www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

Arch. Metall. Mater. **68** (2023), 2, 525-530

DOI: https://doi.org/10.24425/amm.2023.142431

A.M. ȚÎȚU[1*], A.B. POP[2]

IDENTIFICATION OF DATA AFFECTED BY ABERRANT ERRORS IN THE STUDY OF THE MILLING PROCESS ON ALUMINUM ALLOYS

In this scientific paper it is presented the statistical analysis of the experimental data obtained by the study of the influence of the cutting parameters exerted in end-milling process on the surface roughness. The surface roughness parameter is measured in the cutting feed direction and against it. The parameters of the cutting process, the number of levels and their values were established. Based on these parameters, the research was designed on a complete factorial experiment, randomized with seven blocks. The surface roughness values were measured using a roughness tester. The research method used involved the Romanovski test. The aim of this test was to identify the data affected by aberrant errors, to remove them from the samples and then to repeat the tests for the remaining data strings until all values met the conditions imposed by the test.

Keywords: cutting process; surface roughness; aberrant error; Romanovki test

1. Introduction

Experimental sciences base their knowledge on the problems most often addressed on experimental results obtained from measurement processes [1,2]. It is currently accepted that any experimental result is affected by at least random measurement errors and as such, the formulation of conclusions and decisions must be made considering the existence of these errors, so it is implicitly accepted that the statements made will had a certainty lower than 100%, respectively the probability that the predicted event to occur based on the analysis of the experimental results is subunit [3,4]. The fundamental objective of statistics is the problem of decision in case of uncertainty [5,6]. The decision means adopting or rejecting a hypothesis. The hypothesis is statistical because it is statistically justified, based on surveys. It refers to the law of distribution and the parameters of a random variable [7]. Mathematical statistics provide solutions for extracting those values of interest from the entire data set for further processing [8]. Starting from the idea that no statistical method can prevent the occurrence of mistakes, false reasoning, etc., it is necessary and indicated that the one who applies statistical methods to know well the experimental process, its importance and to choose the appropriate methods [9].

2. Description of the research method

The main purpose of the paper is to analyze the experimental values related to the arithmetic mean deviation of the surface profile Ra, resulting from the processing by milling of aluminum alloy 7136. Specifically, we are interested in tracking data affected by aberrant errors to remove them from the sample [10,11]. Reading the results of a measurement also allows highlighting data that are inconsistent with reality. Analyzing a series of experimental data, obtained by repeated measurements under conditions of repeatability, it is possible that one or a few data differ relatively much from most others, ie with abnormally high or low values compared to them. Their influence on the statistical parameters of the survey can be quite high, especially when the number of measurements n is small. These values are likely to be affected by gross errors and it is therefore necessary to determine whether they should be removed from the measured range [12,13]. The elimination of such errors requires in a first phase the analysis of the correctness of the measurement chain, of the measurement method, of the conditions in which the measurement is carried out, etc. if no anomalies were detected in the first phase, these errors are eliminated mathematically, based on some criteria, as it is known or not the quadratic mean

1 "LUCIAN BLAGA" UNIVERSITY OF SIBIU, FACULTY OF ENGINEERING, INDUSTRIAL ENGINEERING AND MANAGEMENT DEPARTMENT, 10 VICTORIEI STREET, 550024, SIBIU, ROMANIA

2 TECHNICAL UNIVERSITY OF CLUJ-NAPOCA, NORTHERN UNIVERSITY CENTRE OF BAIA MARE, FACULTY OF ENGINEERING – DEPARTMENT OF ENGINEERING AND TECHNOLOGY MANAGEMENT, 62A, VICTOR BABES STREET, 430083, BAIA MARE, MARAMURES, ROMANIA

* Corresponding author: mihail.titu@ulbsibiu.ro

error of the measured data. These calculations are made on the assumption that all measurements are performed with the same degree of accuracy and independently of each other [14]. Several tests can be used to study the appropriateness of removing certain values from the range of measured values: Chauvenet test; Romanovski test; Grubbs test; Student test etc.

