Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2026, 20(10), 1–20
https://doi.org/10.12913/22998624/222712
ISSN 2299-8624, License CC-BY 4.0

Received: 2026.04.23
Accepted: 2026.08.13
Published: 2026.09.01

Prediction of tensile strength of PHBV-sunflower husk biocomposites: Challenges and limitations of linear regression models

Grzegorz Janowski[1], Marcin Leszek Olech[2*], Radosław Frącz[2], Jarosław Gawryluk[3], Aneta Leś[4]

[1] Department of Materials Forming and Processing, Rzeszow University of Technology, Powstańców Warszawy 8, 35-029 Rzeszów, Poland
[2] Department of Computer Science, Rzeszow University of Technology, Powstańców Warszawy 8, 35-029 Rzeszów, Poland
[3] Department of Applied Mechanics, Lublin University of Technology, Nadbystrzycka 36, 20-618 Lublin, Poland
[4] Polimarky Sp. z o.o., Bieszczadzka 10 a, 35-082 Rzeszów, Poland
* Corresponding author's e-mail: molech@prz.edu.pl

ABSTRACT

The aim of this study was to develop a model for predicting the properties of biocomposites, based on a PHBV biocomposite with sunflower husks, using statistical optimization methods and regression-based predictive modeling. Examination of the tensile strength value according to the different input production parameters, like mold temperature, cooling time, packing pressure, and injection rate, was subjected to a detailed analysis. Various variants of linear regression models were considered, i.e., effect coding of input variables and the original input parameter values. In developing the predictive models, the significance of the variables used was assessed using linear correlation analysis with heat map generation, Taguchi SNR response matrix analysis, and ANOVA analysis. Additionally, the interactions between input variables were analyzed. The developed predictive models were also validated using the Leave-One-Out-Cross-Validation method. It was determined that mold temperature and cooling time are the key processing parameters that most strongly influence the tensile strength of PHBV–sunflower husk biocomposites.

Keywords: PHBV sunflower husk biocomposites, material property prediction, injection molding, Taguchi method, ANOVA analysis, linear regression.

INTRODUCTION

Modern materials science is increasingly focused on developing materials that align with the concept of a circular economy and searching for sustainable alternatives to petrochemical-derived plastics. A key element of this evolution is the development of biopolymer-based composites, such as poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (PHBV), reinforced with natural fillers, among which sunflower seed hulls (SSH) are one of the most promising and demanding waste raw materials [1]. PHBV, as a microbiologically derived polyester, is characterized by full biodegradability in natural conditions and high biocompatibility, making it an ideal candidate for applications in packaging, agriculture, and even medicine [2]. However, technical barriers related to its high brittleness, narrow processing window, and sensitivity to thermal degradation require a precise approach to designing the composition and manufacturing parameters [3].

The introduction of lignocellulosic fillers from sunflower husk into the PHBV matrix allows for a significant reduction in material costs and an improvement in the stiffness and dimensional stability of products, while maintaining their ecological character [1]. Nevertheless, predicting the properties of such complex systems is a task burdened with a high degree of uncertainty due to the biological heterogeneity of the filler, its variable chemical composition, and complex interfacial interactions [4]. Traditional optimization methods based on the trial-and-error approach are becoming insufficient, which forces the implementation of advanced statistical techniques, such as the Taguchi method and analysis of variance (ANOVA), as well as modern machine learning (ML) algorithms [5].

The sensitivity of biocomposites with plant fillers is multicomponent and encompasses both material and process variability. Natural fibers/fillers are inherently hydrophilic, while typical thermoplastic matrices (including PHBV) are relatively hydrophobic. This incompatibility promotes poor dispersion and poor stress transfer, which can lead to reduced mechanical properties and increased result scatter [6]. The literature on biocomposites emphasizes that adhesion at the interface is critical, as it determines the efficiency of stress transfer and the final benefit of reinforcement [7]. Furthermore, moisture and water sorption in lignocellulosic composites significantly complicate prediction, as they lead to particle swelling, void formation, and microcracks at the interface, interface degradation, and property changes over time. In the case of PHBV–sunflower husk biocomposites, it is explicitly indicated that the addition of natural fillers increases water absorption, and differential swelling of the matrix and the rigid phase initiates microcracks and capillary channels; the absorption mechanism tends to be parallel, and for high filler contents, the phenomenon becomes more complex [1]. Variability of processing parameters and their interactions can also strongly modulate the microstructure (in the context of, among others, porosity, wetting, orientation, crystallinity), and thus the mechanical properties. Even in composites with low filler concentrations, the process can determine the dispersion quality and homogeneity of the molded part [5]. For PHBV, an additional risk factor is the narrow processing "window" and sensitivity to overheating, underscoring the importance of precise modeling of temperature and melt time.

Predicting the properties of PHBV-SSH is uniquely challenging due to the biological nature of the reinforcement. Unlike synthetic fibers (glass or carbon), plant raw materials are characterized by variability resulting from vegetation conditions, climate, and harvest time [4]. The heterogeneity of SSH manifests itself not only in composition but also in morphology. Microscopic studies reveal the presence of both fibrous and lamellar structures originating from different seed coat layers [1]. This geometric diversity makes the stress transfer mechanisms in the composite nonlinear and difficult to describe with simple analytical models. It has been demonstrated that the high lignin and hemicellulose content of SSH promotes the immobilization of polymer chain segments on the filler surface, leading to complete suppression of PHBV crystallization under rapid cooling conditions during injection. From a practical point of view, the SSH filler drastically improves dimensional stability, significantly reducing processing shrinkage, in some cases even by about 80%, which makes this material attractive for the production of precision technical parts [1].

A detailed characterization of PHBV–sunflower husk (SSH) biocomposites, including their mechanical, thermal, and structural behavior, was presented in the study [1]. It was shown that incorporating lignocellulosic filler increases stiffness, while reducing tensile strength and elongation at break due to limited interfacial adhesion and stress concentration effects. The same study also indicated that SSH affects PHBV crystallization and microstructure formation, which confirms that the mechanical performance of PHBV–SSH composites should be analyzed in the context of processing–structure–property relationships [1].

Despite this growing understanding of the material behavior, there is still a lack of comprehensive studies addressing the prediction of mechanical properties of PHBV-based biocomposites, particularly with regard to the combined influence of processing parameters and natural filler variability. In particular, there are few studies using statistical methods and machine learning algorithms to model the properties of PHBV composites with lignocellulosic fillers. Therefore, the aim of this study was to develop a model for predicting the properties of biocomposites, based on a PHBV biocomposite with sunflower husks, using statistical optimization methods and regression-based predictive modeling. However, due to some issues found in the literature, like no consideration of how the coding type of input variables affects the model, presentation of the prediction model's metrics only in terms of the train dataset, with no verification of the generalization aspect of the model, current research was limited according to the initial plans. The analysis and modeling of the measured mechanical properties, as well as the implementation of predictive models using various methods such as Ridge Regression, Lasso Regression, Decision Trees, Random Forests, Artificial Neural Networks, Support Vector Regression, Gradient Boosting Regression, and Extreme Gradient Boosting, are planned for the subsequent article. Therefore, the current paper focuses on the predictive model-building process for tensile strength as a function of selected injection molding parameters: mold temperature, cooling time, packing pressure, and injection rate. The interactions between input variables were also analyzed to identify possible relationships relevant to process optimization. The work extends previous predictive approaches reported in similar research areas [8, 9].

MATERIALS AND MANUFACTURING

Research materials

This work utilized PHBV, a naturally derived polymer matrix, and sunflower husks as the filler. PHBV belongs to the polyhydroxyalkanoates (PHA) group and is a copolymer of poly(3-hydroxybutyric acid) (PHB) and poly(3-hydroxyvaleric acid) (HV). The filler intended for the biocomposite was in the form of sunflower seed husks after hulling sunflower seeds intended for human consumption. The husks were supplied in their original form, meaning they were not physically or chemically modified. In the present study, all experimental runs were performed using a fixed composite formulation consisting of 70 wt% PHBV and 30 wt% ground sunflower husk filler. The filler content was not varied and was not included as a control factor in the Taguchi experimental design. A Chemland FW177 laboratory mill was used to grind the sunflower husks. The milling process was carried out for 180 seconds to obtain short fibers. Thick and large pieces of sunflower husks would have caused problems during extrusion. The husks and their condition after milling are shown in Figure 1.

