

COMPARATIVE METHODOLOGY OF NON-LINEAR MODELS FOR PREDICTING RHEOLOGICAL PROPERTIES OF RUBBER MIXTURES IN INDUSTRIAL LINES

Urraca Valle, R.; Sodupe Ortega, E.; Antoñanzas Torres, J.; Alonso García, E.;
Sanz García, A.; Martínez de Pisón Ascacibar, F. J.

Universidad de La Rioja

Data mining and statistics are applied to predict certain of the properties of rubber-extruded mixtures. These properties are associated to their cure curves using data from the mixing phase at the beginning of the process. The main goal is to automatically obtain the model that provides operators accurate set points to control mixing process. The operators would be able to anticipate possible failures in the quality of vulcanized rubber mixture.

There are several strategies to develop optimum models. This work proposes the following methodology to optimize the information extraction from the available data. First, an initial analysis of database attributes is performed seeking for significant information to future model derivation. Second, a wide comparison of different non-parametric methods is carried out to determine which one is the most appropriate. Instead of directly contrasting prediction errors, an automatic statistical system of comparison is included by using several non-parametric techniques. Third, some alternative strategies are tested taking advantage of the specific attributes of the database.

Keywords: *Data mining; Rubber mixture process; Prediction models; Principal component analysis; Rheological properties*

METODOLOGÍA COMPARATIVA DE MODELOS NO LINEALES PARA LA PREDICCIÓN DE LAS PROPIEDADES REOLÓGICAS DE MEZCLAS DE GOMA IN LINEAS INDUSTRIALES

La minería de datos y la estadística son aplicadas para predecir ciertas propiedades de las mezclas de goma extruida. Estas propiedades están asociadas a sus curvas de curado utilizando datos procedentes de la fase de mezclado al comienzo del proceso. El principal objetivo es obtener automáticamente un modelo que proporcione a los trabajadores puntos de consigna precisos para controlar el proceso de mezcla. Los trabajadores serían capaces de anticiparse a posibles fallos en la calidad de la mezcla de goma vulcanizada.

Existen varias estrategias para desarrollar modelos óptimos. En este trabajo se propone una metodología para optimizar la extracción de información de los datos disponibles. En primer lugar se analizan las características de la base de datos buscando información útil para el entrenamiento de modelos. En segundo lugar, se lleva a cabo una amplia comparativa de diferentes modelos no paramétricos para determinar cuál el más apropiado. En vez de comparar directamente los errores de predicción, se incluye un sistema automático de comparación basado en el uso de varias técnicas no paramétricas. Por último, se analizan otras estrategias alternativas que tratan de aprovechar la información obtenida en la fase de análisis de las variables.

Palabras clave: *Minería de datos; Proceso de mezclado de gomas; Análisis de las componentes principales; Propiedades reológicas*

Correspondencia: enrique.sodupeo@unirioja.es

1 INTRODUCTION AND OBJECTIVES

Rubber extrusion industry requires high quality rubber mixing processes to continue being competitive producing more complex profiles. This is due to the important quality requirements usually demanded by its main client, the automotive industry. In recent decades, many companies have been aware of this need and they have increased their investments for improving the control of extrusion lines, the quality of raw materials, etc.

Much of the actual research focuses on developing model-driven methods to control the extrusion process (Ha, Y.S. et al., 2008, Díaz, J.J. et al., 2008). This is a challenging task due to the rubber behavior as a non-Newtonian fluid at extrusion temperatures and also the intricate geometries of the car profiles required by automotive industry. However, commercial solvers are gradually including new upgrades to deal with these drawbacks, easing the development of more realistic numerical models. As a result, not only die design process can be optimized, but also extrusion parameters can be estimated to provide direct online information about the process. Besides, a future trend is the implementation of semiautomatic extrusion lines that follow a group of predefined extrusion set points.

