



# Efficient improvement for the estimation of the surface of free energy asphalt binder using Machine Learning tools

## Mejora eficiente para la estimación de la energía libre superficial del ligante asfáltico mediante herramientas de Machine Learning

David Sierra-Porta <sup>1</sup>

<sup>1</sup> Departamento de Física, Universidad de los Andes, Colombia. Email: [d.sierrap@uniandes.edu.co](mailto:d.sierrap@uniandes.edu.co)  
Orcid: 0000-0003-3461-1347

Received: 13 October 2020. Accepted: 02 February 2021. Final version: 04 June 2021.

### Abstract

The Surface Free Energy (SFE) of a material is defined as the energy needed to create a new surface unit under vacuum conditions. This property is directly related to the resistance to fracture and recovery of material and the ability to create strong adhesion with other materials. This value can be used as a complementary parameter for the selection and optimal combination of materials for asphalt mixtures, as well as in the micromechanical modeling of fracture and recovery processes of said mixtures. This document describes the results of the implementation of the use of machine learning and Random Forest prediction techniques for the estimation of surface free energy based on data from previous studies. The experimental samples were twenty-three asphalt binders used in a Strategic Highway Research Program (SHRP). A decrease of 54% and 82% in the mean absolute error (MAE) and the mean square error (MSE), respectively was found for the new model built. While the model fits better with a 12% improvement, according to the adjusted determination coefficient, the accuracy and the score of the model also increases notably in 2% and 55%, respectively.

**Keywords:** asphalt cement; surface free energy; asphalt mixtures; machine learning; random forest; strategic highway research plan.

### Resumen

La energía libre de superficie de un material se define como la energía necesaria para crear una nueva unidad de superficie en condiciones de vacío. Esta propiedad está directamente relacionada con la resistencia a la fractura y recuperación de un material y con la capacidad de crear una fuerte adhesión con otros materiales. Este valor puede ser utilizado como parámetro complementario para la selección y combinación óptima de materiales para mezclas asfálticas, así como en el modelado micromecánico de procesos de fractura y recuperación de dichas mezclas. Este documento describe los resultados de la implementación del uso del aprendizaje automático y las técnicas de predicción de bosque aleatorio para la estimación de la energía libre superficial basada en datos de estudios anteriores. Las muestras experimentales fueron veintitrés ligantes de asfalto usados en un Programa de Investigación Estratégica de Carreteras (SHRP). Podemos destacar una disminución de 54% y 82% en el error medio absoluto (MAE) y el error cuadrático medio (MSE), respectivamente. Si bien el modelo encaja mejor con una mejora del 12%, según el coeficiente de determinación ajustado, la precisión y la puntuación del modelo también aumentan notablemente en un 2% y 55% respectivamente.

ISSN Printed: 1657 - 4583, ISSN Online: 2145 - 8456, CC BY-ND 4.0



How to cite: D. Sierra-Porta, "Efficient improvement for the estimation of the surface of free energy asphalt binder using Machine Learning tools," *Rev. UIS Ing.*, vol. 20, no. 3, pp. 179-188, 2021, doi: [10.18273/revuin.v20n3-2021013](https://doi.org/10.18273/revuin.v20n3-2021013)

**Palabras clave:** cemento asfáltico; energía libre de superficies; mezclas asfálticas; aprendizaje automático.

## 1. Introduction

Asphalt mixtures used in pavement structures, are porous materials that result from the combination of multiple aggregates (among many of these you can have crushed rock in various sizes and with a density of different proportions) and asphalt cement. The strength and durability of an asphalt mix depend to a large extent on the quality of the adhesion between the cement and the aggregates. In this way, the adhesion between these two or more materials usually turns out to be a function of their mineralogical and chemical composition, the morphology of the aggregates (shape and textures), and in addition to the conditions in which they are mixed is prepared. Often what happens is that the asphalt mixtures are deficient and therefore their performance in the works for which they were made is also deficient. This could be due to inadequate conditions in the preparation of the mixtures by not ensuring compatibility of the compounds. In addition, very high or very low temperatures, or outside the preparation standards, could also affect the integrity of the mixtures as well as their performance on the work.