To analyze the structure of sunflower husks after milling, the shape and size of the ground particles were examined using a Nikon MM-800 workshop microscope with Nikon E-MAX software. Figure 2 shows typical ground sunflower husk particles. The particles are clearly fiber-shaped. Measurements of the fiber length and width indicate an average length of 1.74 mm and an average width of 0.62 mm.

a) b)

Figure 1. Sunflower husks a) before grinding, b) after grinding

Composite production

The composite was produced using a Zamak EHP 25-E single-screw laboratory extruder. Extrusion was performed at a speed of 30 rpm. The temperatures of the individual heating zones of the plasticization system are presented in Table 1. Granulate was produced from the extruded composite using a Zamak granulator.

Test sample production

Samples for testing mechanical and physical properties were produced using injection molding technology. The molded pieces were injected using Taguchi orthogonal plans. Initial parameter ranges were defined through preliminary tests based on the quality of the molded pieces obtained. The polymer temperature was not changed. It was set at 174 °C, which ensured the correct viscosity of the material visually, without any tendency to excessive flow. All specimens used in the Taguchi experimental plan were produced from the same PHBV–sunflower husk composite formulation, i.e., 70 wt% PHBV and 30 wt% ground sunflower husk. Therefore, the variables considered in the experimental design were limited to injection molding process parameters, whereas the filler content was kept constant. These properties were assessed visually and based on the readings of the polymer pressure sensors mounted in the injection mold cavity.

The Priamus system was used to assess the biocomposite viscosity and monitor the polymer viscosity. Temperature and pressure sensors were placed in two injection mold cavities. The molds were produced in a typical dog-bone shape (Type A according to PN ISO 527-1). The injection process quality and the switch-over point were determined based on the readings of the pressure and temperature sensors using the Priamus system. The adjustable parameters are defined in Table 2.

The chosen mechanical properties included: Young's modulus, tensile strength, and sample elongation were determined. Particular attention was paid to the tensile strength of composites.

The Taguchi method was used to conduct the optimization analysis. An L16 orthogonal table was selected for optimization. Four control factors were determined for four levels of processing parameters variation. Sixteen sets of 10 samples each were produced (over 160 injection cycles in total) for the specified adjustable parameters. Complete molded parts were obtained for all configurations, free from geometrical defects and burns.

APPLIED METHODS DESCRIPTION

Mechanical properties testing method

Basic mechanical properties were tested using a static uniaxial tensile test in accordance

a) b)

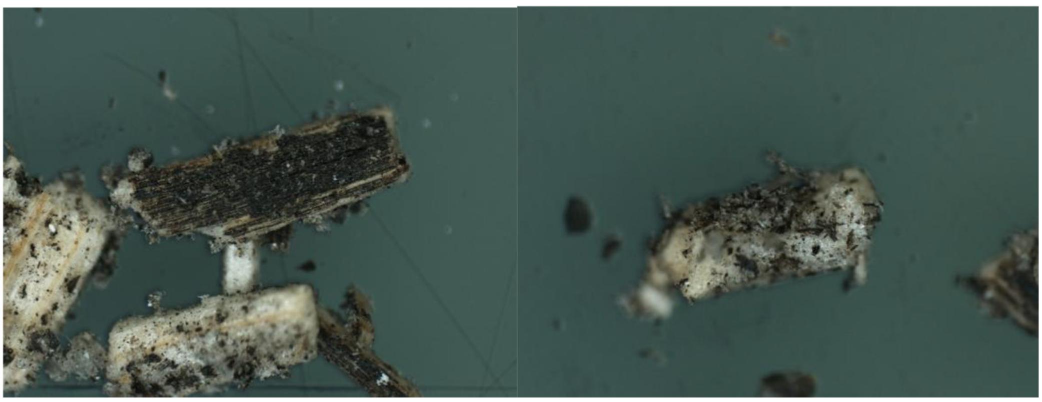

Figure 2. Geometry of typical ground sunflower husk particles. (magnification 30x)

Table 1. Temperature of the extruder heating zones used in the production of the PHBV sunflower husk composite

Head [°C]	Zone 3 [°C]	Zone 2 [°C]	Zone 1 [°C]	Charge [°C]
162	159	158	157	29

Table 2. Adjustment parameters for the injection molding of samples for testing the properties of injection molded parts

Level	Mold temperature [°C]	Cooling time [s]	Packing pressure [MPa]	Injection rate [cm³/s]
1	65	20	35	25
2	70	25	40	35
3	75	30	45	45
4	80	35	50	55

with ISO 527-1 using a Zwick Z030 tensile testing machine. The results of the tensile strength tests are summarized in Table 3.

Linear regression method

The analytical part of the research was conducted mainly by using the linear regression method, which is described briefly in the current section of the paper.

Areas of application

Linear regression remains one of the fundamental predictive tools, which allows building the reference predictive model that the other techniques refer to. It is widely used in different areas of applications, including material engineering, where it is used, e.g., to predictive modeling of Boron Nitride Nanoplatelet-Reinforced PLA Nanocomposites and their mechanical properties [5], to estimate optimal production parameters according to the expected final value of chosen mechanical properties of SS404L Alloy [13], or to model chosen plastic properties in metallic glasses [14]. Moreover, the popularity of applying predictive modeling to established required mechanical properties during the production process can be easily noticed in review papers, e.g., [15], where results of those studies are presented.

Fundamental characteristic

The linear regression belongs to the category of supervised machine learning methods, and it is well described in the literature [16] [12]. It allows for the modeling of a linear relationship between a dependent (output) variable and one or more independent variables [9]. Moreover, it can be applied to examine interactions between chosen input variables and their joint influence

Table 3. Summary of the test results of mechanical properties determined in the static tensile test for individual sets of adjustable parameters

Mold temperature [°C]	Cooling time [s]	Packing pressure [MPa]	Injection rate [cm³/s]	Tensile strength [MPa]
65	20	35	25	25.65
65	25	40	35	25.32
65	30	45	45	25.42
65	35	50	55	24.89
70	20	40	45	25.26
70	25	35	55	24.98
70	30	50	25	24.6
70	35	45	35	24.51
75	20	45	55	24.61
75	25	50	45	24.67
75	30	35	35	24.42
75	35	40	25	24.39
80	20	50	35	24.85
80	25	45	25	24.49
80	30	40	55	24.41
80	35	35	45	24.00

on the output variable. Of course, some assumptions must be complied with, such as linearity between independent and dependent variables, independence of error term, normality of errors, homoscedasticity (constant variance of errors), and no multicollinearity [9]. Depending on the number of independent variables, we can distinguish between simple linear regression (SLR) and multiple linear regression (MLR) [10]. The model of SLR is depicted in Equation 1, while the MLR model is presented in Equation 2, respectively [12].

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

$$y = \beta_0 + \beta_1 x + \cdots + \beta_p x_p + \varepsilon \quad (2)$$

where: β_0 – intercept term, the expected value of y if x = 0, $\beta_1 \ldots \beta_p$ – slope term, the average increase in y if x_p increases by one unit, ε - mean-zero random error term.

The main goal of linear regression is to fit a linear equation according to the real data, which allows for generating possible predictions of output value as well as to examine relationships between input and output variables [11].

The fitting process typically uses the least-squares method to determine the coefficients of the SLR or MLR equation [11]. The method is performed by minimizing the sum of squared residuals (RSS) as depicted in Equation 3 [12].

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \\ = \sum_{i=1}^{n} \begin{pmatrix} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} + \\ -\hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip} \end{pmatrix}^2 \quad (3)$$

where: y_i – real value of the ith observation of the output variable, $\hat{y}_i$ – predicted value for the ith observation of the output variable, $\hat{\beta}_0$ – estimated value of the intercept term by using the least squares method, $\hat{\beta}_1 \ \ldots \ \hat{\beta}_p$ – estimated value of slope terms according to the pth input variable by using the least squares method, $x_{i1} \ldots x_{ip}$ – real value of the pth input variable in the ith observation.