Prior to the extrusion phase, and usually underestimated by researchers, the mixing phase of the rubber compounds is performed. Efforts to control the extrusion process alone are useless, whether the variance of raw material properties is excessively high. Therefore, every extrusion company must establish a tight quality control on the mixing phase before starting extrusion process. The modeling of the mixing phase using data-driven methods has already proved to yield optimum results without being as complex as numerical simulation. Moreover, it takes advantages of the large amount of data collected from processes by companies (Ordieres, J. et al., 2003, Martínez de Pisón, F.J. et al., 2008, Martínez de Pisón, F. Javier et al., 2010). In this context, data mining (DM) appears as a proper solution to extract useful information from the existing databases (DBs), specifically nonparametric models. While nowadays the process of analyzing the mixture quality in the laboratory takes too much time, a correctly tuned nonparametric model can monitor mixture increasing the decision-making ability of the worker. The rheology curve is the method selected to measure the final quality of the mixture obtained.

The article focuses on the development of nonparametric models to obtain the most accurate predictions of the mixture properties. Despite of their lack of the interpretability compared to traditional linear regression models, nonparametric models are more flexible to the unknown nature of numerical regression problems. Furthermore, due to growing computing capacity and new advances in computer science, nonparametric algorithms are continuously being released to solve more complex prediction problems. This increases the possibilities of finding the most feasible algorithm. Finally, a multiple comparison procedure is proposed to automatically compare and then select which model, within a set of nonparametric algorithms, is the most suitable for the particular case studied.

2 CASE OF STUDY AND METHODOLOGY

2.1 DATABASE DESCRIPTION

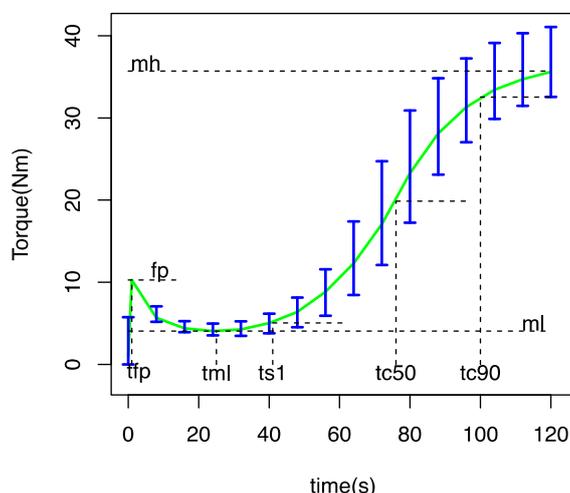
The DB has already been described by González, A. et al. (2007), who cleaned and preprocessed the complete set of raw data. In addition, González, A. et al. successfully implemented artificial neural networks (ANNs) predicting mixture properties of rubber compounds. Although this works keeps the same input and output variables, a brief DB description is presented.

The DB contains 1240 samples of 6 different compound formulas. An unequal number of samples n is available per each compound formula (627, 122, 228, 91, 82, 90). The inputs and outputs of the DB were recorded in different points of the factory:

- Input variables were obtained from the real mixing process in the production line. Two different measures of the same variable are mainly measured. First, the setting points desired by the worker (usually labeled as pre-variables) and second the real value of the variable recorded by the sensors.

Output variables were obtained from the samples analyzed in the laboratory tests. These variables are the parameters of the rheology curves from the samples analyzed. The rheology curve (see Fig. 1) is drawn by measuring torque every 8 seconds, obtaining a total amount of 16 torque values. These 16 measures are named gates. Besides, some representative values of the rheology curve are measured. These values include the maximum and minimum torque (mh and ml), the initial torque peak (fp, t_{fp}) and some critical time values (tml, ts1, tc50, tc90).

Figure 1: Rheology curve



2.2 DATABASE ATTRIBUTES

First, DM techniques are proposed to extract useful hidden knowledge related to the inherent structure of the variables selected in order to ease the following model-training phase.

Analysis of inputs

The final set of inputs is composed by the following 15 variables: vca1, vca2, vca3, vne1, vne2, vne3, vexten, prerci1, prerci2, terein, terefi, pocome, predu1, predu2, dureme.

