The use of discriminant analysis methods for diagnosis of the causes of differences in the properties of resin mortar containing various fillers

Bernardeta Dębska[1,*]

[1]Rzeszow University of Technology, The Faculty of Civil and Environmental Engineering and Architecture, Poznanska 2, 35-082 Rzeszow, Poland

Abstract. Resin mortars belong to the group of concrete-like construction composites. They are obtained by mixing a synthetic resin with a hardener and an appropriately selected aggregate. The latter component is usually as much as 90% of the composite mass and can largely shape the characteristics of the finished product. The fact that the type of filler used can significantly differentiate the values of physical and mechanical parameters of epoxy mortars is confirmed by the results of the exploratory data analysis method used in this article, which is discriminant analysis. This allows us to examine differences between groups of objects based on a set of selected independent variables (predictors). It is used to solve a wide range of classification and prediction problems. The core of discriminant analysis is a model presented in the form of a linear combination of independent variables, which allows classification of observations (e.g. test mortars) into one of the groups that are of interest to the researcher. In discriminant analysis one can distinguish the learning stage (model building), in which classification rules are created based on research results (training set) and the classification stage, i.e. the use of the model, e.g. for testing its prognostic accuracy.

1 Introduction

Polymer mortars are composites in which the cement binder has been completely replaced with synthetic resins, usually chemically hardened ones. In the mortars, their content varies from 8 to 20% by volume. The remaining part is mostly filled with mineral quartz aggregate [1–3]. Aggregate therefore accounts for about 90 percent of the resin composite, which is why it has an enormous impact on its quality. For the production of resin mortars, durable and clean aggregate is used. The mechanism of the aggregate-binder interaction in resinous concrete has not yet been explained, therefore it is difficult to give a detailed description of the aggregate selection process. However, when designing a new composite, one should certainly consider such aggregate features as: type (physical and chemical properties), grain size and degree of contamination and moisture. In practice, natural

* Corresponding author: bdebska@prz.edu.pl

aggregates (gravel aggregates, mineral aggregates from natural resources) and broken aggregates (produced from solid rocks, i.e. basalt, granite, sandstone, dolomite, limestone, etc.) are used for the production of resin concrete [1, 4]. There are also studies [5–7] on the possibility of obtaining resin mortars with the addition of other types of fillers.

Resin mortars are characterized by good chemical resistance and high durability. In addition, they are characterized by a very good seal, excellent adhesion to other building materials, and a short time needed to reach operating efficiency. The results of the tests described in this article prove that the use of aggregates other than quartz sand can significantly affect the parameters characterizing the mortars, including primarily their strength characteristics. In addition, it appears that a lighter composite, characterized by greater thermal and acoustic insulation, can also be obtained. The type of filler used therefore differentiates mortar properties. Deciding which features describing the mortar distinguish (discriminate) two or more naturally separating groups, enables the use of a statistical method called discriminant analysis.

Discriminant analysis allows us to examine differences between groups of objects based on a set of selected independent variables (characteristics, attributes, predictors). The group of statistical techniques referred to as discriminant analysis applies to a whole range of research and prediction problems, among others, due to the use of the fairly simple mathematical model underlying it. At its core is a linear combination of independent variables (also known as discriminatory variables), which allows us to classify observations (for example, of a tested mortar) to one of the groups that are of interest to the investigator [8].

Two main stages can be distinguished in discriminant analysis:
1. The learning stage (model building) in which classification rules are created based on research results (training set).
2. The classification stage (using the model) in which the set of objects whose membership is unknown is classified based on the class characteristics found earlier.

In the first stage of application of discriminant analysis, procedures describing and interpreting intergroup differences are launched. Subsequently, the procedures for classifying cases are carried out, i.e. based on the observation or experiment obtained from the values of the features to which the new case belongs. This task consists in determining the canonical discriminative functions separating the studied groups. In the case of differences between groups, each of them can be treated as a cloud of points in the space with axes that are discriminating variables. These point clouds may overlap slightly, but most of the points are concentrated in centroids spaced apart, i.e. fictitious points whose coordinates are equal to the group mean of each discriminating variable. It is accepted that the centroids are typical representatives of each group.