Applying the fitting process allows establishing intercept and slope terms to make predictions according to Equation 4 [12].

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \cdots + \hat{\beta}_p x_p \quad (4)$$

Of course, because of the data characteristic, the linear regression model should be evaluated by using appropriate metrics according to the model itself as well as to the predicted values.

Linear model and prediction accuracy metrics

The obtained prediction model can be evaluated typically according to the two main metrics R^2 and R^2 adjusted, which was done in the current paper. The first one (R^2 – coefficient of determination) allowed evaluating how much the output variable is determined by the input variables in terms of proportion of variance [17]. The main disadvantage of the R^2 is the fact that adding more features to the model brings the R^2 value higher; however, in some cases, the prediction ability of the model stays the same or even decreases. Here, another important metric can be proposed, which is adjusted R^2. The metric takes into consideration not only the original R^2 value, but also the number of observations in the dataset as well as the number of input variables used in the model [18]. Of course, the quality of the prediction model can also be evaluated according to the generated predictions from the model, if compared to the real output values. Here, another set of metrics can be applied like mean average error (MAE), mean average percentage error (MAPE), or root mean squared error (RMSE), which were applied in the paper. Those metrics allow calculating errors, which are generated by the prediction model in different ways. Their definitions, together with the mathematical background and discussion, can be found in [17] and [19].

Interactions between input variables

An interesting aspect of the linear regression model, especially in terms of material engineering, is connected to the ability to examine if and how interactions between chosen input variables (process parameters) affect the output variable (mechanical property). This can be done by extending the basic regression model of chosen input variables by another predictor called the interaction term. It is constructed by computing the product of chosen input variables. After the regression equation is fitted, the predictions of the output variable can be made, and interaction graphs can be prepared. The interaction graph shows if interaction exists, which means the input variables stay in interaction or graph lines stay parallel to each other, which means the input variables are independent of each other. Detailed information can be found in [12]. Of course, if source data comes from the Taguchi method, the visible interaction can result in noise, so the technologist needs to check if observable potential dependency is possible according to the physics of the process.

PREDICTIVE MODELS DEVELOPMENT PROCESS

Main assumptions

The two main goals of the analytical work presented in the paper are connected to:

- the investigation process, what key parameters of the production process need to be identified,
- predictive model development process, where the most important parameters of the production process are applied to construct the prediction model, which will be used for further prediction of mechanical properties.

The production process itself is performed according to a set of four input variables: mold temperature, cooling time, packing pressure, and injection rate, and considers four output variables: tensile strength, Young's modulus, elongation at break, shrinkage, and molded part weight.

The detailed analysis according to both main goals is presented fully according to the tensile strength variable. Due to the limitations of the paper, other output variables are planned to be examined in the future.

However, according to the similar works that can be found in the literature [5], two important remarks, which explain the context of the current analytical research, have been formulated:

- Remark 1.

It is worth to mention the obtained dataset contains average values of output variables according to the set of predetermined production parameters, which is typical for the DOE approach. It causes, however, that some prediction methods cannot be successfully established due to the limitation of the input dataset with regard to the size of the dataset in terms of a low number of rows and variables, and with regard to the fact that the input variables are predefined parameters (categorical alike). E.g., artificial neural networks seem to be an inappropriate tool, because in this case, the dataset needs to be split into training and verification sets, which is difficult if each row contains unique values.

- Remark 2.

The classic and simple approach, according to the prediction model development process, is typically oriented to linear regression. The obtained linear model can be used as a reference model for testing other techniques, which allow the creation of different prediction models. However, the regression model itself, which can be found e.g. in the literature [5], was evaluated only by metrics like R2 and R2 adjusted. Because of the fact, the regression model is based on the Taguchi orthogonal table, which forms the whole training dataset, evaluations of the generalization aspect of the model are generally not provided. To bridge the gap, that aspect is addressed in the current paper.

To conclude, we have focused on the following issues:

- which variables should be considered as important factors of the regression equation – the identification of the most important input variables and their possible interactions is very important not only because of the regression equation, but also because it can influence the output value of mechanical properties, which should be examined and predicted;
- how input variables should be coded – in Taguchi analysis, effect coding or one-hot encoding is typically considered – that solution causes the specific factors existing in the regression equation to be set on or off; so it is not possible to make the prediction for input parameters that are between planned values of factors. Another possible option is to use input parameters as continuous values, but here the quality of the model needs to be assessed as well as its ability to generalize results (to predict according to new, unknown before input values of input variables);
- applying the model on the whole dataset simply can easily lead to overfitting and poor generalization results; there is a need to verify the generalization aspect of the proposed model. It can be achieved by using data stored in the Taguchi orthogonal table and Leave-One-Out-Cross-Validation (LOOCV) or by making predictions according to the real data before their aggregation to the Taguchi orthogonal table.

Identification of the most important input variables according to the tensile strength

Before the prediction model is established, the key factors identification process has been completed. The identification can be performed by using different methods, but we have focused on: Pearson's correlation coefficient, Taguchi response matrix, and Analysis of variance (ANOVA).

Key factors identification using Pearson's correlation coefficient

The easiest possible way to check if the potential input variables affect the output variable is the calculation and visualization of Pearson's correlation coefficient. The calculated correlation matrix heat-map is presented in Figure 3. As can be seen, Mold temperature has the strongest negative influence on tensile strength, equal to -0.76. The second is the Cooling time with the same relation and value of -0.54. The last two variables (Packing pressure and Injection rate) have almost no influence on the output value. At this point, there is an initial recommendation for choosing the most important variables for the regression equation. Of course, there is a chance that the interactions between those variables will be important.

Key factors identification using "Signal-to-Noise-Rate" (SNR) response matrix

The second method for the identification of key input variables is based on the Taguchi response matrix. In this case, the mean value of Tensile Strength can be analyzed as well as SNR values. The selection criterion applied in the case was the "large-the-better". The response table is presented in Table 4, where mean values of Tensile Strength are presented, while in Table 5, the SNR values are shown according to the proposed levels of each factor. The delta values are marked with a color to help with the identification of the most important variables.

In both cases the analysis has shown that two input variables: Mold temperature and Cooling time are the most important factors affecting the Tensile Strength. The other two variables have

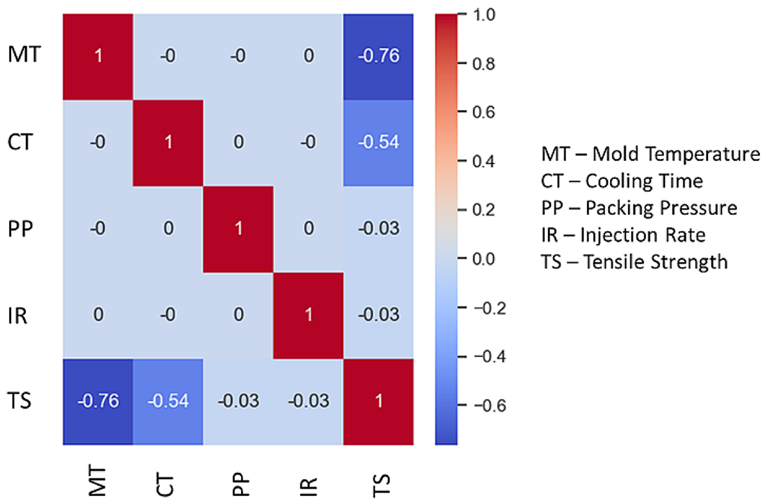

Figure 3. Correlation matrix heat-map

Table 4. Response table for tensile strength according to the mean value

Mean of tensile strength				
Level/Factor:	Mold temperature [°C]	Cooling time [s]	Packing pressure [MPa]	Injection rate [cm^3/s]
1	25.3200	25.0925	24.7625	24.7825
2	24.8375	24.8650	24.8450	24.7750
3	24.5225	24.7125	24.7575	24.8375
4	24.4375	24.4475	24.7525	24.7225
Delta	0.88	0.65	0.09	0.12
Rank	1	2	4	3

a marginal outcome. This confirms the results of Pearson's correlation coefficient analysis. Moreover, the situation can be visually identified according to Figure 4.

