

PAPER • OPEN ACCESS

Estimation of the impact of semiconductor device parameters on the accuracy of separating a mixed production batch

To cite this article: G Sh Shkaberina *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 032088

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

Estimation of the impact of semiconductor device parameters on the accuracy of separating a mixed production batch

G Sh Shkaberina¹, V I Orlov^{1,2}, E M Tovbis¹, E V Sugak¹ and L A Kazakovtsev^{1,3}

¹Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarskiy Rabochiy av., Krasnoyarsk, 660037, Russia

²Testing and Technical Center – NPO PM, 20, Molodezhnaya St, Zheleznogorsk, 662970, Russia

³Siberian Federal University, 79, Svobodny av., Krasnoyarsk, 660041, Russia

E-mail: levk@bk.ru

Abstract. In this paper, we investigate the problem of separation of a mixed production batch of semiconductor devices for the space industry into homogeneous production batches. The method of factor analysis is applied to reduce the dimensionality of the problem. We investigate the impact of measured parameters of semiconductor devices in the accuracy of the separation of the mixed lot, composed several homogeneous batches. It was shown, that with any orthogonal rotations of factor structure as the number of homogeneous batches in the sample increases, the clustering accuracy reduces. Groups of semiconductor device parameters which have the greatest impact on the partition accuracy regardless of the number of homogeneous batches in the sample detected.

1. Introduction

In order to install space equipment with highly reliable electronic components, specialized testing centers conduct a variety of tests of each installed device. Electronic component base (ECB) designed for installation in spacecraft equipment, along with the classical input control is subjected to additional rejection tests, including a selective destructive physical analysis (DPA).

The DPA allows us to confirm the good quality of the batches of ECB or to identify the batches, which have defects due to manufacturing technology not detected during conventional rejection tests and additional non-destructive testing. In order to be able to transfer the results of the DPA of several devices for the entire batch of semiconductor devices, the following requirement is put forward for the ECB intended for installation in space equipment: all devices from the same batch must be made from the same raw materials. Manufacturers for general consumption equipment (not designed solely for use in a spacecraft) cannot guarantee the implementation of this requirement. Therefore, the problem of automatic grouping of semiconductor devices by production batches is relevant.

In paper [1] it is shown, that the problem of allocation of homogeneous batches can be further reduced to the problem of cluster analysis. Each group (cluster) must represent a homogeneous batch made from one type of raw materials. To solve the problem of identifying homogeneous batches, in papers [2,3,4], the application of the k-means clustering algorithm was proposed. In [5], the authors consider the fuzzy clustering method based on the EM algorithm. In [6], the problems of using ensembles of clustering algorithms are considered (k-means, k-medoids, k-medians, EM, as well as



their optimized versions). In [1], the authors consider the application of genetic algorithms with greedy heuristics, as well as modifications of the EM algorithm for the separation of homogeneous batches of electronic devices.

In this paper, we consider the problem of reducing the dimensionality of the original data for the corresponding problems of cluster analysis.

2. Source data

In this paper, we consider a sample consisting of seven different homogeneous batches. The sample is deliberately composed of batches, some of which are extremely difficult to separate by known methods of cluster analysis.

The total number of devices is 3987: batch 1 contains 71 devices, batch 2: 116, batch 3: 1867, batch 4: 1250, batch 5: 146, batch 6: 113, batch 7: 424. Each batch contains information about 205 measured input parameters of the device. Input parameters for which the data vector contains only zero values or for which the number of non-zero values does not exceed 10% were excluded from consideration. For further processing, 67 input parameters remain to be considered.

An analysis of the hit frequency histograms of parameters shows that the nature of the parameter distributions in different batches is identical, mean standard deviations are commensurable. We combined parameters with identical distributions into several groups and marked them as follows: parameters In10-In20 as group 1, parameters In21-In28 as group 2, parameters In39-In46 as group 3, parameters In57-In82 as group 4, parameters 84-91 as group 5, parameters In92-In107 as group 6. All measurements (parameters, dimensions) of nondestructive tests can be divided into three groups:

- A. parameters for which the histograms are represented by a Gaussian distribution: group 2 and group 3 (figure 1(a));
- B. parameters for which the histograms are represented by a Gaussian distribution with frequency gaps: group 5 (figure 1(b));
- C. parameters for which the histogram does not correspond to Gaussian distributions: group 1, group 4, group 6 (figure 1(c)).