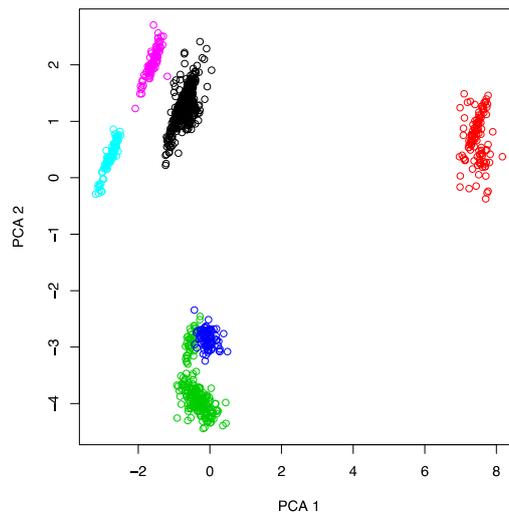
First, the relations between setting values and real values of inputs are studied. Two different patterns are clearly distinguished:

- Formula dependent inputs: Setting and real values of these variables are identical because they represent the amount of each component to create a mixture with a specific formula (vca1, vca2, vca3, vexten, vne1, vne2, vne3, predu1, predu2) or the fixed settings required by the machine, i.e. the working rpm of mixer blades (prerci1, prerci 2). Using the setting values or the real values when predicting the outputs was not making any difference.

- Process dependent inputs: Real values do not tightly follow the setting values. The setting value is just a desired process value that the worker sets; however, the real value varies around the set point. Therefore, these variables add randomness and variance to the process. They represent the desired temperature at the mixing process (terein, terefi), the expected power consumed (pocome) or the expected time of the process (dureme).

Second, the division into groups is analyzed. Prior knowledge provides that the DB studied is composed of samples from 6 different formulas. A mapping technique should be able to split 6 different groups.

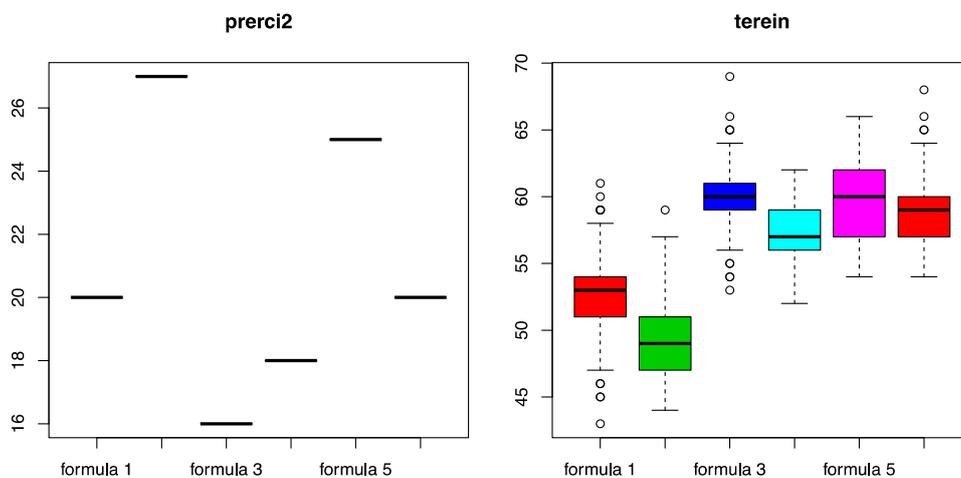
Figure 2: PCA projection



In Figure 2, the linear principal component analysis (PCA) is able to split 4 of the 6 groups, while the other 2 groups division remains not so clear due to some nonlinearities that linear projectors as PCA are not able to capture.

Third, once the group division is set, variables that create group division can be identified. A boxplot depiction of every input variable by splitting data into the 6 known formulas is the proper technique to represent how variables change within a specific group.

Figure 3: Boxplots of variables preci2 (left) and terein (right)