From a physical point of view, adhesion in a mixture of liquid and solid materials (such as asphalt) is defined in terms of the physical surface properties of the materials that allow the liquid to wet or coat the solid component. This phenomenon is known as wettability [1] [2] [3] [4] it is defined as the resistance of a liquid droplet to stay in balance when in contact with a solid body. The ability of liquids to coat solid bodies, and solids to be coated by liquids, is directly related to the surface tension or Surface Free Energy (SFE) of the materials (i.e., the energy required to generate a new unit of area of the material). Adhesion between two materials is only possible if the SFE of the solid body is superior to the ESL of the liquid. SFE is a fundamental property of materials, and its quantification is done through the application of advanced characterization techniques, such as the Wilhelmy Plate Method (WPM) [5] [6], the Sessile Drop Method [7] [8] [9] [10] [11], the Universal Adsorption Method (UAM) [12], among others.

The main motivation to characterize the adhesion in asphalt mixtures is the growing need for better material selection techniques (i.e., the combination of aggregates and asphalt cement) based on fundamental properties of the materials, which guarantee more resistant and durable mixtures. It has been shown that by studying adhesion in aggregate-asphalt cement systems, is possible to identify combinations of materials that

produce high adhesion systems and high resistance to moisture damage. This type of damage in asphalt mixtures is defined as the decrease in adhesion between the asphalt cement and the aggregate or the reduction of the cohesion within the asphalt cement [13]. By determining the SFE of the materials and applying the basic theory of surface physics, it is possible to identify combinations of aggregates and asphalt cement with high adhesion in the dry state and with low susceptibility to moisture damage. From the thermodynamic point of view, the SFE of a material is defined as the work required to create a new unit of area in said material, under vacuum conditions [12].

In previous studies [14], SFE for twenty-three asphalt binders of the Strategic Highway Research Program (SHRP) have been calculated and determined using the contact angle technique (sessile drop). Gray correlation analyses were carried out to determine which chemical components and chemical elements of asphalt binders are most related to surface free energy (SFE) measurements of asphalt binders. The measurement of the contact angle was carried out using a Drop Shape Analysis 10, manufactured by Krüss Co., with three different liquids: distilled water, glycerol, and formamide. The Owens Wendt theory was applied to determine the surface free energy. The experimental procedure, as well as the determination of the surface free energies for these twenty-three asphalt samples, can be consulted in the manuscript [14].

The asphalt identification codes, the contents of four fractions, the wax content, and the elemental analysis of these asphalt samples are given in Table 2. Furthermore, from the analyses performed in [15], a simple and multiple regression analysis was carried out to correlate and obtain parametric mathematical relationships between the free energy of the surface and the chemical compositions of the asphalt binders, including group type analysis (saturated, naphthenic, polar aromatics and asphaltenes), wax content and elemental content, based on published data on chemical composition [14].

The present manuscript explores a different line of action for determining relationships between SFE and the chemical characteristics of asphalt samples. In this case, the main objective of our study is to implement a non-parametric methodology with the use of machine learning tools for the estimation and prediction of SFE in terms of the dependent variables measured in previous studies.

The use of machine learning for these purposes is aimed at improving the quality of predictions and estimates. This is possible thanks to the advantages of having deep and automatic learning methods with algorithms that try to learn from the data, and the more data available to learn and richer and more complete the algorithm will work better.

**2. Data and methods**

The effect of eleven input variables was investigated in this research study, namely: Component Analysis, (% of Saturates Aromatics Resins Asphaltenes), Wax content, and Analysis of elements (carbon content (C%), hydrogen content (H%), oxygen content (O%), nitrogen content (N%), sulfur content (S%), nickel (Ni, ppm) as well as vanadium (V, ppm) over the dependent variable SFE.

Data used in this study for the asphalt identification codes, four fractions' contents, wax content, elemental analysis of these asphalt samples are given in Table 1 [14].

**2.1. Data exploration and statistical analysis**

Generally, when trying to statistically study the behavior of a variable alone, a process of analysis of the distribution of this variable is required. This analysis provides information from the systematic exploration of the properties of each variable under study. Probability density diagrams can visually (as a first step) to study the general behavior of the variable under study. One way to obtain this empirical estimate of density (which is certainly a nonparametric methodology) is by using histograms of individual counts or relative frequencies. Often, this preliminary step can reveal what type of distribution the variable follows and thus characterize the central properties of the entire possible range of variable values. This will determine if the distribution is completely symmetric and if the central measures represent a good estimator, which is particularly useful because many times, some known probability density functions are applicable to be modeled by the data set. In this case, we present scatter plots for each input variable with the output variable.