The dominant influence of mold temperature and cooling time can be explained by considering the mechanisms of structure development during injection molding. In PHBV-based systems, thermal history plays a crucial role in determining the degree of crystallinity and the resulting microstructure. Mold temperature directly affects the cooling rate of the material, influencing the kinetics of crystallization and the relaxation of internal stresses. Higher mold temperatures may lead to slower cooling and structural rearrangement, reducing tensile strength due to changes in crystalline morphology and potential relaxation effects.

Cooling time, on the other hand, determines the duration over which the structure evolves under controlled thermal conditions. Extended cooling time may promote additional structural changes, including stress relaxation and potential development of microstructural heterogeneities, which can negatively affect tensile strength. In PHBV–lignocellulosic composites, this effect is further amplified by the presence of sunflower husk particles, which restrict polymer chain mobility and contribute to heterogeneous structure formation.

Table 5. Response table for tensile strength according to SNR

Signal-to-noise-rate of tensile strength				
Level/Factor:	Mold temperature [°C]	Cooling time [s]	Packing pressure [MPa]	Injection rate [cm³/s]
1	28.0688	27.9898	27.8732	27.8811
2	27.9015	27.9111	27.9034	27.8794
3	27.7912	27.8571	27.8731	27.8999
4	27.7605	27.7640	27.8723	27.8615
Delta	0.31	0.23	0.03	0.04
Rank	1	2	4	3

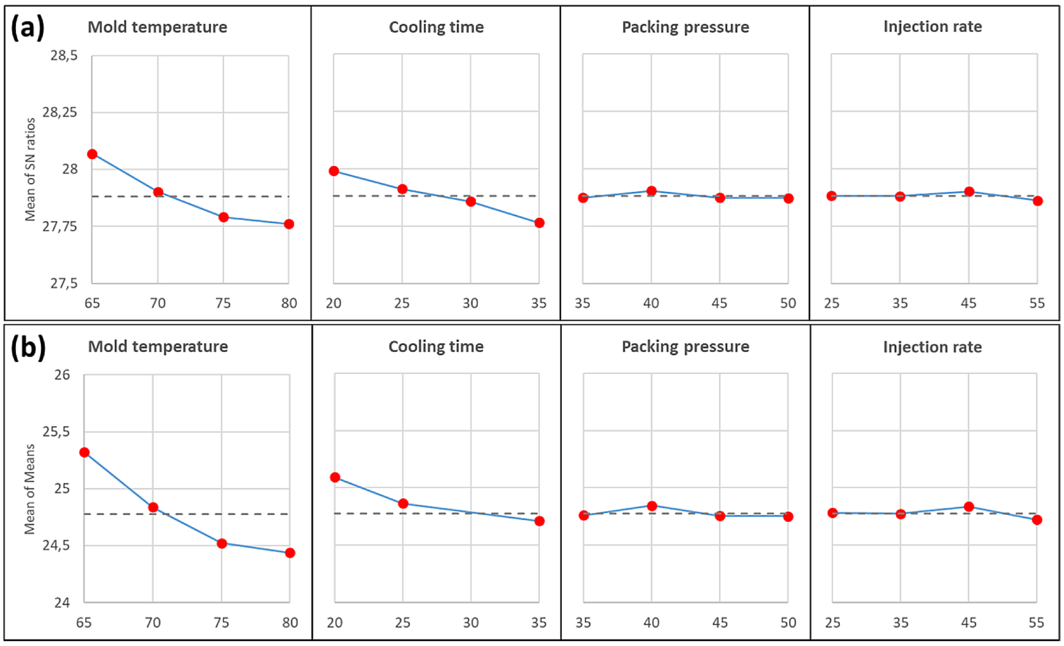

Figure 4. Main effect plot of (a) S/N ratio for Tensile Strength, (b): Mean for tensile strength (large is better)
Note: The graphs indicate that temperature is the most important factor. Both graphs show that as the temperature increases from 65 °C to 80 °C, the tensile strength decreases. The second most important factor is cooling time, and in this case, tensile strength decreases with increasing cooling time. Both the mean signal-to-noise ratio graph and the average-of-averages graph show similar trends. Both graphs indicate that the injection rate and packing pressure have no effect on the tensile strength.

These observations are consistent with previous findings for PHBV–SSH systems [1], where the final mechanical performance was shown to strongly depend on the interplay between crystallization behavior, filler-induced constraints, and processing conditions. Therefore, the dominant role of mold temperature and cooling time identified in the statistical analysis is not only a result of regression modeling, but also reflects the underlying material physics governing structure–property relationships in such biocomposites.

Key factors identification using analysis of variance (ANOVA)

The last step, connected to the identification of the most important input variables, was done by applying ANOVA. The results of the analysis show which factors are statistically important and how large their contribution is to explain the target value of tensile strength. The results of the analysis are presented in Table 6.

The ANOVA analysis confirmed that the Mold Temperature is the most important factor, which allows for explaining almost 64% of the tensile strength value. What is more, it is also statistically important according to the calculated p-value=0.0353, which is less than the threshold = 0.05. Another important parameter is Cooling time, with a contribution of 29%, but here the p-value is slightly higher than the threshold, so it means that the parameter is not statistically important if the threshold is set to 0.05. The same is true for two others: Packing pressure and Injection rate, with the difference that both have quite small contribution values compared to Cooling time. The interesting thing is connected to the R^2 value, which is quite high (almost 95%), while R^2 adjusted is equal to 73% – the situation suggests the regression prediction model will probably be overfitted, which can cause problems with prediction, if new data, which were not included in the original dataset, will be provided in the new data set. Thus, it is recommended to conduct verification of the obtained models.

To complement the statistical analysis, the identified key process parameters are interpreted in the context of material behavior and structure development, as summarized in Table 7. It is worth mentioning that the small effect of the packing pressure and injection rate variables on tensile strength may be related to the selected production process parameters. The use of other process input parameters, where their range will be broader, could make this relationship more visible. Of course, this would require the production process and testing process of additional specimens. However, even using the current production parameters may reveal the influence of these variables on other mechanical properties of the tested biocomponent, because this time only tensile strength was analyzed. A similar analysis was planned to be presented in the next article, where other mechanical properties will be examined: Young's modulus, sample elongation, and mold part weight.

Although no additional SEM or DSC measurements were performed within the present study, the statistical importance of mold temperature and cooling time can be interpreted in relation to the known structure development mechanisms of PHBV–sunflower husk biocomposites. These parameters determine the thermal history of the molded part, including cooling rate, crystallization kinetics of PHBV, stress relaxation, and microstructure formation during solidification. The presence of lignocellulosic sunflower husk particles may additionally restrict polymer chain mobility and promote heterogeneous structure formation, which makes the final tensile strength sensitive to processing conditions. Therefore, the observed statistical trends are

Table 6. Optimization of processing parameters according to tensile strength by using ANOVA

Factor	df	Seq SS	MS	F-value	p-value	SS%
Mold temperature [°C]	3	1.9141	0.6380	12.0292	0.0353	63.74%
Cooling time [s]	3	0.8800	0.2933	5.5303	0.0969	29.30%
Packing pressure [MPa]	3	0.0232	0.0077	0.1456	0.9261	0.77%
Injection rate [cm³/s]	3	0.0266	0.0089	0.1670	0.9122	0.88%
Error	3	0.1591	0.0530			5.30%
Total	15	3.0029				100.00%
	R^2:	94.70%	R^2(adj):	73.51%		

Table 7. Relationship between statistically identified key process parameters and their physical interpretation in PHBV–SSH biocomposites

Parameter	Statistical significance	Observed effect on tensile strength	Physical interpretation
Mold temperature	Highest (ANOVA, correlation, SNR)	Increase → decrease in tensile strength	Controls cooling rate and crystallization kinetics; affects crystalline morphology and internal stress relaxation
Cooling time	High (2nd rank)	Increase → decrease in tensile strength	Governs the duration of structural evolution; influences crystallization development and stress relaxation processes
Packing pressure	Low	Minor influence	Affects material densification and void reduction; secondary role compared to thermal effects
Injection rate	Low	Negligible influence	Limited impact on flow-induced orientation; minor effect on final microstructure under studied conditions

physically consistent with the expected behavior of PHBV-based lignocellulosic biocomposites. Nevertheless, further SEM and DSC analyses of specimens produced under different processing conditions would be valuable to directly confirm the relationship between processing parameters, microstructure, crystallization behavior, and tensile properties.

