ISSN 1392–1320 MATERIALS SCIENCE (MEDŽIAGOTYRA). Vol. 29, No. 1. 2023

Machine Learning Methods for the Prediction of Abraded Butadiene-Styrene Rubber Surface Roughness

Loreta MAČĖNAITĖ [1] *, Dovydas LUGAS [1], Eglė FATARAITĖ-URBONIENĖ [2, 3]

[1] *Kaunas University of Technology, Faculty of Mathematics and Natural Sciences, Studentų st. 50, LT- 51368 Kaunas, Lithuania*
[2] *Kaunas University of Technology, Faculty of Mechanical Engineering and Design, Studentų st. 56, LT- 51424 Kaunas, Lithuania*
[3] *Applied Research Institute for Prospective Technologies, Vismaliukų st. 34, LT-10243 Vilnius, Lithuania*

crossref http://dx.doi.org/10.5755/j02.ms.31288

Received 28 April 2022; accepted 03 May 2022

The aim of this investigation was to evaluate the ability to predict the roughened surface area of soft polymeric materials using the machine learning method. Random forest and extremely random forest machine learning methods have been applied to predict roughened monolithic butadiene styrene rubber surface area based on the grain diameter of the abrasive paper. It was found that there is a very strong negative correlation between surface roughness and abrasive grain diameter (the Pearson correlation coefficient is equal to - 0.83 and the Spearman correlation coefficient is equal to - 0.91). The statistics of Shapiro-Wilk criteria confirmed that independently of the diameter of abrasive grain, the lengths of rubber surface profiles are normally distributed. It was determined that random forest and extremely random forest algorithms are suitable tools for predicting rubber surface area in dependence on abrasive grain diameter. The forest structure in which results obtained best coincidence with the observed (R^2 = 0.95): for the random forest model the number of trees is equal to 20, the leaf size is equal to 4, at the subset of the feature (abrasive paper grain diameter) equal to 1. In the case of an extremely random forest, the number of trees is equal to 350, the size of the subset of the features is equal to 1, and the leaf size is 2. The experimental and generated results agree well. The relative approximation error does not exceed 0.079 %.
Keywords: surface roughness, prediction, machine learning, random forest, extremely random forest.

1. INTRODUCTION

Adhesive joints for soft polymeric materials are widely used in many applications and are subject to extensive research to evaluate their durability [1]. The results of many researchers indicate that not only the chemical nature and compatibility of materials results in the strength of adhesive joints but also a very important parameter is the size of the contact area, which, as a rule, can be increased by mechanical surface roughening [2 – 6]. This kind of treatment is also used to remove surface impurities, but the main effect of surface abrasion is an increase in contact surface area. It is shown, that by variation of surface roughness parameters the strength of adhesive joints for rubber substrate can be increased from 25 % to 90 % compared to those of unroughened samples [7].

The bond contact area can only be found during the experiment. The method of contact or noncontact profilometry is usually used to obtain surface roughness characteristics. The surface is examined by a needle or laser beam perpendicular to the surface to be examined and transmits its image in the form of electrical signals to a digital one.

Since the experiment is usually a time-consuming process, it is important to create a model that would allow predicting the real surface area of the substrate depending on the surface roughness characteristics without performing a large experiment.

One way to determine the surface roughness parameters is based on the theory of random functions. In this case, the rough surface profile is assumed to be a normal stationary random process whose correlation function is continuous and has continuous derivatives. The stationarity of the random process allows the use of the average value and correlation function of the profile independently of the positioning on the surface, and if the condition of ergodicity is met, the parameter values can be calculated based on only one surface profile of sufficient length. Deviations from the stationarity and ergodicity conditions affect only the methodology of receiving experimental data and their processing. However, general probability regularities at the selected conditions are the same. There, the model roughness parameters are the numerical characteristics of a random process. The possibility of applying a rough surface model as a random process has been investigated in [8 – 10].