The following should be considered when applying the tests:

– The measuring device must have a normal distribution.
– The test applied, being a statistical decision, may be affected by decision errors. An erroneous decision can be made to estimate a value that is not wrong and vice versa, a result that is wrong can be kept. For this reason, caution should be exercised when eliminating errors that are supposed to be affected by gross errors.
– The results of statistical data processing are more accurate the higher the number of measurements.
– In all cases where the measurement results are affected by determinable systematic errors, they shall be removed from the gross measurement result and only then verification tests shall be applied.

Too little volume of measurements has a negative effect on their accuracy, and too many loads the research program unnecessarily. Excessive repetition of some measurements can even lead to the measurement of trends in the variation of parameters, because of the inevitable dispersion of data, resulting in incomplete and erroneous conclusions.

3. Conducting experiments

The material of the semi-finished product used in carrying out the research is aluminum alloy 7136. The cutting operation is that of milling. The cutting tool used is a standard one in aluminum processing. Milling was performed on the HAAS VF2 CNC. The collection of experimental data was performed after measurements on the surface roughness using the Mitutoyo SURFTEST SJ-210. Starting from an experimental plan based on the complete factorial experiment, the research strategy was developed. The parameters of the cutting regimes and their corresponding levels have been established. The cutting speed, the cutting depth and the feed per tooth exert their influence on the answer to be analyzed, represented by the arithmetic mean deviation of the surface profile. The values assigned to these parameters are:

- Cutting speed – 6 levels: 495, 530, 570, 610, 660, 710 [m/min].
- Feed per tooth – 5 levels: 0.04, 0.06, 0.08, 0.11, 0.14 [mm/ tooth].
- Cutting depth – 5 levels: 2, 2.5, 3, 3.5, 4 [mm].

After performing the experiments, the measured data are collected. According to Montgomery [15] experiments and measurements can be repeated between 3 and 7 times for each set of values of the input parameters, to determine the constancy of the measurements. In this case, to obtain the most accurate results, a repetition will be performed 7 times. In accordance with this decision, the subsequent statistical analysis of the resulting data will be based on methods that are appropriate in this direction. Therefore, the identification of values affected by aberrant errors will be performed using the Romanovski test. The reason for this choice is that this test can be applied to up to 20 data points, the other methods used for this purpose being suitable for a large amount of data. In this case the number of data is a maximum of 7, represented by the total number of measurements performed for each test. All these tests aim to certify that the values obtained are real values of the end milling process. In this context, it is necessary to perform $6 \times 5 \times 5 = 150$ experiments. The roughness of the processed R_a surfaces will be determined, on the milling direction in the direction of feed as well as transversely on this direction, further marked with R_a longitudinally, respectively R_a transversely. This will result in a total of 300 measurements – performed on each value combination of the parameters of the established cutting regimes, resulting in 300 (= 150×2) measurements. Regarding the repeated experiments (each 7 times), a total number of 2100 measurements results (1050 measurements related to R_a longitudinally, respectively 1050 measurements related to R_a transversely).

4. Identification of data affected by aberrant errors

Identification of data affected by aberrant errors can be performed with Chauvenet or Romanovski tests [16,17]. Own research will be conducted based on the Romanovski test because it can be applied to up to 20 data points. In this case the number of data is a maximum of 7, represented by the total number of measurements performed for each test. Data affected by aberrant errors are usually the minimum or maximum values of the experimental data string. The Romanovski test assumes that an experimental value is considered aberrant if condition (1) is met and, after detecting the aberrant values, they are removed from the string [16,17]:

$$t = \frac{\left|x_e - \bar{x}\right|}{s \cdot \sqrt{\dfrac{n}{n-1}}} \tag{1}$$

where:

t – the test variable;
x_e – the value affected by aberrant errors (string value to be tested);
$\bar{x}$ – the arithmetic mean of the values in the data string, neglecting the tested value x_e;
s – the mean square deviation, neglecting the tested value x_e;
n – the number of measurements.

