

Received: 2020-09-23

Accepted: 2020-11-27

Online published: 2020-12-30

DOI: <https://doi.org/10.15414/meraa.2020.06.02.54-63>*Original Paper*

Regression analysis of expanded polystyrene properties

Dušan Páleš^{1*}, Milada Balková²¹ Slovak University of Agriculture in Nitra, Faculty of Engineering, Department of Machine Design,² Slovak University of Agriculture in Nitra, Faculty of Engineering, Department of Building Equipment and Technology Safety

ABSTRACT

Own measurements examine the tensile strength of expanded polystyrene (EPS) depending on its bulk density. 30 samples were used to calculate the correlation coefficients between these two properties. In addition to the standard Pearson coefficient, we also calculate the rank correlation coefficients, Spearman's and Kendall's. By testing the hypotheses, we verify the correlation of the entire population. After finding a relatively close correlation (0.6 - 0.8), we apply different regression models, especially polynomial, but also exponential. We evaluate the properties of parameters in models, their point estimates and confidence intervals. Based on the characteristics of each of the seven regressions, we found the best exponential form of the dependence, before the linear polynomial. The complexity of a mathematical model does not always mean that it is also a more accurate approximation. On the other hand, a simple model makes it possible, in addition to its ease of use, to more closely reflect the examined dependence.

KEYWORDS: Expanded polystyrene (EPS), bulk density, tensile strength, correlation, regression model**JEL CLASSIFICATION:** C12, C13, C15

INTRODUCTION

Expanded polystyrene (EPS) is a proven heat-insulating, but also sound-insulating material. It is lightweight and has a bulk density, which makes it easy to process [4]. Sometimes its strength properties can be used and therefore it is appropriate to measure its tensile strength [2]. The interdependence between density and strength of EPS is presented by many distributors [8], [11]. To determine the intensity of this relationship, it is advantageous to apply a correlation analysis [1]. A mathematical description of such dependence can be provided by regression analysis, which has a number of characteristics evaluating every regression model [14]. It is then important to evaluate these values correctly, so that we do not

* Corresponding author: Dušan Páleš, Slovak University of Agriculture in Nitra, Trieda Andreja Hlinku 2, 949 76 Nitra, Slovakia, e-mail: dusan.pales@uniag.sk

have to rely on subjective judgment, but we were able to classify the suitability and unsuitability of the chosen mathematical formula by comparing specific numbers.

MATERIAL AND METHODS

Expanded polystyrene (EPS) - properties

Expanded polystyrene (EPS) is an organic matter that belongs to a group of foam plastics. It is made from the chemical substance styrene, which is expanded (foamed) with water vapor and blowing agent (pentane). Industrial styrene and pentane are extracted from petroleum, although they are also commonly found in nature. EPS is one of the most widely used plastics, right after PVC, polyethylene and polypropylene. The highest rated properties of polystyrene are its excellent thermal insulating properties, easy workability due to its low bulk density and affordability.

EPS foam matter consists of about 2% polystyrene and 98% air. It is actually engineered wrapped air and it causes its excellent insulating ability. It has high compressive, tensile and flexural strengths that increase linearly with increasing bulk density. Its low weight reduces the load on the load-bearing structure, the transport costs and the effort in use. EPS is not soluble in water and therefore its cells do not absorb water into its structure, causing its very low water absorption. It can only absorb water vapor to a certain extent and therefore it is important to ensure that the dew point is not inside the polystyrene structure for a long time. Polystyrene products have been used for many years for food packaging, which is evidence of their health safety. The rapid development of the use of EPS as thermal insulation has necessitated the development of so-called self-extinguishing polystyrene, which meets the strict requirements for fire protection of buildings. Polystyrene as thermal insulation reduces fuel consumption, making it very environmentally friendly. Last but not least, EPS production is also ecological and polystyrene is also recyclable.

In construction, EPS is applied primarily for thermal insulation, but also for sound insulation. It can be used to create various decorative or shaping elements, for air filling or for foundation of traffic constructions on soft soils. Everyone has already encountered the purpose of polystyrene as a packaging material, and thus the field of application is far from over.

