Revised: 2 April 2021 | Accepted: 9 April 2021

DOI: 10.1002/pc.26071

RESEARCH ARTICLE

Polymer COMPOSITES WILEY

Green composites based on thermoplastic starches and various natural plant fibers: Impacting parameters of the mechanical properties using machine-learning

Sophie Morin[1] | **Lionel Dumoulin**[1,2] | **Louise Delahaye**[1] | **Nicolas Jacquet**[2] | **Aurore Richel**[1]

[1]Laboratory of Biomass and Green Technologies, University of Liège – Gembloux Agro-Bio Tech, Gembloux, Belgium

[2]Food Science and Formulation, University of Liège – Gembloux Agro-Bio Tech, Gembloux, Belgium

Correspondence
Lionel Dumoulin, Laboratory of Biomass and Green Technologies, University of Liège – Gembloux Agro-Bio Tech, Passage des Déportés, 2, 5030, Gembloux, Belgium.
Email: lionel.dumoulin@uliege.be

Funding information
Walloon Region, Belgium, Grant/Award Number: D31-1395

Abstract

Multivariate analyses on formulation and mechanical behavior of nonwoven and nonoriented natural fibers reinforced thermoplastic starch (TPS) composites were performed. Glycerol and water were considered as TPS plasticizers. Fibers composition (i.e., cellulose, hemicellulose, lignin), fibers morphology (fibers length), starch composition (i.e., amylose/amylopectin ratio) as well as the processing conditions (i.e., temperature, rotor speed, relative humidity during aging) were evaluated for their ability to affect the elastic modulus, tensile strength, and elongation at break of the final materials. Multivariate linear regressions were computed to unveil the importance of each variable on the mechanical behavior. Fibers composition impacted the most the models: cellulose maximization improved the elastic modulus and tensile strength while lignin reduced the elastic modulus and hemicellulose decreased the tensile strength. TPS plasticizers, temperature, and rotor speed of the process were negatively impacting the elastic modulus but in a lesser extent than the fiber composition. Within the range of the created database, the selected variables and attributed coefficients were permitted to explain the variability. The produced models revealed that complex and yet uninvestigated interactions are to be considered within TPS-based biocomposites. Therefore, this work discusses and suggests a “must-have” list of variables for comparable analyses of new TPS-based biocomposites using natural fibers as reinforcement.

KEYWORDS

biocomposites, mechanical properties, multivariate regression, natural fibers, thermoplastic starches

1 | INTRODUCTION

Composites are a class of materials used for their versatile properties in several sectors such as construction, leisure, and transportation. These materials, composed of at least two raw materials, possess the advantage of versatile and hybrid properties of their initial components. Composites and, by extension, biocomposites (BC; i.e., composites with either a biobased or biodegradable plastic matrix or natural fillers like natural plant fibers) are considered for such precited applications as they offer a weight reduction

Sophie Morin and Lionel Dumoulin contributed equally to this work and should be considered as coauthors.

while having interesting specific physicochemical properties.[1] When both matrix and fillers are biobased and biodegradable, the BC belong to the green composites family.[2,3] Thermoplastic starch (TPS) BC are an example of such materials and are particularly under consideration for eco-friendly bags, films, and packaging devices manufacturing.

TPS are studied since the beginning of the 1990s.[4,5] A few years later, BC were investigated, as an improvement of TPS through the addition of natural fibers.[6,7] For 30 years, researches focused on the improvement of the mechanical properties by numerous ways, such as the choice of starch source/fiber type,[8,9] choice of plasticizer,[8,10] starch and fibers modification,[11,12] or use of matrix blends.[13] Independently tuned, all these parameters were demonstrated as impacting the final material properties. At this stage, it is unclear which of these impacting parameters is the most worth to tune among others, to drastically change the final BC properties. Moreover, it is still unclear whether the final TPS and BC mechanical properties are affected independently by the parameter or if the interactions are important as well. Here stands the first objective of this paper: opening a more global vision on these complex materials by attempting to describe their final mechanical properties.

Among the multiple papers available on TPS and BC mechanical properties, no scientific consensus was reached to set a constant list of variables to set or to control in order to describe the BC mechanical behavior. In fact, some of the process conditions are systematically specified in the literature. But, for some other parameters, especially aging conditions (temperature, time, and relative humidity [RH]), few data are referenced, even if they were proved as impacting final TPS and BC mechanical properties.[14,15] As the BC production parameters are presented in the literature depending on the researcher's sensibility basis, comparisons between literature results is problematic, and by extent the transfer to valuable applications (such as industrial applications) uncertain. Therefore, the second objective of this paper consists in proposing a common list of impacting variables on BC and TPS properties.

Harmonized studies about TPS and BC are beneficial both for academics and industrials. The lack of simple description and prediction of these materials mechanical properties is detrimental to be considered at the industrial scale. Only few studies about TPS-related composites materials tried to study the link between mechanical properties or barrier properties and composites formulation parameters.[16,17] In both cases, the range of validity of the obtained model was restricted due to limited database length.

In contrast, many sectors using complex materials, such as the cement industry and by extension the construction field, already implemented their analyses by models obtained thanks to machine learning or artificial intelligence algorithms.[18]

For polymer plastic materials, description data for thousands of materials references exists in specific database (such Campus Plastic or UL Prospector) which allow them to be considered for product design. Decision-making expert systems are thus created and fed to help designer in selecting the best final materials[3] and are coupled to such database and analyses. These materials description helps to save time and financial resources as predictions could be made instead of costly multiple trials.[3]

As complex materials, specific engineering to improve desired properties of the final mixture could be foreseen. But as long as the interactions between the different components are not clear, it is hazardous to anticipate such oriented-engineering.