Applying linear regression to establish a prediction model

The dataset originates from an experiment performed according to the Taguchi approach. It means that the potential regression model can be established by using the direct value of production parameters stored in the Taguchi orthogonal matrix as continuous values or by processing those values through specific coding (transformation). Typically, in this case, effect coding or one-hot encoding is applied. To show differences, both potential solutions have been implemented. Another aspect that needs consideration is connected to the selection process of input variables. As can be seen in the "Key factors identification" section, ANOVA suggests only one variable (Mold temperature), while the Taguchi response table suggests an additional variable (Cooling time); the same is true for linear correlation. The model can also be implemented by using all input variables. To check what solution will be optimal, a total of six regression models have been established where 1, 2, or 4 input variables were considered, as well as direct values or effect coding taking place. The results are presented in Table 8, where R^2, R^2 adjusted, MAE, MAPE, and root mean square error are presented. To make the analysis easier, conditional formatting was applied to help to find the most important values.

As can be seen in Table 8, the highest possible R^2, together with the smallest MAE, MAPE, and RMSE, is the first model, where all input variables were taken under consideration and effect coding has been performed. However, that solution shows a significant difference between R^2 and R^2 adjusted, which can suggest overfitting to the training data. The better solution seems to be the 2nd model, where effect coding was applied, and two variables: mold temperature and cooling time were used during model creation. This time, the R^2 adjusted is closer to R^2, while other metrics are slightly worse. It is interesting that using only one variable suggested by the ANOVA analysis and the p-coefficient brings a poor regression

Table 8. Metrics of implemented regression models according to the type of coding and the set of input variables

Coding type	Input variables	Metrics calculated according to the training dataset						
		R^2	R^2 adj.	MAE	MAE-SD	MAPE	MAPE-SD	RMSE
Effect coding	All input variables	0.9470	0.7351	0.0806	0.0587	0.33%	0.24%	0.0997
	Mold temperature Cooling time	0.9304	0.8841	0.0943	0.0645	0.38%	0.26%	0.1143
	Mold temperature	0.6374	0.5468	0.2119	0.1522	0.85%	0.61%	0.2609
No coding	All input variables	0.8766	0.8317	0.1145	0.1002	0.46%	0.40%	0.1522
	Mold temperature Cooling time	0.8748	0.8555	0.1171	0.0989	0.47%	0.40%	0.1533
	Mold temperature	0.5845	0.5549	0.2359	0.1494	0.95%	0.60%	0.2792

model. The regression equation of the 2nd model is presented in Equation 5.

$$\begin{gathered} \textit{Tensile strength} = 24.779375 + \\ + 0.540625 \cdot MT65 + 0.058125 \cdot \\ \cdot MT70 - 0.256875 \cdot MT75 + \\ 0.313125 \cdot CT20 + 0.085625 \cdot \\ \cdot CT25 - 0.066875 \cdot CT30 \end{gathered} \tag{5}$$

where: MT – mold temperature set at the specified temperature (65, 70, 75[°C]),
CT – cooling time set at a specific time (20, 25, 30 [s]).

Evaluating regression models according to the generalization aspect

Of course, one can ask why models with no effect coding were implemented and presented in the paper, while their performance according to the models based on effect coding is worse. Here, it is important to remember that all presented metrics are calculated according to the full Taguchi matrix, which is the training dataset in this case. So, the presented analysis does not take into consideration the generalization aspect of those models. To check the issue, another analysis must be performed. Two approaches were proposed:

- application of leave-one-out-cross-validation (LOOCV), where a whole set of regression models are built in the following way: each row from the dataset is temporarily removed, and the model is implemented on all other data, then if the model is established, the input values from the temporarily removed row are applied to the regression model and predicted value is compared to the real value of that row – the method is recommended if only the Taguchi orthogonal table is available,
- making a prediction according to the real data from the experiment, where each proposed regression model is used to make the prediction, and the results are compared to the real value.

The results of LOOCV verification are presented in Table 9.

The results are interesting because, if the generalization aspect is taken under consideration, the 2nd model, based on two variables and effect coding transformation, has worse results than a similar model (5th), which is based on the same input variables, but it uses direct values of variables without coding at all. The generalization aspect is important because the 1st model (coding with all input variables) has very weak performance according to the prediction, which is only slightly indicated in Table 7, where the training dataset is considered. On the other hand, a regression model based on all four variables with no coding has ok performance, so the aspect of whether coding is needed and what method should be used, e.g., one-hot encoding or effect coding, needs to be carefully considered, because without the LOOCV verification, the worst model could be chosen. If we choose to use the optimal model connected to the 5th regression model for future prediction, its equation is presented in Equation 6.

$$\begin{gathered} \textit{Tensile strength} = 30.223125 - \\ - 0.05925 \cdot MT - 0.04175 \cdot CT \end{gathered} \tag{6}$$

where: MT – mold temperature [°C],
CT – cooling time [s].

The last way of checking the generalization aspect of the proposed models is prediction according to real input and output data from the experiment. It was done by making predictions based on 16 sets of results, where each set contains data from 4 separate trials. In total, 64 cases were predicted by using the proposed regression models, and their results were compared to the

Table 9. Metrics of implemented regression models for LOOCV

Coding type	Input variables	Metrics calculated according to LOOCV						
		R^2	R^2 adj.	MAE	MAE-SD	MAPE	MAPE-SD	RMSE
Effect coding	All input variables	-0.5072	-6.5361	0.4300	0.3130	1.74%	1.27%	0.5319
	Mold temperature, Cooling time	0.7802	0.6336	0.1676	0.1147	0.68%	0.47%	0.2031
	Mold temperature	0.3554	0.1942	0.2825	0.2029	1.14%	0.82%	0.3478
No coding	All input variables	0.7438	0.6507	0.1666	0.1425	0.67%	0.57%	0.2193
	Mold temperature, Cooling time	0.8092	0.7798	0.1442	0.1225	0.58%	0.49%	0.1893
	Mold temperature	0.4396	0.3995	0.2725	0.1758	1.10%	0.71%	0.3243

real value obtained during the experiment. The results are presented in Table 10.

As can be seen, the best accuracy is connected to the regression models based on effect coding of input variables. The optimal model seems to be the 2nd model, where the two most important variables are applied: Mold temperature and Cooling time. Of course, the model based on effect coding assumes that only production processes based on predefined values of process parameters can be applied to the model, which allows for generating possible values of examined mechanical properties. If it is planned to use input process parameters beyond the predefined set of values, the 2nd model cannot be used at all, and therefore the 5th model is recommended in this case. The 5th model accepts as input parameters any value and needs no transformation of those values to generate a prediction.

It is worth mentioning that the test set dataset is based on input production parameters that take on predetermined values in accordance with the DOE. As a result, the linear regression model developed to handle continuous values operates as if it were receiving discrete input values.

Therefore, a realistic and meaningful assessment of its predictive effectiveness would be possible if continuous data were available, i.e., if the input values of the production parameter fell within the specified minimum and maximum ranges. Such validation would require the production and mechanical testing of an additional set of specimens manufactured at intermediate values of the selected process parameters.

Applying linear regression to the original dataset

The regression models developed based on the Taguchi table, which uses average tensile strength values, were tested using real test data from the experiments. This data – real and unaggregated can, of course, also be used to build regression models. As before, these models were developed in two variants: one using direct values of production parameters and the other after encoding them. Furthermore, for each variant of the regression model, three separate models were developed, differing in the amount of input data. The models were created using all four input variables, the two most significant variables identified by Pearson's linear correlation coefficient analysis and the signal-to-noise ratio matrix analysis, and only one variable identified as statistically significant by the ANOVA method. To ensure the comparability of results and address the model's generalization ability, LOOCV validation was performed, and a set of metrics was calculated in the same manner as in the other regression model cases. The results are presented in Table 11.

An analysis of the results obtained indicates that the most effective regression model is the one that uses the two most significant variables, i.e., mold temperature and cooling time. Models that use all four input variables yield weaker results, while the weakest results come from models that base their predictions on only one variable: mold temperature.