The extensive use of expanded polystyrene was the main motivation for us to determine the tensile strength of EPS and the corresponding density. The tests were performed on a jaw tear machine and a bulk density was determined for each sample [7], [9], [10].

Correlation coefficients and correlability

The correlation analysis examines the relations between random variables, the most important being the intensity of interdependence. The result of this examination is a correlation coefficient r , which takes values from -1 to $+1$. Values close to -1 or to $+1$ indicate a strong correlation and values close to 0 , on the contrary, weak, respectively. no correlation. The Pearson correlation coefficient is calculated from the relationship

$$r = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)} \cdot \sqrt{\text{Var}(y)}} = \frac{\overline{x \cdot y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right)} \cdot \sqrt{\left(\overline{y^2} - \bar{y}^2\right)}} \quad (1)$$

where $\text{Cov}(x, y)$, $\text{Var}(x)$, $\text{Var}(y)$ are covariance and variance of measurements of the variables x , y and the bar above the value of the variable denotes its mean value. An assumption for applying the coefficient (1) is a linear relationship between the variables x , y and their normal probability distributions.

The sensitivity to extreme values does not show rank correlation coefficients. Nor do we need assumptions about the types of measurement distributions or about the linear dependence between variables. The Spearman rank correlation coefficient is calculated from the formula

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n \cdot (n^2 - 1)} \quad (2)$$

where $d_i = R_i^x - R_i^y$, $i = 1, 2, \dots, n$, are the differences between the order numbers R_i^x and R_i^y and n is the number of observations. As can be seen from the relation (2), instead of the values of the measured quantities, we assume their order in the set.

The Kendall rank correlation coefficient is expressed in the form

$$r_k = \frac{c - d}{\sqrt{c + d + x_{\text{extra}}} \cdot \sqrt{c + d + y_{\text{extra}}}} \quad (3)$$

where $c = \sum c_i$ and $d = \sum d_i$, are the sums of the concordant and discordant orders between the independent variable x and the dependent y , x_{extra} and y_{extra} represent the number of variables with the same order.

The sample correlation coefficients r are point estimates of the correlation coefficient ρ of the population. Therefore, nonzero r does not mean that ρ will be nonzero too. The importance of testing the hypothesis H_0 : Random variables are uncorrelated ($\rho = 0$), compared to the alternative hypothesis H_1 : Random variables are correlated ($\rho \neq 0$), comes to the fore especially in the case of a small number of measurements. This test can be performed for all three types of described correlation coefficients (Pearson, Spearman, Kendall) and is most often carried out at a significance level of $\alpha = 0.05$. Refer to the literature for further details [5], [6].

Regression analysis

While the correlation analysis determines the existence and intensity of the dependence between variables, the regression analysis looks for the form of this dependence. The general form of the linear function in the parameters is written

$$y = f(x) = \sum_{j=0}^k a_j \cdot \varphi_j(x) \quad (4)$$

where a_j are parameters, constants and $\varphi_j(x)$ are functions of independent variable that no contain already other parameters [6]. The point estimation of parameters a_j most commonly uses the least squares method, which minimizes the sum of squares for error (SSE)

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{a}_0 \phi_0(x_i) - \dots - \hat{a}_k \phi_k(x_i))^2 \tag{5}$$

$y_i, i = 1 \dots n$, are obtained values of the dependent variable, $\hat{y}_i, i = 1 \dots n$, are the estimated values by the regression function and $\hat{a}_j, j = 1 \dots k$ are the estimated values of the parameters. In the confidence intervals of parameters a_j the Student distribution with $n-k-1$ degrees of freedom is applied.