Test variable t, compares with the critical value t_{crt} which, according to [16], takes the values listed in TABLE 1. Here, the table value t_{crt}, is determined by the number of values in the data string and the confidence interval. If $t > t_{crt}$, then it turns out that

TABLE 1

Critical values t_{crt} for the Romanovski test after [16]

No of string data	Confidence level [%] 0.95	0.98	0.99	No of string data	Confidence level [%] 0.95	0.98	0.99
3	1.41	1.41	1.41	12	2.52	2.66	2.75
4	1.71	1.72	1.73	13	2.56	2.71	2.81
5	1.92	1.96	1.97	14	2.60	2.76	2.86
6	2.07	2.13	2.16	15	2.64	2.80	2.91
7	2.18	2.27	2.31	16	2.67	2.84	2.95
8	2.27	2.37	2.43	17	2.70	2.87	2.98
9	2.35	2.46	2.53	18	2.73	2.90	3.02
10	2.41	2.54	2.62	19	2.75	2.93	3.05
11	2.47	2.61	2.69	20	2.78	2.96	3.08

we are dealing with an aberrant value, and this is removed from the data string. Following the analysis of TABLE 1, the three confidence levels of 95%, 98% and 99% are observed. Given that the last two levels are used for calibrations, the results of our own research are considered satisfactory in the case of 95%. For example, for an accuracy of 95%, the value t_{crt} = 2.18 for a series of 7 data (measurements of the arithmetic mean deviation of the surface profile.

Next, based on the calculation method presented, the Romanovski test will be applied for the 150 experiments. The main topic of interest is the arithmetic mean deviation of the surface profile measured longitudinally in the direction of the forward motion, respectively transversely on it. The calculation method of this test was applied, and a large volume of data was obtained. The following conclusions are presented based on these results.

4.1. R_a longitudinally

The situation of the results obtained, regarding the data affected by aberrant errors in the case of longitudinally R_a measurements, is presented in TABLE 2. As can be seen from this table, in the case of cutting schemes which check the first given condition, equivalent to the string of 7 data, namely: t_{crt} = 2.18, only 49 of the 150 tests have no aberrant values. Therefore, after detecting these errors, they are removed from the sample and the test is repeated, considering the new condition t_{crt} = 2.07, corresponding to the new data string with 6 values.

In this situation, the new strings of 6 values each are tested. The result obtained led to 28 tests out of a total of 150, which meet the given condition. For the remaining difference, the data verification for the remaining 5 measurements was resumed after the errors were removed from the sample. This resulted in 31 tests that retain the number of data, and the difference of 42 experiments was retested with respect to the new condition t_{crt} = 1.71, corresponding to the string with 4 values. Finally, there were no data affected by aberrant errors. Finally, all data were tested and included in the ranges corresponding to the conditions shown in TABLE 2. The total values of the number of tested data, completed in the table: 606, 365, respectively 168, we determined them based on our own calculation method, starting from the previous total (in the first situation from 1050) subtracting the number of data removed from the sample and the number of cutting regimes, which verifies the given condition, multiplied by the number of string data (example: 606 = 1050 – 101 – 49 × 7). Graphic illustration of the statistical results obtained and presented in TABLE 2, we presented it in Figs 1-4.

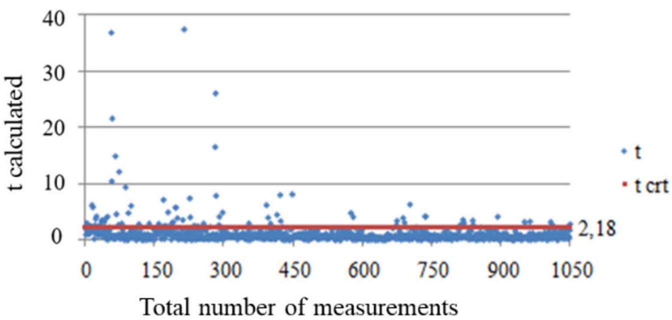

Fig. 1. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the longitudinal direction, relative to the critical value t_{crt} = 2.18

TABLE 2

Romanovski test for the identification of data affected by aberrant errors, in the longitudinal direction

No string data	The given condition	No data tested: Total	The number of data that verifies the condition	The number of data that does not verify the condition – is removed from the sample	The number of cutting regimes that check the given condition
7	t_{crt} = 2.18	1050	949	101	49
6	t_{crt} = 2.07	606	533	73	28
5	t_{crt} = 1.92	365	323	42	31
4	t_{crt} = 1.71	168	168	0	42
Total					150

The graph in Fig. 1, we made it starting from the total number of 1050 measurements. After testing this data, we calculated the test variable t for each measurement. The value obtained, we compared it with the critical value t_{crt}, corresponding to the amount of data in the string, in this case 7. The result is a total of 949 data within the required range, and the difference of 101 data exceeding the threshold of 2.18 is removed from the sample.