Based on these assumptions, several models for calculating the contact area between surfaces have been proposed. As examples can consider Hertz, Johnson-Kendall-Roberts (JKR) [11, 12]; Greenwood-Williamson (GW) [13]; Greenwood-Tripp (GT) models. Predicting the contact area according to Hertz's theory, large errors are obtained. They can be reduced by applying the JKR model, which improves the Hertz model and additionally evaluates the adhesion work between elastic surfaces. The disadvantage of this model is its suitability only for macroscopic surfaces. The disadvantage of the GW and GT

* Corresponding author. Tel.: +370-611-11898.
E-mail address: *loreta.macenaite@ktu.lt* (L. Macenaite)

models is that they assume peaks of irregularities as spherical. It is clear that the real surface irregularities not always are spherical. Fractal models (Majumdar-Bhushan (MB) model [14]) are also used to calculate the contact area. However, the disadvantage of the MB model is that the conditions for a fractal surface are not always met. Therefore, some researchers are skeptical about the application of fractal models to characterize surface roughness [15]. The various shortcomings of the models encourage researchers to develop new models to predict contact area, which mainly depends on the surface roughness of the substrate.

In recent decades a real breakthrough has been made in the prediction of various processes using different machine learning methods. Among them, the random forest method and its modifications are presented as the most advantageous. Random forest is a popular machine learning procedure that can be used to develop prediction models. First introduced by Breiman in 2001, random forest are a collection of classification and regression trees, which are simple models using binary splits on predictor variables to determine outcome predictions. Random forests consistently offer among the highest prediction accuracy compared to other models in the setting of classification [16].

Assuming that, this investigation aimed to evaluate the ability to predict the roughened surface area of soft polymeric materials using random forest and extremely random forest methods.

2. EXPERIMENTAL

2.1. Materials

It was found a very limited amount of information about the application of machine learning methods for the prediction of soft polymer materials surface roughness, and the prediction of the influence of the contact surface area on the strength of adhesive joints. Butadiene-styrene rubber is well known for its poor adhesion properties with different types of adhesives. It is believable, that selection of this material will allow evaluating the effect of mechanical surface treatment of soft polymer materials, decreasing the effect of chemical interaction between adhesive and substrate. For the investigations black ~~leather like~~ butadiene-styrene rubber (BSR) pieces have been used. The density and hardness according to the Shore scale were $\rho = 1.25$ g/cm^3 $H = 75$ a.u., respectively.

For surface roughening, 5 abrasive papers of different grade number have been selected. The relation between the number of abrasive paper and the average grain diameter is presented in Table 1. The larger the number of abrasive papers the smaller diameter of abrasive grain is found.

Table 1. Relation between abrasive paper number and the abrasive grain diameter

Abrasive paper number	Average grain diameter *d*, μm
P24	698
P40	382
P60	260
P100	149

2.2. Methods

2.2.1. Experimental procedure of profiles obtaining and evaluation

BSR rubber strips (100 mm × 100 mm) surface was roughened using abrasive papers of different numbers. During treatment sample feed rate to the abrasion tool was 10 mm/s and the clamping force was $P = 80$ N. For each abrasive paper seven samples have been prepared.

Roughened rubber surface profiles were carried out perpendicular to the abrasion direction using a testing-machine profilograph "Hommelwerke T500" (Germany). Measurement accuracy was ±0.2 μm. 7 profiles for each abrasive grade number treated surface was carried out.

It was assumed that each profile is described by a set of points of the curve (x_n, y_n), $n = 1, 2,\ldots, N$. Then the length of the profile l can be calculated according to the curve arc length formula:

$$l = \sum \sqrt{(x_{n+1} - x_n)^2 + (y_{n+1} - y_n)^2}. \quad (1)$$

Surface roughness characteristic l for each case of the roughened surface was calculated for the same lenght segment equal to 10 mm.

2.2.2. Machine learning methods

Two machine learning methods were used to predict the length of the roughened surface profile: random forest and extremely random forest. The recursive feature removal method was used to determine the best features of each method.

Machine learning was performed using the programming languages *R 4.0.5* and *JupyterLab 3.0.14*. environment. The *R* package *randomForest 4.6-14* was used for the random forest algorithm and *extraTrees 1.0.5* for the extremely random forest algorithm testing.

