

Application of Cluster Analysis for Polymer Classification According to Mechanical Properties

Danijela Pezer

University Department of Professional Studies, University of Split, Croatia
dpezer@oss.unist.hr

Abstract. The selection of the appropriate polymer material represents one phase in the product design, both, in terms of data on the different properties of the material and for detailed assessments and selection of the best material for the selected product. In the material selection phase, it is necessary to define criteria related to the mechanical properties of the material (strength, stiffness, toughness). Another important item is certainly the stress-strain diagram by which the materials can be classified into certain groups, based on their mechanical properties. The paper presents a study of materials classification according to the following groups: *Hard and brittle*, *Hard and strong*, *Hard and tough* and *Soft and tough* materials, using the cluster analysis (k-means algorithm).

Keywords. Cluster analysis, k-means algorithm, polymer materials, mechanical properties.

1. Introduction

Polymer materials are significantly represented in industry, but also in everyday use. Due to the reasonable price, processing and recovery possibilities and the variety of mechanical properties, polymers have become one of the most important technical materials today. In addition to technical applications, polymer materials are used in many other industries such as pharmaceuticals, food, textile, etc. The design can ensure functionality, safety and product life. The role of the engineer in selecting a suitable material for a particular product is significant [1]. The conditions of product application determine a material choice. For this purpose, it is necessary to roughly define the conditions that need to be met, which relate to the characteristics/properties and scope. With this information, engineers can compare technical requirements and make an assessment. In the recommendations, existing information (available through online databases) is compared with technical data and industry specific empirical values. Once the optimal material for an individual application has been determined, during the part/component design phase, it is compared with previously performed calculations. Based on defined/established criteria, it is possible to limit the choice of suitable materials. However, selection can only be a recommendation, it cannot replace practical examination/testing. Basic questions that are asked when choosing materials: Is plastic (polymer materials) suitable for this application at all? Why plastic? Has any other material been used before? If other material has been used before, why make changes?

The choice of materials for different applications is usually based on mechanical properties. Knowledge of the mechanical properties of the material is very important for the design application, and they can be determined by conducting an experimental test on specimens. Mechanical properties determine the behavior of a material due to the action of force. Material testing is performed for

several reasons, for example to determine the suitability of the material for a particular application, or to determine the values of strain (or stress values) by which the stress-strain graph is drawn.

Mechanical properties depend on the environmental temperature and the type of polymeric material. As the temperature increases, the tensile strength and the modulus of elasticity decrease, while the intermittent elongation decreases sharply after the maximum is reached. Amorphous polymer materials change with increasing temperature pass from the glassy to the rubbery state, then pass to the liquid state, while the crystalline plastomers pass from the glassy state to the plastic state and after that into the liquid state.

The paper deals with the classification of polymer materials, i.e. classifying into four groups based on selected mechanical properties. The most commonly used polymer materials from the group of plastomers (amorphous and crystalline structures) and duromers were selected for classification.

Based on the presented curves, from stress-strain diagram, and the associated materials according to the above groups, the paper uses cluster analysis, i.e. using the k-means algorithm, the materials are grouped into 4 defined clusters.

2. Stress and strain analysis for polymer materials

Each material has a specific stress-elongation curve, most often with respect to stiffness and yielding point. When conducting a tensile test on the same material, the values of the yield strength differ, depending on the temperature and the load speed. To better understand the stress-strain curve, the following terms are clarified according to ISO 527-1 [2] which can see on Figure 1:

- Stress – normal force applied to a material, (MPa)
- Strain – material deformation from stress which changes the dimension of material, (%)
- Stress at break – stress at which specimens break, (MPa)
- Yield strain – increasing deformation without increasing stress, (%)
- Yield stress – stress at which material permanently deforms, (MPa)
- Strain at break – strain at last recorded data point before the stress reduction, (%)
- Tensile modulus (Young's modulus) – the relationship between the deformation of material and the force required for its deformation, (MPa).