In this paper, the first step of this process consists in collecting the most studied parameters (called input parameters) to establish a model. Data-mining and multivariate analyses are performed in the objective to unveil the role of these commonly monitored parameters to formulate TPS and, by extension, TPS-based BC.

Here, this paper aims to produce a model that could explain and predict some mechanical properties from formulation data collected in the literature. The scope of this work is green composites made from nonwoven and nonoriented natural plant fibers with TPS matrix. Only glycerol and water were considered as TPS plasticizer. Papers respecting this scope and presenting values of at least one of the output parameters (i.e., mechanical parameters of the final materials, such as tensile strength, elastic modulus or elongation at break) were selected. Analyses of the models are presented to highlight the parameters that are more impactful and that should be further studied to optimize TPS and BC properties depending on the specifications of a targeted application. Then, the models' qualities, as well as their implications on the current knowledge of BC formulation are discussed.

2 | METHOD

2.1 | Software

Models were established on RStudio from Anaconda. The following packages were used: keras, devtools, magrittr, dplyr, psych, tfdatasets, interaction, and ncpen. The following libraries were used to run the algorithm: tibble, caret, and stringr. References can be found in the Supplementary Material section.

2.2 | Algorithm

Data from the literature were studied with multilinear regression algorithms (algorithm presented in the Supplementary Material section). The creation of the linear regressions was made on the basis of:

- validity of the hypothesis of the multivariate linear regressions (no-collinearity, homoscedasticity, linearity of the data, independence of residuals error terms, and normality of the residuals);
- adjusted R-square (R^2_{adj}) maximizing;
- root mean square error (RMSE) minimization; and
- maximum of significant variable coefficients.

Multiple models (Table 1) were tested to find the best fit between output parameters (i.e., the studied mechanical properties) (y) and input parameters (x_i) (i.e., chemical composition of the fibers, starch, and process parameters). For each database, the selected model was the one with the highest R^2_{adj} and lowest RMSE.

Then, to study the importance of each variable coefficient, normalization was performed on each database variable according to Equation (1) (variable normalization).

$$x_{i\,\mathrm{norm}} = \frac{x_i - \mathrm{average}(X)}{\mathrm{standard\ deviation}(X)} \tag{1}$$

with x_i being a value within the variable X.

2.3 | Database description

Three output parameters were studied: Young's modulus (YM, in MPa), tensile strength (TS, in MPa) and elongation at break (EaB, in %).

The selected input values are: TPS chemical composition (starch/water/glycerol content [% w/w]), starch amylose and amylopectin contents (%) (described as the amylose to amylopectine ratio (A/A ratio)), fibers chemical composition (Cellulose/Hemicellulose/Lignin [% w/w of dry matter]), fiber percentage, fiber length (mm), plasticizing temperature (°C), rotor speed (rpm) (0 if no rotor was used), and relative humidity during TPS and BC aging. Plant fibers used in the selected papers were: sisal, sugarcane bagasse, cotton, eucalyptus (*urograndis*), kapok, jute, luffa, cassava bagasse, ramie, pich seed shell, semicrystalline cellulose, nanofibrilated cellulose, and micro cellulose. Starches used in the selected papers were extracted from several plants: corn, wheat, rice, cassava, maize, potato, waxy rice, pea, ahipa, sugarpalm, anchote, and pich tree. To improve the model's quality, the fiber percentage was used as factor of the fiber's properties. Starch percentage was also applied as factor of the A/A ratio. When the amylose and amylopectin contents and/or the fibers chemical composition were missing, standards values from the literature were used. However, when the other parameters were missing, the data were not considered.

Three databases were created to unveil the cross-interactions between fibers and the matrix. Thus, databases only related to TPS (without fibers), to BC (always with fibers) or with all data (TPS and BC) were created. The databases ranges are presented in Tables 2 and 3.

TABLE 1 Models and corresponding equations tested

Model	Equation
Linear	$y = \mathrm{Intercept} + \alpha \times x_1 + \beta \times x_2 + ... + \delta \times x_n$
Square root	$\sqrt{y} = \mathrm{Intercept} + \alpha \times x_1 + \beta \times x_2 + ... + \delta \times x_n$
Inverse	$1/y = \mathrm{Intercept} + \alpha \times x_1 + \beta \times x_2 + ... + \delta \times x_n$
Polynomial	$y^2 = \mathrm{Intercept} + \alpha \times x_1 + \beta \times x_2 + ... + \delta \times x_n$

3 | RESULTS

3.1 | Database exploration

Initial database contained 551 entries (database length, i.e., lines possessing at least one of the output parameters) from 72 references. Rotor speed or fiber length could not be replaced by standard values and papers missing these data were therefore not considered. This database was then reduced to 63 studies and 477 lines after completion of the fibers and TPS classical compositional missing values. Fibers diameters, process duration, or complete TPS aging conditions were barely referenced. Even if these parameters are crucial for describing the final materials properties, we decided to remove them from considerations in order to still possess a database of significant length.

Only 1/3 of the entries possessed information about chemical composition content of the fibers in the related paper (respectively 30% of the database entries for the cellulose, 38% for the hemicellulose, and 34% for the lignin). Considering the A/A ratio, 29% of the references omit to mention this value in any form.

