systems, Edit. Academiei Române, Vol. 5, No. 4, 2010, pp. 249–253.

[7] A.S. Paris, C. Târcolea, *Computer aided selection in design processes with multivariate statistics,* Proceedings of the International Conference on Manufacturing Systems – ICMaS, Vol. 4, 2009, Edit. Academiei Române, pp. 335-338.

[8] A.S. Paris, C. Târcolea, *Anwendungen der multivariaten Verfahren* (Applications of the multivariate processes), Workshop Ökonomische Anwendungen mit der multivariaten Verfahren, Politehnica University Bucharest, FILS, May 2010.

[9] C. Târcolea, A. S. Paris, *The Joreskog technique applied for materials design,* Proceedings of the 17th International Conference on Manufacturing Systems – ICMaS 2008, Edit. Academiei Române, pp. 309–312.

[10] C. Târcolea, A. S. Paris, A. Demetrescu –Târcolea, *Statistical methods applied for materials selection* The International Conference DGDS-2008 & MENP-5 2008, Mangalia, Romania, Applied Sciences (APPS*)* Volume 11, 2009, pp. 145–150.

[11] *** www.xlstat.com

[12] C. Târcolea, A., S., Paris, P., Voicu, *Principal Component Analysis Applied to Agricultural Equipments* 11th International Congress on Mechnization and Energy in Agriculture, 21–23 September 2011, TRAKAGENG 2011, Istanbul

[13] C. Târcolea, A. S. Paris, *Discriminant Analysis and Applied Regression* The International Conference of Differential Geometry and Dynamical Systems (DGDS-2010), August 2010, University Politehnica Bucharest, BSG Proceedings 18, ISSN 1843-2859, Balkan Society of Geometers, Geometry Balkan Press 2011 pp. 95–100.