A random forest algorithm consists of a set of decision trees that are constructed according to the following algorithm [17]:

1 – Selection of the number of trees *m*;
2 – Design a cycle for finding trees:
for *i = 1 to m* **do**
A random sample of data distribution is generated.
According to this sample, the tree is trained:

for *each cros-section* **do**
A random subset of features with *k* elements is formed.
An optimal cross-section is selected from the features of this subset.
end
Tree training is finalized when the stop criterion is reached.
end

The predicted cross-section suitability is estimated by calculating the total sum of squares error *SSE*:

$$SSE = \sum_{i \in S_1} (y_i - \bar{y}_1)^2 + \sum_{i \in S_2} (y_i - \bar{y}_2)^2, \quad (2)$$

where S_1, S_2 are the subsets into which the data are split; y_i is the response values; $\bar{y}_1$ and $\bar{y}_2$ are the averages of the response values of the subsets S_1 and S_2. Cross-section is acceptable if condition $SSE \rightarrow min$ is fulfilled.

Typically, a splitting of the tree leaf is not applied when less than five values remain in it. If there are no more leaves left that can be split, the tree construction is stopped. The optimal values of parameters k and m are determined by cross-validation. The overall forecast is made after calculating the average of all prognoses.

The extremely randomized forest algorithm is a modification of the random forest algorithm [18]. Like a random forest, an extremely random forest consists of a set of trees, but individual trees are trained differently. Firstly, not random data sets as in the case of random forest algorithm, but a full data distribution is used for tree training. Second, this method does not require an optimal cross-section of features. Instead, they are selected at random, and the best among them is selected. Due to that, in this model, one additional parameter, the number of tested random sections for each characteristic, is introduced.

The root mean square error (RMSE) for all predicted models was calculated to evaluate how predicted values coincide with observed ones in a regression analysis. In other words, how concentrated the data around the predicted line of best fit:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}}, \quad (3)$$

where P_i is the predicted value for the *i-th* observation in the dataset; O_i is the observed value of the *i-th* observation in the dataset; n is the size of the dataset.

3. RESULTS AND DISCUSSION

3.1. Surface roughness characteristics of a real roughened leather-like rubber surface

In Fig. 1 typical profiles of abraded BSR surfaces vs abrasive paper grain diameter are presented. It is evident that the shape and number of micro-irregularities are highly dependent on the abrasive paper number. If it will be assumed that roughness is treated as the overall micro-irregularities with a relatively small distance between them, it can be stated that the amount of surface irregularities after treatment with the abrasive paper P24 is less than obtained with the abrasive paper P100. The digital data of profiles were used to calculate the statistical characteristics and length of the profile.

The Shapiro-Wilk criterion with a significance level of 0.05 was applied to confirm/ deny the assumption of normal distribution of profile length. The calculated values of the Shapiro-Wilk statistic W and p-value are presented in Table 2. W values close to 1 and p-value significantly higher from 0.05 confirms the normal distribution of profile length. The R software was used to calculate confidence intervals for the average profile length l_{ave} of seven tested samples. The results of the calculations are presented in Table 3. The relative increase in the surface profile length Δ was calculated to show the effect of surface roughening compared to that of untreated (l_o = 10 mm):

