Heliyon 10 (2024) e25170

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Determination of the most significant rubber components influencing the hardness of natural rubber (NR) using various statistical methods

Lilla Virág [a,*], Attila Egedy [b], Csilla Varga [c], Gergely Erdős [d], Szabolcs Berezvai [e], László Kovács [d], Zsolt Ulbert [b]

[a] *Department of MOL Hydrocarbon and Coal Processing, Research Centre for Biochemical, Environmental and Chemical Engineering, Faculty of Engineering, University of Pannonia, H8200 Veszprém, Egyetem Str. 10, Hungary*
[b] *Department of Process Engineering, Research Centre for Biochemical, Environmental and Chemical Engineering, Faculty of Engineering, University of Pannonia, H8200 Veszprém, Egyetem Str. 10, Hungary*
[c] *Sustainability Solutions Research Lab, Research Centre for Biochemical, Environmental and Chemical Engineering, Faculty of Engineering, University of Pannonia, H8200 Veszprém, Egyetem Str. 10, Hungary*
[d] *ECon Engineering Kft. H1116, Budapest, Kondorosi út 3, Hungary*
[e] *Department of Applied Mechanics, Faculty of Mechanical Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

ARTICLE INFO

Keywords:
Natural rubber
Shore a hardness
statistical analysis
Linear regression
t-test
Ridge regression
Binary decision tree

ABSTRACT

Manufacturers use a large number of components in the production of modern rubber products. The selection of the constituents of the rubber recipe is primarily determined by the purpose of use. The different fields of applications of rubbers require the presence of appropriate mechanical properties. In this respect, it can be useful to know which substances forming the rubber recipe have significant influence on the different mechanical properties. In this study, the statistical analysis of the influence of rubber components on the hardness of natural rubber (NR) is proposed based on literature review. Based on the literature data, various statistical analyses, like *linear regression*, *constrained linear regression, Ridge* regression, *Ridge sparse* regression and *binary classification decision trees* were performed to determine which rubber components have the most significant effect on the hardness. In the statistical analyses, the effect of a total of 42 constituents of rubber compound on hardness was investigated. Most of the applied statistical methods confirmed that the traditional frequently used rubber components, such as carbon black and sulfur, have a primary effect on the hardness. However, the substances forming the rubber compound that are not widely used in practice or newly developed components appear differently in the lists of significant additives obtained by the different statistical methods.

1. Introduction

Rubber is an indispensable product of modern societies. Both natural rubbers (*NR*) and synthetic rubbers such as styrene-butadiene rubber (*SBR*), nitrile rubber (*NBR*), polyisoprene rubber (*PR*), polychloroprene rubber (*PCR*), ethylene-propylene-diene-terpolymer

* Corresponding author.
E-mail address: virag.lilla@mk.uni-pannon.hu (L. Virág).

https://doi.org/10.1016/j.heliyon.2024.e25170
Received 6 October 2023; Received in revised form 11 January 2024; Accepted 22 January 2024
Available online 24 January 2024

(*EPDM*) or silicone rubbers are widely used in the production of consumer goods and structural materials [1]. They are applied as tires, hoses and beltings, mountings, bearings, rolls, seals, energy storages and energy dissipators (twisted strands and columns of rubber, rubber ropes) rubber linings, shoes, sports equipment etc. [2]. Rubbers are often blended to achieve optimum properties and performance for a specific period in a particular environment [3].

The properties of rubber products are both affected by the processing parameters and the different components incorporated into the rubber. The production technology for rubber products involves three main process steps: mixing, shaping, and vulcanization as shown in Fig. 1 [1,4]. In each of these steps there are operating parameters that affect the properties of the product. For example, (incorporation, and dispersion of the ingredients in the elastomer matrix) affects the viscosity, scorch safety, molding, and mechanical properties of the rubber [5]. During the curing step the cross-links between the polymer chains are formed, providing the elastic structure and determining the properties of the final product. The formation of cross-links between polymer chains is determined by the process parameters: the mold pressure, the mold temperature, and the holding time [6,7]. In addition to the operating parameters, the chemical composition of the rubber is very important, as the quantity and quality of the components also have a major impact on the properties of rubber products.

Manufacturers use a wide range of components to make different rubber products. The selection of rubber components and their composition are primarily determined by the purpose of use. The various applications of rubber require the right mechanical properties, so identifying the variables that affect the quality of the rubber and the relationships between these variables and the properties of the vulcanizate is an important design task in the production of rubber products.

The statistical modeling approach proves to be an effective tool for estimating the specific contribution of different filler or additive properties to the characteristics of natural rubber composites. In addition, the modeling approach can be used to predict the properties of different rubber composites based on the specific characteristics of the rubber and filler, processing conditions and compounding formulations [8]. Statistical methods are often used in experimental design. The main objective of statistical experimental design is to determine the variables that affect the quality of the process by collecting information on the properties that affect the performance of the process. It is also used to identify the factors and optimal levels of the factors so that the process can operate under optimal conditions. Experimental design techniques are essential tools for developing new products, improving high-cost processes, and improving the performance characteristics of existing products [9].

Several literatures focus on how different rubber compounds affect the properties of rubber products, with the use of statistical analysis and models [8,10–14]. Several statistical models, such as the Gaussian statistics model, 3-chain model, 4-chain model, 8-chain model and Flory–Erman model, have been proposed since the 1940s to describe the mechanical properties of vulcanized natural rubber (*NR*). These models are, however, too simplified as they do not consider the sulfur content of the polymer chains and the enthalpic elasticity of the polymer matrix [14]. To improve the traditional Gaussian models, Zhang et al. made two modifications. The sulfur content was considered in the model, and the single-chain elasticity (including entropic and enthalpic elasticity) of the two main components of *NR* (*cis*-1,4-polyisoprene (*PI*) and polysulphide) was also incorporated into the model. Zhang et al. find that the new model (called the *TCQMG* model) is significantly better than conventional models in describing the mechanical properties of bulk *NR*, especially at high deformation [14]. Barrera et al. [8] developed practical statistical models to quantify the contributions from the properties of conventional and non-conventional fillers and to predict the resulting mechanical properties of *NR* composites. They used the filler characteristics as explanatory variables in multiple linear regression analyses. The filler surface area and loading were found to be important variables contributing to composite properties, like the tensile strength and the 300 % modulus. Multiple linear regression analysis was used by Lewis et al. [10] to investigate the effects of rubber ratio, carbon black, and accelerator level on the cure characteristics and physical properties of natural rubber/bromobutyl (*BIIR*) rubber blends. They found that the level of carbon black is the most significant factor affecting scorch time, cure time, hardness, tensile properties, and compressive strength. The difference in the reactivity of *NR* and *BIIR* towards sulfur has resulted in vulcanization behavior and physical properties that are determined by the *NR* content in the rubber blend. Krivtsov et al. [11] applied an empirical approach to the causal analysis of the failure of a certain type of automobile tire. In their study, they used Cox survival regression model and standard linear regression to model the initiation and propagation of the failure. The statistical models developed help to identify the elements of tire design that influence the probability of tire failure due to the failure mode in question. S. Srewaradachpisal et al. investigated the effect of filler (carbon black and clay)/plasticizer (paraffinic oil) fractions on the shock absorption of natural rubber compounds. The statistical response surface method (*RSM*) was used to optimize the loading of natural rubber to reduce the peak impact force. The results confirm that high amounts of plasticizer successfully reduce both the peak impact force and the hardness of the rubber. On the other hand, both fillers increase the hardness, resulting in a corresponding increase in the impact force. Based on experimental data, a predictive *RSM*

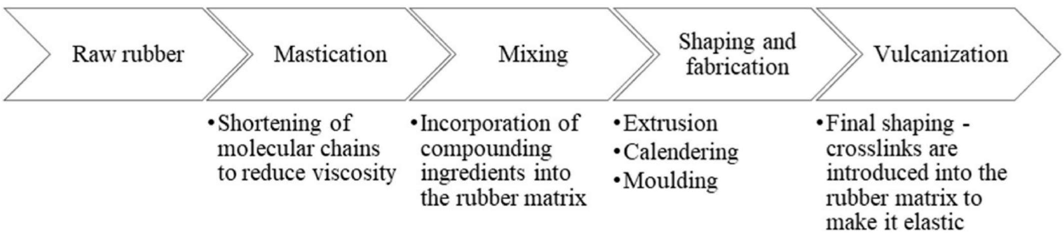

Fig. 1. Flow chart of rubber processing [4].

model was developed. The results show that quadratic regression models in the context of *RSM* give optimal loading levels for carbon black, clay and paraffinic oil in *NR* and predict properties corresponding to low impact forces [15]. The current problem with natural rubber formulation methods used in industry is that they depend primarily on the experience of the formulator. Therefore Román et al. investigated the implementation of a rubber compound formulation methodology that targets material properties relevant to the product's application domain by incorporating predictive models like linear regression, response surface method, artificial neural networks (*ANN*) and Gaussian process regression (*GPR*). By selecting the appropriate viscoelastic properties and prediction methods, a small number of experimental composites captured the highly non-linear behaviour of *NR* compounds, resulting in *GPR* showing the highest cross-validation prediction accuracy of 100 % with five-fold cross-validation. *GPR* was able to accurately predict short-term behaviour, long-term viscoelastic behaviour, transient and dynamic properties. *RSM*, *ANN* and *GPR* result in prediction accuracy of 90 %, 97 % and 100 % respectively [16].

The hardness of the rubber is a key component in the design of rubber products. Hardness is the property of a material that allows it to resist permanent distortion, penetration, indentation, and scratching. Accordingly, it is essential to understand which rubber components have the greatest impact on the hardness of the product. Shore hardness is a measure of the resistance of a material to penetration or indentation. There are several scales for measuring the hardness of elastomers, of which Shore A is used to measure the hardness of softer materials such as rubber. In this study, statistical investigations have been carried out to determine how rubber components can affect the hardness of natural rubber (*NR*). Literature sources reporting mechanical tests in addition to the rubber components used for rubber sample preparation were reviewed. Using the literature data various statistical analyses (*linear regression*, *constrained linear regression, Ridge* regression, *Ridge sparse* regression and application of *binary classification decision trees*) were performed to determine which rubber components have the most significant effect on Shore A hardness value of natural rubber. In the statistical analyses, we investigated the effect of a total of 42 rubber components on hardness of rubber.

2. Factors affecting rubber properties

2.1. General

The properties and performance of a rubber product depend on many factors involving the chemical composition of the rubber, the amount, and type of ingredients like fillers and additives incorporated into the rubber compound, and the processing and vulcanizing conditions [17]. In the production of rubber products, an important design task is to identify the variables that affect quality of the rubber and to explore the correlations between these variables and properties of the vulcanizate.