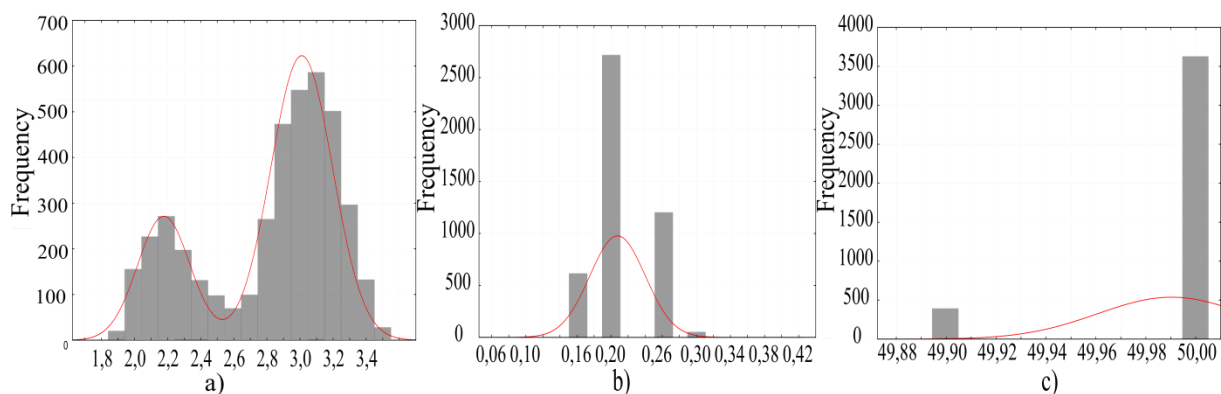


Figure 1. Histogram of observed frequencies and graphs of the distributions: a) Gaussian distribution (parameter In21); b) Gaussian distribution with frequency gaps (parameter In90); c) non-Gaussian distribution (parameter In64).

Apparently from table 1, the kurtosis criterion [7] allows us to separate group C parameters from the others. For such parameters, the values of this criterion are high (more than 10). For a normally distributed random variable, this criterion has zero expectation.

Table 1. Kurtosis criterion for parameter groups (average value).

| Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 |
|---------|---------|---------|---------|---------|---------|
| 30,3174 | -0,7848 | -0,6161 | 11,3501 | 1,1031 | 22,0853 |

3. Factor analysis

Factor analysis is based on the definition of the factor model (1).

$$X_i = \sum_{j=1}^m a_{ij}F_j + u_i \quad (1)$$

where X_i is vector of values of measured parameter ($i=1..n$), F_j - primary factors ($j=1..m$), a_{ij} are coefficients named factor loadings, u_i are characteristic (specific) factors describing the part of the parameter not included in any primary factor. When $m < n$ the reduction of the dimensionality of the original problem taking place.

Assuming the orthogonality of the factors, we obtain

$$R = A \cdot A^T \quad (2)$$

where R is a correlation matrix, A is the factor loadings matrix.

In [8] was shown, that reducing the dimension of the data vectors can be achieved by applying factor analysis without reducing the accuracy of the clustering, and, in some cases, with increasing accuracy.

To extract factors we used the principal components method, principal factor with multiple R-square method, principal axes method, maximum likelihood factors method, iterated communalities method (MINRES) and centroid method [9]. In table 2 values of total variance given by all extracted factors are presented. The eigenvalues of the factors obtained using these methods for full mixed lot are given in table 3.

In further consideration we used principal components method since it describes the maximum variance of input parameters. Various rotations of the factor structure were considered: varimax, quartimax and unrotated structure.