The main purpose of discriminant analysis is to predict the group to which the classified case belongs. All classification procedures use a case-by-case comparison with each calculated centroid to find the closest one. The classification process is associated with the creation of one or more functions, classifying the analysed cases into appropriate groups. This is the so-called linear discrimination consisting of a linear combination of simple separating classified groups. This was first put forward by Ronald Fisher, who introduced a separate linear character combination for each i-th group (1):

$$K_i = a_{io} + a_{i1}*x_1 + \ldots + a_{ij}*x_j \tag{1}$$

where: $a_{ij,\ j} = 0.1, \ldots , n$ they are coefficients calculated from discriminatory variables for each classification function. There are as many functions as there are groups ($i = 1, 2, ..., g$). With the functions defined in this way, the case is classified into the group for which K_i assumes the highest value.

To assess the usefulness of classification equations, usually a set of data is divided into two subsets: the learner and tester (if the sample is large), or new data should be collected to confirm the accuracy of the classification.

Discriminant analysis can be effectively applied in many fields of science. Most examples concern psychology, sociology, economics or medicine. Also, in the field of technical sciences, discriminant analysis is a diagnostic method applied. For example, M. Hajigholizadeh and A. M. Melesse used methods of discriminant analysis to assess water quality and to assess its spatial and temporal changes [9]. S.R. Oro et al. described a multidimensional statistical analysis of displacements of a concrete dam in relation to environmental conditions, using various statistical methods, including discriminant analysis [10]. Gabriela Vítková et al. used this method to classify bricks [11]. In the available literature, however, there are no articles describing the possibility of using discriminant analysis as an effective tool to solve the problems of designing new building materials using this method of data mining. This article presents the possibility of using the discriminant analysis method for testing mortars obtained using three different types of aggregates, i.e. perlite, expanded clay and rubber waste granulate, which are a partial substitute for quartz sand.

2 Materials and Methods

Epidian 5 epoxy resin was used to obtain resin mortars. Z-1 hardener (triethylenetetramine) in the amount of 10% (by weight) compared to the amount of resin, was used to cure the resin.

The main aggregate was quartz sand of a 0–2 mm grain size in accordance with the PN-EN 196-1 specification. Mortar modification consisted in substituting individual sand fractions at 0, 10, 20, 30%, 40% and 50% by volume, respectively with perlite (P), expanded clay (LECA) and rubber waste granulate (RW).

Based on the available literature data [12] and our own findings on resin mortars [13–15], a fixed ratio of resin to aggregate of 0.22 was established.

2.1 Sample preparation

Adequate amounts of epoxy resin were weighed and mixed thoroughly with a hardener of 10% by weight of the resin mass until a homogeneous structure of the mixture was obtained. The resin compositions thus prepared were transferred to the bowl of the laboratory mixer and mixed with standard sand previously weighed and mixed with an appropriate amount of modifier, maintaining the same mixing time and constant rotation of the mixer. The finished mortar was placed in steel moulds with dimensions 40×40×160 mm for the purpose of strength tests and to determine the bulk density. For the curing process to take place, the samples were left for 7 days under laboratory conditions.

2.2 Testing method

2.2.1 Flexural and compressive strength

The flexural strength and compressive strength tests were carried out in strength machines equipped with appropriate inserts, in accordance with the PN-EN 196-1: 2006 standard. For compressive strength tests, the bar halves remaining after the flexural strength tests were used.

2.2.2 Bulk density

The bulk density determination was carried out in accordance with PN-85/B-04500:1985, for samples with dimensions 40 x 40 x 160 mm. The weight of the bars was determined on technical scales. The volume of the samples was calculated based on their dimensions. The value of the bulk density was determined according to the formula (2):

$$\rho = m/V \tag{2}$$

where:
ρ - bulk density, g/cm^3,
m - sample mass, g,
V - sample volume, cm^3.

3 Results and Discussion

Results of this study were summarized in the table, which then served as a spreadsheet with the data necessary to carry out the analysis in the program Statistica 12. A fragment of this database is presented in the table shown in Figure 1. The data set contains values for three input variables: bulk density, flexural strength and compressive strength; epoxy mortars were obtained using three different fillers, i.e. perlite, expanded clay and rubber waste. The input file contained 5 columns. The first column gives information about the type of aggregate used to obtain the mortars (type of modifier – marked as P for perlite, LECA for expanded clay and RW for rubber waste). It was the type of aggregate that was the so-called grouping variable identifying the type of mortar. However, in the next columns of the table, the values of bulk density, flexural strength and compressive strength determined for the samples made were compiled. The final column of the table (Stage) contains data on the basis of which will be created and evaluated a system for classification of mortars into three groups, highlighted in the first column of the table. This variable has the character of the sample identifier and allows us to distinguish the sample intended for analysis (Training) and the sample intended for cross-checking (Test) allowing us to assess the quality of the classifier.