2.2. Effect of the vulcanization system

Many different types of vulcanization systems have been developed, of which sulfur-based and peroxide-based systems are the most widely used [6,17]. The type of vulcanization system used determines the cross-linking [18]. The final structure, mechanical performance and thermal stability of vulcanized elastomers formulated with sulfur and accelerator-based systems are highly dependent on the type and extent of cross-linking that occurs in the compound [19]. Kruželák et al. [18] prepared rubber compounds based on *NR* and *NBR* by applying of mixed sulfur and peroxide curing systems to investigate the influence of curing system composition on cross-linking and physical-mechanical properties of rubbers. The sample cured only with peroxide system showed the highest crosslink density. The modulus and hardness of the vulcanizates are in a close correlation with the change in the cross-link density. Different types of correlation were observed for tensile strength and elongation at break, depending on the composition of the curing system. Both properties of *NR*-based vulcanizates were improved by increasing the amount of sulfur and decreasing the amount of peroxide.

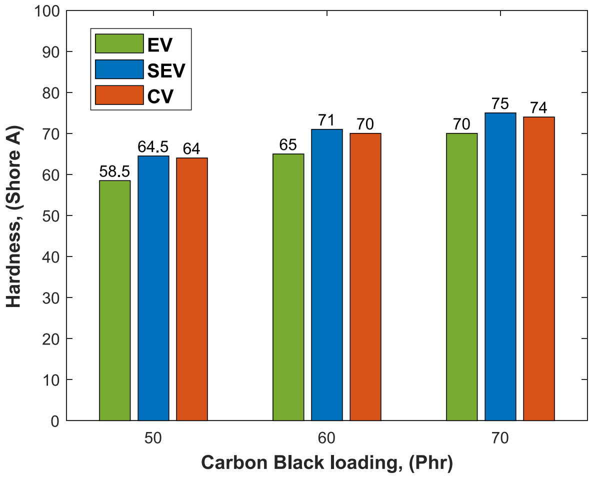

Fig. 2. Effect on carbonblack content on Shore A hardness at different ratio of accelerators to sulfur [22].

Similar relations were found by Nah et al. [20] and Nabil et al. [21]. Nah et al. [20] also found that higher tensile strength, modulus and hardness and lower elongation at break were achieved by peroxide-based curing than sulfur-based vulcanization because of higher crosslink density. The properties of rubber composites are significantly determined by the ratio of the accelerator to the vulcanizing agent. Sulfur vulcanization systems can be classified by the ratio of accelerators to sulfur as follows: efficient (*EV*), semi-efficient (*SEV*) and conventional (*CV*) [22]. An *EV* system is characterized by a high ratio of accelerators to sulfur (>2.5), resulting in the formation of monosulphides during cross-linking. The *CV* system has a low ratio of accelerators to sulfur (0.1–0.6), which result in a high percentage of polysulfides in the formation of crosslinking, while the *SEV* system is in between the two, characterized by a moderate ratio of accelerators to sulfur (0.7–2.5). Mayasari and Yuiniari [22] showed that the *EV* system resulted in faster vulcanization time than *SEV* and *CV* systems. *EV* system also performed the highest tensile strength, elongation, and tear strength, while *SEV* system resulted in the highest hardness as presented in Fig. 2. Furthermore, the conventional vulcanization system resulted in the lowest swelling index.

2.3. Effect of accelerators

The type of accelerator also has an influence on the curing characteristics, mechanical and thermal properties of vulcanizates. The effect of various vulcanization accelerators such as N-*tert*-butyl-2-benzothiazyl-sulphonamide (*TBBS*), N-cyclohexyl-benzothiazyl-sulfenamide (*CBS*), tetramethylthiuram disulfide (*TMTD*), and 2-mercaptobenzothiazol (*MTB*) and two ratios of vulcanization accelerator/sulfur (2:1 and 1:2 systems) on curing characteristics, mechanical properties and thermal properties of rubber were investigated by Formela et al. [23]. They found that the scorch time was mostly influenced by the type of accelerator, but much less by the accelerator/sulfur ratio. The highest values of the scorch time were observed with benzothiazole sulfonamide derivatives, *TBBS* and *CBS*. Rubber vulcanized with a conventional vulcanization system based on *TBBS* and *CBS* had the best processing and mechanical properties. Nabil el al [24]. investigated the effect of four types of commercial accelerators (*TBBS*, *CBS*, *TMTD* and *MTB*) on curing characteristics, tensile properties of *NR*/recycled *EPDM* blends. It was found that the tensile strength of the blends cured in the presence of *CBS* was relatively higher than the other three accelerators. The *NR/*(*R-EPDM*) mixtures cured with *TMTD* showed the highest cross-linking density, followed by *TBBS*, *CBS* and *MBT*, respectively. Kim et al. [25] found a distinct correlation between the chemical crosslink structures and the mechanical properties of the *NR* rubber. The tensile strength of rubber slightly increased when the accelerator content was higher as presented in Fig. 3.

2.4. Effect of temperature

Another important parameter of the cure process of rubbers and elastomers is the vulcanization temperature because it also has a significant influence on the structure and mechanical properties [19]. The typical temperature used for vulcanizing rubbers is 150–180 °C. Mukhopadhyay et al. [26] found that the scorch time, the optimum cure time, and reversion time decreases as the vulcanization temperature increases. In terms of mechanical properties, it was found that the values of tensile strength, elongation at break and Shore A hardness decreased with increasing vulcanization temperature because the lower crosslinking level as shown in Fig. 4 [26,27].

2.5. Effect of fillers

An important parameter in the production of rubber composites is what and how much additives and fillers are added. Carbon black is a universal reinforcing filler [28,29], but clay [29,30], silica [29], calcium carbonate [31], and other inorganic [32] or organic

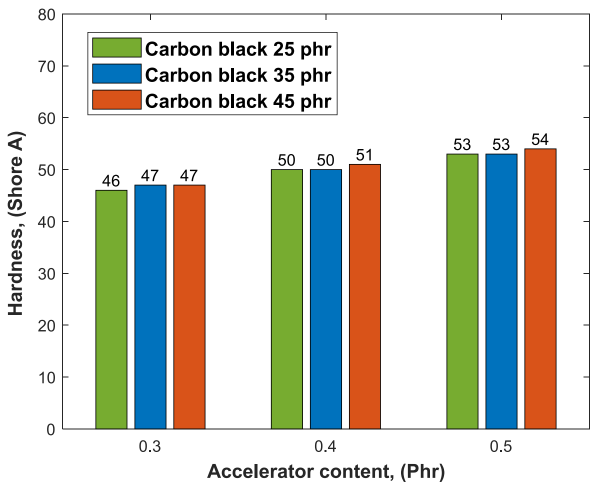

Fig. 3. Effect of the accelerator content (phr) on rubber hardness (Shore A) at different Carbon Black (CB) content [25].

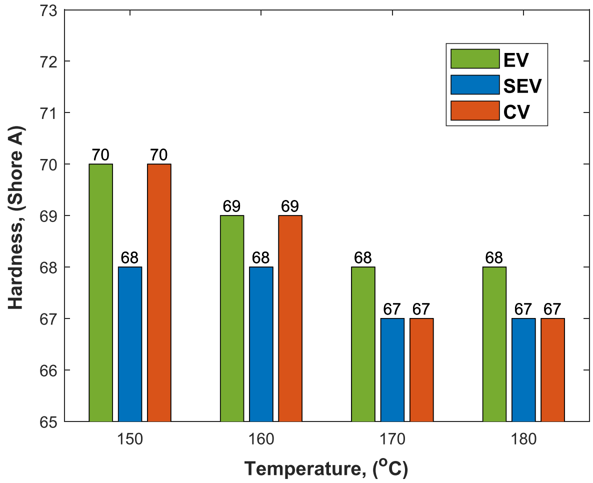

Fig. 4. Effect of the vulcanization temperature (°C) on rubber hardness (Shore A) at different ratio of accelerators to sulfur [27].

(fiber) [33] materials can also be used as fillers in rubber products. Tan and Isayev [29] investigated the vulcanization behavior and mechanical performance of silica-, nanoclay-, and carbon black-filled (0–30 phr) *EPDM* rubber. The hardness, modulus, elongation at break, and tensile strength of all the vulcanizates was found to increase with increasing filler content. However, an increase in the elongation at break of carbon black-filled *EPDM* vulcanizates was not significant. The increase in tensile strength, elongation at break and hardness of silica-filled *EPDM* vulcanizates with increasing filler content was greater than that of carbon black-filled *EPDM* vulcanizates. Al-maamori et al. [34] investigated the effect of carbon black filler on the mechanical and physical properties of *NBR*. The amount of filler was varied between 0 and 80 phr. Their results showed that the mechanical properties such as hardness, elastic modulus and tensile strength increase with the amount of filler up to a certain loading level (60 phr) and then decrease slightly. The effect of CB content (phr) on rubber hardness is shown in Fig. 5.

Not only the amount of filler but also the particle size of the filler is important for the properties of the rubber. Chuayjulit et al. [31] investigated the effect of carbon black filler particle size (26–30 nm, 40–48 nm, and 49–60 nm) and quantity (30, 45 and 60 phr) on the properties of vulcanized natural rubber. Increased Mooney viscosity and shortened vulcanization and scorch time was observed with decreasing particle size of the carbon black. The same effect can be obtained by increasing the amount of the fillers. Rattanasom and Prasertsri [35] used calcined clay filler for natural rubber with carbon black of different type (N330, N550 and N774). Carbon black fillers differed in their particle size and its distribution. At similar hardness, the strengths and thermal ageing resistance of vulcanizates increased with the addition of various carbon black fillers, as carbon black has a higher strengthening efficiency than clay. Among the carbon black types tested, the vulcanizate containing N330 (particle size 26–30 nm) gave the highest tensile strength and the lowest elasticity. Although the N330 filler has the lowest volume fraction in the sample, its smaller particle size probably resulted in better dispersion in the rubber.