Table 1. SHRP core asphalts and surface free energy results

Asph.Cod	SFE(mJ/m <sup>2</sup> )	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>
AAB-1	15.38	8.6	33.4	38.3	17.3	3.85	82.3	10.6	0.80	0.54	4.7	56	220
AAB-2	16.09	10.8	36.5	35.7	16.7	5.05	85.7	10.59	1.06	0.54	5.4	36	163
AAD-1	16.51	8.6	25.1	41.3	20.5	1.94	81.6	10.8	0.90	0.77	6.9	145	310
AAD-2	16.79	10.0	26.7	40.1	21.3	1.41	81.9	10.3	1.17	0.9	8.3	135	266
AAF-1	16.26	9.6	37.7	38.3	13.3	4.19	84.5	10.4	1.10	0.55	3.4	35	87
AAF-2	14.81	11.9	34.6	38.7	13.0	4.20	84.8	10.2	0.82	0.28	4.6	22	102
AAG-1	25.79	8.5	32.5	51.2	5.0	1.13	85.6	10.5	1.10	1.1	1.3	95	37
AAG-2	27.09	6.6	35.3	51.0	5.0	1.11	87.0	10.5	1.93	1.15	2.9	11	33
AAH	14.48	13.5	28.6	41.4	15.9	4.41	86.3	10.1	1.00	0.8	2.8	43	84
AAK-1	17.02	5.1	30.0	41.8	20.1	1.17	83.7	10.2	0.80	0.7	6.4	142	1480
AAK-2	18.09	7.5	30.6	39.4	19.2	1.14	83.2	10.3	1.14	0.7	6.9	117	1165
AAL	16.27	12.1	30.3	37.3	18.9	1.23	83.4	10.1	1.00	0.6	5.5	98	244
AAM-1	24.11	1.9	41.9	50.3	4.0	4.21	86.8	11.2	0.50	0.55	1.2	36	58
AAN	17.05	10.3	40.1	33.9	15.7	2.74	84.5	10.2	0.80	0.7	4.3	65	157
AAP	20.44	13.2	36.4	36.9	12.6	4.77	85.9	10.9	0.80	0.6	1.7	68	128
AAS-2	16.27	6.4	46.4	30.0	17.1	2.85	83.1	9.8	0.96	0.44	6.76	37	133
AAS-3	17.27	3.8	39.9	37.7	17.3	3.53	81.7	10.1	0.83	0.48	6.21	40	137
AAT	15.86	7.7	32.3	42.5	17.3	2.55	83.9	10.1	0.70	0.6	5.1	80	201
AAV	17.94	10.9	38.9	39.5	9.7	3.13	86.4	10.5	1.10	0.8	2.4	41	92
AAW	15.52	9.3	37.1	35.7	17.9	4.20	84.5	10.1	0.90	0.7	4.5	80	334
ABC	13.77	7.1	44.0	23.2	25.6	2.90	83.2	9.9	0.40	0.3	6.4	25	37
ABD	25.95	10.4	28.4	52.7	7.0	0.81	86.8	10.7	1.20	1.2	1.6	123	62
ABM-1	24.93	9.0	29.6	52.4	7.1	1.10	86.3	10.3	1.22	1.09	1.28	111	63

In head we denote X<sub>1</sub>=Saturates, X<sub>2</sub>=Aromatics, X<sub>3</sub>=Resins, X<sub>4</sub>=Asphaltenes, X<sub>5</sub>=Wax, X<sub>6</sub>=%C X<sub>7</sub>=%H, X<sub>8</sub>=%O, X<sub>9</sub>=%N, X<sub>10</sub>=%S, X<sub>11</sub>=Ni(ppm) and X<sub>12</sub>=V(ppm).

Source: [14] [15].

Since the data studied here are non-Gaussian, the Spearman rank correlation coefficient can be used to obtain a statistical metric concerning the strength of association of each input variable with the output. The Spearman rank correlation coefficient can characterize general monotonic relationships and is in the range of -1 to 1, where the negative sign indicates that it is inversely proportional and the positive sign indicates a proportional relationship, while the magnitude denotes what is very strong in this relationship. In addition, we evaluate if this relationship is statistically significant with the p-values and verify the importance at the 0.01 level.

## 2.2. Multivariate analysis methods

When the variable studied is properly interrelated (or intends to be related) with another set of variables, which we call predictors, the multivariate factorial analysis is convenient to establish and expose the underlying structure in a data matrix that precisely measures this degree of relationship. The first is to determine all the relationships between each pair of these variables without making a priori distinction of which is the dependent or independent variable, or in other words, which is the predictor variable, and which is predicting. Using this information, we can calculate a set of dimensions, known as FACTORS, that seek to explain these interrelationships. Therefore, it is a data reduction technique, the information contained in the data matrix can be expressed, without much distortion, in a smaller number of dimensions represented by said FACTORS.