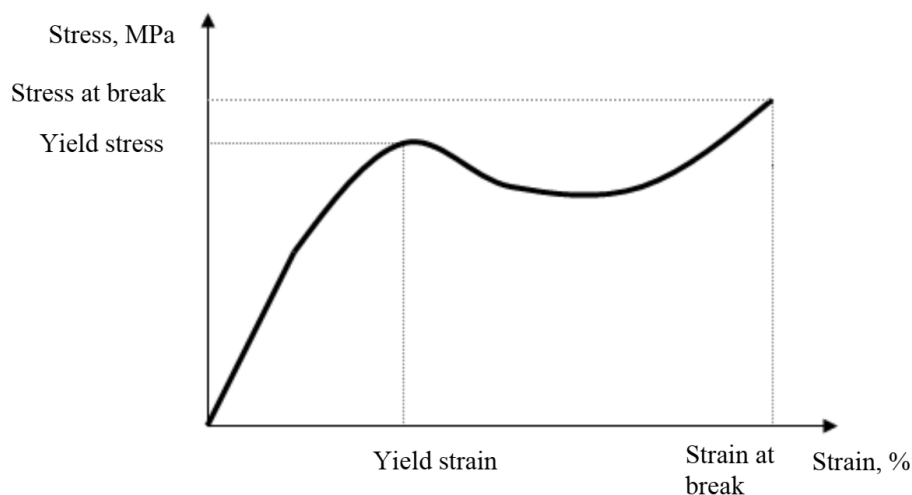


Figure 1. Stress-strain diagram

In this paper, will be consider a widely used four groups for polymer materials classification (shown on Figure 2), which not consider elastomers:

Hard and brittle material (curve a) - characteristic of this group of polymers is a high modulus of elasticity and tensile strength, low strain at break, have no pronounced yield strength (yield before breaking).

Hard and strong material (curve b) - characteristic of this group of polymers is high modulus of elasticity, high yield stress and tensile strength and low strain at break.

Hard and tough material (curve c) - characteristic of this group of polymers is high modulus, high yield stress, high tensile strength and high strain at break.

Soft and tough material (curve d) - characteristic of this group of polymers is low modulus, low yield stress, high strain (elongation) and high stress at break.

In the Table 1 are presented the general stress-strain curve characteristics.

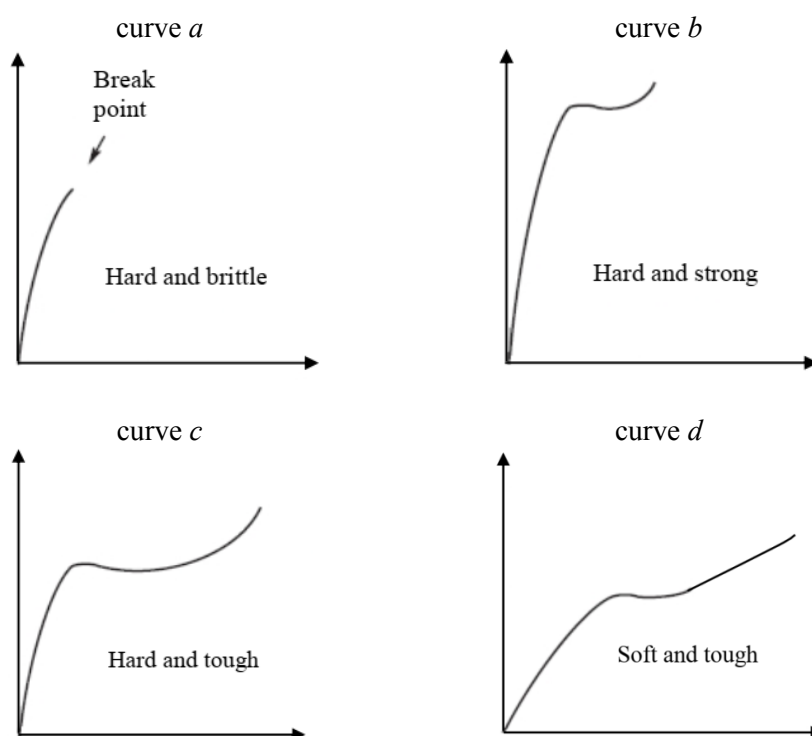


Figure 2. Stress – strain curves (*a – d*) for polymer materials classification

Table 1. Stress-strain curve characteristics

Polymer	Stress-strain curve characteristics		
	Tensile modulus	Yield stress	Strain at break
Hard and brittle	High	--	Low
Hard and strong	High	High	Medium
Hard and tough	High	High	High
Soft and tough	Low	Low	High

3. Methods

3.1. Cluster analysis

The quality of clustering data into clusters depending on the type of methods and data structure. Different methods can be of different complexity and separateness of clusters. The basic classification of clustering method is to the hierarchical and non-hierarchical (partition).