By means of SSE it is possible to determine the mean squared error (MSE) and simultaneously root mean squared error (RMSE)

$$MSE = \frac{SSE}{n - k - 1} \quad RMSE = \sqrt{MSE} \tag{6}$$

When we use the least squares method, the relationship must hold

$$SST = SSR + SSE \tag{7}$$

where

$$\begin{aligned} SST &= \sum_{i=1}^n (y_i - \bar{y})^2 \\ SSR &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ SSE &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned} \tag{8}$$

SST is the total sum of squares and SSR is the sum of squares for residuals, further denotation is explained above. According to (7) it must therefore apply

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{9}$$

Coefficient of determination R^2 is called the ratio

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{10}$$

which takes the values $0 \leq R^2 \leq 1$. It determines that part of the total variability of the observed values that can be explained by the regression model. For small sample sizes n , a modified coefficient of determination R^{*2} is used due to the bias of the estimate

$$R^{*2} = R^2 - (1 - R^2) \cdot \frac{k}{n - k - 1} \tag{11}$$

The regression model as a whole can be tested for the significance of the coefficient of determination R as well as of all regression coefficients with the exception of the absolute coefficient a_0 . At the selected significance level α we test the hypothesis $H_0: a_1 = a_2 = \dots = a_k = 0$, i. e. the regression model is statistically insignificant compared to the alternative hypothesis H_1 : at least one of the regression coefficients $a_1, a_2, \dots, a_k \neq 0$, i. e. the regression model is statistically significant.

RESULTS AND DISCUSSION

Obtained values of bulk density and tensile strength are given in the Tab. 1. We performed 30 measurements, which we statistically evaluated on the boxplots in the Fig. 1. The interdependence of both sets was estimated by different types of correlation coefficients in the Tab. 2. Their values are positive and relatively high, which represents a direct dependence between quantities. As mentioned above, it is appropriate to verify the mutual correlability with the hypothesis $H_0: \rho = 0$, at the significance level $\alpha = 0.05$. Values of p in the second line of the Tab. 2 came out less than α , which makes it possible to reject H_0 and assume $\rho \neq 0$.

Table 1 Values of bulk density x_i [$\text{kg} \cdot \text{m}^{-3}$] and corresponding tensile strength y_i [MPa]

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	23.53	0.316	11	33.82	0.378	21	44.41	0.449
2	27.94	0.223	12	33.47	0.405	22	43.65	0.476
3	28.23	0.276	13	35.58	0.391	23	43.12	0.452
4	33.52	0.212	14	40.88	0.374	24	36.76	0.449
5	32.64	0.288	15	47.65	0.473	25	41.47	0.394
6	36.47	0.232	16	36.06	0.428	26	40.29	0.396
7	33.20	0.308	17	37.65	0.415	27	40.85	0.405
8	35.88	0.309	18	44.52	0.511	28	42.06	0.441
9	36.90	0.309	19	44.12	0.500	29	41.99	0.449
10	31.17	0.318	20	39.70	0.469	30	40.35	0.408