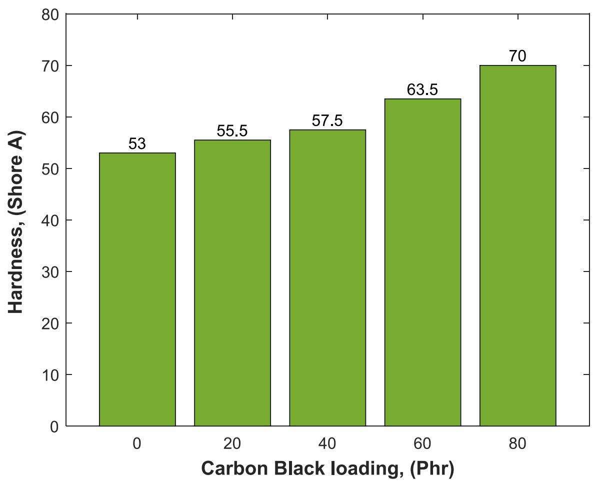

Fig. 5. Effect of the CB content (phr) on rubber hardness (Sh A) [34].

2.6. Effect of additives

In addition to vulcanizers, accelerators, activators, and various fillers, rubber composites also contain other additives such as plasticizers [36–38], peptizes, dispersants and lubricants, homogenizers, adhesion promoters [38], antioxidants [38–40], deactivators, ozone depletion inhibitors [38], which also affect the properties of rubber products. For example, various processing oils such as paraffinic oils, naphthenic oils, aromatic oils, vegetable oils are used to reduce the viscosity of rubber composites and thus improve their process [36,37]. The use of plasticizers reduces the hardness and tensile strength of the rubber composite [36,37].

3. Analysis data

Reviewing the related literature, it can be seen that various fillers and additives are used in the rubber recipes to improve the mechanical properties of the rubber. Each component has a different effect on each mechanical property. The purpose of the statistical analyses is to find a relationship between the recipe components, the vulcanization parameters, and the hardness of rubber. In the literature, there are number of components used by the authors to produce the rubber samples. In this respect, it is important to obtain any information about which components or vulcanization parameters have a more significant effect on mechanical properties among hardness.

The literature review identified 20 publications containing a recipe for production of natural rubber sample and measurements for the *Shore* A hardness of the produced rubber samples. The processed publications provided a total of 142 recipes and their corresponding *Shore* A hardness values. Each recipe used different components to produce natural rubber. They contain the percentages of components in relation to the weight of the natural rubber. In total, 42 different components were found in the published recipes, which are summarized in Table 1. In the table, for each component, publications where the component was included in the published

Table 1
Operational parameters and rubber components used as predictor variables in the statistical analyses.

	Operational parameters and rubber components	Predictor variable name	Function	References
1	Temperature of vulcanization	Temp	operational parameter	[18,25–27,35,41–52,52,54,55]
2	Caolin	Caolin	filler	[55]
3	Clay	Clay	filler	[35,41,54]
4	Multi-walled carbon nanotubes	MWCNT	filler	[53]
5	Carbon nanotubes	CNT	filler	[52]
6	Bamboo multi-walled carbon nanotubes	B-CNT	filler	[52]
7	Natrium-montmorillonite	Na-MMT	filler	[45,46]
8	Organo-modified montmorillonite	OMMT	filler	[45,46]
9	Carbon black	Carbblack	filler	[25,27,35,41,45,49,50,54]
10	SiO_2	SiO2	filler	[35,41,48,54]
11	Alkylamide	Alkylamide	filler	[48–50]
12	Palygorskite	Palyg	filler	[42]
13	Porous carbon fiber	PCF	filler	[51]
14	Surface-acetylated cellulose powder	SACP	filler	[51]
15	Polyethylene glycol	PEG	filler	[41]
16	Sulfur	Sulfur	accelerator	[18,25–27,35,41–52,52,54,55]
17	N-Cyclohexyl-2-benzothiazole sulfenamide	CBS	accelerator	[18,27,41,47,51,52]
18	Tetramethylthiuram disulfide	TMTD	accelerator	[27,41–43,46,47]
19	2-morfolinoditio-benzotiazol	MDB	accelerator	[26,43],
20	N-*tert*-butyl-benzothiazole sulfonamide	TBBS	accelerator	[44,53]
21	1-phenyl-2,4-thiobiuret	DTB	accelerator	[44]
22	4,4′-dithiodimorpholine	DTDM	accelerator	[27]
23	dicumyl peroxid	DCP	accelerator	[18,55]
24	Ethylene glycol diacetate	EGDA	accelerator	[18]
25	Mercaptobenzothiazole	MBT	accelerator	[52]
26	Benzothiazole disulfide	MBTS	accelerator	[25,45,46,48–50]
27	Diphenylguanidine	DPG	accelerator	[25]
28	N-Isopropyl-N′-phenyl-1,4-phenylenediamine	IPPD	accelerator	[48–50]
29	Accelerator Acc-G	Acc-G	accelerator	[35,54]
30	Accelerator Acc-S	Acc-S	accelerator	[35,54]
31	Accelerator Acc-T	Acc-T	accelerator	[35,54]
32	Zinc-diethyldithiocarbamate	ZDEC	accelerator	[42]
33	Accelarator Accel-M	Accel-M	accelerator	[55]
34	Softener Soft-HH	Soft-HH	softener	[53]
35	Lubricant	Lubricant	softener	[25]
36	N-Phenyl-2-naphtylamine	PBN	antioxidant	[45]
37	Antioxidant 6PPD	6PPD	antioxidant	[27]
38	Antioxidant	Antiox	antioxidant	[25,42,55]
39	Nano-ZnO	nanoZnO	activator	[47]
40	Micro-ZnO	microZnO	activator	[47]
41	Zinc-oxide	Zinc-oxide	activator	[18,26,27,35,41–52,52,54,55]
42	Stearic acid	Stearicacid	activator	[18,26,27,35,41–52,52,54,55]

rubber recipe are also indicated.

4. Statistical methods

The statistical analyses used in this study to process the literature experimental data were the *t*-test based on *linear regression*, the *constrained linear regression*, the *Ridge* regression, the *Ridge sparse* regression, and the application of *binary classification decision.*

In the application of linear regression, a linear model describing the relationship between the predictor variables and the response variable is defined. The number of predictor variables considered is *p*. Let X be an $n \times (p+1)$ matrix whose first column contains constant 1 values, and columns $2, \ldots, p+1$ contain the values of the predictor variables. *n* is the number of response values. Let y denote the vector containing the responses to the predictor variables, which is a column vector containing *n* response values. In the case of linear regression, the following linear function is assumed between the predictor variables and the response variable.

$$y_i = f\left(x_{i,1}, x_{i,2}, \ldots, x_{i,p}\right) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_p x_{i,p} + \varepsilon, \nearrow i = 1, \ldots, n \tag{1}$$

where $\beta_0, \ldots, \beta_p$ are the coefficients, ε is the error. The coefficients are determined based on the method of least squares by minimizing the following objective function using **X** and **y**.

$$\widehat{\beta}^{LS} = \min_{\beta}\left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \right\} \tag{2}$$

Solving the minimum search problem formulated in equation (2) provides the following equation for the coefficients.

$$\widehat{\beta}^{LKN} = \left(X^T X\right)^{-1} X^T y \tag{3}$$

Using the data collected from recipes, the predicator variable matrix X and the response variable vector y were created as the starting data structures for the statistical analyses. The matrix X, which is a 142 × 42 matrix, contains the amount of components used in each recipe. The vector**y**, which is a column vector with 142 elements, contains the Shore A hardness values of the rubber samples applying the recipes. Since each recipe does not use all 42 rubber components, the elements of the matrix X that were empty, i.e., in the case where the recipe did not contain the given component, were filled with 0 values.

In the application of linear regression a linear model describing the relationship between the predictor variables and the response variable is defined. Implementing the linear regression the ***regstats*** function of MATLAB software package version R2021b was used, which has the option to return the *p* values of the *t*-tests for the coefficients of the linear model. The *p* values for the coefficients indicate whether the relationship between the independent variables and the dependent variable is statistically significant, and they were used to decrease the number of predictor variables.

In the case of constrained linear regression, limits can be defined for the model coefficients to be determined. The constrained linear regression was applied using the ***lsqin*** function of MATLAB. In the analyses using this method, a non-negativity criterion for the model coefficients belonging to the filler, accelerators, cross-linkers, and activator components was defined.

The *Ridge regression* is one of the so-called shrinkage methods. Its application ensures not only the decrease of the number of predictor variables but also provides model coefficients with a smaller standard deviation which are closer to each other and closer to zero. The application of the Ridge method was implemented in the MATLAB environment by using the MATLAB ***ridge*** function. When decreasing the number of the predictor variables,. first the variables with the smallest absolute coefficient are removed.

The *Ridge sparse* regression method is also one of the shrinkage methods available in the MATLAB environment. The *Ridge sparse* regression uses so-called sparse matrices. The data used in the statistical analyses typically correspond to such data, i.e. many rubber components are available, but only a small number of components are used in the compositions. The Ridge sparse method was used with the ***fitrlinear*** function of MATLAB. The use of a ***fitrlinear*** function, in addition to the application of a sparse matrix and the shrinking of coefficient values, allows cross-validation.

In determining the significant predictor variables, an analysis method that includes the application of so-called binary classification decision trees was also applied. Application of this method results in paths that represent a set of predictor variables and their intervals, if applied, we can get a classification value of the response variable. MATLAB's ***fitctree*** function was used to grow the classification decision trees. Based on the generated classification decision tree, MATLAB's ***predictorImportance*** function was used to estimate the importance values of the predictor variables.

More details of applied mathematical methods are given in the chapter 7. Appendix.

5. Results and discussion

5.1. General

In this section the detailed outcome of the statistical analysis of the large amount of processed experimental data is presented in a structured form. The analyzed rubber constituents were put in the order of importance as far as the influence on Shore A hardness is concerned. The outcome is investigated in detail for all implemented statistical evaluation methods and then general conclusions are also made focusing on the most important constituents.

The investigation of the outcome focusses on those substances that are very general and typically are supposed to influence the mechanical response of the rubber the most such as.

1. Filler components that impact the spatial distribution of generated cross links such as carbon black, quartz sand (SiO_2), clay, PCF etc.,
2. Parameters that directly influence the generation of cross-links by either:
a. Providing the molecules to react (sulfur) or,
b. Activating or accelerating the chemical reaction to form these links (e.g. vulcanization temperature or activators etc.).
3. Additives that directly impact the hardness/softness of the blend being mixed before vulcanization (such as lubricants).