$$\Delta = \frac{|l_{ave} - l_o|}{l_o} \times 100\%. \quad (4)$$

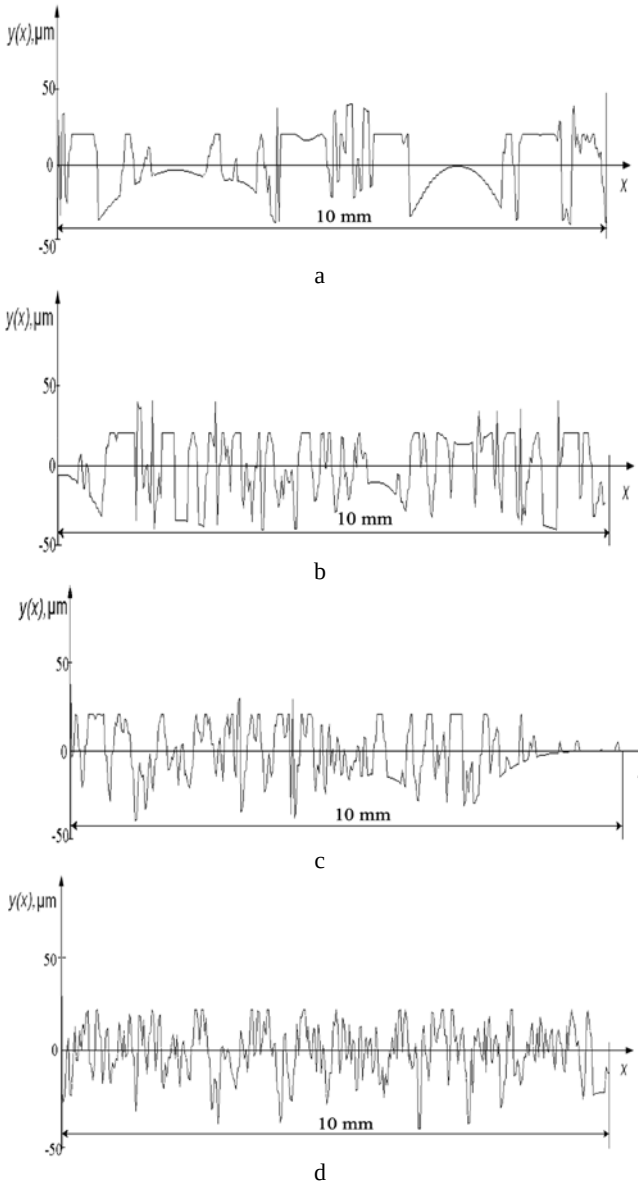

Fig. 1. BSR surface profiles *vs* abrasive paper diameter *d*, μm: a – 698; b – 382; c – 260; d – 149

Table 2. Shapiro-Wilk statistics *W* and *p*-value for the profiles lengths vs abrasive paper grain diameter

Abrasive grain diameter *d*, μm	Statistics	
	W	*p-value*
698	0.913	0.414
382	0.844	0.108
260	0.967	0.871
149	0.934	0.584

As is evident from the data presented in Table 3, the decrease in abrasive grain diameter results in increase in the total profile length. The length of the abraded surface profile of the highest number of abrasive paper (smallest grains diameter) increases by 18.32 % compared to the length of the untreated one. The Pearson correlation coefficient between the abrasive grain size and the profile length is equal to - 0.83 and the Spearman correlation coefficient is

equal to - 0.91. Confirmation of a strong negative correlation between the size of the abrasive grain and the length of the roughened surface profile is presented in Fig. 2.

Table 3. Real profile characteristics vs. abrasive paper number

Abrasive grain diameter *d*, μm	l_{ave}, μm	Confidence interval (α = 0,05), μm	*Δ*, %
698	10470.35	(10359; 10582)	4.49
382	10644.25	(10460; 10829)	6.05
260	11384.43	(11224; 11545)	12.16
149	12243.58	(12086; 12402)	18.32

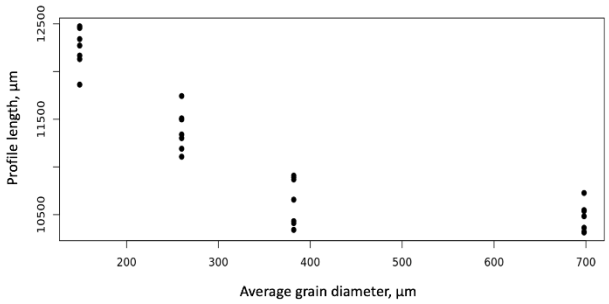

Fig. 2. Correlation between the length of the roughened surface profile length and the diameter of the abrasive grain

That allows us to state that abrasive grain size can be used as a variable to predict profile length in the machine learning process. This presumption was verified by methods of random forest and extremely random forest.

3.2. Prediction of surface roughness using machine learning methods

To predict the relationship between surface area and abrasion paper characteristics, the lengths of the roughened surface profiles were generated using two machine learning algorithms: random forest and extremely random forest.

Machine learning for each model was performed by five-time repeating multiple (10 times) cross-validation. Two assumptions were made: I – the test area of the sample is S_o = 10 mm × 10 mm = 100 mm^2 (10^8 μm^2); II – the surface without roughness is smooth and its true surface area is equal to 100 mm^2. The surface area of the roughened surface *S* is obtained by multiplying the generated profile length by the width of the sample (w = 10 mm).