To evaluate the significant differences between the sites for all the water quality variables, the data were analyzed through the analysis of variance. The multivariate analysis of the water quality data sets was done through hierarchical group analysis (HCA) and principal components analysis (PCA) [16]. The objective of clustering is to divide the objects into homogeneous groups so that the similarities within the group are large compared to the similarities between groups. The Principal Components, on the other hand, are extracted to represent the patterns that encode the highest variance in the data set and not to maximize the separation between groups of samples directly. The statistical package used in this case is R version 3.4.4 (2018-03-15) [17] [18] [19]. The software was used for both the HCA and the PCA.

## 2.3. Classification using random forests

In many practical applications, the inputs may show a complicated functional relationship to determine the output. The classification and regression tree method

(CART, for its acronym in English of its Classification and Regression Tree) is a method conceptually simple, although powerful nonlinear, which often provides reasonable results [20] [21]. CART works by successively dividing the space of the input entity into smaller and smaller subregions.

This procedure can be visualized as a tree that is divided into successively smaller branches, each of which represents a subregion of the ranges of the input variables. The tree grows until it is not possible to divide it further or a certain criterion has been fulfilled. A natural extension of CART is the methodology of its random forests (RF), which is simply a collection of many trees [22]. The training procedure is the same as in CART with the difference that a subset of candidate variables chosen at random can be used to select the optimal variable for each division; the practice has shown that the RF algorithm works extremely well in many different applications [20] [21] [22]. In addition, RF has the desirable ability to promote the most important input variables to predict the output variable as part of its inherent learning strategy [21]. We emphasize that the importance of the variable is not evaluated independently for each variable; instead, it is evaluated jointly for the subset of characteristics used in the RF, making use of the concepts of relevance (strength of association of variable and response), redundancy (strength of association between variables) and complementarity (force of joint association of variables with the answer). Effectively, this means that highly correlated variables (which show high correlations between/among the variables) are penalized and, therefore, redundant variables are not assigned great importance even though they can be highly correlated with the response.

## 3. Results and discussion

The results of the Principal Component Analysis revealed that the first two independent variables that result from the decomposition study add up to 64.8% of the variability in the influence of composition parameters in all samples. The analysis of the correlation between the four fractions contents, wax content, elemental analysis of these asphalt samples, and their contributions to these variables (Figure 1) shows that Wax and Aromatics content is positively correlated with dimensions 2 and negatively correlated with dimensions 1; Carbon and Hydrogen contents are positively correlated with dimensions 1 and 2.

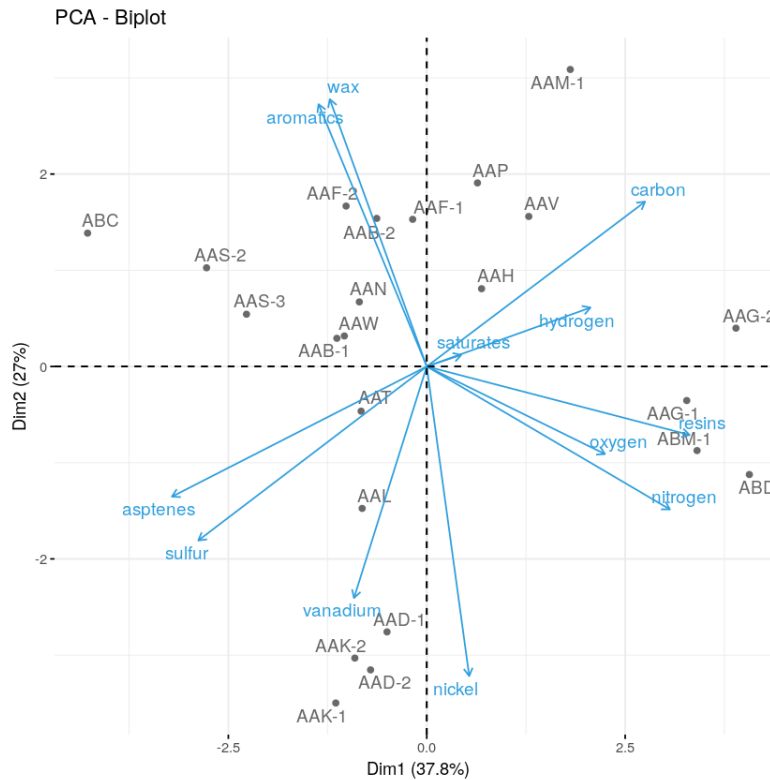


Figure 1. Projection of asphalt binder samples and independent variables on factorial variables determined for principal component analysis.