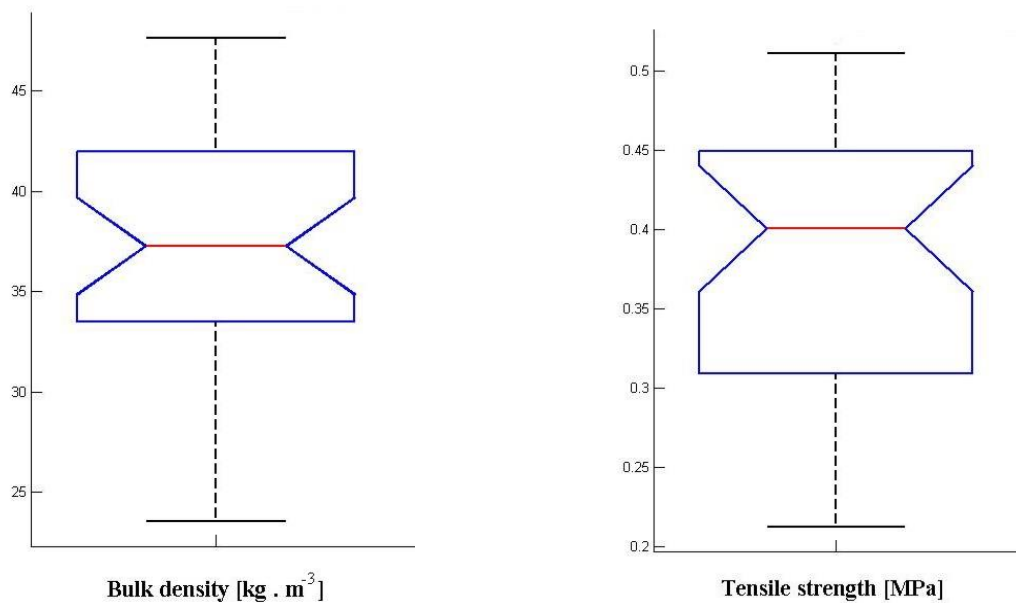


Figure 1 Boxplots of bulk density and tensile strength

Table 2 Correlation coefficients calculated by different methods and their correlability at the significance level $\alpha = 0.05$

	Pearson	Spearman	Kendall
Correlation coefficient	0.7585	0.7996	0.6104
Correlability - p	1.194e-6	1.152e-7	2.648e-6

For regression analysis we have chosen 7 types of functions. The first six are polynomials up to the sixth degree and the seventh is an exponential function. Shapes of regression functions states (12)

$$\begin{aligned}
 \text{Polynomial 1}^{\text{st}}: & \quad y = a_1 \cdot x + a_0 \\
 2^{\text{nd}}: & \quad y = a_2 \cdot x^2 + a_1 \cdot x + a_0 \\
 3^{\text{th}}: & \quad y = a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0 \\
 4^{\text{th}}: & \quad y = a_4 \cdot x^4 + a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0 \\
 5^{\text{th}}: & \quad y = a_5 \cdot x^5 + a_4 \cdot x^4 + a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0 \\
 6^{\text{th}}: & \quad y = a_6 \cdot x^6 + a_5 \cdot x^5 + a_4 \cdot x^4 + a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0 \\
 \text{Exponential:} & \quad y = a_0 \cdot \exp(a_1 \cdot x)
 \end{aligned}
 \tag{12}$$

In the Fig. 2, we plotted the measured values and fitted the first three polynomials and the exponential function. The polynomials 4th, 5th and 6th order were not shown for clarity on the picture. However, there is an apparent tendency that the higher the degree of polynomial, the better it adapts to the measured values. Noteworthy is a 3rd degree polynomial, in green, that curves down on the right side, which may indicate its poor ability to predict tensile strengths for higher bulk weights, as it is likely to go to zero values.

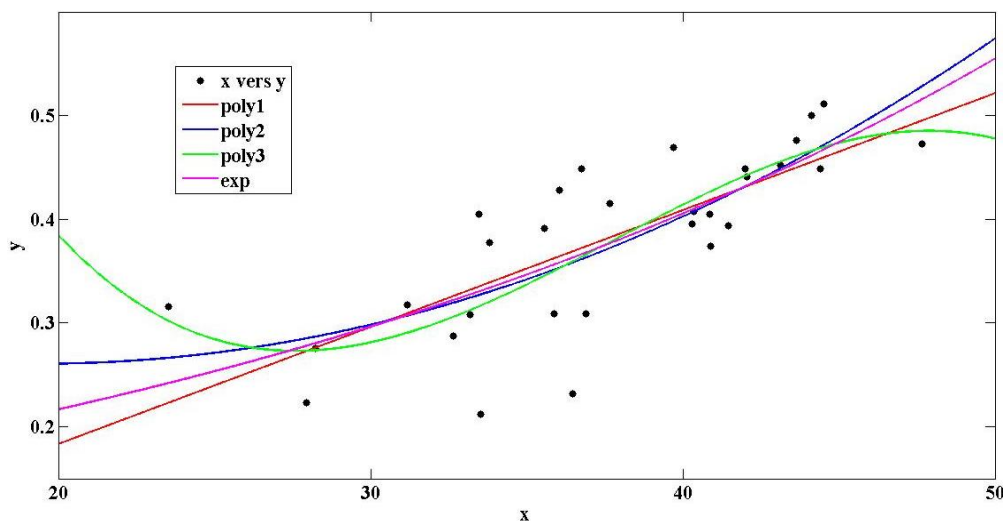


Figure 2 Different types of regression functions between bulk density x [$\text{kg} \cdot \text{m}^{-3}$] and tensile strength y [MPa]