5.2. Linear regression and t-test

In the first statistical study, linear regression and *t*-test were used to find a relationship between the predictor variables and *Shore* A hardness values, and to determine which predictor variables have the greatest influence on the hardness value. The principle of stepwise regression was used to construct a reduced set of predictor variables. As a result of the statistical analysis, 19 significant predictor variables were optained. For each variable, the *p*-value was below 0.05, i.e. they can be considered such variables that can significantly affect the *Shore* A hardness value. The values of -log(*p*) were calculated for each predictor variables and depicted in Fig. 4. It shows the -log(*p*) values in descending order. The higher this value, the more significant the component is considered. It can be seen from the values in Fig. 6 that in addition to the large number of variables, the narrowed set of predicator variables contains 8 filler components, 6 accelerator components, 2 activator components, 1 softener component, 1 antioxidant component and the vulcanization temperature as operational parameter. The first three components with the smallest *p* values are filler components. The large number of variables makes it difficult to prepare an experimental design for the production of rubber samples. One possible way to further narrow the significant variable set may be to select the most significant one, two or three variables from each component group.

It can be concluded from the derived order of importance that the large influence of carbon filler is not surprising. This is also true for the additional fillers (PCF and SACP). The impact on hardness of SiO_2 and softeners according to this evaluation is only moderate and even a number of accelerators were ranked higher that was not expected. The influence of vulcanization temperature is small which might be explained by the fact that there is equivalence between vulcanization temperature and time to get to a ca. 100 % cured situation that all published vulcanization processes were supposed to optimized for. Thus, if cured for enough time and the temperature is above the activation limit, the effect of that is not expected to be significant on the cross-link density of the vulcanizate. That is also somewhat surprising that Sulfur does not appear in the first 20 places. It is expected that the amount of added sulfur fundamentally determines the number of cross-links and thus, the internal structure and mechanical resistance of the vulcanized rubber.

A linear regression was performed using the significant predictor variable set and the measured *Shore* A hardness values. Fig. 7 shows the comparison of the measured and estimated *Shore* A hardness values obtained with the narrowed set of predictor variables. It contains the measured and the connected estimated *Shore* A hardness values for each rubber samples. The R^2 value of linear regression was found to be 0.89. As shown in Fig. 7, the trend of the estimated values by the fitted linear model matches well with the trend of the values obtained by the experiments.

5.3. Constrained linear regression

For the coefficients of the linear model, non-negativity constraints were determined in the case of filler, accelerator, cross-linker, and activator components.

The ***lsqin*** MATLAB function has limited options due to that it does not provide information on the significance of the predictor

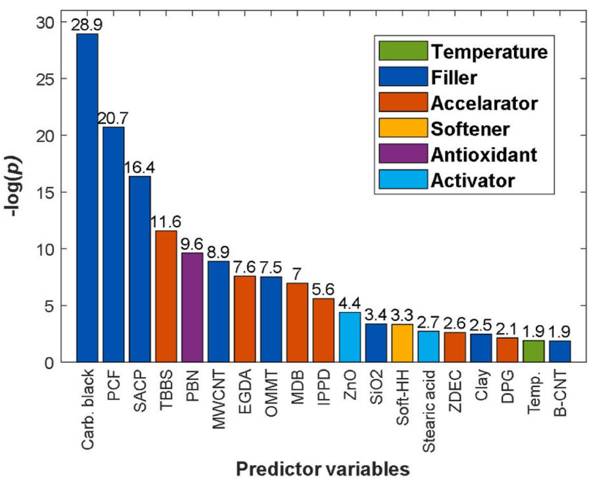

Fig. 6. Significance of predictor variables based on their *p*-values.

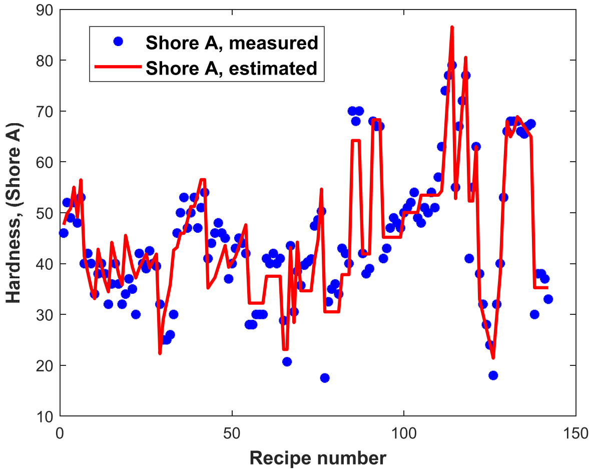

Fig. 7. Measured and estimated *Shore* A hardness values for each rubber sample.

variables (*p* values) and the *Ridge* or *Lasso* technique cannot be applied in parallel either. For this reason, linear regression was performed with normalized variables, and the resulting coefficients were arranged in descending order.

Fig. 8 shows the results of the linear regression obtained by the application of constrained linear regression using all the predictor variables. It can be seen that the values of the coefficients of the filler, accelerator and activator components resulted in positive values. Since normalized variables were used, the values of coefficients can provide some information about the significance of the rubber component. The higher the absolute value of the coefficient, the more significant the component is.

Fig. 9 compares the estimated values obtained with the regression model and the measured hardness values. The value of R^2 is 0.925. as it can be seen the obtained regression model describes the measured hardness values well.

The exhibited results show that the antioxidant and lubricant as softener additive have highly the biggest influence on hardness. This is in line with the expectations as far as the lubricant is concerned; however, the weight of this parameter is higher compared to the others than anticipated. The very high level of significance of antioxidant contradicts to some extent to the expectations and it is also hard to explain the mechanism that leads to the experienced high influence on Shore A hardness. As a result of this regression model carbon and sulfur are also high ranked which agrees with the initial anticipations based on the argumentation above. The role of quartz sand and processing temperature is low similarly to the results from linear regression.

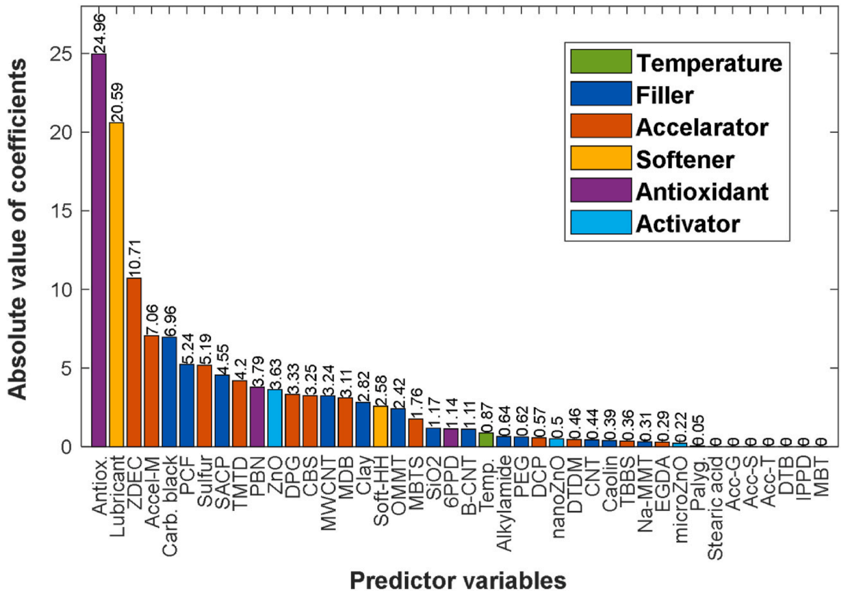

Fig. 8. Values of coefficients of linear model fitted.

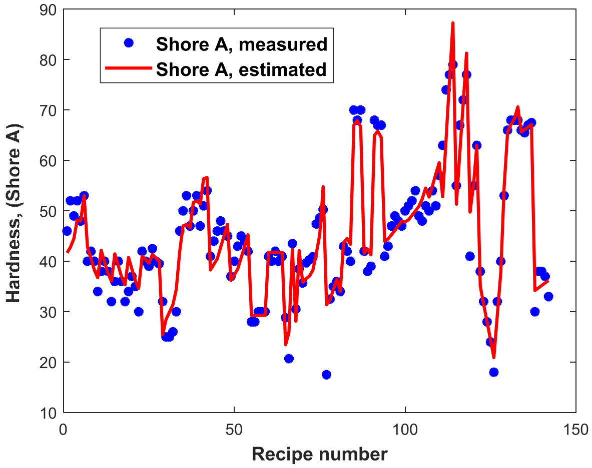

Fig. 9. Estimation results by the linear model obtained by the constrained linear regression.

5.4. Ridge regression

In the application of Ridge regression, first the effect of the shrinkage parameter λ on the coefficients of the fitted linear model was investigated. When $\lambda = 0$, the penalty term has no effect, and ridge regression will produce the classical least square coefficients. However, as λ increases to infinite, the impact of the shrinkage penalty grows, and the ridge regression coefficients will get close zero.

In the following analyses, the normal Ridge regression was investigated with different values of parameter λ shown in Table 2. The entire data set (42 predictor variables, 142 rubber compositions and their corresponding Shore A values) was used in the analyses. As it can be seen from Table 2, increasing the value of parameter λ decreases the quality of the regression since the value of the objective function increases with the parameter λ and the coefficients of the linear model shrink.

Based on the results in Table 2 the value $\lambda = 10$ is chosen for further investigations. In this case, the value of R^2 does not decrease significantly and the lambda parameter λ remains so large that its shrinking effect can be effective. A normalized Ridge regression was performed using the full set of predictor variables, and then the predictor variables were ordered by the fitted coefficients. The results are shown in Fig. 10. It can be concluded that the first three components with the largest absolute value are filler component.

The results of Ridge regression with regards to the key parameters agree well with the outcome of linear regression: carbon black and additional fillers (PCF, SACP) perform well whilst quartz sand and vulcanization temperature turned out to be less-effective parameters. One main difference is the predicted significance of sulfur which is ranked 7th. This outcome is closer to the baseline chemical and mechanical expectations. The high influencing effect of antioxidant is also present (PBN) that agrees with the trends seen from constrained linear regression method. The physical background of this phenomenon is not well established.

Fig. 11 compares the estimated values obtained with the regression model and the measured hardness values. The value of R^2 resulted in 0.921.

5.5. Ridge sparse regression

The *Ridge sparse* regression uses the so-called sparse matrices. The data used in the statistical analyses typically correspond to such data, i.e. many rubber components are available, but only a small number of components are used in the compositions.

In the analyses, the normalized Ridge sparse regression was investigated with different values of parameter λ shown in Table 3. In this case too, the entire data set was used for the analyses. The value $\lambda = 0.0001$ corresponding to the maximum R^2 was considered to

Table 2
R^2 values obtained by Ridge regression with different values of λ.