After splitting the original database depending on the final mechanical property to study, three subdatabases were obtained. The database studying the elastic modulus contained 50 references; the TS database was built from 57 references and the EaB included 56 references. Length of each database is described in Tables 2 and 3.

TABLE 2 Databases description for the elastic modulus, mean (min–max); DM, dry matter

	TPS	BC	TPS and BC
Cellulose (%, w/w, DM)	-	70.47 (0.00–100.00)	35.14 (0.00–100.00)
Hemicellulose (%, w/w, DM)	-	9.08 (0.00–89.90)	4.53 (0.00–89.90)
Lignin (%, w/w, DM)	-	8.81 (0.00–48.40)	4.40 (0.00–48.40)
Fiber percentage (%, w/w)	-	20.16 (0.30–100.00)	10.06 (0.00–100.00)
Fiber length (mm)	-	5.55 (0.00–150.60)	2.77 (0.00–150.60)
Starch percentage (%, w/w)	57.68 (2.88–86.96)	55.73 (0.00–75.00)	56.71 (0.00–86.96)
Amylose (%)	38.86 (0.00–87.00)	19.76 (0.00–28.00)	29.33 (0.00–87.00)
Amylopectin (%)	60.84 (13.00–100.00)	69.32 (0.00–100.00)	65.07 (0.00–100.00)
Glycerol (%, w/w)	19.26 (0.00–60.00)	20.43 (0.00–50.00)	19.84 (0.00–60.00)
Water (%, w/w)	22.24 (0.00–96.10)	13.91 (0.00–96.10)	18.09 (0.00–96.10)
Temperature (°C)	139.90 (85.00–180.00)	119.50 (0.00–200.00)	129.70 (0.00–200.00)
Relative humidity (%)	55.76 (7.00–95.00)	50.83 (7.00–83.00)	53.30 (7.00–95.00)
Rotor speed (rpm)	163.20 (0.00–2000.00)	200.70 (0.00–2000.00)	181.90 (0.00–2000.00)
Cellulose × fiber percentage	-	1230.50 (0–8700.00)	613.7 (0–8700.00)
Hemicelluloses × fiber percentage	-	244.80 (0–2800.00)	122.08 (0–2800.00)
Lignin × fiber percentage	-	289.00 (0–4840.00)	144.13 (0–4840.00)
Fiber percentage × fiber length	-	381.67 (0–15000.00)	190.34 (0–15000.00)
Starch percentage × (*A*/*A* ratio)	-	16.98 (0–29.17)	49.79 (0–463.34)
Elastic modulus (MPa)	229.62 (0.12–3204.00)	907.76 (0.50–15000.00)	567.80 (0.12–15000.00)
Database length	192	191	383

3.2 | Parameters affecting mechanical properties

The equation proposed was as described in Equation (2) (calculating the output parameters value), following how the literature assume how those parameters impact BC mechanical properties.

$$\mathrm{OP} = I + F\% \times (C + H + L + FL) + S\% \times (A/A\,\mathrm{ratio}) + G + W + T + S + \mathrm{RH} \quad (2)$$

with OP is output parameter, $F\%$ is fiber percentage, C is cellulose content, H is hemicellulose content, L is lignin content, FL is fiber length, $S\%$ is starch pourcentage, A/A ratio is amylose/amylopectin, G is glycerol amount, W is water amount, T is temperature, S is rotor speed, and RH is relative humidity during aging.

Glycerol amount, water amount, temperature, rotor speed, and relative humidity were considered to apply themselves to the whole materials while fiber percentage and starch percentage are affecting the global content of respectively the fibers composition or starch composition of the mixture.

The TPS database was tested according to the same equation except the part concerning fibers (i.e., "$F\% \times (C + H + L + Fl)$"). The intercept of the equation should represent the gluing effect between the fiber and the surrounding matrix, and all more complex interactions between parameters. This parameter should contain, beside others, all the surface interaction between the fibers and the matrix.

For the TPS database, the tested equations failed to obtain a model for the elastic modulus, tensile strength, and elongation at break. The study of the elongation at break also failed to obtain a fit with the data of all databases.

Models were obtained for the elastic modulus and tensile strength for the BC and the whole database (TPS + BC) using the square root transformation on the output value. While the statistical parameters, such as R^2_{adj} and RMSE, were low for the TPS + BC database (R^2_{adj} [elastic modulus] = 0.7615, R^2_{adj} [tensile strength] = 0.6649), they sensibly increased with the database containing only BC (R^2_{adj} [elastic modulus] = 0.8594, R^2_{adj} [tensile strength] = 0.7915) (Tables 4, 5). Most of the variables studied were significant using these models. Simple linear regression permitted also to obtain equivalent R^2_{adj} but the RMSE and residuals were larger than using square root transformation. The

TABLE 3 Databases description for tensile strength, mean (min–max), DM, dry matter