The characteristics of regression analysis were explained in the theoretical part and are presented for individual models in the Tab. 3. It can be seen that with increasing degree of polynomial the sum of squares for error (SSE) decreases, the exponential model fits between linear and quadratic function. The root mean squared error (RMSE) decreases only to the 3rd degree polynomial, then begins to grow. The coefficient of determination R^2 improves with increasing degree of polynomial, but its modified form R^{*2} improves the regression model only to the 3rd degree of the polynomial, then begins to decrease. The meaning of SSR and SST characteristics are explained above, DFE represents the number of degrees of freedom in the model. The test of significance of the regression model as a whole is done at the level $\alpha = 0.05$ and since p values are less than this value, we can conclude that the linear, quadratic, cubic and 4th degree models significantly contribute to the estimation of the dependent variable. We did the test only to the 4th degree polynomial, because at higher degrees numerical problems arose.

Table 3 Regression characteristics of different mathematical models, SST = 0.2039

Deg. pol.	SSE	RMSE	R^2	R^{*2}	SSR	p	DFE
1 st	0.0866	0.0556	0.5753	0.5602	0.1173	1.19e-6	28
2 nd	0.0817	0.0550	0.5994	0.5697	0.1222	4.33e-6	27
3 rd	0.0762	0.0541	0.6261	0.5830	0.1277	9.43e-6	26
4 th	0.0760	0.0551	0.6275	0.5679	0.1279	3.86e-5	25
5 th	0.0729	0.0551	0.6423	0.5677	0.1310		24
6 th	0.0727	0.0562	0.6434	0.5504	0.1312		23
exp.	0.0832	0.0545	0.5918	0.5773	0.1207		28

In the Tab. 4, we present point estimates of the coefficients in regression models. For polynomials, the index of the coefficient a_j also means its position at the degree of the independent variable x^j . Since these are only values calculated from our particular realization of random variables X, Y, it is also appropriate to make confidence intervals of the coefficients, which are given in the Tab. 5. However, except for the linear and exponential model, all intervals at the higher degrees of the polynomial intersect zero, thereby degrading the regression model.

Table 4 Point estimates of the coefficients in regression models

Deg. pol.	a_6	a_5	a_4	a_3	a_2	a_1	a_0
1 st						1.127e-2	-4.184e-2
2 nd					3.345e-4	-1.298e-2	0.3865
3 rd				-5.081e-5	5.753e-3	-0.2014	2.519
4 th			1.737e-6	3.002e-4	1.896e-2	-0.5064	5.105
5 th		-1.04e-6	1.88e-4	-1.347e-2	0.4775	-8.371	58.2
6 th	-6.645e-8	1.333e-5	-1.093e-3	4.679e-2	-1.098	13.34	-64.86
exp.						3.134e-2	0.1158

Table 5 Confidence intervals (95%) of the coefficients in regression models

Deg. pol.	a ₆	a ₅	a ₄	a ₃	a ₂	a ₁	a ₀
1 st from						7.521e-3	-0.1843
1 st to						1.502e-2	0.1006
2 nd from					-2.041e-4	-5.219e-2	-0.3176
2 nd to					8.731e-4	2.624e-2	1.091
3 rd from				-1.275e-4	-2.44e-3	-0.4884	-0.7725
3 rd to				2.586e-5	1.395e-2	0.0856	5.81
4 th from			-1.011e-5	-2.003e-3	-7.145e-2	-2.606	-12.85
4 th to			1.358e-5	1.402e-3	0.1094	1.593	23.06
5 th from		-3.195e-6	-1.981e-4	-4.079e-2	-0.4766	-24.8	-53.25
5 th to		1.114e-6	5.741e-4	1.386e-2	1.432	8.055	169.6
6 th from	-5.641e-7	-9.432e-5	-1.07e-2	-0.4054	-12.94	-150.1	-993.5
6 th to	4.312e-7	1.21e-4	8.511e-3	0.4989	10.74	176.8	863.8
exp. from						2.088e-2	6.779e-2
exp. to						4.179e-2	0.1638