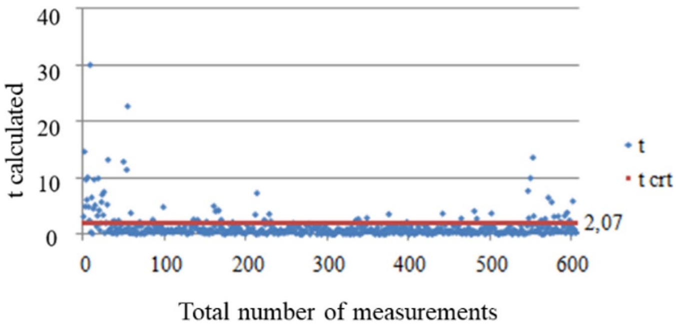

Fig. 2. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the longitudinal direction, relative to the critical value t_{crt} = 2.07

Similarly, in Fig. 2 we presented the situation of the tests redone on the remaining 606 data out of 1050. Thus, the test variables t recalculated, this time are compared with the new condition t_{crt} = 2.07. As can be seen in the graph, some of the data meet the required condition and their number is 533 measurements, and the difference exceeding the threshold will be removed from the sample.

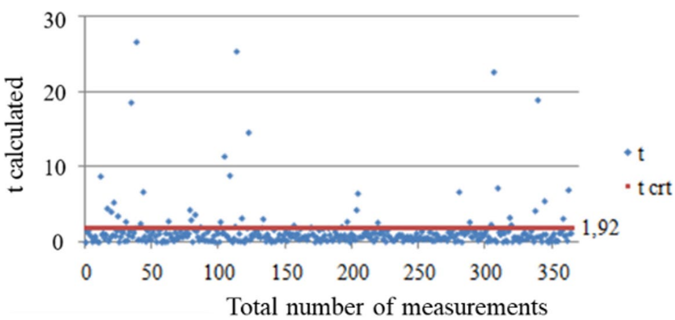

Fig. 3. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the longitudinal direction, relative to the critical value t_{crt} = 1.92

In the case of Fig. 3, the total number of data tested is 365. After comparing the results of the test variable t with t_{crt} = 1.92, the calculations showed a total of 323 data within the required range, eliminating the difference of 42 measurements that exceeds it.

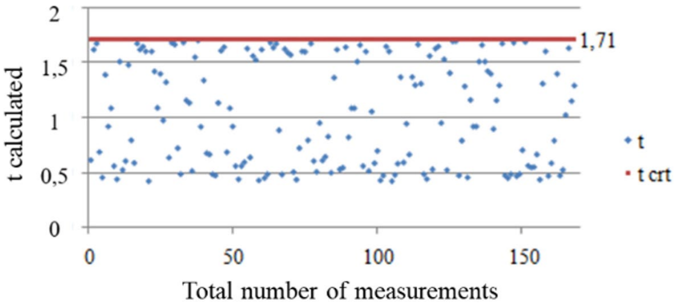

Fig. 4. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the longitudinal direction, relative to the critical value t_{crt} = 1.71

Finally, in Fig. 4, we plotted the last aberrant error check and tested the remaining 168 data. Thus, the last tested data fully satisfy the condition $t > t_{crt}$, when t_{crt} = 1.71.

4.2. R_a transversely

In the following we would like to present the situation of the obtained results, regarding the data affected by aberrant errors, in the case of the measurements of the arithmetic mean deviation R_a of the surface profile, measured transversely in the direction of the feed movement. In TABLE 3 we summarize the total number of cutting regimes that check the given conditions for each data set.

Thus, the first condition t_{crt} = 2.18, is verified by 60 tests t_{crt} = 2.07 of 26 tests t_{crt} = 1.92 of a total of 21 tests and, finally, the last condition t_{crt} = 1.71 is verified by the difference of 43 tests. The statistical results obtained and presented in TABLE 3 are graphically represented by Figs 5-8.