Initially, the dependence of the profile length on the abrasive grain size was generated using the random forest algorithm. The random forest model was trained in three stages: Stage I was the determination of the number of trees at which RMSE reaches its lowest value; Stage II was the determination of the optimal subset of features, and Stage III was the determination of the size of the leaf at which RMSE reaches its lowest value.

To find a number of trees in the forest that meets the requirement for learning Stage I, various numbers of trees were tested, and for each case the RMSE was calculated. The dependence between the calculated tree number and the RMSE is presented in Fig. 3. It seems that the dependence is non-linear with the lowest error at the number of trees equal to 20.

Usually, the optimal value of a subset of characteristics is determined in the same way. Since only one variable (abrasive grain diameter) is included in the model, the only possible value for this parameter is 1. Therefore, the size of a subset of variables is simply set to 1.

The last parameter to be defined is the smallest tree leaf size. The value of this parameter and the RMSE was estimated by testing various leaf sizes. The relationship between RMSE and leaf size during random forest learning is presented in Fig. 4. According to the results presented, it can be stated that the leaf size at which the lowest RMSE value is reached is 4. Thus, modeling the dependence of the profile length on the abrasive grain size by the random forest method gives the best results when the number of trees in the random forest is equal to 20, the size of the subset of the variable is equal to 1, and the leaf size is equal to 4.

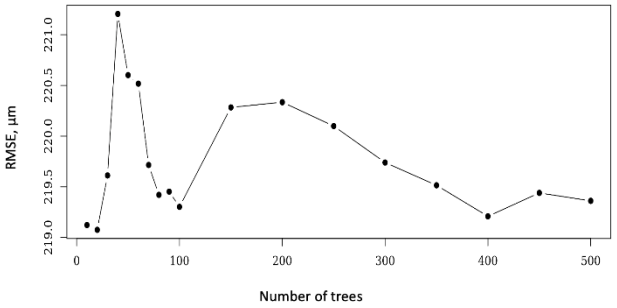

Fig. 3. Random forest model error vs. the number of trees

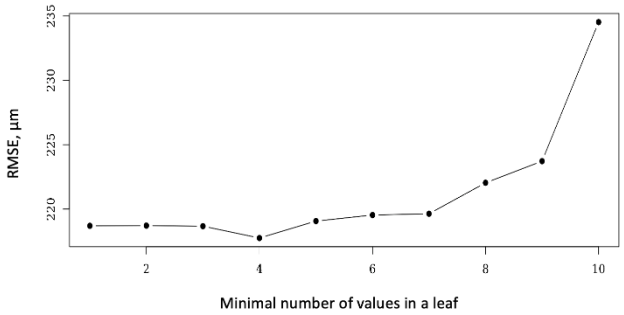

Fig. 4. Dependence between random forest error RMSE and leaf size

The extremely random forest algorithm is trained in the same way. First, the number of trees in the forest was determined by testing the various numbers of trees and estimating the model error RMSE. The dependence between RMSE and the number of trees during extremely random forest model learning is presented in Fig. 5. Based on these results, it can be stated that the number of tested trees needs to be 350.

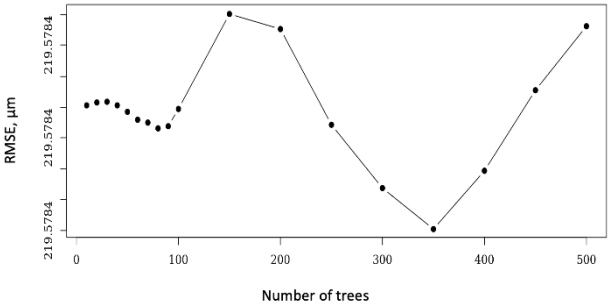

Fig. 5. Relation between an extremely random forest algorithm error and the number of trees

Since the model includes only one variable, the same as for the random forest algorithm, the only possible value for the size of a subset of features is set to 1.