	TPS	BC	TPS and BC
Cellulose (%, w/w, DM)	-	44.70 (0.00–100.00)	36.43 (0.00–100.00)
Hemicellulose (%, w/w, DM)	-	8.98 (0.00–89.90)	4.57 (0.00–89.90)
Lignin (%, w/w, DM)	-	4.25 (0.00–48.40)	4.34 (0.00–48.40)
Fiber percentage (%, w/w)	-	18.52 (0.30–100.00)	9.41 (0.00–100.00)
Fiber length (mm)	-	5.41 (0.00–150.60)	2.75 (0.00–150.60)
Starch percentage (%, w/w)	58.43 (2.88–95.00)	59.47 (0.00–95.00)	58.96 (0.00–95.00)
Amylose (%)	37.31 (0.00–87.00)	20.18 (0.00–28.00)	28.61 (0.00–87.00)
Amylopectin (%)	62.32 (13.00–100.00)	70.16 (0.00–100.00)	66.30 (0.00–100.00)
Glycerol (%, w/w)	18.11 (0.00–50.00)	19.41 (0.00–50.00)	18.77 (0.00–50.00)
Water (%, w/w)	22.72 (0.00–97.00)	12.51 (0.00–96.10)	17.54 (0.00–97.00)
Temperature (°C)	139.40 (25.00–180.00)	126.30 (25.00–175.00)	131.90 (0.00–200.00)
Relative humidity (%)	53.78 (7.00–90.00)	52.82 (7.00–83.00)	53.29 (7.00–90.00)
Rotor speed (rpm)	171.70 (0.00–2000.00)	175.10 (0.00–2000.00)	173.40 (0.00–2000.00)
Cellulose × fiber percentage	-	1143.20 (0–8700)	580.90 (0–8700)
Hemicelluloses × fiber percentage	-	221.60 (0–2800)	112.60 (0–2800)
Lignin × fiber percentage	-	259.20 (0–4840)	131.70 (0–4840)
Fiber percentage × fiber length	-	338.19 (0–15,000)	171.85 (0–15,000)
Starch percentage × (*A*/*A* ratio)	-	18.30 (0–36.94)	4078.00 (0–7600)
Tensile strength (MPa)	7.86 (0.10–53.50)	17.04 (0.18–550)	12.53 (0.10–550)
Database length	211	218	429

coefficients obtained using the square root transformations are described in the Tables 4 and 5.

All parameters marked with one or several "*" or "°" are significantly different from zero and could therefore be considered (α > 10%). Apart from hemicelluloses × fiber percentage and the temperature, all parameters were marked as significant for the TPS-BC database for the Young's modulus. For the BC database hemicelluloses × fiber percentage was again not significant, with the starch percentage × A/A. In the case of the study of the tensile strength, water, rotor speed, and fibers composition (except cellulose) were not significant for the TPS-BC database. Only rotor speed, lignin amount and fibers' length were not significant in the BC database. Cellulose × fibers' length is positively affecting the Young's modulus and tensile strength in the two databases. TPS-related variables affect significantly the Young's modulus and tensile strength models in most of the models.

In order to compare the coefficients importance, the databases were normalized. Coefficients after normalization are presented in Tables 6 and 7.

Fibers length and cellulose amount, coupled with the fiber's percentage, are the more important positive parameters in the elastic modulus TPS-BC analyses. On the contrary, lignin content and glycerol amount affected

TABLE 4 Models' coefficients for TPS-BC database

	Elastic modulus	Tensile strength
Model	Square root	Square root
R^2_{adj}	0.7615	0.6649
RMSE	584.49	25.36
(Intercept)	32.9406***	5.1245***
Glycerol	−0.7301***	−0.0631***
Water	−0.1512***	−0.0031
Temperature	−0.0018	−0.0076**
RH	−0.1217***	−0.0198***
Speed	−0.0018°	−0.0001
Cellulose × fiber percentage	0.0084***	0.0012***
Hemicelluloses × fiber percentage	0.0023	−0.0003
Lignin × fiber percentage	−0.0047***	0.0002
Fiber percentage × fiber length	0.003***	0.00004
Starch percentage × (*A*/*A* ratio)	0.019***	0.0042***

Note: Significance is expressed as ****p*-value < 0.001, ***p*-value < 0.01, **p*-value < 0.05, °*p*-value < 0.1.

TABLE 5 Models' coefficients for BC database

Model	Elastic modulus Square root	Tensile strength Square root
R^2_{adj}	0.8594	0.7915
RMSE	554.67	30.65
(Intercept)	40.0297***	5.8355***
Glycerol	−0.6288***	−0.0618***
Water	−0.1635**	−0.0184*
Temperature	−0.0554°	−0.0115***
RH	−0.1107**	−0.0104°
Speed	−0.0046***	−0.0003
Cellulose × fiber percentage	0.0088***	0.0013***
Hemicelluloses × fiber percentage	0.0009	−0.0006*
Lignin × fiber percentage	−0.0053***	0.0001
Fiber percentage × fiber length	0.0025***	−0.00008
Starch percentage × (*A*/*A* ratio)	−0.2235	−0.0462*

Note: Significance is expressed as ****p*-value < 0.001, ***p*-value < 0.01, **p*-value < 0.05, °*p*-value < 0.1.

TABLE 6 Coefficients for TPS-BC after normalization

	Elastic modulus	Tensile strength
Glycerol	−0.1415***	−0.0433***
Water	−0.0496***	−0.0033
Temperature	−0.05	−0.0527**
RH	−0.0653***	−0.038***
Speed	−0.0075°	−0.007
Cellulose × fiber percentage	0.6048***	0.4351***
Hemicelluloses × fiber percentage	0.0334	−0.0973
Lignin × fiber percentage	−0.3008***	−0.0339
Fiber percentage × fiber length	0.3773***	−0.0265
Starch percentage × (*A*/*A* ratio)	0.04676***	0.0434***

Note: Significance codes are on the basis on non-normalized coefficients and provided as reminded.