λ	R^2
0.00001	0.9365
0.0001	0.9365
0.001	0.9365
0.01	0.9365
0.1	0.9364
1	0.9346
10	0.9210
100	0.8460
1000	0.6901
10000	0.6207

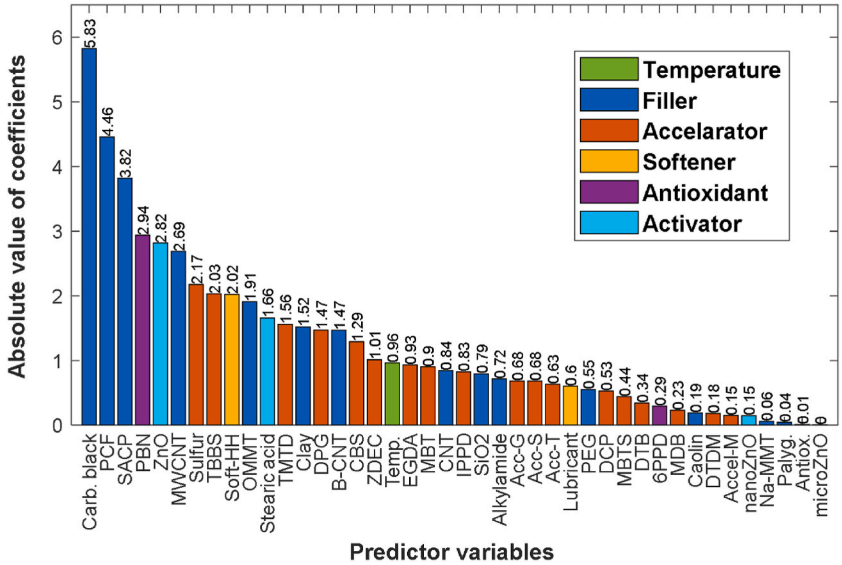

Fig. 10. Values of coefficients of linear model fitted.

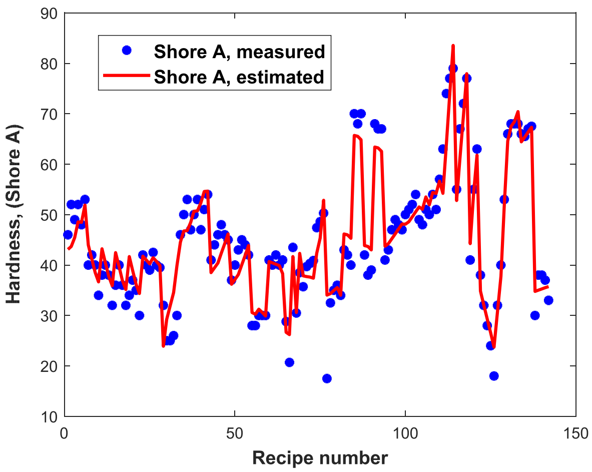

Fig. 11. Estimation result of the linear model obtained by the normalized *Ridge* regression ($\lambda = 10$).

fit the linear model. The sorted predictor variables based on their fitted coefficients are pictured in Fig. 12, while Fig. 13 shows the result of the estimation with the fitted linear model. The value of R^2 was found as 0.9258.

With regards to the obtained order of influence of the predictor parameter one can conclude that the results are close to what was

Table 3
R^2 values obtained by Ridge regression with different values of λ.

λ	R^2
0.00001	0.925770
0.0001	0.925835
0.001	0.925834
0.01	0.925835
0.1	0.925820
1	0.925848
10	0.741079
100	0.702938
1000	0.634481
10000	0.114832

generated with the constrained linear regression: high influence of lubricant and sulfur (ranked on 7th place) whilst SiO_2 and temperature do not play a significant role in determining hardness. What is very unexpected, however, is the ranking of carbon black filler (32nd). This is the only regression method that did not output carbon black being in one of the first 7 places. This contradicts to industrial experience as well which makes the results from this regression model questionable with the λ parameter used. The positive effect of PBN antioxidant on Shore A hardness is apparent based on the results of this regression method as well. Since 4 out of 5 methods do suggest the antioxidant to be in the 1st 5 most influencing factors, this outcome is expected to be rather physical than the artefact of the statistical processing technique.

5.6. *Statistical analysis based on binary classification decision trees*

The result obtained by applying the binary classification decision trees is shown in Fig. 14. The method provides an order of importance between the components.

This analysis resulted the sulfur as being the most influencing chemical contributor. This meets the engineering expectations given that almost all regression methods (apart from the linear) returned Sulfur as being in the 1st quarter of the list. According to this method carbon performed slightly worse than in the other cases (8th place), however, its position still suggests strong influence on hardness. As for the other models SiO_2 underperformed the expectations and exhibits moderate influence on mechanical properties. Two surprising outcomes can be observed in this case: temperature took the 6th place which is far better result than in any other cases and contradicts to the assumption made about time and temperature equivalence during vulcanization. The role of lubricant is hard to forecast since in this case it shows moderate effect on hardness whilst in case of other regression methods applied (such as constrained linear regression and ridge sparse method) it clearly exhibits higher impact on Shore hardness.

5.7. *Comparison of methods*

The previous sections presented the analysis of natural rubber recipes collected from literature using different statistical and linear regression methods. The aim of the analyses was to find the significant components that have a considerable effect on the *Shore* A hardness value of natural rubbers. Based on the literature reviewed, the shore A hardness of compounds typically varies between 30 and 80 depending on the rubber composition.

The application of the different methods typically resulted in different sets of significant variables. Thus, purely based on a preselected method it is hard to draw general conclusions that are equally true for all regression approaches. In general, in light of the expectations given at the beginning of this section the following observations can be made.

- Carbon black and sulfur are significant constituents of the blend before vulcanization that does determine the hardness of the vulcanizate.
- The aforementioned statement is true for a number of other fillers as well (e.g. PCF, SACP, clay etc.).
- In contrast, one of the most typical fillers, quartz sand did not come out as an important parameter from mechanical aspect.
- The majority of regression analyses also also confirm the expectation that lubricant and softening constituents: they directly affect the rubber material's hardness.

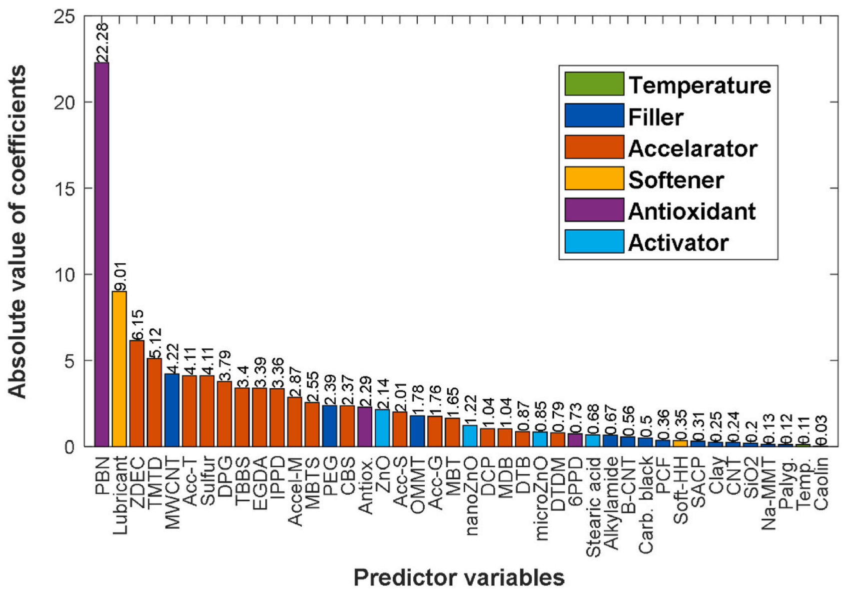

Fig. 12. Values of coefficients of linear model fitted.

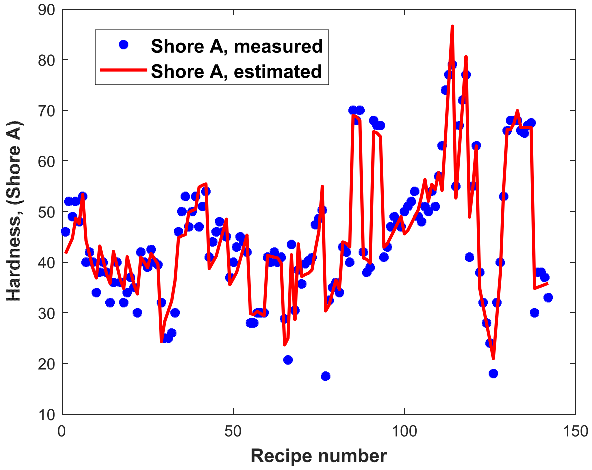

Fig. 13. Estimation result obtained by applying the Ridge sparse regression model.

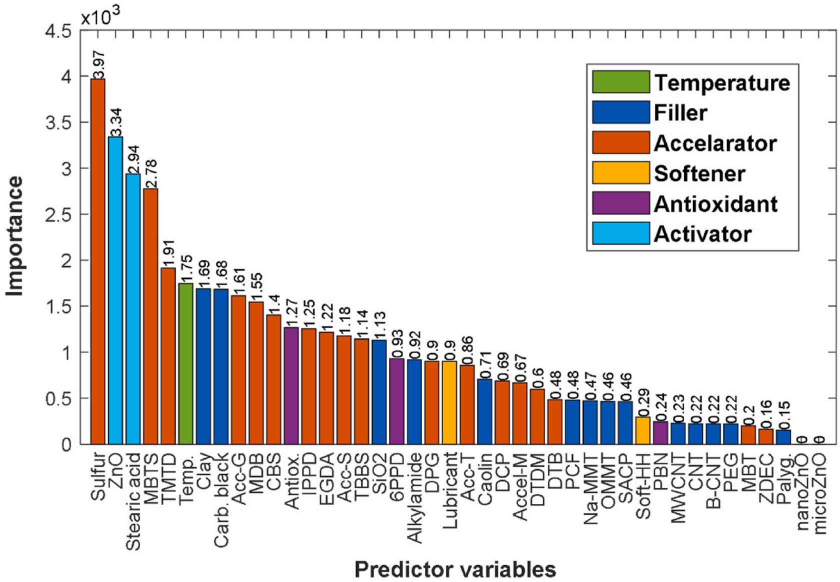

Fig. 14. Order of importance of predictor variables obtained by binary classification decision tree.