In the final Figures 3 - 6, we plot 95% confidence intervals of regression models by 1st, 2nd and 3rd degree polynomials as well as by the exponential model.

After reviewing all results, the linear and exponential forms of regression appear to be the best choices, while the exponential shows slightly more favorable values.

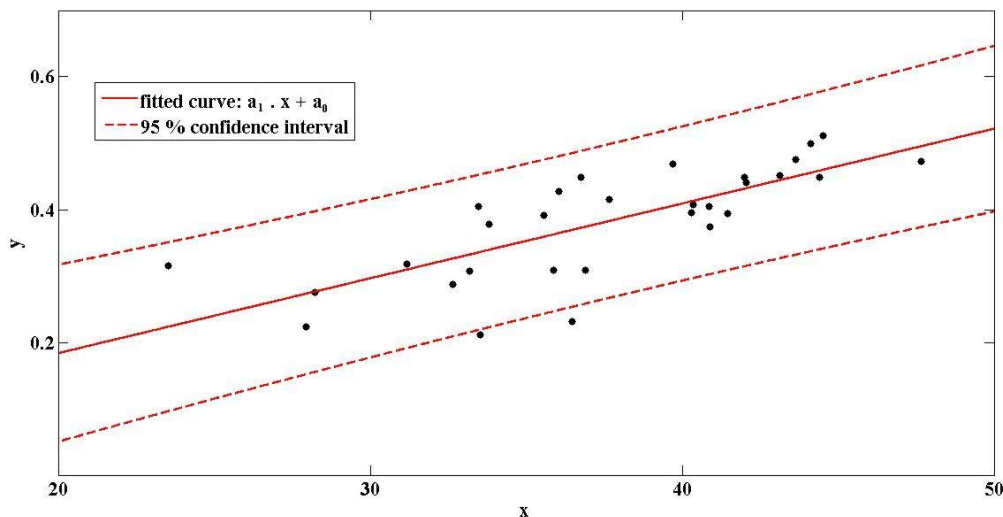


Figure 3 95% confidence interval for linear function: $y = a_1 \cdot x + a_0$

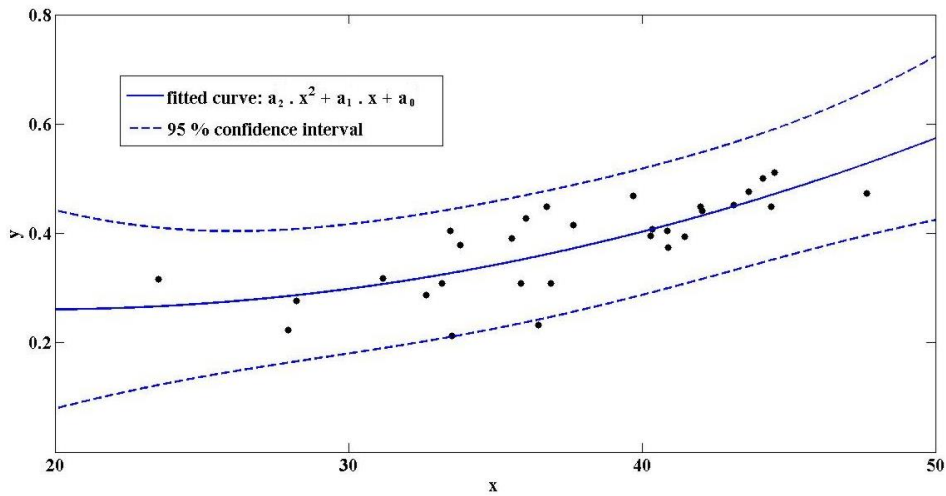


Figure 4 95% confidence interval for quadratic function: $y = a_2 \cdot x^2 + a_1 \cdot x + a_0$