On the other hand, Resin, Oxygen, and Nitrogen are positively correlated with dimensions 1 and negatively correlated with dimensions 2; and finally, Sulfurs, Aspens, and Vanadium content are negatively correlated with dimensions 1 and 2.

The analysis of the projection of the different groups of samples (Figure 2) on the dimensions of independent variables shows that there are three very well-identified groups. The first group identified as cluster 2 for the asphalt types AAK-1, AAK-2, AAD-1, AAD-2, and AAL, with very high values of Wax and Aromatics content; a second group identified as cluster 5 for the asphalt types ADB, ABM-1, AAG-1, and AAG-2 correlating with Carbon, Hydrogen, Resin, Oxygen, and Nitrogen content. Finally, a third group was identified as cluster 3 for the asphalt types AAT, AAB-1, AAB-2, AAW, AAN, AAH, AAF-1, AAF-2, AAP, and AAV, associated with the main presence of Asphaltene, Sulfur, Vanadium, and Nickel contents.

The correlation coefficient (see Table 2) between the dependent variable (SFE) and the independent variables shows that the most important variables are X3, X4, X6, X9, and X10 with correlation coefficients 0.826013, -0.848026, 0.615821, 0.745924, and -0.673967 respectively. The corresponding p-value for each variable shows that at a level of significance of 1%

these variables have the most contribution to the value of the SFE.

In the reference [15], the authors developed a single regression and multiple regression analysis were applied to correlate the relationships between chemical composition and surface free energy of asphalt binders. Several regressions were constructed by the authors to examine the behavior by groups of separate variables, however, one of the most important estimates that could be made is precisely that which provides the relationship between the SFE and all the measured variables. This found relationship takes the mathematical form:

$$y_{wei} = -47,4678 - 0,3279X_1 + 0,1779X_2 + 0,0688X_3 - 0,0543X_4 - 1,0217X_5 + 0,3789X_6 + 2,8364X_7 + 3,8659X_8 - 0,7983X_9 - 0,5736X_{10} + 0,0256X_{11} - 0,0018X_{12} \quad (1)$$

Our approach is different, and in this case, we find a model built through machine learning tools using Random Forest estimators. In this strategy, several decision trees have been used to find the best model that fits the SFE data in terms of the twelve variables measured. For the evaluation of the method and the good approximation of the solution have been found several parameters and metrics that allow measuring the efficiency of the model (see Table 2 and 3).

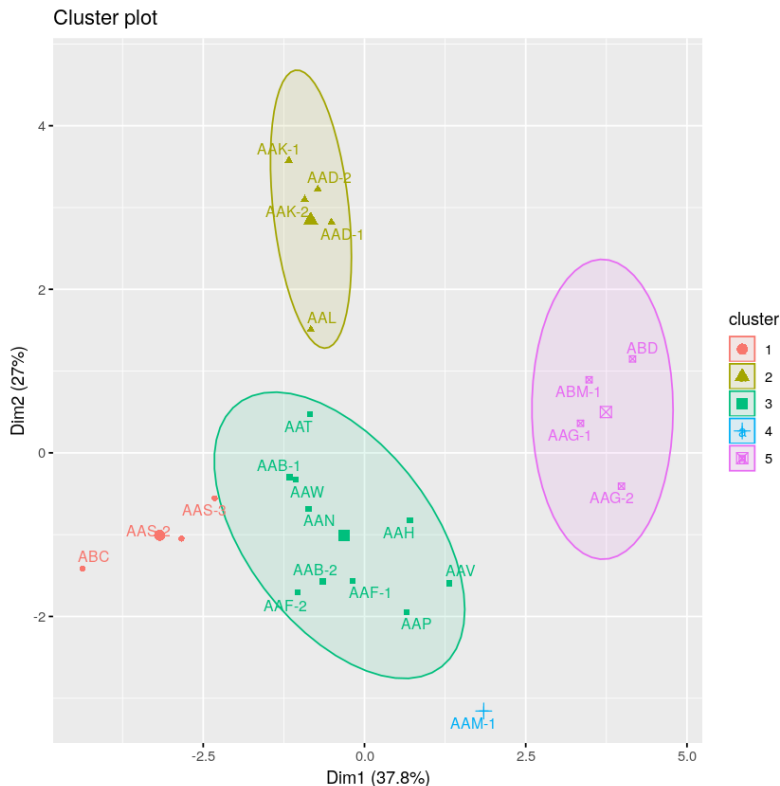


Figure 2. Projection of groups on factorial variables determined for principal component analysis.