	1 Type of modifier	2 Bulk density, g/cm^3	3 Flexural strength, MPa	4 Compressive strength, MPa	5 Stage
1	P	2,03	27,39	97,65	Training
2	P	2,038	29,14	106,8	Training
3	P	2,043	28,55	105,9	Training
4	P	1,981	25,96	95,7	Test
5	P	2,029	27,01	107,6	Training
6	P	2,009	30,48	107,2	Training
7	P	1,887	25,58	100,5	Training
8	P	1,88	26,08	100,75	Training
9	P	1,928	25,33	103,55	Test
10	P	1,845	24,6	95,9	Training
11	P	1,869	22,41	94,9	Training
12	P	1,856	24,7	99,05	Training
13	P	1,821	21,78	91,45	Training

Fig. 1. Fragment of the table of results for the determination of the physical and mechanical parameters of the mortars.

The creation of the diagnostic system proceeded in two stages:

- Step 1. Building a classifier using the cases that make up the training set, marked in the table as Training – a total of 45 cases. At this stage, the analysis of discriminatory functions is used to decide which variables allow the best way to divide a given set of cases into naturally occurring groups.
- Step 2. Validating the operation of the classifier using cases labelled Test. For this purpose, 3 cases for each type of mortar were randomly selected from the entire data set, which created a 9-piece test set that allowed for the assessment of the prognostic correctness of the designated discriminant functions by the cross-analysis method.

Data analysis was started from the calculation of descriptive statistics such as: mean, medians, standard deviations. The average values of individual variables and the number of important cases (N) determined in the Statistica program for each group and in total for all groups are summarized in Table 1. The graph showing the ranges of the variables examined (summary mean values with measures of dispersion) is shown in Figure 2.

Table 1. Results of calculation of parameters describing the variables.

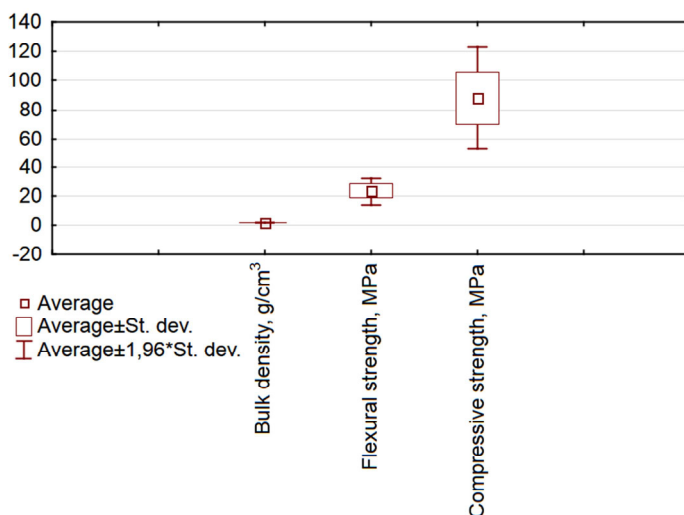| Type of aggregate	Bulk density, g/cm^3	Flexural strength, MPa	Compressive strength, MPa	N important
P	1.893778	25.33500	97.59444	18
LECA	1.899222	26.24111	94.96111	18
RW	1.857889	18.98833	71.56389	18
Total	1.883630	23.52148	88.03982	54

Fig. 2. Ranges of variables identified in studies.

Based on Figure 2, it can be concluded that the variables are significantly different in size, which calls for the application of standardization procedures during the construction of the classification model by means of discriminant analysis. For the discriminant analysis, the *Multidimensional exploratory techniques* module included in the Statistica program was used. The variable grouping was the variable Type of modifier, which clearly defines the membership of cases (test mortars) to the three groups P, LECA and RW. The independent variables are the parameters determined during mortar tests.