Table 2. Total variance given by all extracted factors (%).

| | Principal components | Multiple R-square | Principal axes | Maximum likelihood | Minres | Centroid |
|----------------|----------------------|-------------------|----------------|--------------------|--------|----------|
| Full mixed lot | 76.593 | 71.593 | 72.132 | 71.998 | 72.176 | 72.012 |
| Four batches | 66.131 | 61.204 | 61.024 | 60.304 | 61.025 | 63.332 |
| Three butches | 79.916 | 73.196 | 73.247 | 73.230 | 73.256 | 30.276 |
| Two butches | 76.031 | 70.221 | 70.104 | 69.719 | 70.111 | 70.841 |

Table 3. Eigenvalues for full mixed lot.

| Factors | Principal components | Multiple R-square | Principal axes | Maximum likelihood | Minres | Centroid |
|---------|----------------------|-------------------|----------------|--------------------|----------|----------|
| 1 | 14.00009 | 13.79306 | 13.80443 | 13.62651 | 13.80426 | 11.99882 |
| 2 | 12.44429 | 12.15745 | 12.18109 | 10.73055 | 12.18097 | 11.00877 |
| 3 | 8.68898 | 8.44518 | 8.47136 | 9.80307 | 8.47160 | 8.23812 |
| 4 | 5.37063 | 5.02919 | 5.07701 | 4.47814 | 5.07841 | 5.18165 |
| 5 | 4.06157 | 3.85231 | 3.87730 | 4.45746 | 3.87763 | 5.60322 |
| 6 | 3.50069 | 3.08871 | 3.12985 | 3.34283 | 3.12996 | 3.30473 |
| 7 | 2.04150 | 1.58570 | 1.67960 | 1.75964 | 1.68737 | 2.84382 |
| 8 | 1.01583 | 0.56490 | 0.60103 | 0.59965 | 0.60144 | 0.50902 |
| 9 | 0.95981 | 0.16670 | 0.22819 | 0.16071 | 0.24781 | 0.28027 |

4. Computational experiments

Before conducting experiments with clustering, the following hypotheses were put forward:

1) the use of factor analysis using input parameters with normal Gaussian distribution and normal Gaussian distribution with frequency gaps improve the accuracy of clustering of a mixed batch consisting of homogeneous batches;

2) input parameters that do not correspond to the Gaussian distribution, do not have a significant impact on the clustering accuracy of the mixed lot.

Various variants of the mixed lot consisting of 2, 3, 4 and 7 homogeneous batches were considered to confirm the hypotheses. Different groups of parameters were consistently excluded from the initial set of input parameters.

Clustering was performed by EM algorithm and by self-organized Kohonen maps (SOM) with Deductor Studio Academic tool. EM algorithm [10] applied with lower bound of likelihood = 0,2, level of accuracy = 10^{-5} , maximum of iterations=100. Self-organizing Kohonen maps (SOM) [11] applied with linear initialization with eigenvalues, bubble neighborhood function, significance level = 0,1%. The clustering accuracy for considered mixed lots with different orthogonal rotations is presented in table 4. Clustering accuracy is calculated as a total percentage of exact hits of the algorithm among all clusters. In some cases (as a rule, for 2 and 3 batches) the separating could not be carried out because only one cluster was found.