Table 2. Correlation coefficients for variables and importance of variable for the random Forest model

Variable	Corre. Coef.	p-value	N	Importance RF
X <sub>1</sub>	-0.231	0.288	23	0.010
X <sub>2</sub>	-0.128	0.559	23	0.004
X <sub>3</sub>	0.826	1.206×10 <sup>-6</sup>	23	0.213
X <sub>4</sub>	-0.848	3.229×10 <sup>-7</sup>	23	0.285
X <sub>5</sub>	-0.458	0.0279	23	0.095
X <sub>6</sub>	0.615	0.001	23	0.037
X <sub>7</sub>	0.527	0.009	23	0.028
X <sub>8</sub>	0.516	0.011	23	0.029
X <sub>9</sub>	0.745	4.388×10 <sup>-5</sup>	23	0.041
X <sub>10</sub>	-0.673	0.421×10 <sup>-3</sup>	23	0.144
X <sub>11</sub>	0.125	0.568	23	0.010
X <sub>12</sub>	-0.196	0.369	23	0.098

We denote X<sub>1</sub>=Saturates, X<sub>2</sub>=Aromatics, X<sub>3</sub>=Resins, X<sub>4</sub>=Asphaltenes, X<sub>5</sub>=Wax, X<sub>6</sub>=%C, X<sub>7</sub>=%H, X<sub>8</sub>=%O, X<sub>9</sub>=%N, X<sub>10</sub>=%S, X<sub>11</sub>=Ni(ppm) and X<sub>12</sub>=V(ppm).

The calculations are made in the computer from a code designed for the implementation of Random Forest in this situation. After examining several alternatives, the model is found and saved on the hard disk so that it can be used later in future applications for data of the same species. In particular, the used metrics in this study are the Mean Absolute Error (MAE), Mean Square Error (MSE), Multiple R 2, R 2 coefficient, Adjusted R 2, Accuracy and Prediction score.

For this case recall

$$\begin{aligned}
 MAE &= \sum_i \frac{|y_i - \hat{y}_i|}{N}, \quad MSE = \sum_i \frac{(y_i - \hat{y}_i)^2}{N}, \\
 ACCURACY &= \left( 1 - \frac{1}{N} \sum_i \frac{|y_i - f_i|}{y_i} \right) \times 100, \quad (2) \\
 Score &= \sum_i \frac{1 \times (y_i = y_{pred})}{N}
 \end{aligned}$$

Table 3. Multiple metrics used to evaluate the accuracy and good performance of the Random Forest model in comparison with linear regression

Parameter	Wei regression	Random Forest	Improvement (%)
Multiple $R$	0.972	0.995	2.45
$R^2$	0.941	0.989	5.07
Adjusted $R^2$	0.871	0.976	12.05
MAE	0.704	0.324	53.9
Accuracy	96.072	98.239	2.26
Score	47.826	73.913	54.55

Source: [15].

A summary of these parameters can be found in Table 3. As can be seen from Table 2, the variables with the highest correlation assignment with SFE generally retain proportionally greater importance in the model developed by Random Forest. This is precisely true for the variables  $X_3$ ,  $X_4$ , and  $X_{10}$ , however, for the variables  $X_6$  and  $X_9$ , there is a loss of importance for these last variables, which is compensated with a gain of importance in the variables  $X_5$ ,  $X_{11}$ , and  $X_{12}$ , which acquire relevance in the machine learning model despite the little initial correlation they had with the SFE measure. This implies that the Wax, Nickel, and Vanadium content are variables that should not be neglected, and their weight is very useful to estimate the dependent variable more adequately. This can be confirmed in the principal component analysis observing that most of the samples report high contents of these variables (see Figures 1 and 2).

On the other hand, the analysis in Table 3 shows that there is considerable improvement in all the metric parameters to evaluate the performance of the model. Significantly, we can emphasize a decrease of 54% and 82% in the mean absolute error (MAE) and the mean square error (MSE), respectively. While the model fits better with a 12% improvement, according to the adjusted determination coefficient, the accuracy, and the score of the model (2% and 55% respectively) also increases notably (understood as the amount of data whose error concerning the model is zero, see equation 2). All these factors and parameters determine a better performance of the machine learning tools and, the estimation using Random Forest, in the approximation (in this case not parametric) for the calculation of the surface free energy for asphalt samples.

Finally, a visual representation of the behavior of the model that reported in previous studies can be seen in Figure 3.