Cluster analysis allows object groupings which are based on the characteristic features and it is also possible grouping variables. The objects, that formed clusters, are as similar as possible within the clusters, and between the clusters the differences are large as possible.

Basic principle of the clustering problem the elements of the set $\mathcal{A} = \{a_1, \dots, a_m\}$ with $m \geq 2$ elements, where is $a_i \in \mathbb{R}^n$, $i = 1, \dots, m$, into disjoint subsets π_1, \dots, π_k , $1 \leq k \leq m$, so that [3-6]:

- $\bigcup_{i=1}^k \pi_i = \mathcal{A}$,
- $\pi_i \cap \pi_j = \emptyset$, $i \neq j$,
- $m_j := |\pi_j| \geq 1$, $j = 1, \dots, k$.

The set \mathcal{A} subsets (π_1, \dots, π_k) will be labeled as $\Pi = \{\pi_1, \dots, \pi_k\}$ and called the set \mathcal{A} partition while (π_1, \dots, π_k) will be called clusters.

When defining the distance between objects, proximity measures will be applied, which are divided into measures of similarity and difference. Two objects are closer if the difference or distance between them is smaller and the similarity is greater [6].

The following distances measures can be used to calculate differences between objects [7]:

- The Euclidean distance

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \|x - y\|_2, \quad (1)$$

- The Squared Euclidean distance

$$d(x, y) = \sum_{i=1}^n (x_i - y_i)^2, \quad (2)$$

- Manhattan distance

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| = \|x - y\|_1, \quad (3)$$

- The Chebychev distance

$$d(x, y) = \max |x_i - y_i|. \quad (4)$$

It is necessary to define objective criterium function [3-5], when measuring the quality of the obtained clusters and the most often used objective function is the sum of the squared error which can be defined as:

$$\mathcal{F}(\Pi) = \sum_{j=1}^k \sum_{a_i \in \pi_j} \|c_j - a_i\|_2^2, \quad (5)$$

where $\|\cdot\|_2$ denotes the Euclidean norm and $c_j = \frac{1}{|\pi_j|} \sum_{a_i \in \pi_j} a_i$, $j = 1, \dots, k$ represents the centre of the cluster π_j .

The main objective of classification by cluster algorithm is to find the optimal partition Π^* . Optimal partition has a feature that the cluster elements sum of leaving the center is minimal. That is how is made to make the distance between the clusters larger and within the clusters smaller. In the case when we do not have a good initial approximation, k-means algorithm with various random initializations is recommended [6].

3.2. K-means algorithm

Unlike hierarchical, non-hierarchical method (k-means algorithm), allows moving objects from earlier formed clusters. K-means method or algorithm, requires pre-select the number of clusters k and the center (centroid) for each cluster. Then it is necessary to determine the distance of the object from the center of the cluster and join the object to the nearest cluster. It is necessary to calculate the new centroids within the clusters (for cluster to which is assigned an object, as for the cluster from which

object leave), and repeat the steps until the criterion of optimality are not fulfilled [4, 7, 8]. K-means algorithm is the most commonly used algorithm due to its simplicity and speed, and also processing capabilities the larger data number.

Steps of a standard k-means algorithm:

Step 1: Read in m, k , cluster \mathcal{A} and select initial centroids c_1^0, \dots, c_k^0 ;

Step 2: Determine the initial partition applying the principle of minimal distances, calculate the centroids c_1, \dots, c_k of the cluster π_1, \dots, π_k and initial value of the goal function $\mathcal{F}_0 = \mathcal{F}(\Pi)$;

Step 3: Form a new partition $\mathcal{N} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ with new clusters, their centroids ζ_1, \dots, ζ_k and the new value of the goal function $\mathcal{F}_1 = \mathcal{F}(\mathcal{N})$;

Step 4: If $\mathcal{F}_1 < \mathcal{F}_0$, put $c_j = \zeta_j$; $j = 1, \dots, k$; $\mathcal{F}_0 = \mathcal{F}_1$ and go to Step 2. Otherwise, STOP.

The k-means algorithm basic limitations [4,5]:

- The number of clusters must be predetermined (it can be defined from the nature of data or based on a previously conducted hierarchical method of cluster analysis)
- The k-means algorithm result depends considerably on initial selection of cluster centroids [3].