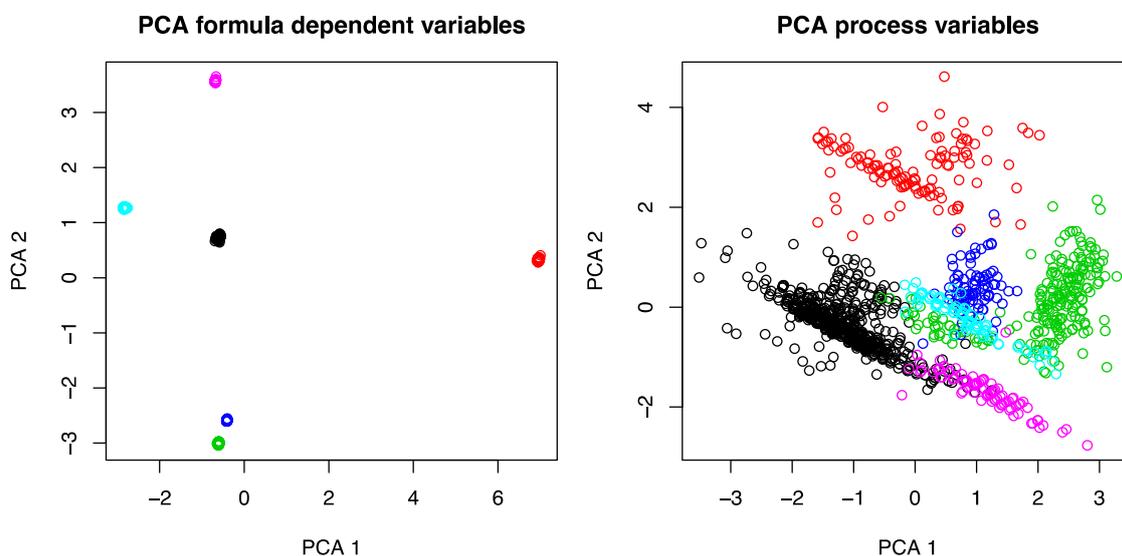


After analyzing all 15 boxplot a different pattern is observed between the defined formula dependent variables and the process variables. An example of the different behavior of a mixture composition variable (prerci2) and a process variable (terein) is shown in figure 2.

Formula dependent variables remain constant for a specific rubber, because they quantify the 'ingredients' to create a specific rubber mixture. Values of different groups do not overlap in the boxplot depiction; hence, these variables are the main cause of group division. On the other hand, process variables do not remain constant for each mixing, the values of one specific group overlaps with other groups.

Representing two additional PCA projections, one only using the formula dependent variables and another one using the process variables, this hypothesis is proved. Figure 4 shows as how group division is emphasized when using only the formula dependent variables and how it almost vanishes when using process variables.

Figure 4: PCA of formula dependent variables (left) and process variables (right)



Analysis of outputs

Two different are feasible to select the prediction variables. The straightest idea would be to predict all the gates; making possible to monitor the exact shape of the rheology curve. However, a first simpler approach is to only predict some of the representative values described in section 2.1 related to crucial information about the mixture behavior, postponing the curve monitoring to future works. Thus, the final set of outputs to be predicted is composed of 5 variables: the maximum and minimum torque (mh, ml) and some critical events of the curve represented by the time they take place (ts1, tc50, tc90).

2.3 MODEL DEVELOPMENT METHODOLOGY

The principal conclusions obtained from section 2.2 are the existence of a strong group division and two clear patterns in input variables. This suggests that the following two different strategies when choosing the most accurate model may be balanced:

1. Generalist models: Models trained with samples from all different formulae. They have optimum generalization capacity, being able to predict samples properties of not previously known formulas.
2. Specialist models: Models trained with samples from only a specific formula. This technique is useful when more accuracy is desired in only one specific formula. Only those variables varying within a group are needed: process variables.

Besides, when dealing with generalist models, a third strategy, DB stratification, may solve the problem of an unbalanced datasets. This procedure can even improve the generalization ability of the model.

Finally, once the performance measurements of the models are computed, a multiple statistical comparison is carried out to automatically determine which model is significantly more accurate than the rest. The techniques employed for this task are the following:

Data preprocessing

Data normalization is firstly carried out because nonparametric models are quite sensitive to the scale of different input variables.

Data stratification is a widely used technique when unbalanced datasets are not obtained by the own nature of the problem studied. For example, when the unequal number of samples per groups is due to random measures availability, balancing the DB would generate a more reliable situation. Two basic approaches used to balance dataset are up-sampling and down-sampling (Ranawana, R. & Palade, V, 2007)

Overview of the models

The following nonparametric techniques are implemented:

- I. Support Vector Machine (SVM) involves a new generation learning system with the capacity to deal with nonlinearities resulting in complex mathematical equations (Vapnik, V. & Lerner, A., 1963). The goal of the SVM is to identify a hyperplane that separates a particular subset of data from the rest in an n-dimensional space. There are many hyperplanes but the idea for finding the best model is to maximize a particular mathematical function with respect to the specified input data.
- II. Multilayer Perceptron Neural Network (MLP) is a feedforward ANN that uses a learning algorithm back propagation to set up the MLP weights (Haykin, S., 1999). One hidden layer is only considered in regression tasks due to any continuous function can be approximated with only one hidden layer if the number of connection weights is enough. The criterion selected for measuring the goodness of fit is the least mean square (LMS) error.