- Temperature is not an influencing technology parameter on the mechanical properties of the vulcanized rubber. The likely reason for that is the time-temperature equivalence above the activation temperature. In that temperature domain full curing can take place if enough time is spent with vulcanization. In practice this fully vulcanized state can be inferred from the vulcanization curve where the change of shear resistance is registered as a function of time. It can be assumed that for all analyzed experiments these curves were recorded and at the selected temperature the corresponding dwell time was chosen to ensure the generation of nearly all possible cross-links.
- Antioxidants (mainly PBN) do show an unexpectedly considerable influence on the Shore A hardness level of the vulcanizate. The physical/chemical reason if this is not well established. This is that major finding that would not have been possible to spot without rigorous statistical evaluation and processing the available large number of experimental information with means of data science.
- The sometimes-contradicting outcome of the individual statistical techniques may imply that there are a bunch of parameters that are not independent, thus, there are some coupled terms in that linear model. Therefore the applicability of the assumed underlying linear approximations may need to be revised.

To define an overall set of significant predictor variables for *Shore A,* an average of the results obtained by all the statistical methods was prepared as follows. Scores ranging from 10 to 1 were assigned to the first ten significant components obtained in the case of each method, and then the scores obtained in the case of different methods were summed up for each component. The result of the averaging is shown in Table 4.

6. Conclusions

In this study we elaborated a data-driven technique to predict mechanical properties of a rubber vulcanizate based on the type and amount of the main ingredients. For this, large number of relevant publications were processed with known rubber recipe and post-vulcanization mechanical properties. With the aid of this technique, design rules can be established that connect the chemical composition of rubber with physical and mechanical properties. These design rules can support and accelerate the preliminary design process of any rubber components subjected to significant mechanical load since there is no need to obtain and manufacture samples for detailed mechanical testing.

Based on the literature data, various statistical analyses, like *linear regression*, *constrained linear regression, Ridge* regression, *Ridge sparse* regression and *binary classification decision trees* were performed to determine which rubber components have the most significant effect on the hardness. Statistical tests related to the regression of predictor variables and response variable are suitable for establishing the significance of the rubber constituents on the hardness. Regression methods can be used to establish the order of the components and offers the opportunity to select significant predictor variables.

The application of the different methods typically resulted in different sets of significant variables. In order to define an overall set of significant predictor variables for Shore A we prepared an average of the results obtained by all the statistical methods. The result of the statistical method shows good agreement with the industrial practice, namely that the widely applied components (e.g. carbon black) has the greatest influence on the Shore A harness, while the less frequently used components have only minor effect.

The statistical evaluation method introduced in this paper is novel for this kind of application. It can be used as an independent

Table 4
Scoring of natural rubber (NR) components based on the results of statistical and regression methods.

	Predictor variable name	Component type	Linear regression *t*-test	Constrained linear regression	Ridge regression	Ridge sparse regression	Binary decision tree	Sum
1	Carbblack	filler	10	6	10		3	29
2	PBN	antioxidant	6	1	7	10		24
3	PCF	filler	9	5	9			23
4	Sulfur	accelerator		4	4	4	10	22
5	SACP	filler	8	3	8			19
6	Lubricant	softener		9		9		18
7	MWCNT	filler	5		5	6		16
8	ZDEC	accelerator		8		8		16
9	TMTD	accelerator		2		7	6	15
10	Zinc-oxide	activator			6		9	15
11	TBBS	accelerator	7		3	2		12
12	Antiox	antioxidant		10				10
13	Stearicacid	activator					8	8
14	MBTS	accelerator					7	7
15	Accel-M	accelerator		7				7
16	Temp	operational parameter					5	5
17	EGDA	accelerator	4			1		5
18	Acc-T	accelerator				5		5
19	Clay	filler					4	4
20	OMMT	filler	3		1			4
21	MDB	accelerator	2				1	3
22	DPG	accelerator				3		3
23	Acc-G	accelerator					2	2
24	Soft-HH	softener			2			2
25	IPPD	accelerator	1					1
26	Caolin	filler						0
27	CNT	filler						0
28	B-CNT	filler						0
29	Na-MMT	filler						0
30	SiO2	filler						0
31	Alkylamide	filler						0
32	Palyg	filler						0
33	PEG	filler						0
34	CBS	accelerator						0
35	DTB	accelerator						0
36	DTDM	accelerator						0
37	DCP	accelerator						0
38	MBT	accelerator						0
39	Acc-S	accelerator						0
40	6PPD	antioxidant						0
41	nanoZnO	activator						0
42	microZnO	activator						0

verification technique of the industrial experience as far as the role of the individual constituents is concerned when it comes to making a blend for rubber manufacturing. The main assets of the derived regression models are as follows.

- Assists in engineering new rubber materials with mechanical properties tailored for purpose.
- The proposed regression techniques offers a data-driven approach for material engineering instead of the trial and error based approach. This method can further be combined with Artificial Intelligence based methods where the training dataset can be defined e.g. by the presented regression techniques. This strategy could be applied for the automatic detection the necessary constituents and their amount to match a series of criteria set against the mechanical behavior of the target vulcanizate.
- In the case of existing rubber materials with known constituents and amount the proposed regression models help predicting those mechanical properties of the vulcanized rubber that are not available on data sheets.
- Typically, such missing mechanical properties for rubbers are the parameters of hyperelastic constitutive models that are used in Finite Element Analyses of rubber parts. The presented regression techniques, can also be applied to support the hyperelastic parameter identification process, namely by providing initial values for the material parameters or constraints between them. These information can significantly improve the accuracy of the fitted material model, and therefore, increasing fidelity in virtual product development supported by numerical simulation.

The industrial experience about strengthening constituents of rubbers was well demonstrated and as such, validated with our analyses. The main conclusions in this regard are as follows.

- Fillers are essential ingredients from hardness aspect, especially carbon black.
- Thus, as a first but most effective approach the carbon content can be varied in the initial blend if the hardness of the vulcanizate needs changing.
- Quartz sand (SiO_2) does not contribute to hardness.
- Lubricants and softeners not only help dispersing and mixing up all constituents better but can directly soften the vulcanized rubber as well.
- Sulfur is a key contributor as well since the cross-links are formed from that substance (for sulfur-based rubbers).
- It is not advised varying processing temperature to achieve hardness change in a certain blend. A well-defined vulcanization curve is suggested to record instead that ensures full cure of the material by maximizing the cross-link density.
- The very high influence of antioxidants (almost all statistical methods predicted so) on the hardness of the final vulcanizate is unexpected. The chemical/physical background of this hardening mechanism, introduced by antioxidants needs further investigation in the future.

Although the conclusions made here do not contradict with industrial experience and the number of recipes feeding the regression models is deemed to be representative, further examination of representativeness is suggested by applying the generated models on other experiments (blends) not included in model derivation and compare the measured and predicted Shore hardness values.

Declarations

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors confirm that the work described has not been published before; it is not under consideration for publication anywhere else. All authors have approved the manuscript and agree with its submission to this journal.

Data availability statement

Data related to the investigation were not deposited in a publicly available repository. Data will be made available on request.

CRediT authorship contribution statement

Lilla Virág: Writing – original draft, Methodology, Conceptualization. **Attila Egedy:** Writing – review & editing, Supervision, Funding acquisition. **Csilla Varga:** Validation, Methodology, Investigation. **Gergely Erdős:** Project administration, Funding acquisition, Conceptualization. **Szabolcs Berezvai:** Writing – review & editing, Supervision, Conceptualization. **László Kovács:** Writing – review & editing, Supervision, Funding acquisition. **Zsolt Ulbert:** Writing – review & editing, Software, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. Acknowledgement

This research was supported by the National Research, Development and Innovation Office of Hungary under the project „Precision characterization of nonlinear mechanical behavior of reinforced and unreinforced polymer materials for engineering simulations" (project code: 2020–1.1.2-PIACI-KFI-2021-00314). Szabolcs Berezvai has been supported by the Hungarian National Research, Development and Innovation Office (NKFIH-PD-137806) and by János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Appendix

Linear regression and t-test

Among the statistical functions of the MATLAB software package, there are many built-in linear regression functions that can perform the linear regression and determine the model coefficients, and in addition, calculate *t*-statistic test for each model coefficients of the linear model and thus can estimate the significance of the predictor variables. In the case of *t*-tests, the null hypothesis is that the value of the coefficients of the linear model is 0, i.e. the parameter is not significant. If we examine the result of *t*-test (*p* value) returned by the linear regression functions and find that its value is close to 1, then, the value of the given coefficient is most likely zero, on the other hand, the smaller the value of the *t*-test, the more significant predictor variable for the response value.

In our statistical analyses, we used the ***regstats*** function of MATLAB, which has the option to return the *p* values of the *t*-tests for the coefficients of the linear model. In the data analysis, we use the *p* values to decrease the number of the predictor variables in such a way that we started from an empty set of predictor variables, and in each iteration step, the elements of the set are extended with another predictor variable (rubber component). For each of the variables that are not yet in the set, we check that if we insert them into the predictor variable set obtained in the previous step and perform a linear regression with the new set, then what value of *p* we get for the coefficient of the new variable. Then, in the basic case, the variable whose coefficient had the smallest *p* value is included in the set, i.e. it is considered the most significant.

The iteration is continued until we found a predictor variable whose coefficient *p*-value was not greater than 0.05 (the chosen significance level). If we only get test values greater than 0.05, i.e. we cannot add more significant predictor variables to the set, the iteration continues by examining the *p*-values for the coefficients of the predictor variables in the set and selecting the largest among those whose value is greater than 0.1. The corresponding predictor variable is deleted from the set and at the same time from the linear model. The iteration ends when we cannot add another variable to the set, or we cannot remove a variable from the set. During the iteration, the *p* value of the coefficients of the variables already in the set can change due to the effect of the variables entering the set.

Constrained linear regression

Rubber components can be classified into several component groups. In most cases, the values of the mechanical properties of rubber (hardness, modulus, tensile strength) show an increasing trend in terms of the amount of individual components. These include e.g. fillers, accelerators, cross-linkers, and activators.

In the statistical methods discussed above we allowed the coefficients of the individual component variables to take on negative values. In the case of constrained linear regression, limits can be defined for the model coefficients to be determined. The constrained linear regression was applied using the ***lsqin*** function of MATLAB. In the analyses, we specified a non-negativity criterion for the model coefficients belonging to the filler, accelerators, cross-linkers, and activator components.