TABLE 7 Normalized coefficients for BC

	Elastic modulus	Tensile strength
Glycerol	−0.0825***	−0.0586***
Water	−0.1093**	−0.066*
Temperature	−0.1435°	−0.072***
RH	−0.0794**	−0.039°
Speed	−0.0077***	−0.0089
Cellulose × fiber percentage	0.6406***	0.4642***
Hemicelluloses × fiber percentage	−0.0253	−0.1385*
Lignin × fiber percentage	−0.3438***	−0.0527
Fiber percentage × fiber length	0.2859***	−0.0931
Starch percentage × (*A*/*A* ratio)	−0.0974	−0.0663*

Note: Significance codes are on the basis on non-normalized coefficients and provided as reminded.

the most negatively the model. Cellulose content affected as well positively the tensile strength model for the TPS-BC. Negative important coefficients were more tedious and only considered to a lesser extent. The TPS-related parameters are less impacting the output variables. For the TPS-BC databases, glycerol amount is the most impacting parameters of them, followed equally by water amount and starch percentage × A/A ratio for the elastic modulus and only starch percentage × A/A ratio for the tensile strength.

Considering only BC, fibers length and cellulose were again impacting the most positively the elastic modulus model. Lignin was also again impacting negatively this model. For tensile strength cellulose coupled with the fiber percentage was positively impacting the model while the hemicellulose was the most negatively impacting coefficient. Cellulose optimization while reducing the other biopolymers in the fibers seems to permit a better resistance behavior of BC according to these models and database. Once again, TPS-related parameters have a limited impact on the output variable, when only BC is considered, with the highest impact by water content, in this case.

4 | DISCUSSION

4.1 | Data-mining

The R^2_{adj} statistical parameter is often used in the scientific community. It expresses the closeness of the experimental value to the equation generated by the model, and so the part of the sample variance that is explained by this same model. To complete the description of the model accuracy, the RMSE is needed, and express the deviation of the empirical value from the regression tendency. Therefore, a satisfactory model should possess an R^2_{adj} converging toward 1 and a RMSE converging toward 0 (RMSE is expressed in the unit and range of the experimental values). Most of the models studied in the literature about predicting mechanical properties are above 0.68.[18] On this base, in our study, only the databases containing TPS values were unable to produce multivariate regression models with sufficient quality.

From our observations in the scientific literature, aging (through the RH parameter) is clearly affecting the final behavior of the material when formulating TPS.[14,15,19] The fact that the other obtained models (containing BC-related data) exhibited both a better regression coefficient with lower prediction errors suggested a limited effect of aging of the BC.[20] As the fibers rigidify the structure and is highly interacting with the TPS matrix (hydrophilic–hydrophilic interactions), aging of the TPS through crystallinity changes[19] would be struggled.

The entries within our database were also unable to produce a model considering the elongation at break parameter. This subdatabase contained, in fact, one study which particularly focused on the aging of the composite and its effect (27 lines).[19] As the aging time was not considered due to missing values, this study could have disturbed the analysis. No conclusions could therefore be formulated, except a highly probable effect of aging on the elongation at break of TPS and TPS-based BC.

4.2 | Validity of the model

Multilinear regression method is multipurpose. Descriptive or predictive goal could be achieved on the basis of the created model. Here, models were essentially built to apprehend the effect of each commonly studied variables on the final properties. To reflect the best possible reality, the RMSE and R^2_{adj} were optimized to select one of the models based on different mathematical transformation of the output variable.

A great part of the variability was explained for the elastic modulus and tensile strength in the range of the input variables (Tables 4 and 5). No extrapolation can be done outside the studied scope. The significance, as well as the sign and the importance of the coefficients may vary when changing the range.

Analysis of interactions would help to optimize two interdependent parameters, such as temperature and rotor speed during extrusion, because of technical limitations. The two parameters are chosen together, as for example a too high temperature and too high speed would lead to fibers and matrix degradation.

Quality of the models could be improved both by improving the data collected (such as the chemical composition of the fibers and the TPS, and a better description of also impacting parameters but poorly referenced) and by mathematically transform the input variables which may be linked nonlinearly to the output parameter. Therefore, several mathematical transformations were tested in an attempt to linearize the output variable. Square transformations reached the best these objectives on the basis of the error minimization and R^2_{adj} maximization. Increasing the database will limit the impact of outliers due to manipulation mistakes reported in papers.

4.3 | TPS-related coefficients

TPS composition and A/A ratio are significantly impacting final properties, except for the water amount in the tensile strength TPS-BC database and starch × A/A ratio interaction for the elastic modulus BC database. It is well known that changing plasticizer content and type (such as water and glycerol in this work) is impacting the final properties by changing crystallinity and linkages between starch chains. This can have major impacts on mechanical properties. It was demonstrated that a glycerol content higher than 50% would destroy the interactions between starch molecules and prevent the formation of double helix in β-type starches.[19] Plasticizer impact is supposed to be dependent on the starch crystallinity and A/A ratio. As a perspective of this work, supplementing the database with other plasticizers would determine the importance of changing the plasticizer.