While defining the discriminant function, the stepwise analysis option was applied, thanks to which the program introduced variables to the discriminant function model one after another, always choosing the variable that made the most significant contribution to discrimination. Table 2 shows the contribution of each variable to the general discrimination of mortars containing various types of aggregates. The Wilks' lambda parameter shown in Table 2 is a standard statistic used to assess the statistical significance of the discriminatory power of the current model. Its value ranges from 1.0 (no discriminatory power) to 0.0 (excellent discriminatory power). On the other hand, the partial Wilks' lambda parameter defines the specific contribution of a given variable to the process of group discrimination. The partial Wilks' lambda value indicates that the bulk density variable makes the greatest contribution, the compressive strength variable less, and the flexural strength variable – the smallest contribution to the overall discrimination. Therefore, it can be concluded that bulk density is the main variable that allows the discrimination of mortars obtained using different types of aggregates.

Table 2. Assessment of the suitability of individual variables in discriminant analysis.

N=54	Summary of the discriminant function analysis: N var. in the model: 3; Grouping: Modifier type (3 groups) Wilks' lambda: .20836 approx. F (6.98)=19.449 p< .0000					
	Wilks' lambda	Partial Wilks	F removed (2.49)	p	Toler.	1-Toler. (R^2)
Flexural strength, MPa	0.264144	0.788824	6.55890	0.002992	0.186143	0.813857
Bulk density, g/cm^3	0.447912	0.465188	28.16689	0.000000	0.192125	0.807875
Compressive strength, MPa	0.319290	0.652583	13.04311	0.000029	0.113812	0.886188

In the next step, Statistica generated a sheet containing the results presented in Table 3, which show how many discriminatory functions can be interpreted. In this case, both discriminant functions are statistically significant ($p < 0.05$). Therefore, it is necessary to consider two separate conclusions (interpretations) on how much the values of bulk density and flexural and compression strength allow discrimination between types of mortar. The determination of canonical discriminant functions was based on the standardized coefficients of these functions listed in Table 4. These coefficients relate to standardized variables and refer to comparable scales, so they can be used for interpretation (this is particularly important when, as shown in Figure 2, the variables are significantly different in size). In the first discriminating function, the bulk density and compressive strength are the most important. Flexural strength also contributes to this function. The second function is determined mainly by strength variables, and to a lesser extent by bulk density. In the last two rows of table 4 there are eigenvalues (roots) for each discriminant function and the cumulative ratio of the explained variance corresponding to each function. These data allow us to conclude that the first function is responsible for over 93% of the explained variance, i.e. 93% of the total discriminative power is explained by this function. Thus, the number one function is clearly "the most important".

Table 3. The results of a chi-square test of the following roots.

Roots removed	Chi-square tests of the following roots					
	Eigenvalue	Canonical R	Wilks' lambda	Chi-square	df	p
0	2.993157	0.865778	0.208363	78.42359	6	0.000000
1	0.201883	0.409844	0.832028	9.19448	2	0.010080

Table 4. Standardized coefficients of the discriminant functions.

Variable	Standardized coefficients for canonical variables	
	Root 1	Root 2
Flexural strength, MPa	-0.64124	2.21790
Bulk density, g/cm^3	1.91555	0.44495
Compressive strength, MPa	-1.64650	-2.46484
Eigenvalue	2.99316	0.20188
Cum. prop.	0.93681	1.00000

Based on the data shown in Table 4, it was possible to state the share of variables in the discrimination of mortars with different fillers. It is possible to determine the nature of this discrimination for each canonical root after generating table 5 containing the so-called canonical means. On the basis of that, it can be concluded that the first discriminatory function differentiates mainly mortars obtained with the use of rubber waste. The canonical mean in this case is different from that calculated for the remaining fillers. The same is true for the second function, but the amount of this discrimination is much smaller, which is in line with the results presented in Table 4.

Table 5. Average values of the discriminant functions.