Table 4. Clustering accuracy, %

| Number of homogeneous batches in the mixed lot | EM unrotated | EM varimax | EM quartimax | SOM unrotated | SOM varimax | SOM quartimax |
|--|--------------|------------|--------------|---------------|-------------|---------------|
| Full mixed lot | | | | | | |
| 7 | 43 | 43 | 45 | 14 | 15 | 34 |
| 4 | 97 | 97 | 97 | 79 | 92 | 67 |
| 3 | 97 | 91 | 93 | 83 | 68 | 82 |
| 2 | 98 | 94 | 100 | 98 | 83 | 100 |
| Without group of parameters In10-In20 (group 1) | | | | | | |
| 7 | 36 | 41 | 38 | - | 14 | 14 |
| 4 | 97 | 76 | 76 | 74 | 91 | 72 |
| 3 | 97 | 93 | 93 | 85 | 71 | 70 |
| 2 | 98 | 80 | 100 | 98 | 83 | 81 |
| Without group of parameters In57- In82 (group 4) | | | | | | |
| 7 | 41 | 24 | 26 | - | 17 | 14 |
| 4 | 69 | 74 | 76 | - | 88 | 88 |
| 3 | - | - | - | - | - | - |
| 2 | - | - | - | - | - | - |
| Without group of parameters In84- In91 (group 5) | | | | | | |
| 7 | 33 | 33 | 38 | - | 14 | 14 |
| 4 | 81 | 79 | 80 | 67 | 63 | 82 |
| 3 | 66 | 89 | 64 | 60 | 77 | 75 |
| 2 | 98 | 80 | 100 | 98 | 80 | 97 |
| Without group of parameters In92- In107 (group 6) | | | | | | |
| 7 | 45 | 40 | 38 | 14 | 27 | 22 |
| 4 | 98 | 42 | 72 | 77 | 91 | 61 |
| 3 | 100 | 100 | 100 | 78 | 78 | 89 |
| 2 | - | - | - | - | - | - |
| With normal distribution only (without groups 1,4,6) | | | | | | |
| 7 | 43 | 35 | 38 | 7 | 7 | 7 |
| 4 | 98 | 69 | 69 | 98 | 99 | 92 |
| 3 | 100 | 100 | 100 | 74 | 67 | 74 |
| 2 | - | - | - | - | - | - |

In addition, the variants of excluding each homogeneous batch from the full mixed lot are considered with all input parameters (figure 2). The factors obtained by variants of orthogonal rotations with cumulative variance 60% and 70% were used as input data. Clustering was performed by EM algorithm.

Thus, EM algorithm worked better than the SOM. For the SOM algorithm clustering performed with higher accuracy for varimax rotation, while for EM algorithm varimax has lower accuracy, and higher accuracy was achieved by unrotated factor structure.

Clustering accuracy increased when number of batches in the mixed lot reduced. Parameters exclusion had the least impact on clustering accuracy in case of 2 and 3 batches. For 4 and 7 batches parameters exclusion led to reducing clustering accuracy.

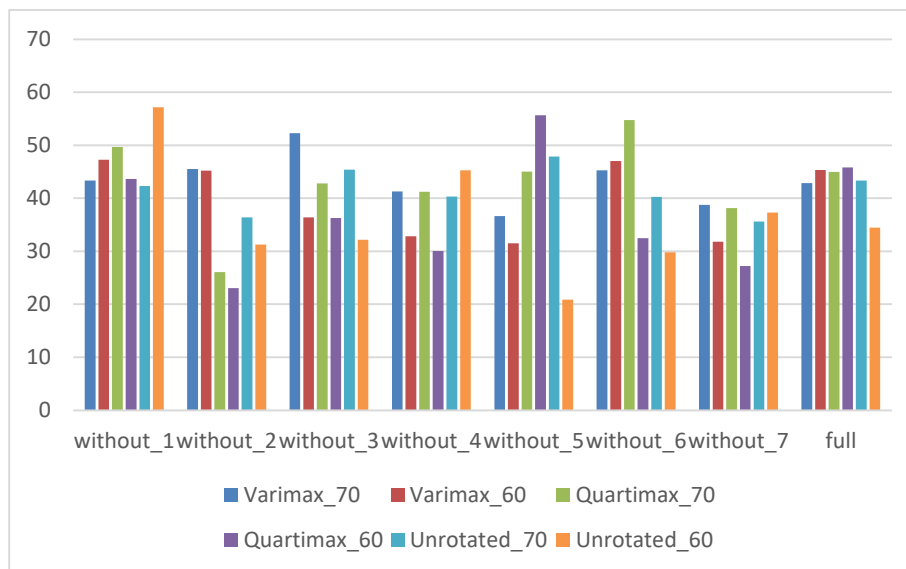


Figure 2. Clustering accuracy with excluding batches 1-7 and full mixed lot respectively, variants of rotations with 60% and 70% of total variance, %.

5. Conclusion

The strongest impact on the clustering results was the exclusion of a group of parameters with a normal distribution with frequency gaps from the factor model, the accuracy decrease was 9.9% in average. The least impact was the exclusion of a group of parameters, which do not correspond to Gaussian distribution, the accuracy decreased in average for 3.3%. Exclusion of homogeneous batches from a mixed lot led to frequency gaps in frequency histograms, as a result clustering accuracy reducing.