As for the impact of input variable encoding on the results, both types of models achieve fairly similar results, with a slight preference for the model without encoding (using actual values of input production parameters) because this model generally has slightly better metrics; the most noticeable difference concerns the adjusted R^2 measure: 0.6139 vs. 0.5841.

It is interesting to note that the overall results obtained for regression models developed using actual data are worse than those for regression models developed based on the Taguchi table in the context of LOOCV validation (Table 9). This

Table 10. Metrics of implemented regression models based on the test dataset

Coding type	Input variables	Metrics calculated according to the test dataset						
		R^2	R^2 adj.	MAE	MAE-SD	MAPE	MAPE-SD	RMSE
Effect coding	All input variables	0.7155	0.6472	0.2159	0.1560	0.87%	0.63%	0.2664
	Mold temperature, Cooling time	0.7027	0.6708	0.2170	0.1645	0.88%	0.67%	0.2723
	Mold temperature	0.4811	0.4548	0.2886	0.2147	1.16%	0.87%	0.3597
No coding	All input variables	0.6615	0.6381	0.2278	0.1803	0.92%	0.73%	0.2905
	Mold temperature, Cooling time	0.6602	0.6488	0.2294	0.1792	0.93%	0.73%	0.2911
	Mold temperature	0.4421	0.4330	0.3010	0.2203	1.21%	0.89%	0.3730

Table 11. Metrics of implemented regression models based on a real dataset

Coding type	Input variables	Metrics calculated according to LOOCV						
		R2	R2 adj.	MAE	MAE-SD	MAPE	MAPE-SD	RMSE
Effect coding	All input variables	0.5503	0.4424	0.2719	0.1954	1.10%	0.79%	0.3349
	Mold temperature, Cooling time	0.6243	0.5841	0.2445	0.1841	0.99%	0.75%	0.3061
	Mold temperature	0.4085	0.3784	0.3082	0.2291	1.24%	0.93%	0.3841
No coding	All input variables	0.6008	0.5732	0.2479	0.1952	1.00%	0.79%	0.3155
	Mold temperature, Cooling time	0.6264	0.6139	0.2409	0.1875	0.97%	0.76%	0.3052
	Mold temperature	0.4037	0.3939	0.3110	0.2280	1.25%	0.92%	0.3856

is due to the fact that LOOCV validation for the model based on a set of 64 elements involves testing a total of 64 regression models, whereas testing the model based on the DOE matrix uses only 16 regression models. This imbalance means that if a given regression model makes a specific error, in the case of a model based on a DOE matrix, that model is used once, and the next one is based on a different set of input production parameters. In the case of models created based on a set of 64 elements, each model is applied 3 times, because the set contains data for 3 samples produced using the same production parameters. Consequently, such a model is tested 3 times against 3 measured tensile strength values, which may lead to the aggregation of error measures.

However, it should be noted that the general trend indicating which model is optimal in this case does not change when switching from Taguchi table-based models to the models based on actual data. In both cases, it is recommended to use a model that does not employ effect coding and is based on the two most significant variables, namely mold temperature and cooling time.

To ensure better comparability of the results from the developed models, it would likely be advisable to set aside a specific subset of real-world data from the dataset obtained during the research and use it as test data to assess the effectiveness and generalization ability of each developed model. This should allow the models to be tested in a single, repeatable manner, without having to consider whether the verification method depends on how the model was created.

This approach will be used in further research, which will focus on modeling the remaining mechanical properties of the analyzed biocomposite and, for this purpose, will also employ selected machine learning methods, rather than relying solely on the classical linear regression method.

Examination of interactions between input variables and its influence to tensile strength

An important aspect of applying the linear regression method is connected to the possibility of study interactions, which can occur between chosen input variables. This can be done according to the process described in [12]. In the current paper, six additional regression models have been established to check the possible interaction between each pair of input variables in terms of output tensile strength value. Each regression model is based on two chosen input variables and an additional term, which is calculated as a product of those variables. When the model was fitted, it was further used to perform the prediction of tensile strength value. The results of prediction according to the values of the input variables are presented in the form of interaction graphs in Figure 5.

The interaction plots were created as 2D line charts because this type of representation makes it easier to see whether there are potential interactions between the variables. Of course, one could consider using 3D contour or surface plots, but given the small number of data points (16 for each pair of variables), these plots appear less readable and useful.

The interaction effects were interpreted in relation to the physical mechanisms occurring during injection molding. Mold temperature and cooling time mainly control the thermal history of the molded part, including cooling rate, PHBV crystallization, stress relaxation, and structure development during solidification. Therefore, their influence on tensile strength is physically justified. The lack of a clear interaction between these two parameters suggests that, within the studied processing window, their effects are mainly additive and largely independent.

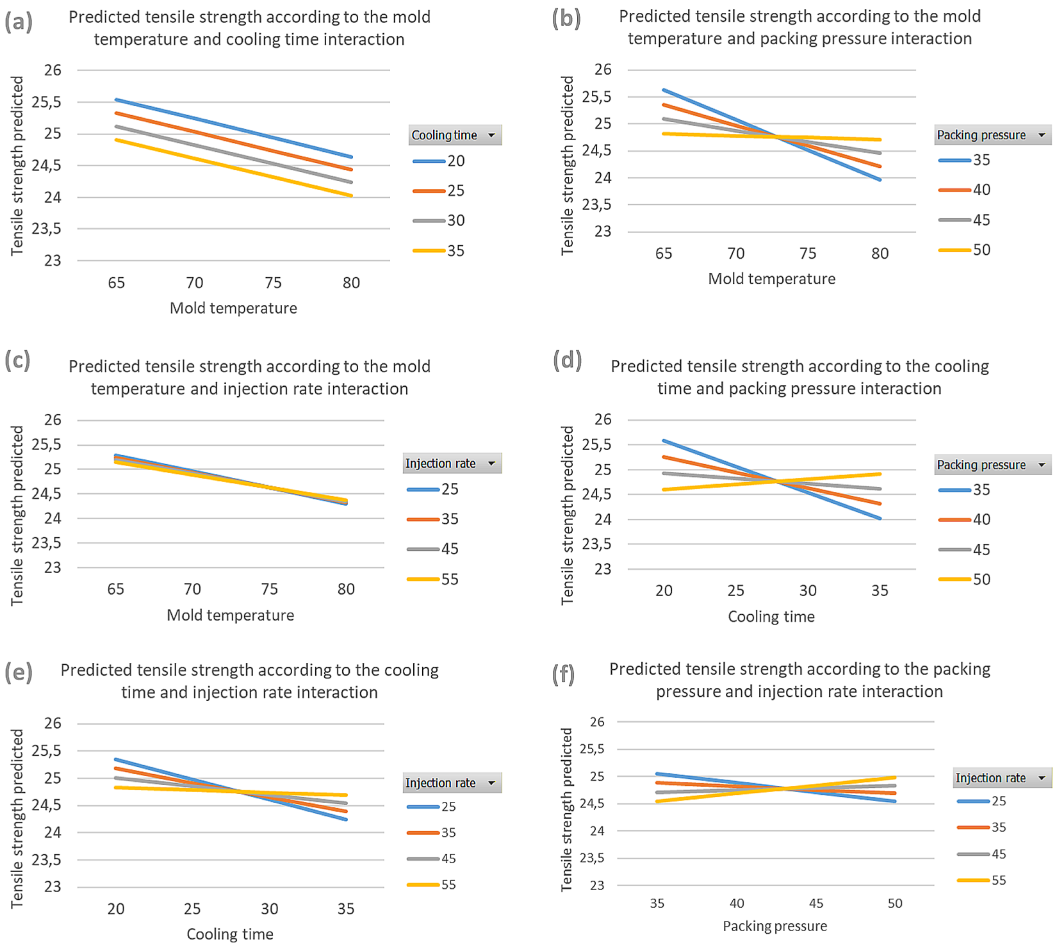

Figure 5. Interaction graphs of pairs of input variables according to predicted tensile strength

The interactions involving packing pressure and injection rate may be related to melt flow, pressure transmission inside the mold cavity, filler orientation, local densification, and void formation. However, their influence on tensile strength was limited within the selected parameter ranges, which is consistent with the ANOVA, Pearson correlation, and Taguchi S/N results. No additional experimental verification was performed at this stage; therefore, independent injection molding trials at intermediate parameter values are recommended for future work to confirm the observed interaction trends.