Concerning the starch composition effect, it has been identified as mainly significant. This result highlights the importance to adapt starch content and source when formulating TPS and BC. Starch description may be improved through the study of multiple descriptive parameters such as: A/A ratio, crystallinity type (A, B, and C) and content and polymerization and ramification degrees.[21,22] Natural starches present variable combinations of these independent parameters, which could be therefore individually optimized.

These parameters are impacting the presence and strength of interactions inside de TPS matrix (between starch molecules) and between the TPS matrix and the fiber's surface. In our model, it was demonstrated that A/A ratio is significantly impacting the final properties. All the TPS-related coefficients were negative when significant, except starch percentage × A/A ratio for TS modelling with the TPS-BC database. It is impossible to decrease simultaneously water amount, glycerol amount, and starch content for optimization and therefore TPS composition should be a balance of these components.

Our model considers each parameter as impacting individually the output parameters. Interactions between component (cross-effect of having for example a high waxy starch content and high temperature) could be considered but made the model understanding more complicated. Interactions between TPS components and starch descriptive parameters as well as with fibers-related variables are possible. In the presented work, the overall impact of starch-related parameters is divided between the selected variables and the intercept, hiding possible positive effects of TPS-related variables. "Hiding" those parameters in the Intercept could also have impacted negatively the R^2_{adj} and RMSE.

4.4 | Fibers-related coefficients

The study of the database highlighted the importance of the fibers composition on the final BC properties. Cellulose was significantly affecting all the models while hemicelluloses were taken into account in the models using the BC database. Cellulose affected positively the model, with the highest normalized coefficient (due to its intrinsic resistance). Hemicellulose was found nonsignificant in the study of the elastic modulus. Such output value is surprising: hemicellulose is supposed to be the most ramified biopolymer in the complex matrix of the fibers. Therefore, this polymer has a high potential of interaction with the TPS matrix, through hydrophilic–hydrophilic interactions. Fibers' composition is an overall composition, and does not reflect the fibers surface presented to the TPS matrix, which is a hemicelluloses-rich surface. As the surface effects are considered to be included within the Intercept of the model, hemicelluloses effect could be "hidden" in this parameter. In fact, fibers' surface plays a crucial role in the interaction with their surrounding chemical environment.[23]

Moreover, the database covers a large number of entries where the fibers composition data were missing, and had to be replaced by generic fibers composition on the basis of the studied specie. Our analysis clearly showed that fibers composition is crucial in the understanding and description of the obtained mechanical properties. This indicates that the "filled" data have major impact on the model. Generic fibers compositions being not precise enough, or not corresponding to the fibers studied in the related papers, could have led to a decreased R^2_{adj} and increased RMSE.

4.5 | TPS–fibers interactions

As TPS–fibers interaction is supposed to have an importance here, further study of starch and fibers surface (with or without modification) and their impact on the final properties is to be made. To do so, impartial parameters have to be determined to well compare modifications, such as type and amount of chemical functions presented at the surface, surface hydrophilicity for further studies about TPS-BC.

Table 8 suggests a list of variables representing the physical and chemical variety among the fibers and the matrix. These parameters will help to proper compare results from diverse independent studies for future meta-analyses and increase the interaction comprehension of these materials.

Fibers morphology is impacted during the plasticization process by reduction of the fibers size, especially if a shearing stress is present (such as in extrusion processes).[24,25] Fibers morphology (length and diameter) will affect the way fibers surface is exposed and the bonding opportunities with the matrix. Global composition of the fibers should really represent the studied materials. Fibers composition is dependent on the botanical origin and used variety, environmental growing conditions (i.e., climatic factors), retting conditions (if applicable), storing conditions, preparation (i.e., decortication). Taking individually each of these parameters in account is far more complex than considering the final chemical composition. Cellulose, hemicellulose, lignin content, and fibers moisture should be at least mentioned. We also suggest to more deeply evaluate the proportion of crystalline and amorphous cellulose globally present in the fibers as this biopolymer is supposed to mainly participate to the mechanical resistance of the BC.[20,26] Hemicelluloses are also complex polymers and determining its intrinsic composition will extremely probably be valuable in the comprehension of the final BC behavior. In fact, hemicelluloses components are ramified, and differently exposed at the fibers surface.[1] We could therefore expect a various reactivity toward hydrophilic–hydrophilic interactions with the matrix dependent on the

TABLE 8 Parameters describing the materials that should be considered for biocomposites studies

Fibers	Starch	TPS and BC
• Morphology (length, diameter) • Global chemical composition • Surface composition • Cellulose crystallinity	• A/A ratio • Crystallinity (type and content) • Starch surface characterization (functions, properties) • Starch polymerization and ramification degrees	• Plasticizer (type and content) • Process temperature and duration • Extrusion speed and flow • Aging parameters (relative humidity, duration, temperature)

polysaccharides presents. Galactan, mannan, xylan, arabinan, glucan, and rhamnan are quite common to study and reflect already the hemicelluloses composition variety.[27] Some studies evaluate by infrared spectroscopy the final BC composition. We suggest using this analysis to highlight qualitatively and quantitatively which chemical groups are available at the fibers surface, as they will participate to the interface interactions.

Starch characterization can be done at different levels of complexity. *A*/*A* ratio is simple to analyze thanks to commercialized kits or lab procedures.[28] Crystallinity type is dependent on the starch source, and is referenced in reviews such as presented by [21,22], and therefore easy to report. Crystallinity content (as well as its type) can be determined by XRD analysis.[19] Starch polymerization can be determined by several methods, such as size-exclusion chromatography,[29] while the ramification degree can be analyzed by ^{1}H-NMR[30] or titration.[31] Concerning starch surface analysis, it is more complex, as unprocessed starch is found as structured granules. The evaluation of the functions found at the starch surface through FTIR analysis and quantification of degree of substitution after modification would also be possible with minimal difficulty.