Thus, it was shown experimentally that the use of factor analysis using the input parameters with Gaussian distribution and Gaussian distribution with frequency gaps does not increase the accuracy of the clustering of the mixed lot consisting of homogeneous batches. However, the clustering accuracy decreases slightly.

References

- [1] Orlov V I, Kazakovtsev L A, Masich I S and Stashkov D V 2017 *Algorithmic Support of Decision-making on Selection of Microelectronics Products for Space Industry*. Krasnoyarsk: Siberian State Aerospace University)
- [2] Ackermann M R *et al* 2012 StreamKM: A clustering algorithm for data streams. *J. Exp. Algorithmics*. **17** doi: 10.1145/2133803.2184450
- [3] Kanungo T, Mount D M, Netanyahu N S, Piatko C D, Silverman R and Wu A Y 1999 Computing nearest neighbors for moving points and applications to clustering *Proc. of the tenth annual ACM-SIAM symp. on Discrete algorithms (Society for Industrial and Applied Mathematics)* 931-2

- [4] Kazakovtsev L A, Stupina A A and Orlov V I 2014: Modification of the genetic algorithm with greedy heuristic for continuous location and classification problems. *Control systems and information technologies* **2(56)** 31-41
- [5] Orlov V I, Stashkov D V, Kazakovtsev L A and Stupina A A 2016 *IOP Conf. Ser.: Mater. Sci. Eng.* **155** 012026 doi: 10.1088/1757-899x/155/1/012026
- [6] Rozhnov I, Orlov V and Kazakovtsev L 2018 Ensembles of clustering algorithms for problem of detection of homogeneous production batches of semiconductor devices. *CEUR Workshop Proceeding* **2098** 338-48 <http://ceur-ws.org/Vol-2098/paper29.pdf>
- [7] Hazewinkel M 2001 *Excess Coefficient Encyclopedia of Mathematics* (Springer)
- [8] Shkaberina G, Orlov V, Tovbis E and Kazakovtsev L 2019 Factor analysis with use of the Spearman matrix in the problem of automatic grouping of electronic radio components on production batches *Control systems and information technologies* **1(75)** 91-6
- [9] Uberla K 1977 *Factorenanalyse* (Berlin: Springer-Verlag) doi: 10.1007/978-3-642-61985-4
- [10] Dempster A, Laird N and Rubin D 1977 Maximum likelihood estimation from incomplete data *J. Royal Stat. Soc.*, **39** 1–38
- [11] Kohonen T 1980 Self-organized formation of topologically correct feature maps *Biological Cybernetics* **43(1)** 59–69 doi:10.1007/bf00337288
- [12] Kaiser H F and Dickman K 1962 Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix *Psychometrika* **27** 179-82 doi: 10.1007/bf02289635
- [13] Cattell R B 1966 The scree test the number of factors *Multivar. Behave. Res.* **1** 245-76 doi: 10.1207/s15327906mbr0102_10
- [14] Harman H 1967 *Modern Factor Analysis* (Chicago: The university of Chicago press) doi: 10.1002/bimj.19700120119
- [15] Thurstone L L 1961 *Multiple Factor Analysis* (Chicago)
- [16] Afifi A and Azen S 1973 Statistical analysis. A computer oriented approach *Biometrics* **29(3)** 603
- [17] Lawley D and Maxwell A 1963 *Factor Analysis as a Statistical Method* (London: Butterworths)
- [18] Ledesma R D, Valero-Mora P and Macbeth G 2015 The Scree Test and the Number of Factors: a Dynamic Graphics Approach *Span. J. Psychol* **18** 1-10 doi: 10.1017/sjp.2015.13
- [19] Brown T 2006 *Confirmatory Factor Analysis for Applied Research* (New York: The Gilford press)
- [20] De Roover K, Timmerman M E, De Leersnyder J, Mesquita B and Ceulemans E 2014 What's hampering measurement invariance: detecting non-invariant items using clusterwise simultaneous component analysis *Front. Psychol.* **5** 604 doi: 10.3389/fpsyg.2014.00604