Group	Average canonical variables	
	Root 1	Root 2
P	-1.30600	-0.516036
LECA	-1.06778	0.551755
RW	2.37378	-0.035719

This interpretation is also confirmed by the scatterplot obtained for both discriminant functions, shown in Figure 3. This is a graph of non-standardized values for Root 2 relative to Root 1. Mortars containing rubber wastes are placed on the chart much more to the right, so the first discriminant function mainly distinguishes this type of mortar from the other two. The second function marginally better discriminates mortars containing expanded clay – for them the second canonical function takes the most positive values (greater than -0.5, while for other mortar samples with perlite filler, these values are less than -0.5). Discrimination in this case, however, is much less pronounced (0.551755<<2.37378) than in the case of the first function separating the collection of mortar samples obtained as a rubber composite.
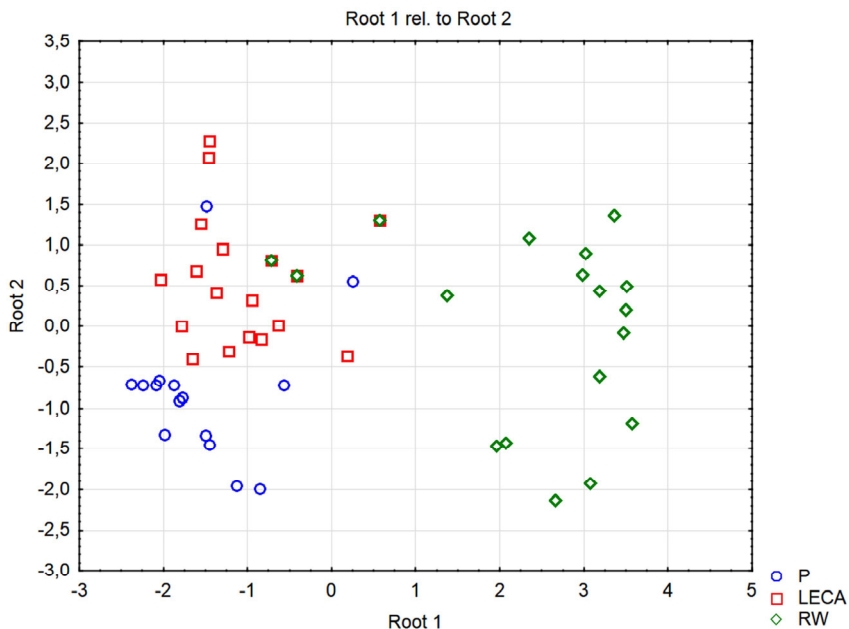
Fig. 3. Scatterplot for both discriminant functions.

One of the purposes of discriminant function analysis is to allow the investigator to classify cases. The Statistica program makes it possible to generate a table with so-called coefficients of classification functions (table 6), which are not the same as discriminatory functions.

Table 6. Parameters defining classification functions K1, K2 and K3.

Variable	Classification functions; Grouping variable: Modifier type		
	P	**LECA**	**RW**
Flexural strength, MPa	-0.018	0.617	-0.389
Bulk density, g/cm^3	448.906	456.971	511.790
Compressive strength, MPa	-2.859	-3.078	-3.384
Constant (free term)	-286.428	-296.983	-351.736

Classification functions are calculated for each group and can be used directly to classify cases. A given case can be classified in the group for which it has the highest classification value. The calculated coefficients contained in Table 6 were used to create the linear classification functions K1, K2 and K3. Equations presenting classification functions take the form:

$$K1 = -286.428 - 0.018* \text{ Flexural strength} + 448.906* \text{ Bulk density} + \\ - 2.859* \text{ Compressive strength} \tag{3}$$

$$K2 = -296.983 - 0.617* \text{ Flexural strength} + 456.971* \text{ Bulk density} + \\ - 3.078* \text{ Compressive strength} \tag{4}$$

$$K3 = -351.736 - 0.389* \text{ Flexural strength} + 511.790* \text{ Bulk density} + \\ - 3.384* \text{ Compressive strength} \tag{5}$$

8

The results of the accuracy of the classification for the training set (45 cases) are summarized in Table 7. The classification matrix presented in this table contains information on the number and percentage of objects (cases) correctly classified in each group.

Table 7. Classification matrix of cases in the training set.

Class	Classification matrix Classification: Rows (Observed) Columns (Predicted) (Test analysed)			
	Percentage correct	P p=0.3333	K p=0.3333	G p=0.3333
P	73.33333	11.00000	4.00000	0.00000
LECA	73.33333	4.00000	11.00000	0.00000
RW	80.00000	0.00000	3.00000	12.00000
Total	75.55556	15.00000	18.00000	12.00000

The designated functions K1, K2 and K3 make it possible to classify new cases. For each case, the values of three classification functions were calculated. The mortar containing the filler (case) is included in the group for which the value of the classification function is the largest. To check how well the designated classification functions work, the cases from the drawn test set were classified (column 5 in Figure 1), i.e. those that were not used to calculate the coefficients of the K1, K2 and K3 functions (cross-checking). The results obtained are summarized in Table 8.