Finally, when processing both materials together, many physical parameters could affect the biopolymers. The temperature first: lignocellulosic biopolymers (i.e., cellulose, hemicellulose, and lignin) and starches degrade upon high temperature. Hemicellulose starts degrading at 180°C,[32] cellulose above 260°C and starch around 300°C.[33] It becomes obvious that processes letting fibers and starch above these temperatures will create new chemical phenomena especially if the residence time is long enough to permit degradation. Glucose based polymers (i.e., starch and cellulose) will submit ongoing chain size reductions until monosaccharides are created. In extreme case, humin could be created.[34] Amylopectin ramification $\alpha(1 \rightarrow 6)$ linkages might also be broken and turned onto amylose. So, starch composition will change, as will the final material properties, as predicted by the previously proposed models. This highlights the need to study interactions such as compositional factors-temperature. Extrusion speed is necessary to homogenize the material but it also creates shearing stress and local heating which affect both fibers and matrix. Collecting all these data prior to study BC made from TPS and natural plant fibers will permit homogeneous data comparisons and field improvement.

5 | CONCLUSION

Computation of the TPS and BC literature produced regression models of output variables (elastic modulus and tensile strength) with a 86% variability description (corresponding to R^2_{adj}), but with a low predictability power (RMSE <554 for elastic modulus and <25.36 for tensile strength). A square root transformation of the output variables is the transformation giving the results in all cases (higher R^2_{adj} and lower RMSE). No model of elongation at break could be computed, because of the lack of highly affecting parameters in papers, such as the aging (temperature, time, and relative humidity).

Significance of parameters affecting mechanical properties (such as fibers cellulose and lignin content, fibers length, starch composition, and plasticizers content for TPS-BC) versus nonimpacting parameters (process temperature and fibers hemicelluloses content for TPS-BC) was studied. Still, as a part of the mechanical properties' variability could not be explained by the proposed models, it was hypotheses that nonreferenced or poorly referenced parameters (or interactions) were impacting those properties.

For each part of the BC plasticization (fibers, starch, and process parameters), a list of parameters to referenced systematically was presented, with their method of measurement.

The use of multivariate analysis is of great interest for implementation in the design of multiple materials. This would save time and money to companies and researchers developing new materials for the utilities of the future.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Interreg France-Wallonie-Vlaanderen Call V, which granted this work as part of the COMPOSENS project. The authors would also like to thank the Walloon Region, Belgium, for supporting this work as part of the FEVERPRO project (File D31-1395).

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest.

ORCID

Sophie Morin https://orcid.org/0000-0002-0903-2549
Lionel Dumoulin https://orcid.org/0000-0003-2865-5431
Aurore Richel https://orcid.org/0000-0002-4352-6542

REFERENCES

[1] S. Morin, B. Lecart, T. Istasse, C. Bailly Maître Grand, F. Meddeb-Mouelhi, M. Beauregard, A. Richel, *Int. J. Biol. Macromol.* **2020**, *148*, 851. http://dx.doi.org/10.1016/j.ijbiomac.2020.01.232.

[2] H. P. S. Abdul Khalil, A. H. Bhat, A. F. Ireana Yusra, *Carbohydr. Polym.* **2012**, *87*, 963. http://dx.doi.org/10.1016/j.carbpol.2011.08.078.

[3] M. Jawaid, S. M. Sapuan, O. Y. Alotman, *Green Biocomposites*, 1st ed., Green Energy and Technology, Springer International Publishing, Cham, Switzerland **2017**. https://doi.org/10.1007/978-3-319-46610-1.