The ***lsqin*** function has limited options because it does not provide information on the significance of the predictor variables (*p* values) and the *Ridge* or *Lasso* technique cannot be applied in parallel either. In order to somehow rank the predictor variables, they were normalized, and then the coefficients obtained in this way were sorted. Normalization was performed based on the expected value (μ) and standard deviation (σ) of the predictor variables:

We replaced each variable value with the normalized value $z_j = \frac{x_j - \mu_j}{\sigma_j}$, which gives the matrix $\boldsymbol{Z}$ of the normalized values of predictor variables. In this case, the regression estimates the coefficients $\beta_1^{scaled}, \ldots, \beta_1^{scaled}$ based on the following linear model.

$$y - \mu_y = \beta_1^{scaled} z_1 + \beta_2^{scaled} z_2 + \ldots + \beta_p^{scaled} z_p + \varepsilon \tag{4}$$

If we perform the regression with normalized variables $z_j = \frac{x_j - \mu_j}{\sigma_j}$, the coefficients of the linear model written with non-normalized variables can be obtained with the following relations.

$$\beta_0 = \mu_y - \sum_{j=1}^{p} \frac{\beta_j^{scaled} \mu_j}{\sigma_j} \tag{5}$$

$$\beta_j = \frac{\beta_j^{scaled}}{\sigma_j}, \swarrow j = 1, \ldots, p \tag{6}$$

The regression was performed with the normalized variables, and the resulting coefficients were arranged in descending order.

Ridge regression

The objective of statistical analyses is to determine an accurate linear model between the rubber components and hardness. To get a suitable linear model it is necessary to find the predictor variables that have the most significant effect on the hardness of the rubber. The so-called shrinkage methods (e.g. *Lasso*, *Ridge*) can help to decrease the number of predictor variables used. Their application ensures not only the decrease of the number of predictor variables but also provides model coefficients with a smaller standard deviation which are closer to each other and closer to zero.

Ridge regression uses equation (4) to calculate the regression coefficients by modifying equation (3).

$$\widehat{\beta}^{Ridge} = \left(X^T X + \lambda I\right)^{-1} X^T y \tag{7}$$

where λ is the Ridge parameter and $\boldsymbol{I}$ is a unit matrix. In the case of small values, the conditionality of the task improves, and the standard deviation of the estimated coefficients decreases, i.e. their values shrink. The shrinking effect of the parameter λ is also clearly visible in the case of the objective function used in *Ridge* regression, in which the last term is called the penalty term:

$$\widehat{\beta}^{Ridge} = \min_{\beta}\left\{ \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 \right\} \tag{8}$$

Since we are looking for a minimum, the larger the value of the parameter, the closer the coefficients are to each other and to zero. The solutions of *Ridge* regressions are not independent of the scaling of the predictor variables, it is worth to normalize them before calculating the coefficients. The normalization results in coefficients with a smaller range of values, and the decrease of the number of predictor variables can be based on coefficient values with a smaller standard deviation and closer to each other.

The application of the Ridge method was implemented in the MATLAB environment by using the MATLAB ***ridge*** function in form of ***ridge*(y,X,*scaled*)**. The value of the ***scaled*** parameter can be either 0 or 1. In processing the experimental data, we use the **ridge** function with the ***scaled*** = 1 parameter which entails the internal normalization of the variables.

When decreasing the number of the predictor variables, we first remove the variables with the smallest absolute value of the coefficient.

Ridge sparse regression

The *Ridge sparse* regression method is also one of the shrinkage methods available in the MATLAB environment. It is similar to the normal Ridge method presented above, which ensures the decrease of the number of the predictor variables so that we get model coefficients with a smaller standard deviation and closer to each other and zero.

The *Ridge sparse* regression uses so-called sparse matrices. The data used in our statistical analyses typically correspond to such data, i.e. many rubber components are available, but only a small number of components are used in the compositions. In the case of the normal Ridge method, where a component is not present in a composition, the cell of the component in the data table was filled with a zero value. However, the representation of the sparse matrix is not matrix-like MATLAB actually stores only those cell data where data is available.

The Ridge sparse method was used with the ***fitrlinear*** function of MATLAB. The use of the ***fitrlinear*** function, in addition to the use of the sparse matrix and the shrinking of the coefficient values, enables the use of cross-validation. Applying cross-validation, we use a part of the available data only for regression of the linear model and use the remaining part of the data set to test the fitted linear model. During cross-validation, the range of data used for regression and testing changes continuously and covers the entire data set.

Statistical analysis based on binary classification decision trees

In determining the significant predictor variables, we also used an analysis method that includes the application of binary classification decision trees. In these methods, a binary classification decision tree is grown based on the data containing the values of the predictor and the response variables. Each internal node of the decision tree represents a predictor variable, while its leaves represent response variable values. A binary decision tree starts with a node which branches in two directions. Each internal node can branch in two directions by moving down the binary decision tree. The interval of the predictor variable is split into two parts for each internal node. The direction in which we can proceed on the node is determined by the split interval which includes the value of the predictor variable. Following the selected branch we can reach either a new internal node or a leaf that represents a specific value of the response variable. In this way, we get a path that represents a set of predictor variables and their intervals, if applied, we can get a classification value of the response variable.

The binary decision tree was applied using the ***fitctree*** function of MATLAB. We used it with default parameters to grow a classification decision tree. The ***fitctree*** function allows to specify a maximum value for the number of internal nodes, but in this case the error of the model may increase.

Based on the classification decision tree grow we can applied the ***predictorImportance*** function of MATLAB to estimate the importance values of predictor variables for the tree. This function sums the mean squared error due to splits on each predictor and divides the sum by the number of branch nodes.