As shown, there is no interaction between mold temperature and cooling time variables (a). The predicted values of tensile strength are parallel if each value of cooling time is taken into consideration. Of course, those variables are independent of each other, and each of them directly affects the tensile strength value, because those variables are key elements of the generated regression Equations 5 and 6. The lack of interaction between mold temperature and cooling time suggests that both parameters influence tensile strength in an additive and independent manner. This is consistent with the physical interpretation of the process, in which both parameters are directly related to thermal history but act on different stages of structure development. Mold temperature primarily affects the cooling rate and crystallization onset, while cooling time governs the duration of structural evolution. As a result, their effects remain largely separable within the investigated parameter range.

Interaction can be found if mold temperature and packing pressure interaction are examined (b). The general dependency says that higher mold temperature decreases tensile strength value, it is possible to limit that negative relation by using higher packing pressure. On the other hand, using a higher packing pressure if the mold temperature is low, causes the tensile strength value to be smaller compared to a lower packing pressure at the same mold temperature. The observed interaction between mold temperature and packing pressure may be related to the combined influence of thermal and pressure conditions on material densification and microstructure formation. Higher packing pressure can improve consolidation of the material and reduce the presence of voids, partially compensating for the negative effect of elevated mold temperature. This suggests that pressure-induced densification may play a secondary but relevant role in determining tensile strength, especially under less favorable thermal conditions.

No interaction can be found, if mold temperature and injection rate variables are taken under consideration (c) according to the tensile strength value, because lines are located parallel to each other and they lie almost in the same place of plane. The lack of interaction between mold temperature and injection rate indicates that flow-related effects have a limited influence on tensile strength compared to thermally driven phenomena. This supports the conclusion that, in PHBV–SSH systems, the final mechanical performance is predominantly governed by thermal history rather than flow-induced orientation effects within the investigated processing window.

If the cooling time and packing pressure interaction are examined (d), the interaction is clearly visible. The general tendency shows that the longer cooling time affects tensile strength negatively. It can be potentially overcome by applying higher packing pressure. On the other hand, applying higher packing pressure along with shorter cooling time will result in a lower tensile strength value. The interaction between cooling time and packing pressure may reflect a competition between structural stabilization and densification processes. Longer cooling time allows for extended structural evolution, while increased packing pressure enhances material consolidation. The observed trends suggest that appropriate combinations of these parameters may partially mitigate adverse effects related to prolonged cooling, indicating a complex interplay between thermal and pressure-driven mechanisms.

The interaction of cooling time and injection rate is visible (e) and similar to the previous interaction described. A longer cooling time leads to a decrease in tensile strength, which can be compensated for by using a higher injection rate. In general, using a higher injection rate at a specified cooling time leads to a decrease in tensile strength. The interesting thing is that an injection rate of 55cm^3/s stabilizes tensile strength value between 24.6 and 24.8 [MPa]. The interaction between cooling time and injection rate may be associated with the influence of flow conditions on the initial distribution and orientation of the filler within the polymer matrix. Although the injection rate alone has a limited effect, its combination with cooling conditions may influence the evolution of microstructure during solidification. This suggests that flow-induced effects can become relevant when coupled with thermal conditions that control structure stabilization.

The last interaction graph illustrates the existence of interaction between packing pressure and injection rate variables (f). The main dependency shows that the higher packing pressure causes decreased tensile strength, which can be overcome by using a higher injection rate. What is more, applying a higher injection rate while the packing pressure is equal to 35 MPa or 40 MPa causes lower tensile strength. However, if equilibrium according to both variables is reached, a higher injection rate leads to higher tensile strength if the packing pressure increases. In this case, the differences between the obtained tensile strength values according to both input parameters are smaller than in the case of other pairs of variables. The interaction between packing pressure and injection rate may be related to the combined effect of flow and pressure on material compaction and internal structure formation. While both parameters individually show limited influence, their interaction suggests that the balance between filling dynamics and pressure application may affect local material density and defect formation, which in turn can influence tensile strength.

The proposed relations and dependencies should be interpreted with caution and further examined by production technologists in the context of the underlying physical phenomena. Due to the limited size of the dataset and the Taguchi-based experimental design, some of the observed interaction effects may be influenced by experimental noise. Therefore, the identified relationships should be treated as physically motivated hypotheses rather than definitive conclusions.

The influence of interactions of chosen input variables on predicted tensile strength

Found important interactions can be applied to the basic regression model to enhance the model's prediction ability. Of course, after that, the verification process needs to be performed, and the chosen metrics should be calculated. Due to the limitation of the length of the paper, detailed results are available from the authors, while only conclusions are presented below.

In our case, even the best model with interaction according to LOOCV had worse accuracy than the optimal regression model with no interaction at all. The optimal model is based on "Mold temperature" and "Cooling time" (eq. 6) and has the following metrics:

R^2 = 0.8092; R^2 adjusted = 0.7798; MAE = 0.1442; MAPE = 0.5815%; RMSE = 0.1839.

The best model with interactions modeled is based on "Mold temperature" and "Packing pressure" variables and has the following characteristics:

R^2 = 0.7482; R^2 adjusted = 0.6853; MAE = 0.1791; MAPE = 0.72%; RMSE=0.2174.

Thus, even if the interaction between chosen variables is identified and may provide useful insights for production process optimization, it should not be automatically incorporated into the regression equation, because it can make further predictions less accurate. In our case, we have chosen to incorporate some interactions into the best base regression equation (Equation 6), but it didn't improve the results of the base model. The results obtained during LOOCV validation of that regression equation are presented in Table 12.

Analysis process conclusions

The observed dependence of tensile strength on mold temperature and cooling time can be further interpreted in the context of structure development and crystallization behavior of PHBV-based biocomposites. It is well established that the mechanical performance of PHBV is strongly influenced by its crystalline structure, which is highly sensitive to thermal history during processing. Variations in cooling conditions directly affect crystallization kinetics, leading to differences in crystalline morphology, degree of ordering, and internal stress distribution. In the case of PHBV–lignocellulosic systems, such as the PHBV–sunflower husk composite analyzed in this study, these effects are further intensified by the presence of the filler. Sunflower husk particles restrict polymer chain mobility and introduce heterogeneities that can suppress or modify crystallization processes. As a result, even relatively small changes in process parameters related to heat transfer and solidification can lead to measurable differences in mechanical properties. This interpretation is consistent with previous findings reported in [1], where the crystallization behavior of PHBV–SSH composites was shown to be strongly dependent on both filler content and processing conditions. Therefore, the statistical relationships identified in the present study should be understood as a macroscopic manifestation of underlying structure–property mechanisms governed by crystallization and microstructure evolution.

To sum up, the way the regression model is built is not a trivial task. The performance of the model is affected mainly by the input variable selection process and the way those variables are used, in their original form, like continuous values or coded in a specific way. It is worth mentioning that calculating metrics, if only the training dataset is taken under consideration, can cause the not optimal model to be selected. It is especially true if the model is planned to be used for the prediction of mechanical properties – in the case of LOOCV, the validation technique is recommended, mainly because of the small number of cases in the Taguchi orthogonal table. The interesting aspect that can be addressed during the analysis process is the interaction between input variables. Incorporating additional variables into the original dataset can improve the model's metrics better, and moreover, can suggest that the examined mechanical property is affected by the interaction of specific variables, where their influence cannot be seen if we analyze them separately.

To conclude, it is recommended to choose the model expressed in Equation 6, especially if production parameters will be set to any value, which can be out of the original set of parameters. However, if the input production parameters are set to the originally planned values, the regression model presented in Equation 5 should bring more accurate results.