Evaluation and performance procedure

Resulting models are evaluated by computing a set of performance measurements and then, comparing these by a multiple statistical comparison. The performance measurements selected to evaluate model accuracy is the mean average error (MAE). These values are calculated in both training and testing folds of the validation method, which is the repeated k-

fold cross validation (CV). This generates a larger number of estimates so a more reliable performance of models is obtained (Kohavi, R., 1995, Molinaro et al., 2005).

In addition, it is required to determine whether a group of algorithms has significantly better performance than others. Non parametric techniques are widely used to perform multiple statistical comparison due to the nature of the estimates obtained from a repeated k- fold CV. (Derrac et al. ,2011, García et al., 2010). The strategy of the comparison is divided as follows:

- 1) Nonparametric Friedman test is used to determine whether there is any significant difference within the whole group of regression models. Friedman test requires the evaluation of a number of M_c different algorithms in N_c different datasets. This procedure strongly requires that M_c has to be less or equal than N_c .
- 2) $1 \times N$ comparison using Finner post hoc contrasts a control method against $M_c - 1$ left algorithms. The control method is the model that seems to perform better against the rest. The issue of the $1 \times N$ comparison is to roughly quantify how significant is the difference between the algorithms compared. This difference is measured in terms of the p-value where a value of 0% means totally different algorithms, while 100 % means statistically equivalent algorithms.

3 RESULTS AND DISCUSSION

Results obtained are structured in two blocks: First, generalist models are trained using all available samples. A multiple statistical comparison is carried out in order to automatically select the significantly most accurate methods. Two examples are included, one looking for the best non parametric model and another seeking for the best settings for a unique regression algorithm. Second, two different alternatives to generalist models are presented: specialist models and generalist models using stratified inputs.

The following parameters are set:

- Performance MAE measurements are computed using 100 x 10-fold CV, generating 1000 performance values per model and output variable.
- The average of 1000 MAE testing measurements is chosen to make the comparison. The performance of the model is tested in $N_c = 5$ datasets, where each output prediction is considered as a particular dataset.
- A threshold p-value of 5% is selected to establish whether two algorithms are significantly different.

These evaluations are carried out with the statistical software R-project 2.15 running on a dual quadcore Opteron server with Linux SUSE 11.2.

3.1 SEARCHING THE MOST ACCURATE GENERALISTS ALGORITHM

Two different types of algorithms with different settings are trained to look for the most accurate model, making a total amount of $M = 9$ different regression models.

- SVM: 2 models. Auto tuning function is used to adjust cost and gamma parameters. A linear and a nonlinear kernel function are tested.
- MLP: 7 models. ANNs are trained with early stopping using an additional validation set (15% of the total training subset). The number of neurons in the hidden layer varies (3, 5, 7, 9, 11, 13), but always using a *tansig* activation function. Besides, a linear activation function is also tested.

The results obtained are depicted in Table 1

Table 1: MAE testing error of the generalist models

	ml	ts1	tc50	tc90	mh
SVM	0.1706 (0.0127)	1.4498 (0.1057)	1.6038 (0.1193)	1.0925 (0.0790)	1.1531 (0.0904)
SVM linear	0.1797 (0.0126)	1.5690 (0.1085)	1.7678 (0.1297)	1.1637 (0.0815)	1.2391 (0.0854)
MLP 3	0.1834 (0.0143)	1.5799 (0.1184)	1.7400 (0.1404)	1.1461 (0.0890)	1.2498 (0.0911)
MLP 5	0.1841 (0.0148)	1.5828 (0.1240)	1.7413 (0.1469)	1.1513 (0.0916)	1.2532 (0.0925)
MLP 7	0.1846 (0.0148)	1.5878 (0.1278)	1.7446 (0.1503)	1.1552 (0.0939)	1.2550 (0.0926)
MLP 9	0.1847 (0.0150)	1.5907 (0.1270)	1.7516 (0.1533)	1.1583 (0.0954)	1.2557 (0.0929)
MLP 11	0.1847 (0.0149)	1.5973 (0.1294)	1.7568 (0.1545)	1.1609 (0.0960)	1.2570 (0.0925)
MLP 13	0.1847 (0.0150)	1.6006 (0.1296)	1.7601 (0.1566)	1.1623 (0.0958)	1.2571 (0.0926)
MLP linear	0.1827 (0.0132)	1.6200 (0.1121)	1.8275 (0.1393)	1.1936 (0.0860)	1.2562 (0.0873)
CUBIST	0.1651 (0.0133)	1.4260 (0.1193)	1.6329 (0.1489)	1.0625 (0.0865)	1.1390 (0.0985)