4. Analysis results

Based on the presented curves in Figure 2 and the associated materials according to the above groups, in this paper was conducted a cluster analysis, i.e. it was used the k-means algorithm, for 16 selected polymer materials which are grouped into 4 defined clusters.

Classification is defined through the four clusters according to the following material characteristics:

- *Hard and brittle* materials (curve *a* at Figure 2 – represents cluster A)
- *Hard and strong* materials (curve *b* at Figure 2 – represents cluster B)
- *Hard and tough* materials (curve *c* at Figure 2 – represents cluster C)
- *Soft and tough* materials (curve *d* at Figure 2 – represents cluster D)

For classification, the most commonly used polymer materials from the group of elastomers (amorphous and crystalline structures) and diureters were selected.

The stress-strain curve for amorphous elastomers and diureters is very similar, and these materials belongs to the *Hard and Brittle* group of materials. The most common criterion (mechanical property) in phase of choosing polymer materials is tensile strength, elastic modulus and yield stress.

In this paper, materials are grouped according to the following characteristics (variables): tensile modulus (MPa), yield stress (MPa), yield strain (%), stress at break (MPa) and strain at break (%).

In the Table 2, mechanical properties (measured at 23°C) for the selected materials mentioned above, are shown. The data from Table 2 were taken from the CAMPUS database [9], available online, and the tests were performed according to ISO 527-1/-2 standard. The tests on specimens according to ISO 527-1/-2 standard are performed on mechanical test rippers, where a force-elongation diagram is drawn, and based on that a stress-strain diagram is obtained.

Table 2. Mechanical properties of selected polymer materials and classification to the defined clusters

No.	Material	Properties	Tensile modulus, MPa	Yield stress, MPa	Yield strain, %	Stress at break, MPa	Strain at break, %	curve a-d (figure 2)	k-means algorithm
1.	LDPE	tough, flexible	130	11	15	35	550	d	cluster D
2.	HDPE	hard, tough	1200	26	11	30	600	d (c)	cluster D
3.	PP	tough	1800	38	8	20	150	c	cluster C
4.	PS	rigid, strong	3300	-	-	59	3	a	cluster A
5.	SAN	rigid	3700	-	-	70	2	a	cluster A
6.	PA66	strong, hard, tough	3500	85	3,5	65	150	b (c)	cluster B
7.	PA12	strong, tough	1900	55	11	38	180	c	cluster C
8.	PPA	strong, stiff	3200	86	6	80	15	b	cluster B
9.	POM	rigid, tough	3000	62	10	66	40	b	cluster B
10.	PBT	strong, (tough)	2550	55	3,5	45	8	b (c)	cluster C
11.	PMMA	rigid, strong	3300	-	-	54	2,5	a	cluster A
12.	PC	strong, (tough)	2400	66	6	60	120	b (c)	cluster B
13.	PSU	rigid, semi-tough	2500	76	5,9	76	50	b	cluster B
14.	PET	strong, stiff	2500	55	4	57	30	b	cluster B
15.	ABS	rigid, tough	2500	46	2,6	38	20	c	cluster C
16.	UP	rigid, strong	4800	-	-	60	0,8	a	cluster A

Cluster analysis, using the k-means algorithm (for 4 defined clusters), was performed using the software package STATISTICA, on the data given in Table 2. Since the k-means algorithm converges to the local minimum, and the choice of initial cluster centers plays a significant role [6], the above algorithm was run 150 times, with various random initial approximations (available through the specified software) to achieve a solution as close as possible to the global optimum. Euclidean and Squared Euclidean Distance, Manhattan and Chebychev distance was used as measures of distance. The solution with the smallest square error, that is, the smallest deviation of the cluster elements around its center, was obtained using the Squared Euclidean Distance.

The analysis of variance is a good indicator of the convenience of classification, which indicates the persistence of statistically significant differences between the formed clusters. If the variability within the cluster is as small as possible, and between the clusters large as possible, classification is satisfactory. The size F is used as an indicator of how much a single, analyzed variable makes a difference (allows the separation) [6] among the clusters (Table 3).

Thus, it is evident that the variable *Stress at break* has the greatest effect on the separation between clusters, i.e. makes a difference between them. The variables *Yield strain* and *Strain at break* have the

least influence on the separation between clusters. *Tensile modulus* and *Yield stress* variables have approximately equal impact on the separation of polymer materials, i.e. classification into individual groups.