[4] A.-L. Ollett, R. Parker, A. C. Smith, *J. Mater. Sci.* **1991**, *26*, 1351. http://dx.doi.org/10.1007/bf00544476.
[5] W. Wiedmann, E. Strobel, *Starch - Stärke* **1991**, *43*, 138. http://dx.doi.org/10.1002/star.19910430404.
[6] H. G. Fritz, W. Aichholzer, T. Seidenstücker, B. Widmann, *Starch - Stärke* **1995**, *47*, 475. http://dx.doi.org/10.1002/star.19950471206.
[7] A. Bledzki, *Prog. Polym. Sci.* **1999**, *24*, 221. http://dx.doi.org/10.1016/s0079-6700(98)00018-5.
[8] Y. Zhang, C. Rempel, D. McLaren, in *Innovations in Food Packaging*, 2nd ed., Elsevier, Plano, TX, USA **2014**, Chapter 16 - Thermoplastic Starch, p. 391. https://doi.org/10.1016/B978-0-12-394601-0.00016-3.
[9] O. V. López, M. A. García, *Mater. Sci. Eng. C.* **2012**, *32*, 1931. http://dx.doi.org/10.1016/j.msec.2012.05.035.
[10] Y. Zhang, J. H. Han, *J. Food Sci.* **2006**, *71*, E109. http://dx.doi.org/10.1111/j.1365-2621.2006.tb08891.x.
[11] A. Mohammadi Nafchi, M. Moradpour, M. Saeidi, A. K. Alias, *Starch - Stärke* **2013**, *65*, 61. http://dx.doi.org/10.1002/star.201200201.
[12] X. Li, L. G. Tabil, S. Panigrahi, *J. Polym. Environ.* **2007**, *15*, 25. http://dx.doi.org/10.1007/s10924-006-0042-3.
[13] F. P. La Mantia, M. Morreale, *Compos. Part A: Appl. Sci. Manuf.* **2011**, *42*, 579. http://dx.doi.org/10.1016/j.compositesa.2011.01.017.
[14] H. Angellier, S. Molina-Boisseau, P. Dole, A. Dufresne, *Biomacromolecules* **2006**, *7*, 531. http://dx.doi.org/10.1021/bm050797s.
[15] J. J. G. Van Soest, D. B. Borger, *J. Appl. Polym. Sci.* **1997**, *64*, 631. http://dx.doi.org/10.1002/(sici)1097-4628(19970425)64:4<631::aid-app2>3.0.co;2-o.
[16] P. H. Manepalli, S. Alavi, *J. Food Eng.* **2019**, *261*, 60. http://dx.doi.org/10.1016/j.jfoodeng.2019.04.004.
[17] G. S. Zeng, C. Hu, S. Zou, L. Zhang, G. Sun, *Polym. Compos.* **2019**, *40*, 3923. http://dx.doi.org/10.1002/pc.25252.
[18] W. Ben Chaabene, M. Flah, M. L. Nehdi, *Constr. Build. Mater.* **2020**, *260*, 119889. http://dx.doi.org/10.1016/j.conbuildmat.2020.119889.
[19] R. Shi, Q. Liu, T. Ding, Y. Han, L. Zhang, D. Chen, W. Tian, *J. Appl. Polym. Sci.* **2007**, *103*, 574. http://dx.doi.org/10.1002/app.25193.
[20] C. M.O. Müller, J. B. Laurindo, F. Yamashita, *Carbohydr. Polym.* **2009**, *77*, 293. http://dx.doi.org/10.1016/j.carbpol.2008.12.030.
[21] R. F. Tester, J. Karkalas, X. Qi, *J. Cereal Sci.* **2004**, *39*, 151. http://dx.doi.org/10.1016/j.jcs.2003.12.001.
[22] X. Qi, R. F. Tester, *Starch - Stärke* **2018**, *70*, 1700304. http://dx.doi.org/10.1002/star.201700304.
[23] S. Morin, B. Lecart, M. Lang, A. Richel, *Carbohydr. Polym.* **2021**, *254*, 117412. http://dx.doi.org/10.1016/j.carbpol.2020.117412.
[24] F. Berzin, T. Amornsakchai, A. Lemaitre, E. Di Giuseppe, B. Vergnes, *Polym. Compos.* **2018**, *39*, 4115. http://dx.doi.org/10.1002/pc.24475.
[25] R. Castellani, E. Di Giuseppe, J. Beaugrand, S. Dobosz, F. Berzin, B. Vergnes, T. Budtova, *Compos. Part A: Appl. Sci. Manuf.* **2016**, *91*, 229. http://dx.doi.org/10.1016/j.compositesa.2016.10.015.
[26] T. Nishino, K. Takano, K. Nakamae, *J. Polym. Sci. Part B Polym. Phys.* **1995**, *33*, 1647. http://dx.doi.org/10.1002/polb.1995.090331110.
[27] Q. Schmetz, G. Maniet, N. Jacquet, H. Teramura, C. Ogino, A. Kondo, A. Richel, *Ind. Crops Prod.* **2016**, *94*, 308. http://dx.doi.org/10.1016/j.indcrop.2016.09.003.
[28] R. Hoover, W. S. Ratnayake, *Curr. Protoc. Food Anal. Chem.* **2001**, *00*, http://dx.doi.org/10.1002/0471142913.fae0203s00.
[29] S. E. Harding, G. G. Adams, R. B. Gillis, *Starch - Stärke* **2016**, *68*, 846. http://dx.doi.org/10.1002/star.201600042.
[30] L. Wang, X. Zhang, J. Xu, Q. Wang, X. Fan, *Process Biochem.* **2020**, *97*, 176. http://dx.doi.org/10.1016/j.procbio.2020.07.015.
[31] A. J. Kortstee, A. M. S. Vermeesch, A. M. S. de Vries, E. Jacobsen, R. G. F. Visser, *Plant J.* **1996**, *10*, 83. http://dx.doi.org/10.1046/j.1365-313x.1996.10010083.x.
[32] M. E. Mngomezulu, M. J. John, V. Jacobs, A. S. Luyt, *Carbohydr. Polym.* **2014**, *111*, 149. http://dx.doi.org/10.1016/j.carbpol.2014.03.071.
[33] X. Liu, Y. Wang, L. Yu, Z. Tong, L. Chen, H. Liu, X. Li, *Starch - Stärke* **2013**, *65*, 48. http://dx.doi.org/10.1002/star.201200198.
[34] I. Thibaut, R. Aurore, Mechanistic aspects of saccharide dehydration to furan derivatives for reaction media design *RSC Adv.* **2020**, *10*, 23720. http://dx.doi.org/10.1039/d0ra03892j.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Morin S, Dumoulin L, Delahaye L, Jacquet N, Richel A. Green composites based on thermoplastic starches and various natural plant fibers: Impacting parameters of the mechanical properties using machine-learning. *Polymer Composites*. 2021;1–10. https://doi.org/10.1002/pc.26071