The last parameter to be determined is the smallest tree leaf size. Various leaf numbers were tested to determine at which leaf number minimal RMSE is found. The results obtained are presented in Fig. 6. The lowest RMSE value is found when the leaf size is equal to 2. In summarizing, it can be stated that by applying an extremely random forest machine learning method for the prediction of profile length vs. the abrasive grain diameter the best results can be found at the number of trees equal to 350, the size of the subset of the variable is equal to 1, and the leaf size is equal to 2.

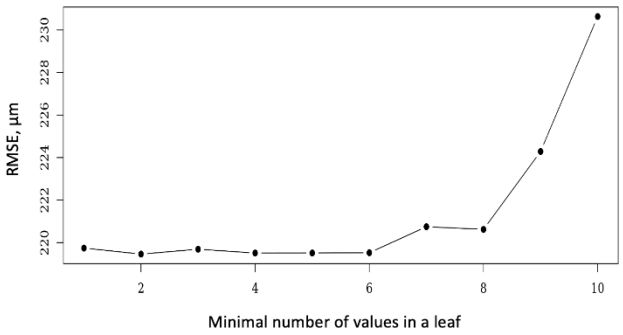

Fig. 6. Relation between extremely random forest algorithm leaf size and model RMSE

In Table 4 RMSE values and determination coefficient R^2 calculation results are presented which confirm the adequacy of the generated results to the experimentally observed values.

Table 4. RMSE and R^2 for models trained only with abrasive grain size

Model	*RMSE*, µm	R^2
Random forest (trees number is 20, variable is equal to 1, and the leaf size is 4)	219.36	0.9495
Extremely random forest (trees number is 350, the variable is equal to 1, and the leaf size is 2)	219.58	0.9493

In Table 5 generated roughened surface area values in dependence on abrasive grain diameter are presented.

Table 5. Predicted results in dependence on machine learning method and the abrasive grain diameter

Algorithm	Abrasive grain diameter d, µm	Surface area S, µm^2	δ_s, %
Random forest	698	104762922.978	0.056
	382	106379009.724	0.059
	260	113835602.677	0.008
	149	122369344.256	0.054
Extremely random forest	698	104739905.794	0.035
	382	106358282.475	0.079
	260	113844600.414	0.0003
	149	122409807.909	0.021

In this table relative approximation errors between real S_r and generated S_g surface area values are also presented. δ_s was calculated according to the following formula:

$$\delta_s = \frac{|S_g - S_r|}{S_r} \times 100\%. \quad (5)$$

Calculated δ_s values show that the experimentally obtained and generated data coincide well.

Obtained results confirm the presumption that random forest and extremely random forest algorithms can be successfully used for the prediction of the roughened rubber surface area when the grain diameter of abrasive paper is known.

4. CONCLUSIONS

Random forest and extremely random forest machine learning methods have been applied to predict roughened monolithic butadiene styrene rubber surface area based on the grain diameter of the abrasive paper. It was found that there is a very strong correlation between surface roughness and the diameter of abrasive grain. The statistics of Shapiro-Wilk criteria confirmed that independently of the diameter of abrasive grain, the lengths of rubber surface profiles are normally distributed. It was determined that random forest and extremely random forest algorithms are suitable tools for predicting rubber surface area in dependence on abrasive grain diameter. The forest structure in which results obtained best coincides with the observed ($R^2 = 0.95$): for the random forest model the number of trees is equal to 20, the leaf size is equal to 4, at the subset of the feature (abrasive paper grain diameter) equal to 1. In the case of an extremely random forest, the number of trees is equal to 350, the size of the subset of the features is equal to 1, and the leaf size is 2. The experimental and generated results agree well. The relative approximation error does not exceed 0.079 %.