Table 3. Analysis of variance for defined variables

Variable	Analysis of variance				F	signif. p
	sum of squares between clusters	degrees of freedom	sum of squares within clusters	degrees of freedom		
Tensile modulus	125324939	3	3621825	12	138,418	0,000000
Yield stress	40869	3	1119	12	146,1124	0,000000
Yield strain	699	3	81	12	34,6415	0,000003
Stress at break	47229	3	1085	12	174,1832	0,000000
Strain at break	673987	3	38995	12	69,1358	0,000000

The Figure 3 represents a graph of means for observed variables. Clusters A, B, C and D were determined based on the characteristic stress-strain diagram of the basic groups of polymer materials (curve *a-d* from Figure 2) at tensile (short-term statistical) load. Cluster A represents a group which contain *Hard and brittle* polymer materials. *Hard and strong* materials are classified into the cluster B. *Hard and tough* materials are classified in cluster C, while *Soft and tough* materials are classified in cluster D.

There was deviation of the PBT material from the originally envisaged group, however, depending on the mechanical properties and the amount of additives (as well as the selected manufacturer), in practice it would not be wrong to classify PBT material into group C (*Hard and tough*). For that reason, that possibility is indicated in parentheses in Table 2.

Cluster A includes four materials (PS, SAN, PMMA, UP). Cluster B includes six materials (PPA, POM, PSU, PET, PA66, PC). Four materials (PP, PA12, PBT, ABS) are classified in cluster C. Cluster D includes two materials (LDPE, HDPE).

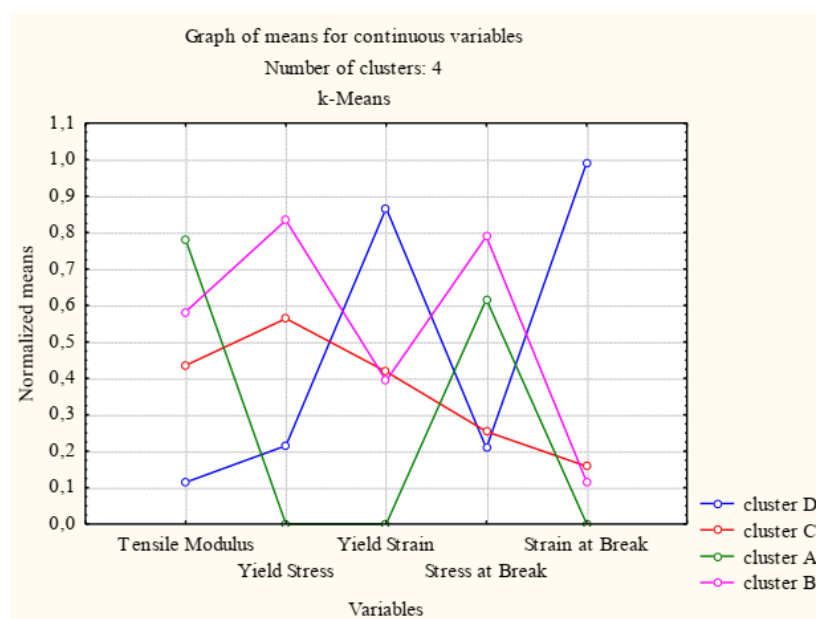


Figure 3. Graph of means for five continuous variables

Variable 1 - Tensile Modulus (Figure 3)

Based on the diagram (Figure 3) of the mean input variables for each cluster, it can be seen that the *tensile modulus* of cluster D (*Soft and tough*) is significantly lower than the *tensile modulus* of the remaining materials arranged in clusters A, B and C, which is defined by the basic characteristic stress-strain diagram. Tensile modulus is most pronounced for materials from group A (*Hard and brittle*) followed by the materials representing the clusters B and C.

Variable 2 - Yield Stress (Figure 3)

Materials which belonging to cluster B and C have a pronounced *yield stress*. Materials with a less pronounced *yield stress* are classified into cluster D, while materials from the cluster A do not have a *yield stress* (*Hard and brittle*).

Variable 3 - Yield Strain (Figure 3)

Yield strain is most pronounced for materials from the cluster D (*Soft and tough*) and they have the highest value of tensile elongation. Half less elongation is characteristic for materials from the cluster B (*Hard and strong*) and C (*Hard and tough*), while materials from the group A (*Hard and brittle*) do not have pronounced the *yield strain value* (that variable is 0).