Table 8. Classification Matrix cases in the test set.

Class	Classification matrix Classification: Rows (Observed) Columns (Predicted) (Test for cross-check)			
	Percentage correct	P p=0.3333	K p=0.3333	G p=0.3333
P	66.6667	2.00000	1.00000	0.00000
LECA	66.6667	1.00000	2.00000	0.00000
RW	100.0000	0.00000	0.00000	3.00000
Total	77.7778	3.00000	3.00000	3.00000

The data contained in table 8 show that the average percentage of correctly classified mortars is slightly larger than the result obtained for the training set and amounts to almost 78% of the total mortar tested. Usually, the correctness of the classification is better for the "training" set than for the "tester" set. In this case, the obtained results point to the opposite conclusion, which is associated with the verification of the prognosis capability of the generated classification functions on the set of data drawn from the research sample.

4 Conclusions

The discriminant analysis method described in the article makes it possible to successfully classify mortars obtained using three different types of aggregates, i.e. perlite, expanded clay and rubber waste granulate, which are a partial substitute for quartz sand.

The analysis allows us to draw the following conclusions:

- Bulk density is the main variable that allows discrimination of mortars with different types of filler.
- Two discriminant functions can be interpreted (both are statistically significant ($p < 0.05$)).
- The first discriminant function is influenced most by bulk density and compressive strength, although the contribution of flexural strength to this function is also important. The second function is determined mainly by strength variables, and to a lesser extent its volume is influenced by bulk density.
- The first function is certainly "the most important", because 93% of the entire discriminatory power is explained by it.
- Mortars containing rubber waste constitute a well-isolated collection, and its roots are placed on the scatterplot much more to the right, so the first discriminant function distinguishes mainly this type of mortar from the other two. The second function discriminates better in the case of mortars containing expanded clay – and the values of the second canonical function calculated for them are mostly positive, whereas the calculated values for perlite mortars are negative.
- The designated functions K_1, K_2, K_3 allow classification of new cases belonging to the test set that was not used to calculate the function coefficients.
- The average percentage of correctly predicted mortar class is slightly larger than the result obtained for the training set and amounts to almost 78% of the total mortar tested. This is the result of receiving a sample for cross-checking using a random method.

References

1. L. Czarnecki, Cem. Lime Concrete 2, 63–85 (2010)
2. B. Dębska, L. Lichołai, Period. Polytech. Civ. Eng. **60**, 281–287 (2016)
3. B. Dębska, L. Lichołai, Constr. Build. Mater. **124**, 11–19 (2016)
4. I. Baic, W. Kozioł, Ł. Machniak, E3S Web of Conferences **8**, 01068 (2016)
5. M. Y. Dinberu, K. Baek-Joong, M. Ji-Yeon, Y. Chongku, Materials **10**, 246–256 (2017)
6. J.M.L. Reis, Mater. Struct. **48**, 3929–3936 (2015)
7. B. Dębska, L. Lichołai, JCEEA, **63** (4/16), 87–96 (2016)
8. P. Radkiewicz, Psychologia Społeczna (Social Psychology) **5**, 2-3(14), 142–161 (2010)
9. M. Hajigholizadeh, A. M. Melesse, Catena **151,** 247–258 (2017)
10. S. R. Oro, A. Ch. Neto, T. R. Mafioleti, S. Ribeiro Pardo Garcia, C. Neumann Júnior, Indep. J. Manag. Prod. **7**, 526–545 (2016)
11. G. Vítková, L. Prokeš, K. Novotný, P. Pořízka, J. Novotný, D. Všianský, Spectrochim. Acta, Part B **101** 191–199 (2014)
12. J.M.L. Reis, E.P. Carneiro, Constr. Build. Mater. **27**, 107–111 (2012)
13. B. Dębska, L. Lichołai, J. Ecol. Eng. **19**, 204–212 (2018)
14. B. Dębska, Poly(ethylene Terephthalate) Based Blends, Composites and Nanocomposites, in: P.M. Visakh, M. Liang (Eds.) Reed Elsevier Inc., 195–212 (2015)
15. L. Lichołai, B. Dębska, Arch. Civil Mech. Eng. **14**, 466–475 (2014)