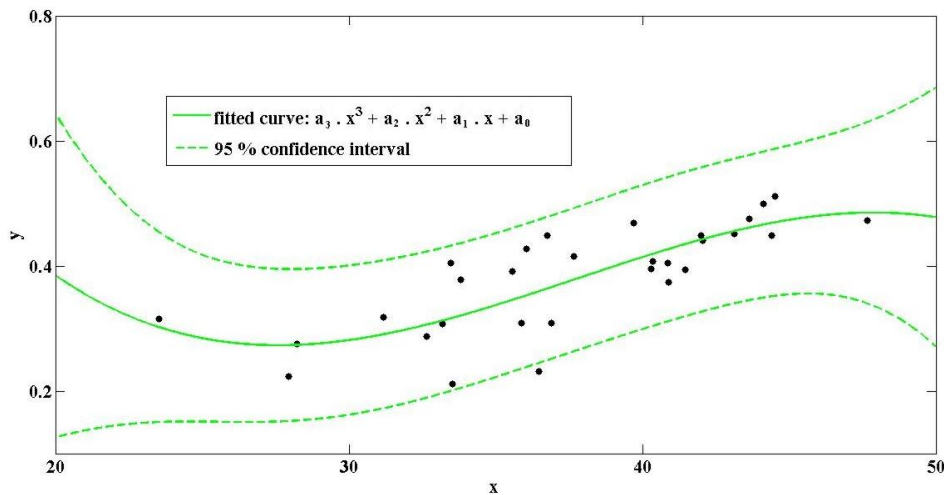


Figure 5 95% confidence interval for cubic function: $y = a_3 \cdot x^3 + a_2 \cdot x^2 + a_1 \cdot x + a_0$

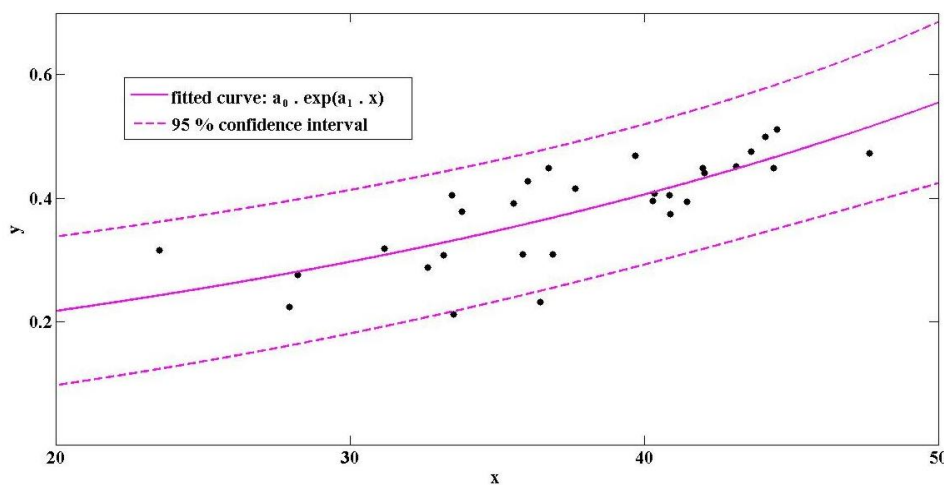


Figure 6 95% confidence interval for exponential function: $y = a_0 \cdot \exp(a_1 \cdot x)$

CONCLUSIONS

Expanded polystyrene (EPS) is a material that is concurrently environmental, practical and reliable [4]. Its use as a building material is miscellaneous and therefore deserves attention to determine its properties [3]. As EPS bulk density increases, tensile strength increases too [8]. It is important to as accurately as possible determine this functional dependence [11]. The intensity of the dependence is specified by various correlation coefficients, and their values in the range of about 0.6 - 0.8 indicate a relatively strong correlation found from the measured values [13]. The seven types of regression models used made it possible to calculate the characteristics of the corresponding model [12]. As we can see from the results, it is not necessary to complicate mathematical formulations of regression, it is sufficient to have a linear polynomial, which shows the best parameters and is also easier to calculate. Better came only exponential function, which also has a relatively uncomplicated formula.