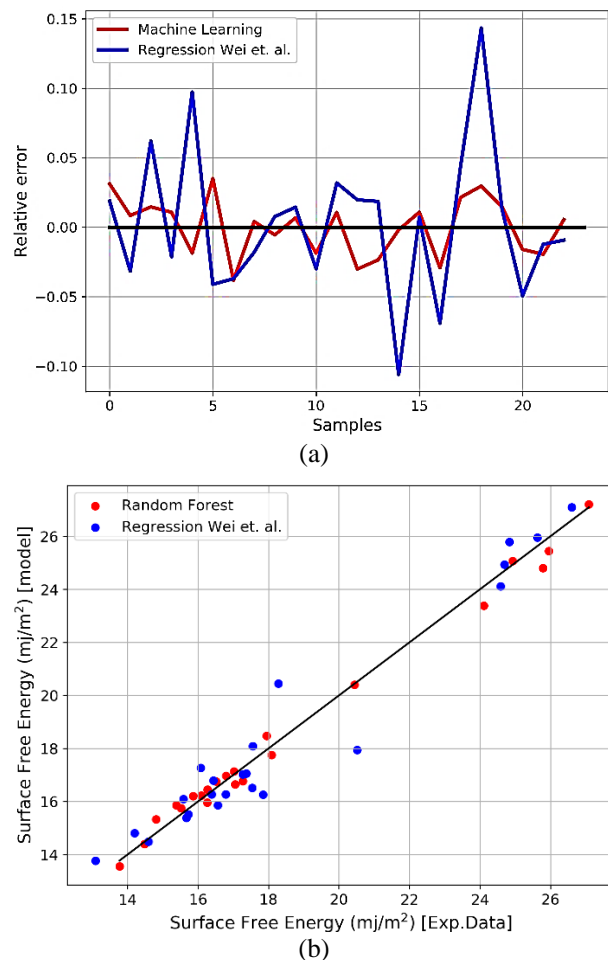


Figure 3. (a) A comparison of the relative error ( $f_i - y_i$ , with  $f_i$  is the model estimation and  $y_i$  is experimental data) between the original data and the two constructed models (multiple linear regression and random forest). (b) Scatter plot for adjusting the original data for the two models studied.

#### 4. Conclusions

We presented methods for performing Random Forest optimization for hyperparameter selection of general machine learning algorithms for the estimation of Surface Free Energy for twenty-three asphalt binders' experimental samples used in an SHRP. We introduced full Random Forest treatment algorithms to make a comparison with previous results for this dataset getting good effectiveness of our approaches. Considering the metrics used in this study we can say that the model determined by multiple linear regression estimates the SFE variable with an error of  $0.9518 \text{ mJ/m}^2$ , whereas the model predicted by Random Forest only an error of  $0.3936 \text{ mJ/m}^2$  (this is  $\text{RMSE}=(\text{MSE})^{1/2}$ ), which represents an 82% improvement over the work of Wei et. al.

The model developed by Random Forest also rescue importance to variables that had a lower weight and correlation in the approach with multiple linear regression, this is a great improvement of the use methods based on machine learning tools. In addition to this, a reported improvement of 52% in the degree of accuracy of the model (score) is recoverable to make the individual data errors as close to zero as possible.

While it is true that the size of the data and the sample is small for the selection of machine learning techniques for the resolution of this problem and the analysis of the study variable in terms of the predictor variables, the same argument also applies to their study using multivariate analysis, so that for equal conditions of data, the best method used for the analysis will always be the one that provides the least errors in the estimation of the study variable.

However, the purpose of this article is to establish criteria that allow us to affirm that machine learning can, and indeed improves, a better estimation of surface energy for the study of asphalt aggregates normally used in construction. With this first scenario, a future work (which is carried out at this time for author and collaboration), is to increase the database to include other types of asphalt binders and aggregates with other predictor contents, and perhaps more study variables.

Based on the results of this research, it can be affirmed that the technique and methodology used will be able to establish very accurate and adequate models for the study of aggregates and asphalt binders used in highway construction.

#### Acknowledgment

The data, as well as the Python code developed for the treatment, organization, analysis, and visualization of the data is available for use by anyone from the author's Github [https://github.com/sierraporta/asphalt\\_binder](https://github.com/sierraporta/asphalt_binder). No funds have been received for the development of this project from any institution.