Before performing multiple statistical comparison, $M_c \leq 5$ models from the total $M=9$ models trained must be chosen. Based on the average MAE magnitude and based on the parsimony of the models the nonlinear SVM and the MLP with 3 neurons in the hidden layer are selected. Friedman ranked test results are listed in Table 2.

Table 2: Friedman rank test

	SVM	MLP 3
dataset1	1	2
dataset2	1	2
dataset3	1	2
dataset4	1	2
dataset5	1	2
avg. rank	1	2

A Friedman statistic of 5 and p-value of 0.025347 are obtained (see Table 3). There are strong differences within the algorithms compared. It clearly seems that SVM beats MLP, but in order to quantify this difference a 1xN comparison where SVM is chosen as control method. The 1xN comparison yields that there is a 2.5% probability of the two algorithms being the same. Thus, according to the threshold value set, SVM can be considered as the significantly most accurate regression models.

Table 3: 1xN comparison

	z value	unadj_P	APV_Finner
MLP 3	2.236067977	0.025347319	0.025347319

This comparison methodology can be used not only to compare different types of algorithms, but also to automatically decide which settings for a specific algorithm are better. For instance, in the previous comparison MLP 3 has been manually selected as the best performing MLP-based on the average errors obtained. However, using multiple statistical comparison is a more automatic robust procedure. In order to compare this $M_c=7$ different settings of the MLP, $N_c > 7$ independent DB are required. Hence, the use of different output

results as different datasets is no longer feasible. More independent datasets are needed so the following strategy is applied:

First, performance measurements are computed with 10 CV alone, without repetitions, obtaining a total amount of 10 folds/output. Later, all 10 different independent test results per outputs are joined together in order to create a total amount of 50 independent datasets. Friedman ranked test results are listed in Table 4.

Table 4: Average rank of Friedman rank test of different MLP

	MLP 3	MLP 5	MLP 7	MLP 9	MLP 11	MLP 13	MLP linear
avg. rank	3.26	3.16	3.88	3.74	4.6	4.64	4.72

Friedman rank yields a statistic of 28.10571 and a p-value of 8.975156e-05 strongly showing the existing difference between the several settings. MLPs of 5 ANNs seem to be the most accurate setting. Nevertheless MLP 3 is chosen as control algorithm because it proved to be the more robust setting when 1000 MAE measures were computed (see Table 1). Besides, if MLP 3 does not significantly lose against MLP 5, MLP 3, it can be selected as the best setting due to its parsimony.

Table 5: Values of the 1xN comparison using different MLP settings

	z value	unadj_P	APV_Finner
MLP linear	3.379243364	0.000726856	0.004353221
MLP 13	3.194079344	0.001402776	0.004353221
MLP 11	3.101497334	0.001925446	0.004353221
MLP 7	1.435021155	0.151281069	0.218109685
MLP 9	1.11098412	0.266575185	0.310670729
MLP 5	0.231455025	0.816961322	0.816961322

Table 5 shows that MLP 3 significantly beats the linear MLP and the MLP with 13 and 11 hidden neurons. It beats MLP 7 and MLP 9 but not under the significance threshold. Besides it loses against MLP 5 but not under the 5% threshold too.

As MLP 3 is the simpler than MLP 5, MLP7 and MLP 9 and all of them are statistically equivalent, choosing MLP 3 as the best setting is the statistically best approach.