References

[1] T.R. Vijayaram, A technical review on rubber, Int. J. Des. Manuf. Technol. 3 (1) (2009) 25–37, https://doi.org/10.18000/ijodam.70043.
[2] J.G. Sommer, Engineered Rubber Products: Introduction to Design, Manufacture and Testing, Hanser Gardner Pubns., 2009.
[3] A.V. Kaliyathan, K.M. Varghese, A.S. Nair, S. Thomas, Rubber–rubber blends: a critical review, Prog. Rubber Plast. Recycl. Technol. 36 (3) (2020) 196–242, https://doi.org/10.1177/1477760619895002.
[4] A. bin Samsuri, Degradation of natural rubber and synthetic elastomers, in: J.A. Richardson, et al. (Eds.), Shreir's Corrosion, vol. 3, Elsevier, 2010, pp. 2407–2438, https://doi.org/10.1016/B978-044452787-5.00117-7, 2010.
[5] K. Sisanth, M. Thomas, J. Abraham, S. Thomas, General introduction to rubber, in: Thomas, H.J. Maria (Eds.), Progress in Rubber Nanocomposites, Elsevier, 2017, pp. 1–39, https://doi.org/10.1016/B978-0-08-100409-8.00001-2.
[6] F. Ignatz-Hoover, B.H. To, Vulcanization, in: B. Rodgers (Ed.), Rubber Compounding Chemistry and Applications, CRC Press, Taylor & Francis Group, 2016, pp. 461–522, https://doi.org/10.1201/b18931.
[7] F.A. Yudha, B.O. Soepangkat, V. Ratnasari, M. Suef, R. Norcahyo, Determination of optimum vulcanization process parameters using Taguchi GRA for reducing quality Loss cost, AIP Conf. Proc. 2114 (2019) 020014, https://doi.org/10.1063/1.5112398.
[8] C.S. Barrera, A.B. Soboyejo, K. Cornish, Quantification of the contribution of filler characteristics to natural rubber reinforcement using principal component analysis, Rubber Chem. Technol. 91 (1) (2018) 79–96, https://doi.org/10.5254/rct.82.83716.
[9] Z.B. Yıldırım, M. Karacasu, Modelling of waste rubber and glass fibber with response surface method in hot mix asphalt, Construct. Build. Mater. 227 (2019) 117070, https://doi.org/10.1016/j.conbuildmat.2019.117070.
[10] C. Lewis, R. Buanpa, S. Kiatkamjornwong, Effect of rubber ratio, carbon black level, and accelerator level on natural rubber/bromobutyl rubber blend properties, J. Appl. Polym. Sci. 90 (11) (2003) 3059–3068, https://doi.org/10.1002/app.13036.
[11] V.V. Krivtsov, D.E. Tananko, T.P. Davis, Regression approach to tire reliability analysis, Reliab. Eng. Syst. Saf. 78 (3) (2002) 267–273, https://doi.org/10.1016/S0951-8320(02)00169-2.
[12] William Amoako Kyei-Manu, Charles R. Herd, Mahatab Chowdhury, James J.C. Busfield, Lewis B. Tunnicliffe, The influence of Colloidal properties of carbon black on static and dynamic mechanical properties of natural rubber, Polymers 14 (6) (2022) 1194, https://doi.org/10.3390/polym14061194.
[13] M. Destrade, G. Saccomandi, I. Sgura, Methodical fitting for mathematical models of rubber-like materials, Proc. R. Soc. A A. 473 (2017) 20160811, https://doi.org/10.1098/rspa.2016.0811.
[14] Fa Zhang, Shuxun Cui, A two-component statistical model for natural rubber, Polymers 242 (2022) 124462, https://doi.org/10.1016/j.polymer.2021.124462.
[15] S. Srewaradachpisal, C. Dechwayukul, S. Chatpun, R.J. Spontak, W. Thongruang, Optimization of the rubber formulation for Footwear applications from the response surface method, Polymers 12 (9) (2020) 2032, https://doi.org/10.3390/polym12092032.
[16] Allen Jonathan Román, Shiyi Qin, Julio C. Rodríguez, Leonardo D. González, Victor M. Zavala, Tim A. Osswald, Natural rubber blend optimization via data-driven modeling: the implementation for reverse engineering, Polymers 14 (2022) 2262, https://doi.org/10.3390/polym14112262.
[17] A. Basfar, M. Abdel-Aziz, S. Mofti, Influence of different curingsystems on the physico-mechanical properties and stability of SBR and NR rubbers, Radiat. Phys. Chem. 63 (2002) 81–87, https://doi.org/10.1016/S0969-806X(01)00486-8.
[18] J. Kruželák, R. Sýkora, I. Hudec, Sulfur and peroxide curing of Rubber Compounds based on NR and NBR. Part I: cross-linking and physical-mechanical Properties, KGK rubberpoint 1–2 (2017) 27–33, https://doi.org/10.1515/polyeng-2014-0034.
[19] M. Mansilla, A. Marzocca, C. Macchi, A. Somoza, Influence of vulcanization temperature on the cure kinetics and on the microstructural properties in natural rubber/styrene-butadiene rubber blends prepared by solution mixing, Eur. Polym. J. 69 (2015) 50–61, https://doi.org/10.1016/j.eurpolymj.2015.05.025.
[20] C. Nah, S.G. Kim, G.S. Shibulal, Y.H. Yoo, B. Mensah, B.H. Jeong, Effects of curing systems on the mechanical and chemical ageing resistance properties of gasket compounds based on ethylene-propylene-diene-termonomer rubber in a simulated fuel cell environment, Int. J. Hydrogen Energy 40 (2015) 10627–10635, https://doi.org/10.1016/j.ijhydene.2015.07.003.
[21] H. Nabil, H. Ismail, A.R. Azura, Properties of natural rubber/recycled ethylene–propylene–diene rubber blends prepared using various vulcanizing systems, Iran. Polym. J. (Engl. Ed.) 23 (2014) 37–45, https://doi.org/10.1007/s13726-013-0197-4.
[22] H.E. Mayasari, A. Yuniari, Effect of vulcanization system and carbon black on mechanical and swelling properties of EPDM blends, Majalah Kulit Karet dan Plastik 32 (1) (2016) 59–64, https://doi.org/10.20543/mkkp.v32i1.706.
[23] K. Formela, D. Wąsowicz, M. Formela, A. Hejna, J. Haponiuk, Curing characteristics, mechanical and thermal properties of reclaimed ground tire rubber cured with various vulcanizing systems, Iran. Polym. J. (Engl. Ed.) 24 (4) (2015) 289–297, https://doi.org/10.1007/s13726-015-0320-9.
[24] H. Nabil, H. Ismail, A.R. Azura, Optimization of accelerators on curing characteristics, tensile, and dynamic mechanical properties of (natural rubber)/(recycled ethylene-propylene-diene-monomer) blends, J. Vinyl Addit. Technol. 21 (2) (2015) 79–88, https://doi.org/10.1002/vnl.21358.
[25] D.Y. Kim, J.W. Park, D.Y. Lee, K.H. Seo, Correlation between the crosslink characteristics and mechanical properties of natural rubber compound via accelerators and reinforcement, Polymers 12 (9) (2020), https://doi.org/10.3390/polym12092020.
[26] R. Mukhopadhyay, S.K. De, S.N. Chakraborty, Effect of vulcanization temperature and vulcanization systems on the structure and properties of natural rubber vulcanizates, Polymer 18 (12) (1977) 1243–1249, https://doi.org/10.1016/0032-3861(77)90287-7.
[27] R.L. Fan, Y. Zhang, F. Li, Y.X. Zhang, K. Sun, Y.Z. Fan, Effect of high-temperature curing on the crosslink structures and dynamic mechanical properties of gum and N330-filled natural rubber vulcanizates, Polym. Test. 20 (2001) 925–936, https://doi.org/10.1016/S0142-9418(01)00035-6.
[28] R. Ismail, Z.A. Mahadi, I.S. Ishak, The effect of carbon black filler to the mechanical properties of natural rubber az a base isolation system, IOP Conf. Series: Earth and Environmental Science. 140 (2018) 012133, https://doi.org/10.1088/1755-1315/140/1/012133.
[29] H. Tan, A.I. Isayev, Comparative study of silica-, nanoclayand carbon black-filled EPDM rubbers, J. Appl. Polym. Sci. 109 (2008) 767–774, https://doi.org/10.1002/app.28130.
[30] S.J. Ahmadi, Y. Huang, W. Li, Fabrication and physical properties of EPDM–organoclay nanocomposites, Compos. Sci. Technol. 65 (2005) 1069–1076, https://doi.org/10.1016/j.compscitech.2004.10.028.
[31] S. Chuayjljit, A. Imvittaya, N. Na-Ranong, P. Potiyaraj, Effects of particle size and amount of carbon black and calcium carbonate on curing characteristics and dynamic mechanical properties of natural rubber, Journal of Metals. Materials and Minerals 12 (2002) 51–57.
[32] A. Hussain, I. Abadir, S.E. Marsafy, Effect of inorganic additives on some properties of NBR vulcanizates, Pigment Resin Technol. 39 (2010) 84–88, https://doi.org/10.1108/03699421011028662.
[33] K. Prukkaewkanjana, S. Thanawan, T. Amornsakchai, High performance hybrid reinforcement of nitrile rubber using short pineapple leaf fiber and carbon black, Polym. Test. 45 (2015) 76–82, https://doi.org/10.1016/j.polymertesting.2015.05.004.
[34] M.H. Al-Maamori, A.A. Ali, A.A. Subeh, Effect of carbon black on mechanical and physical properties of acrylonitrile butadiene rubber (NBR) composite, Acad. Res. Int. 6 (2) (2015) 28–37.
[35] N. Rattanasom, S. Prasertsri, Relationship among mechanical properties, heat ageing resistance, cut growth behaviour and morphology in natural rubber: partial replacement of clay with various types of carbon black at similar hardness level, Polym. Test. 28 (2009) 270–276, https://doi.org/10.1016/j.polymertesting.2008.12.010.
[36] C. Bergmann, J. Trimbach, Influence of plasticizers on the properties of natural rubber compounds, KGK rubberpoint 7–8 (2014) 40–49.
[37] M.M. Rahman, K. Oßwald, K. Reincke, B. Langer, Influence of bio-based plasticizers on the properties of NBR materials, Materials 13 (9) (2020) 2095, https://doi.org/10.3390/ma13092095.
[38] H.G. Moneypenny, K.H. Menting, F.M. Gragg, General compounding, in: B. Rodgers (Ed.), Rubber Compounding Chemistry and Applications, CRC Press, Taylor & Francis Group, 2016, pp. 333–378, https://doi.org/10.1201/9781420030464.
[39] L. Lu, K. Luo, W. Yang, S. Zhang, W. Wang, H. Xu, S. Wu, Insight into the anti-aging mechanisms of natural phenolic antioxidants in natural rubber composites using a screening strategy based on molecular simulation, m RSC Advances 10 (2020) 21318–21327, https://doi.org/10.1039/D0RA03425H.

[40] J. Nie, X. Huang, C. Xu, J. Ding, Y. Chen, Antioxidant effects on curing/processing and thermo-oxidative aging of filled nitrile rubber, Mater. Chem. Phys. 253 (2020) 123403, https://doi.org/10.1016/j.matchemphys.2020.123403.

[41] N. Rattanasom, S. Prasertsri, T. Ruangritnumchai, Comparison of the mechanical properties at similar hardness level of natural rubber filled with various reinforcing-fillers, Polym. Test. 28 (1) (2009) 8–12, https://doi.org/10.1016/j.polymertesting.2008.08.004.

[42] K.C. Lee, N.A. Yosuff, N. Othman, N.A. Mohamad Aini, Effect of vulcanization temperature on curing characteristic, physical and mechanical properties of natural rubber/palygorskite composites, IOP Conf. Ser. Mater. Sci. Eng. 223 (2017) 012017, https://doi.org/10.1088/1757-899X/223/1/012017.

[43] R. Mukhopadhyay, A.K. Bhowmick, S.K. De, Effect of vulcanization temperature and synergism of accelerators on the network and technical properties of efficiently vulcanized natural rubber mixes, Polymer 19 (10) (1978) 1176–1180, https://doi.org/10.1016/0032-3861(78)90067-8.

[44] A.S. Aprem, K. Joseph, T. Mathew, V. Altstaedt, S. Thomas, Studies on accelerated sulphur vulcanization of natural rubber using 1-phenyl-2, 4-dithiobiuret/tertiary butyl benzothiazole sulphenamide, Eur. Polym. J. 39 (7) (2003) 1451–1460, https://doi.org/10.1016/S0014-3057(02)00382-8.

[45] M. Arroyo, M.A. López-Manchado, B. Herrero, Organo-montmorillonite as substitute of carbon black in natural rubber compounds, Polymer 44 (8) (2003) 2447–2453, https://doi.org/10.1016/S0032-3861(03)00090-9.

[46] N. Tabsan, S. Wirasate, K. Suchiva, Abrasion behavior of layered silicate reinforced natural rubber, Wear 269 (5–6) (2010) 394–404, https://doi.org/10.1016/j.wear.2010.04.018.

[47] Y.H. Lee, M. Cho, J.-D. Nam, Y. Lee, Effect of ZnO particle sizes on thermal aging behavior of natural rubber vulcanizates, Polimer Degradation and Stability 148 (2018) 50–55, https://doi.org/10.1016/j.polymdegradstab.2018.01.004.

[48] I. Surya, H. Ismail, A.R. Azura, Alkanolamide as an accelerator, filler-dispersant and a plasticizer in silica-filled natural rubber compounds, Polym. Test. 32 (8) (2013) 1313–1321, https://doi.org/10.1016/j.polymertesting.2013.07.015.

[49] I. Surya, H. Ismail, A.R. Azura, The effect of alkanolamide loading on properties of carbon black-filled natural rubber (SMR-L), epoxidised natural rubber (ENR), and styrene-butadiene rubber (SBR) compounds, Polym. Test. 42 (2015) 208–214, https://doi.org/10.1016/j.polymertesting.2014.12.009.

[50] I. Surya, H. Ismail, The effect of the addition of alkanolamide on properties of carbon black-filled natural rubber (SMR-L) compounds cured using various curing systems, Polym. Test. 50 (2016) 276–282, https://doi.org/10.1016/j.polymertesting.2016.01.014.

[51] W. Zhang, X. Zhang, M. Liang, C. Lu, Mechanochemical preparation of surface-acetylated cellulose powder to enhance mechanical properties of cellulose-filler-reinforced NR vulcanizates, Compos. Sci. Technol. 68 (12) (2008) 2479–2484, https://doi.org/10.1016/j.compscitech.2008.05.005.

[52] G. Sui, W.H. Zhong, X.P. Yang, Y.H. Yu, Curing kinetics and mechanical behavior of natural rubber reinforced with pretreated carbon nanotubes, Material Science and Engineering: A. 485 (1–2) (2008) 524–531, https://doi.org/10.1016/j.msea.2007.09.007.

[53] S.S. Banerjee, I. Arief, R. Berthold, M. Wiese, M. Bartholdt, D. Ganguli, S. Mitra, S. Mandal, J. Wallaschek, A. Raatz, G. Heinrich, A. Das, Super-elastic ultrasoft natural rubber-based piezoresistive sensors for active sensing interface embedded on soft robotic actuator, Appl. Mater. Today 25 (2021) 101219, https://doi.org/10.1016/j.apmt.2021.101219.

[54] N. Rattanasom, S. Prasertsri, Mechanical properties, gas permeability and cut growth behaviour of natural rubber vulcanizates: influence of clay types and clay/carbon black ratios, Polym. Test. 31 (5) (2012) 645–653, https://doi.org/10.1016/j.polymertesting.2012.04.001.

[55] W. Wu, L. Tian, Formulation and morphology of kaolin-filled rubber composites, Appl. Clay Sci. 80–81 (2013) 93–97, https://doi.org/10.1016/j.clay.2013.06.025.