Variable 4 - Stress at Break (Figure 3)

The lowest values of *Stress at break* are for the materials from group D (*Soft and tough*), while the highest value is for materials from group B (*Hard and strong*) followed by materials from group A (*Hard and brittle*) and materials from group C (*Hard and tough*) which also corresponds to the characteristic stress-strain diagram for the specified groups of polymer materials.

Variable 5 - Strain at Break (Figure 3)

Strain at break values are the lowest for the materials that belongs to the first group – cluster A (*Hard and brittle* materials), then follow the materials belonging to the cluster B (*Hard and strong*) and C (*Hard and tough*). *Strain at break* value is the most pronounced for materials which belongs to the group D (*Soft and tough*).

All mentioned above confirmed also the already known cognitions about polymer materials also presented in a graphical representation - four characteristic curves (Figure 2), which shows that the k-means algorithm is an appropriate method for polymer materials classification based on mechanical properties.

5. Conclusion

Suitable design of plastic products implies meeting product specifications according to the required conditions, taking into account the cost of production, and ultimately the possibility of recycling. Choosing a suitable material is a very challenging task given the wide range of polymer materials available. When choosing a material, of course, the function of the product plays a very important role, which determines whether the material should be rigid, flexible, elastic, etc.

Depending on the desired mechanical properties, obtained based on the stress-strain curve, it is possible to group materials based on of the following characteristics: tensile modulus, yield strain, yield stress, stress at break and strain at break. In this way, the materials can be classified into the four the most represented groups: *hard and brittle*, *hard and tough*, *hard and soft* and *soft and tough*.

In this paper, materials are classified according to the previously mentioned characteristics, where the most common polymer materials are grouped into 4 clusters. The k-means algorithm was used. The applied method, proved to be effective (with high accuracy) for grouping polymer materials based on the mechanical properties.

Cluster analysis can play a significant role when comparing materials characteristics in the product development process, speeding up the process, as well as simplification to young engineers who face with many challenges in this area. It would certainly be interesting to observe the characteristics of the materials on a significantly large number of materials, i.e. for a significantly larger database.

References

- [1] D. PEZER: Plastic Products Design: Today's Challenges from the Aspect of Environmental Protection. In: Contemporary issues in economy & technology - CIET2020, *Split*, June 2020, 453-461.
- [2] ISO 527-1, International Standard, Plastics – Determination of tensile properties, Part 1: General principles, International Organization for Standardization, Reference number: ISO 527-1:2012(E), 15 February 2012.
- [3] S. BANDYOPADHYAY, U. MAULIK: An evolutionary technique based on K – Means algorithm for optimal clustering in R^N . *Information Sciences*, ISSN: 0020-0255, 146 (1), 221-237 (2002). [https://doi.org/10.1016/S0020-0255\(02\)00208-6](https://doi.org/10.1016/S0020-0255(02)00208-6).
- [4] A.M. BAGIROV: Modified global k-means algorithm for minimum sum-of-squares clustering problems. *Pattern Recognition*, ISSN: 0031-3203, 41 (10), 3192-3199 (2008). <https://doi.org/10.1016/j.patcog.2008.04.004>.
- [5] M.E. CELEBY, H.A. KINGRAVI, P.A. VELA: A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, ISSN: 0957-4174, 40 (1), 200-210 (2013). <https://doi.org/10.1016/j.eswa.2012.07.021>.
- [6] T. ŠARIĆ, K. ŠIMUNOVIĆ, D. PEZER, G. ŠIMUNOVIĆ: Inventory classification using multi – criteria ABC analysis, neural networks and cluster analysis. *Technical Gazette*, ISSN: 1330-3651, 21 (5), 1109-1115 (2014).
- [7] P. TREBUNA, J. HALČINOVA: Experimental modelling of the cluster analysis processes. *Procedia Engineering*, ISSN: 1877-7058, 48, 673-678 (2012). doi: 10.1016/j.proeng.2012.09.569.
- [8] A.K. JAIN: Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, ISSN: 0167-8655, 31 (8), 651-666 (2010). <https://doi.org/10.1016/j.patrec.2009.09.011>.
- [9] CAMPUS (Computer Aided Material Preselection by Uniform Standards) database, Frankfurt, Germany 2021. [Online]. Available: <https://www.campusplastics.com/>. [Accessed April 2021].