The next step in this procedure would be to include both the p-value results and a parsimony parameter in the same weighted function. This function will automatically select the best model based on both, accuracy and simplicity.

3.2 Alternative models

Specialist model

Specialist models are just trained with samples from group 1 because it is the only group with enough data to obtain accurate results. Besides, only process variables (terein, terefi, pocome, dureme) are used as inputs.

SVM are implemented with a gamma parameter of 0.1 and a cost of 10, both obtained from an automatic tuning. MLP settings remain similar to section 3.1.

Table 6: MAE test errors of group 1 model

	ml	ts1	tc50	tc90	mh
SVM	0.1400 (0.0142)	1.3358 (0.1203)	1.4898 (0.1451)	0.9640 (0.0932)	1.0391 (0.1036)
SVM linear	0.1503 (0.0152)	1.4287 (0.1263)	1.5796 (0.1496)	1.0254 (0.1045)	1.1070 (0.1050)
MLP 3	0.1524 (0.0149)	1.4417 (0.1260)	1.6003 (0.1512)	1.0394 (0.1018)	1.1119 (0.1054)
MLP 5	0.1525 (0.0150)	1.4419 (0.1257)	1.6005 (0.1513)	1.0404 (0.1027)	1.1123 (0.1055)
MLP 7	0.1525 (0.0149)	1.4431 (0.1260)	1.6005 (0.1513)	1.0409 (0.1035)	1.1138 (0.1058)
MLP 9	0.1525 (0.0149)	1.4433 (0.1262)	1.6003 (0.1520)	1.0414 (0.1031)	1.1137 (0.1047)
MLP 11	0.1525 (0.0149)	1.4442 (0.1258)	1.6005 (0.1517)	1.0412 (0.1030)	1.1141 (0.1046)
MLP 13	0.1524 (0.0148)	1.4448 (0.1265)	1.6004 (0.1511)	1.0414 (0.1026)	1.1146 (0.1045)
MLP linear	0.2007 (0.0187)	1.7889 (0.1536)	1.9011 (0.1957)	1.3584 (0.1305)	1.4724 (0.2325)

Comparing errors in Table 6 with those in Table 1, specialist models prove to enhance the prediction accuracy. Training a model just with samples from a specific formula allows increases its learning process of the existing patterns within that formula. Besides the number of inputs is reduced from 15 to 4 because the 11 left formula dependent inputs do not change for a constant formula. Consequently, when a generalist algorithm tries to grasp the slight changes in a specific group, these 11 useless variables are just adding noise to the problem, hindering the output prediction.

Data stratification

Despite of up-sampling and down-sampling are the simplest techniques to carry out a stratification of the data, DB structure does not allow to implement these techniques due to the big gap between the group with the smallest and the highest number of samples (627 vs 82). If down-sampling is selected, not enough number of samples is available to train the model. On the other hand, if up-sampling is desired, the remaining samples up to 627 in the smaller groups are obtained by sampling with replacement. In the smallest group, 627-82 samples are obtained by replacement of the 82 original, creating a final DB strongly dependent where the results obtained may not be reliable.

Consequently, up-sampling and down-sampling are merged together into a intermediate stratification process where X samples with (82 < X < 627) are chosen. X = 250 and X = 350 strategies are tested.

Table 7 Stratification of MAE results

	ml 250	ml 250
SVM	0.1066 (0.0197)	0.1102 (0.0182)
SVM linear	0.1622 (0.0158)	0.1535 (0.0210)
MLP 3	0.1034 (0.0502)	0.1167 (0.0433)
MLP 5	0.0980 (0.0533)	0.1107 (0.0468)
MLP 7	0.0984 (0.0540)	0.1106 (0.0474)
MLP 9	0.0997 (0.0543)	0.1110 (0.0477)
MLP 11	0.1012 (0.0540)	0.1122 (0.0477)
MLP 13	0.1026 (0.0537)	0.1128 (0.0477)
MLP linear	0.1632 (0.0184)	0.1551 (0.0277)

Table 7 depicts the errors obtained in the two different strategies by predicting just the ml output. Apparently, stratification widely improves section 3.1 procedure. However, a deeper

analysis on the results obtained from CV show that the error of those groups with small number of samples almost vanishes if a big number of samples are required to be taken with replacement. The reason is that the test the training fold samples are strongly dependent. The algorithm has already been trained with the samples used to evaluate its performance. Stratification is not a suitable technique to this problem because the few number of samples in the smaller group.