REFERENCES

- [1] Ching, J., Phoon, K.K., & Li, D.Q. (2016). Robust estimation of correlation coefficients among soil parameters under the multivariate normal framework. *Structural Safety*, 63, pp. 21-32, doi: <https://doi.org/10.1016/j.strusafe.2016.07.002>
- [2] Chung, S. Y., Abd Elrahman, M., & Stephan, D. (2018). Effects of expanded polystyrene (EPS) sizes and arrangements on the properties of lightweight concrete. *Materials and Structures*, pp. 51-57, doi: <https://doi.org/10.1617/s11527-018-1182-3>
- [3] Dafalla, M., Al Shuraim, M., & Shaker A. (2019). Influence of density and thickness on polystyrene insulated concrete block walls. *Int. J. of GEOMATE*, Vol. 17, No. 59, pp. 62-67, doi: <https://doi.org/10.21660/2019.59.8108>
- [4] ISOVER Saint-Gobain, *EPS Expanded polystyrene – properties* (in Slovak). Retrieved 2020-01-29 from <https://www.isover.sk/eps-expandovany-polystyren>
- [5] Novák, D., Teplý, B., & Shiraishi, N. (1993). Sensitivity analysis of structure: A review. *Civil Comp '93*, Edinburgh, Scotland, pp. 201-207.
- [6] Ostertágová, E. (2015). *Applied statistics in the computer environment MATLAB*. Technical University Košice, 175 p. (in Slovak).
- [7] Páleš, D., Lendelová, J., & Švec, O. (2000). Calculation of the rank correlation coefficients. *Workshop: Construmat 2000*, Slovak University of Agriculture, Nitra, pp. 148-152 (in Slovak).
- [8] POLYFORM Podolíneč, *Expanded polystyrene* (in Slovak). Retrieved 2020-01-29 from <https://polyform.sk/wp-content/uploads/2016/04/eps.pdf>
- [9] Prieceľ, J. (1982). *Application of plastics for cover and roof constructions of agricultural buildings*. Candidate thesis, Slovak University of Agriculture in Nitra, Nitra, 143 p. (in Slovak).
- [10] Prieceľ, J., & Páleš, D. (1998). Physical-mechanical properties of expanded polystyrene depending on the bulk density. *Workshop: Construmat '98*, University of Žilina, Súčov, pp. 191-195. (in Slovak).
- [11] Sdružení EPS ČR, *Properties of expanded polystyrene (EPS)* (in Czech). Retrieved 2020-01-29 from <https://stavba.tzb-info.cz/tepelne-izolace/8482-vlastnosti-expandovaneho-penoveho-polystyrenu-eps>
- [12] Sepah Vand, A., Sihag, P., Singh, B., & Zand, M. (2018). Comparative Evaluation of Infiltration Models. *KSCE Journal of Civil Engineering*, 22(10), pp. 4173-4184. doi: <https://doi.org/10.1007/s12205-018-1347-1>
- [13] Vakharia, V., & Gujar, R. (2019). Prediction of compressive strength and Portland cement composition using cross-validation and feature ranking techniques. *Construction and Building Materials*, 225, pp. 292-301. doi: <https://doi.org/10.1016/j.conbuildmat.2019.07.224>
- [14] Worden, K., Sohn, H., & Farrar, C.R. (2002). Novelty detection in a changing environment: Regression and interpolation approaches. *Journal of Sound and Vibration*, 258 (4), pp. 741-761. doi: <https://doi.org/10.1006/jsvi.2002.5148>