#### References

- [1] K. L. Mittal, *Advances in contact angle, wettability and adhesion*. Hoboken, NJ, USA: John Wiley & Sons, 2015.
- [2] K. L. Mittal, *Contact Angle, Wettability and Adhesion, Volume 3*. Boca Ratón, FL, USA: CRC Press, 2003.
- [3] P. G. De Gennes, "Wetting: statics and dynamics," *Reviews of modern physics*, vol. 57, no. 3, pp. 827, 1985. doi: 10.1103/RevModPhys.57.827
- [4] O. Voinov, "Dynamics of a viscous liquid wetting a solid via van der Waals forces," *Journal of Applied Mechanics and Technical Physics*, vol. 35, no. 6, pp. 875-890, 1994. doi: 10.1007/BF02369581
- [5] E. Ramé, "The interpretation of dynamic contact angles measured by the Wilhelmy plate method," *Journal of colloid and interface science*, vol. 185, no. 1, pp. 245-251, 1997. doi: 10.1006/jcis.1996.4589
- [6] L. M. Lander, L. M. Siewierski, W. J. Brittain, E. A. Vogler, "A systematic comparison of contact angle methods," *Langmuir*, vol. 9, no. 8, pp. 2237-2239, 1993.
- [7] H. Wu, A. Shen, Z. He, T. Cui, "Study on adhesion between asphalt and steel slag based on surface free energy," in *20th COTA International Conference of Transportation Professionals*, 2020, pp. 1851-1864.
- [8] Y. Yuan, T. R. Lee, "Contact angle and wetting properties," in *Surface science techniques, vol. 51*. Springer, 2013, pp. 3-34. doi: 10.1007/978-3-642-34243-1\_1
- [9] C. Maze, G. Burnet, "A non-linear regression method for calculating surface tension and contact angle from the shape of a sessile drop," *Surface Science*, vol. 13, no. 2, pp. 451-470, 1969. doi: 10.1016/0039-6028(69)90204-0



- [10] J. Bachmann, R. Horton, R. Van Der Ploeg, S. Woche, “Modified sessile drop method for assessing initial soil–water contact angle of sandy,” *Soil Science Society of America Journal*, vol. 64, no. 2, pp. 564-567, 2000. doi: 10.2136/sssaj2000.642564x
- [11] L. Susana, F. Campaci, A. C. Santomaso, “Wettability of mineral and metallic powders: applicability and limitations of sessile drop method and washburn’s technique,” *Powder technology*, vol. 226, pp. 68-77, 2012. doi: 10.1016/j.powtec.2012.04.016
- [12] A. Bhasin, D. N. Little, “Characterization of aggregate surface energy using the universal sorption device,” *Journal of Materials in Civil Engineering*, vol. 19, no. 8, pp. 634-641, 2007. doi: 10.1061/(ASCE)0899-1561(2007)19:8(634)
- [13] B. M. Kiggundu, F. L. Roberts, “Stripping in hma mixtures: state-of-the-art and critical review of test methods,” National Center for Asphalt Technology, Tech. Rep. NCAT Report 88- 02, 1988.
- [14] J. Wei, Y. Zhang, “The application of grey system theory to correlate chemical composition and surface free energy of asphalt binders,” *Petroleum Science and Technology*, vol. 28, no. 17, pp. 1807-1817, 2010. doi: 10.1080/10916460903226098
- [15] J. Wei, F. Dong, Y. Li, Y. Zhang, “Relationship analysis between surface free energy and chemical composition of asphalt binder,” *Construction and Building Materials*, vol. 71, pp. 116-123, 2014. doi: 10.1016/j.conbuildmat.2014.08.024
- [16] I. Jolliffe, “Principal component analysis,” *International Encyclopedia of Statistical Science*. Heidelberg, Berlín: Springer, 2011, pp. 1094-1096. doi: 10.1007 / 978-3-642-04898-2\_455
- [17] R. C. Team et al., “R: A language and environment for statistical computing,” 2013.
- [18] A. G. Bunn, “A dendrochronology program library in r (dplr),” *Dendrochronologia*, vol. 26, no. 2, pp. 115-124, 2008. doi: 10.1016/j.dendro.2008.01.002
- [19] A. G. Bunn, “Statistical and visual crossdating in r using the dplr library,” *Dendrochronologia*, vol. 28, no. 4, pp. 251-258, 2010. doi: 10.1016/j.dendro.2009.12.001
- [20] A. Tsanas, M. A. Little, P. E. McSharry, L. O. Ramig, “Accurate telemonitoring of parkinson’s disease progression by noninvasive speech tests,” *Nature Precedings*, pp. 1, 2010. doi: 10.1038/npre.2009.3920.1
- [21] H. Trevor, T. Robert, and F. Jerome, *The elements of statistical learning: data mining, inference, and prediction, 2da. Ed.* Stanford, CA, USA: Springer, 2009.
- [22] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001. doi: 10.1023/A:1010933404324