The important issue noted during making predictions on the sample test dataset was connected to the values of the input variables and the prediction results. As can be seen in Table 11, there were four specimens tested according to the 1st set of parameters, which are the same for each specimen. However, measured real tensile strength values are slightly different from each other in the given set of the same parameters. The generated regression equation will give the same prediction if the input variable is the same as presented in the example (Table 13). The listed predicted

Table 12. Regression equation E2 with variants of interactions according to LOOCV validation results

Model:	MT+CT (eq. 6)	MT+CT+ CT*IP	MT+CT+ MT*IP	MT+CT+ MT*IP+CT*IP	MT+CT+ MT*CT
R2:	0.8092	0.7872	0.7812	0.7391	0.7665
R2 adjusted:	0.7798	0.7340	0.7265	0.6442	0.7081
MAE:	0.1442	0.1539	0.1553	0.1824	0.1587
MAE - SD:	0.1225	0.1275	0.1302	0.1254	0.1366
MAPE:	0.58%	0.62%	0.63%	0.74%	0.64%
MAPE-SD:	0.49%	0.51%	0.52%	0.50%	0.55%
RMSE:	0.1893	0.1998	0.2026	0.2213	0.2094

values come from a regression equation where all four variables were taken into consideration, and no coding was done.

The problem can possibly be solved by adding more input variables, which are measured and gathered during the production process and can have a potential influence on tensile strength value – of course, the production technologist must select those variables and measure them to incorporate new data into the final dataset. If this solution is not implemented, no methods will be able to distinguish differences between predicted values (green column), which are observable between real values (blue column) of the examined mechanical property, according to the same set of input production parameters (white columns).

The observed discrepancy between identical input parameters and varying experimental tensile strength values highlights an important limitation of the applied modeling approach. The regression model, based solely on a limited set of process parameters, is inherently unable to capture the full complexity of the system. In practice, even when nominal processing conditions remain unchanged, variations in material structure and properties may occur due to factors not included in the model. In PHBV–lignocellulosic composites, this effect can be attributed to the intrinsic heterogeneity of the natural filler, including variability in particle size, morphology, and composition, as well as differences in moisture content and local dispersion within the polymer matrix. Additionally, microstructural factors such as local fiber orientation, interfacial adhesion quality, and the presence of microvoids or defects may vary between samples produced under nominally identical conditions. As a result, the regression model produces identical predictions for identical input parameters, while the real system exhibits variability that cannot be explained without incorporating additional process or material descriptors. This observation indicates that further improvement of predictive models would require the inclusion of additional

Table 13. Example of prediction of tensile strength by using a regression equation based on four input variables with no coding

Set no.	Mold temperature	Cooling time	Injection pressure	Injection speed	Real tensile strength	Prediction of tensile strength
1	65	20	35	25	25.32882	25.5721
1	65	20	35	25	25.70086	25.5721
1	65	20	35	25	25.70874	25.5721
1	65	20	35	25	25.85876	25.5721
2	65	25	40	35	25.04797	25.3399
2	65	25	40	35	25.66558	25.3399
2	65	25	40	35	25.3488	25.3399
2	65	25	40	35	25.22608	25.3399
3	65	30	45	45	25.67023	25.1076
3	65	30	45	45	25.07406	25.1076
3	65	30	45	45	25.2519	25.1076
…	…	…	…	…	…	…

variables, such as real-time process monitoring data or material state indicators, to better capture the stochastic nature of the system.

CONCLUSIONS

Based on the conducted analysis, the following conclusions can be drawn:

1. Mold temperature and cooling time were identified as the dominant process parameters influencing the tensile strength of PHBV–sunflower husk biocomposites, which is consistent with their key role in controlling thermal history and crystallization behavior.
2. The developed regression models demonstrated that models based on effect coding provide higher accuracy for predefined parameter levels, while models using continuous variables offer better generalization capabilities for arbitrary input values.
3. The application of LOOCV revealed that models with the best performance on the training dataset are not necessarily optimal in terms of prediction, highlighting the importance of model validation in small DOE-based datasets.
4. Analysis of interactions between process parameters indicated that, although some interactions may exist, their incorporation into regression models does not always improve prediction accuracy and should be carefully evaluated.
5. The observed variability of tensile strength for identical input parameters confirms that the regression model cannot fully capture the complexity of PHBV–lignocellulosic systems, which are influenced by microstructural heterogeneity, filler variability, and process-induced effects.
6. The results demonstrate that predictive modeling of biocomposites should be interpreted in the context of material behavior, and further model development should include additional descriptors related to material structure and processing conditions.

REFERENCES

1. Janowski, G., Wójcik, M., Krešić, I., Frącz, W., Bąk, Ł., Gajdoš, I., Spišák, E. The influence of sunflower seed hull content on the mechanical, thermal, and functional properties of PHBV-based biocomposites. Materials, 2026; 19(2), 268.
2. Modi, S. J., Cornish, K., Koelling, K. W., Vodovotz, Y. Mechanical and rheological properties of PHBV bioplastic composites engineered with invasive plant fibers. Transactions of the ASABE, 2016; 59(6), 1883–1891.
3. Grgurević, K., Nikolić, M. M., Kučić Grgić, D., Bulatović, V. O. Thermal, mechanical, and barrier properties of PHBV nanocomposites via TiO_2 incorporation for sustainable food packaging. Polymers, 2025; 18(1), 11.
4. Boey, J. Y., Lee, C. K., Tay, G. S. Factors affecting mechanical properties of reinforced bioplastics: A review. Polymers, 2022; 14(18), 3737.
5. Harishbabu, S., Djuansjah, J., Sreekanth, P.S.R., Kumar, A.P., Louhichi, B., Sahu, S.K., Lee, I.E., Wali, Q. Data-driven AI approach for optimizing processes and predicting mechanical properties of boron nitride nanoplatelet-reinforced PLA nanocomposites. Polymers 2026; 18, 185.
6. Zwawi, M. A review on natural fiber bio-composites, surface modifications and applications. Molecules, 2021; 26(2), 404.
7. Gurunathan, T., Mohanty, S., Nayak, S. K. A review of the recent developments in biocomposites based on natural fibres and their application perspectives. Composites Part A: Applied Science and Manufacturing, 2015; 77, 1–25.
8. Malguarnera, S. C., Manisali, A. The effects of processing parameters on the tensile properties of weld lines in injection molded thermoplastics. Polymer Engineering & Science, 1981; 21(10), 586–593.
9. Kulkarni P. A Theoretical and practical study of linear regression. International Journal of Latest Technology In Engineering, Management & Applied Science, 2025; XIV(XI) https://doi.org/10.51583/IJLTEMAS
10. Maulud D. H., Abdulazeez A. M. A review on linear regression comprehensive in machine learning. Journal Of Applied Science And Technology Trends. 2020; 1(2). https://doi.org/10.38094/jastt1457
11. Roustaei N. Application and interpretation of linear-regression analysis. Med Hypothesis Discov Innov Ophthalmol. 2024; 13(3). https://doi.org/10.51329/mehdiophthal1506
12. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning with Applications in R. Springer Science+Business Media. New York. 2015. https://doi.org/10.1007/978-1-4614-7138-7
13. Arunadevi, M., Avinash, L., Vinayakumar, R. Development of a prediction model for material properties of SS404L alloy in additive manufacturing using KNN and linear regression. J. Inst. Eng. India Ser. D., 2025. https://doi.org/10.1007/s40033-024-00854-x

14. Amigo, N., Palominos, S., Valencia, F.J. Machine learning modeling for the prediction of plastic properties in metallic glasses. Sci Rep., 2023; 13, 348. https://doi.org/10.1038/s41598-023-27644-x
15. Stergiou K., Ntakolia C., Varytis P., Koumoulos E., Karlsson P., Moustakidis S. Enhancing property prediction and process optimization in building materials through machine learning: A review. Computational Materials Science, 2023; 220, 112031, https://doi.org/10.1016/j.commatsci.2023.112031
16. Kutner M. H., Nachtsheim C. J., Neter J. Applied Linear Regression Models. McGraw-Hill. New York. 2008.
17. Chicco D., Warrens M.J., Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 2021; 7, e623. https://doi.org/10.7717/peerj-cs.623
18. Plevris V., Solorzano G., Bakas N., Ben Seghier M. Investigation of performance metrics in regression analysis and machine learning-based prediction models. The 8th European Congress on Computational Methods in Applied Sciences and Engineering ECCOMAS Congress 2022. https://doi.org/10.23967/eccomas.2022.155
19. Hodson, T. O. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, Geosci. Model Dev., 2022; 15, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022