The graph in Fig. 5, we made it starting from the total number of 1050 measurements. After testing this data, we calculated the test variable t for each measurement. We compared the value obtained with the critical value t_{crt}, corresponding to the amount of data in the string, in this case 7. A total of 960 data falling

TABLE 3

Romanovski test for the identification of data affected by aberrant errors, in the transversely direction

No string data	The given condition	No data tested			The number of cutting regimes that check the given condition
		Total	The number of data that verifies the condition	The number of data that does not verify the condition – is removed from the sample	
7	t_{crt} = 2.18	1050	960	90	60
6	t_{crt} = 2.07	540	461	79	26
5	t_{crt} = 1.92	305	277	28	21
4	t_{crt} = 1.71	172	172	0	43
Total					150

within the required range resulted, and the difference of 90 data, exceeding the threshold of 2.18 is removed from the sample.

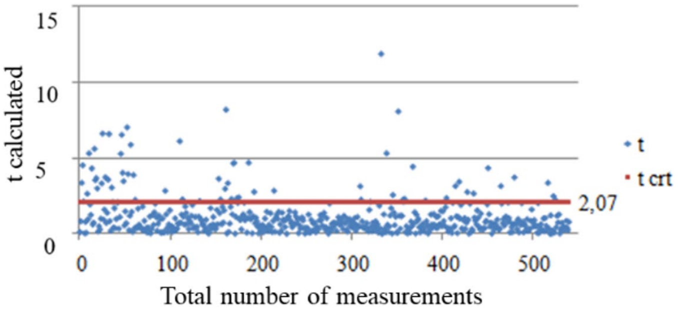

Fig. 6. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the transverse direction, relative to the critical value t_{crt} = 2.07

In Fig. 6 we presented the situation of the tests redone on the 540 data, remaining from the 1050. Thus, the test variables t recalculated, this time we compared them with the new condition, t_{crt} = 2.07. As can be seen in the graph, some of the data meet the required condition, and their number is 461 measurements. The difference above the threshold is removed from the sample.

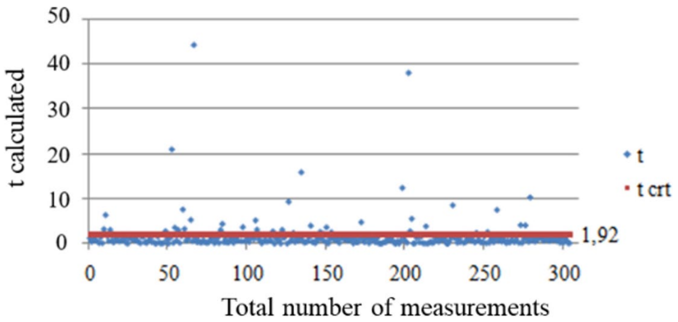

Fig. 7. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the longitudinal direction, relative to the critical value t_{crt} = 1.92

In the case of Fig. 7, the total number of data tested is 305. Thus, we compared the results of the test variable t with t_{crt} = 1.92, and the calculations obtained showed a total number of 277 data, which fall within the required range, and we eliminated the difference of 28 measurements that exceeds this range.

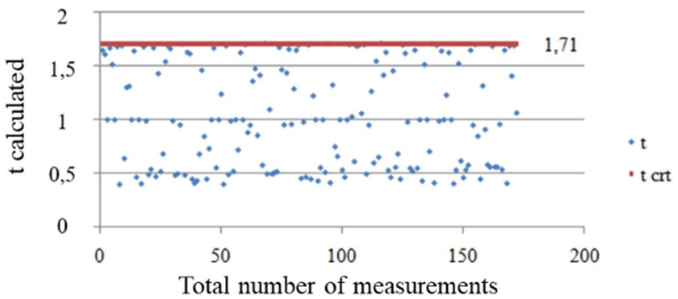

Fig. 8. Graphical representation of the verification of the existence of aberrant data on R_a measurements in the transverse direction, relative to the critical value t_{crt} = 1.71

Finally, in Fig. 8 we graphically represented the last verification of the aberrant errors testing the remaining 172 data. Thus, all experimental data satisfy the condition $t > t_{crt}$, when t_{crt} = 1.71. Based on the results obtained, in the second stage of testing the experimental data which involves checking the randomness of the data, I will use sets of 7, 6, 5 and 4 measurements, respectively, for the 150 data sets.