4 CONCLUSIONS

DM can be applied to improve rubber industrial process where many data is available. By predicting the final properties in the mixing phase, the variability of the process is reduced. This improves the quality of rubber compounds before extrusion and enables the application of similar techniques in the extrusion process obtaining more confident results.

In a first exploratory phase, the mapping techniques and the boxplot representation show the existence of two different patterns in the input variables of the process, formula dependent input and the process inputs. When studying different mixture compositions, formula dependent are the critical variables while if the study is focused on a unique composition, process variables are the main cause of variability.

In a second prediction phase, a model comparison procedure based on non-parametric techniques quantifies the difference between the wide ranges of non-parametric models available. This procedure is an automatic model selection technique that is based in more than just comparing the average error of the models. This makes it a powerful tool to automatically obtain more accurate models. This procedure sets that Support Vector Machines is the regression model that better adapts to the mixing process DB. These techniques can be also applied to set the parameters of a specific model. MLP parameters are tuned using this methodology and 3 hidden neurons with a *tansig* activation function are selected as the most appropriate setting combining both accuracy and simplicity. In future works, an automatic function that balances simplicity (quantified in a complexity parameter of a model) and accuracy (quantified in the p-value of the comparisons) can be developed.

Besides, two additional strategies are compared with the obtained most accurate model using the information obtained in the exploratory phase. Data stratification shows that this procedure is not suitable for DBs where the difference of samples between groups is too wide. On the other hand, specialist models prove to be a more accurate solution when the process is just focused in one single rubber mixture due to the reduction of input variables, showing how the information of the DB analysis can be crucial in model development.

REFERENCES

- De la Coz Díaz, J.J, García Nieto, P.J., Bello García, A., Guerrero Muñoz, J. & Ordieres Meré, J.(2008). Finite volume modeling of the non-isothermal flow of a non-Newtonian fluid in a rubber's extrusion die. *Journal of Non-Crystalline Solids*, 354.4, 5334-5336.
- Derrac, J., García, S., Molina, D. & Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation* 1.1, 3-18.
- García, S., Fernández, A., Luengo, J. & Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180.10, 2044-2064.

- González, A., Pernía, A., Alba, F. & Forcada, A. (2007). A neural network-based approach for optimising rubber extrusion lines. *International Journal of Computer Integrated Manufacturing*, 20.8, 828-837.
- Ha, Y.S., Cho, J.R., Kim, T.H. & Kim, T.H.(2008). Finite element analysis of rubber extrusion forming process for automobile weather strip. *Journal of Materials Processing Technology*, 2011, 168-173.
- Haykin, S. (1999). *Neural networks: a comprehensive foundation*. Upper Saddle River, NJ, USA, Prentice Hall.
- Kohavi,R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International joint Conference on artificial intelligence* (pp. 1137-1145)
- Martínez de Pisón, F.J., Barreto,C., Pernía, A. & Alba, F.(2008). Modelling of an elastomer profile extrusion process using support vector machines (SVM). *Journal of Materials Processing Technology*, 197.1, 161-169.
- Martínez de Pisón, J.P., Pernía, A., Blanco, J., González Marcos, A. & Lostado, R.(2010). Control Model for an Elastomer Extrusion Process Obtained via a Comparative Analysis of Data Mining and Artificial Intelligence Techniques. *Polymer-Plastics Technology and Engineering* 49.8, 779-790.
- Molinaro, A.M., Simon, R. & Pfeiffer, R.M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, 21. 15, 3301-3307.
- Ordieres, J., López, L.M., Bello, A. & García, A.(2003).Intelligent methods helping the design of a manufacturing system for die extrusion rubbers. *International Journal of Computer Integrated Manufacturing*, 16.3,173-180.
- Ranawana, R. & Palade,V. (2007) Genetic Algorithm Approach to Construction of Specialized Multi-Classifer Systems: Application to DNA Analysis. *Frontiers in the Convergence of Bioscience and Information Technologies* (pp. 341-346).
- Vapnik, V. & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Autom. Remote Control* 24, 774–780.