REFERENCES

1. **Dayss, E., Leps, G., Meinhardt, J.** Surface Modification for Improved Adhesion of a Polymer–Metal Compound *Surface and Coatings Technology* 116 1999: pp. 986 – 990. https://doi.org/10.1016/S0257-8972(99)00184-X
2. **Uehara, K., Sakurai, M.** Bonding Strength of Adhesives and Surface Roughness of Joined Parts *Journal of Materials Processing Technology* 127 (2) 2002: pp. 178 – 181. https://doi.org/10.1016/S0924-0136(02)00122-X
3. **Shahid, M., Hashim, S.A.** Effect of Surface Roughness on The Strength of Cleavage Joints *International Journal of Adhesion and Adhesives* 22 (3) 2002: pp. 235 – 244. https://doi.org/10.1016/S0143-7496(01)00059-8
4. **Sargent, J.P.** Adherend Surface Morphology and its Influence on the Peel Strength of Adhesive Joints Bonded with Modified Phenolic and Epoxy Structural Adhesives *International Journal of Adhesion and Adhesives* 14 (1) 1994: pp. 21 – 30. https://doi.org/10.1016/0143-7496(94)90017-5
5. **Kwon, J.W., Lee, D.G.** The Effects of Surface Roughness and Bond Thickness on the Fatigue Life of Adhesively Bonded Tubular Single Lap Joints *Journal of Adhesion Science and Technology* 14 (8) 2002: pp. 1085 – 1102. https://doi.org/10.1163/156856100743095
6. **Niem, P.I.F., Lau, T.L., Kwan, K.M.** The Effect of Surface Characteristics of Polymeric Materials on the Strength of Bonded Joints *Journal of Adhesion Science and Technology* 10 (4) 1996: pp. 361 – 372. https://doi.org/10.1163/156856196X00760

7. **Lyashenko, I.A., Pohrt, R.** Adhesion Between Rigid Indenter and Soft Rubber Layer: Influence of Roughness *Frontiers in Mechanical Engineering* 6 2020: pp. 49. https://doi.org/10.3389/fmech.2020.00049
8. **Borodich, F.M., Pepelyshev, A., Savencu, O.** Statistical Approaches to Description of Rough Engineering Surfaces at Nano and Microscales *Tribology International* 103 2016: pp. 197 – 207. https://doi.org/10.1016/j.triboint.2016.06.043
9. **Borodich, F.M., Brousseau, E., Clarke, A., Pepelyshev, A., Sánchez-López, J.C.** Roughness of Deposited Carbon-Based Coatings and Its Statistical Characteristics at Nano and Microscales *Frontiers in Mechanical Engineering* 5 2019: pp. 24. https://doi.org/10.3389/fmech.2019.0002
10. **Pepelyshev, A., Borodich, F.M., Galanov, B.A., Gorb, E.V., Gorb, S.N.** Adhesion of Soft Materials to Rough Surfaces: Experimental Studies, Statistical Analysis and Modelling *Coatings* 8 (10) 2018: pp. 350. https://doi.org/10.3390/coatings8100350
11. **Popov, V.L., Heß, M., Willert, E.** Handbook of Contact Mechanics: Exact Solutions of Axisymmetric Contact Problems. Springer Nature, 2019: pp. 347.
12. **Johnson, K.L., Kendall, K., Roberts, A.** Surface Energy and the Contact of Elastic Solids *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 324 (1558) 1971: pp. 301 – 313. https://doi.org/10.1098/rspa.1971.0141
13. **Greenwood, J.A., Williamson, J.P.** Contact of Nominally Flat Surfaces *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 295 (1442) 1966: pp. 300 – 319.
14. **Majumdar, A., Bhushan, B.** Fractal Model of Elastic-Plastic Contact between Rough Surfaces *Journal of Tribology* 113 (1) 1991: pp. 1 – 11.
15. **Borodich, F.M., Jin, X., Pepelyshev, A.** Probabilistic, Fractal, and Related Techniques for Analysis of Engineering Surfaces *Frontiers in Mechanical Engineering* 6 2020: pp. 64.
16. **Speiser, J.L., Miller, M.E., Tooze, J., Ip, E.** A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling *Expert Systems with Applications* 134 2019: pp. 93 – 101. https://doi.org/10.1016/j.eswa.2019.05.028
17. **Breiman, L.** Random Forests *Machine Learning* 45 (1) 2001: pp. 5 – 32. https://doi.org/10.1023/A:1010933404324
18. **Geurts, P., Ernst, D., Wehenkel, L.** Extremely Randomized Trees *Machine Learning* 63 (1) 2006: pp. 3 – 42. https://doi.org/10.1007/s10994-006-6226-1