5. Conclusion

The experimental research carried out aims at maximizing the use of the set of experimental data, minimizing measurement errors, ensuring a maximum speed of experimentation, the minimum cost of research and maximum efficiency.

After applying the Romanovski test – we identified the values affected by aberrant errors and removed them from the sample, we repeated the testing for the remaining data strings until all values meet the conditions imposed by the test. Thus, the 150 cutting regimes, as verified by the test conditions, are grouped as follows:

- In case of arithmetic mean deviation of the surface profile, measured longitudinally in the direction of the feed movements:
  - 7-data strings – 49 cutting regimes;
  - 6 data strings – 28 cutting regimes;
  - 5 data strings – 31 cutting regimes;
  - 4 data strings – 42 cutting regimes.
- In case of the arithmetic mean deviation of the surface profile, measured transversely in the direction of the feed movements:
  - 7 data strings – 60 cutting regimes;
  - 6 data strings – 26 cutting regimes;
  - 5 date strings – 21 cutting regimes;
  - 4 data strings – 43 cutting regimes.

As further research directions, following the analysis of the identification of data affected by aberrant errors, the next steps are to verify the random nature of the experimental data and to verify the normal distribution of the experimental data.

After performing the statistical analysis of the experimental data, it will be possible to perform a regression analysis.

REFERENCES

[1] J. Wang, J. Zhang, X. Wang. IEEE Transactions on Semiconductor Manufacturing **31**, 1, 173-182 (2018).

[2] J. Li et al., IEEE Trans. Ind. Informat. **11**, 3, 612-619 (2015).

[3] A.K.Y. Jain, Y. Shrivastava. Mater. Today Proceed. **21**, 1680-1684 (2019).

[4] G. Bolar, S.N. Joshi. Mech. Eng. B – J. Eng. Manuf. (January 2017).
DOI: https://doi.org/10.1177/0954405416685387

[5] M. Wan, X.B. Dang, W.H. Zhang, Y. Yang. Mech. Syst. Signal. Process. **103**, 196-215 (2018).

[6] M. Wan, J. Feng, Y.C. Ma, W.H. Zhang. Int. J. Mach. Tools. Manuf. **122**, 120-131 (2017).

[7] I. Hanif, M. Aamir, N. Ahmed, S. Maqsood, R. Muhammad, R. Akhtar, R. Hussain, I. Int. J. Adv. Manuf. Technol. **100**, 1893-1905. (2019).

[8] A.B Pop. M. Ţîţu, New Technoloies, Development and Application II, NT 2019, Lecture Notes in Networks and Systems; Karabegovi'c, I., Ed.; Springer: Cham, Switzerland, 76 (2020).

[9] M. Aamir, M. Tolouei-Rad, K. Giasin; A. Vafadar, Machinability of Al2024, Al6061, and Al5083 alloys using multi-hole simultaneous drilling approach. J. Mater. Res. Technol. **9**, 10991-11002 (2020).

[10] D.P. Singh, R.N. Mall, Int. J. Technol. Res. Eng. **2**, 2782-2787 (2015).

[11] M.C. Santos, A.R. Machado, M.A.S. Barrozo, M.J. Jackson, E.O. Ezugwu, Int. J. Adv. Manuf. Technol. **76**, 1123 (2015).

[12] A. Abdallah, R. Bhuvenesh, A. Embark, IOSR J. Eng. **4**, 1-10 (2014).

[13] S.V. Alagarsamy, N. Rajakumar, Int. J. Appl. Res. Stud. **3**, 1-8 (2014).

[14] S. Zahoor, N.A. Mufti, M.Q. Saleem, M.P. Mughal, M.A.M. Qureshi, Int. J. Adv. Manuf. Technol. **89**, 3671-3679. (2017).

[15] D. Montgomery, Design and Analysis of Experiments. Eighth Edition ed. Hoboken: John Wiley & Sons, Inc, (2013).

[16] W. Laurenzi, Processing of experimental data, part 1. ProLigno **6** (4), 55-63 (2010).

[17] T. Mihăilescu, Operational Reserch Methods and Computer Programming in Wood Industry. Universitatea Transilvania din Braşov, (1984).