

Article

Tire Wear Sensitivity Analysis and Modeling Based on a Statistical Multidisciplinary Approach for High-Performance Vehicles

Guido Napolitano Dell'Annunziata , Giovanni Adiletta , Flavio Farroni , Aleksandr Sakhnevych * 
and Francesco Timpone 

Department of Industrial Engineering, University of Naples Federico II, 80125 Naples, Italy; guido.napolitanodellannunziata@unina.it (G.N.D.); giovanni.adiletta@unina.it (G.A.); flavio.farroni@unina.it (F.F.); francesco.timpone@unina.it (F.T.)

* Correspondence: ale.sak@unina.it

Abstract: One of the main challenges in maximizing vehicle performance is to predict and optimize tire behavior in different working conditions, such as temperature, friction, and wear. Starting from several approaches to develop tire grip and wear models, based on physical principles, experimental data, or statistical methods available in the literature, this work aims to propose a novel tire wear model that combines physical and statistical analysis on a large number of high-performance vehicle telemetries, tracks, and road data, as well as tires' viscoelastic properties. Another contribution of this multidisciplinary study is the definition of the functional relationships that govern the tire–road interaction in terms of friction and degradation, conducting a thorough analysis of the car's telemetry, the track and asphalt features, and the viscoelastic properties of the tires.

Keywords: tire wear; statistical approach; regression technique; multivariate database; vehicle experimental data



Citation: Napolitano

Dell'Annunziata, G.; Adiletta, G.; Farroni, F.; Sakhnevych, A.; Timpone, F. Tire Wear Sensitivity Analysis and Modeling Based on a Statistical Multidisciplinary Approach for High-Performance Vehicles. *Lubricants* **2023**, *11*, 269. <https://doi.org/10.3390/lubricants11070269>

Received: 26 May 2023

Revised: 18 June 2023

Accepted: 19 June 2023

Published: 21 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The main aim of this research work is to develop a straightforward and analytically effective tire wear model for motorsport applications, able to describe wear in terms of the reduction in tread thickness, starting from vehicle experimental data collected from completely different sources. In the literature, the wear rate of the tire tread is often described by a function proportional to the frictional power dissipated during tire–road contact, where the frictional power is directly dependent on the local sliding speed, the local contact pressure, and the coefficient of dynamic friction [1,2]. These parameters are a direct consequence not only of the tire construction and the maneuvers which the tire undertakes [3,4], but also of the road characteristics and boundary conditions such as the temperature, pressure, and velocity distributions within the instantaneous contact patch, as well as of kinematic and dynamic quantities transmitted via the tire tread and materials' properties in mutual contact [5,6].

The scientific community completely agrees that the wear phenomenon is local, the function of the local pressure, temperature, and velocity distributions within the contact patch, and these are not so easy to obtain [7–9], being currently not directly measurable in the real road scenario [10–12], and their determination usually relies on mathematical formulations, frequently involving nonlinear finite element analysis, brush theories, and flexible multibody-based approaches [13,14]. In addition, the evaluation of the compound characteristics is usually evaluated through destructive approaches, forcing the prototyping engineers to feed the preliminary models with nominal, and not usually reliable, viscoelastic data [15,16].

The first studies related the consumption of a solid body to the work computed by the friction forces [17,18]. This concept was then extended to the tire field by Schallamach, proposing the proportionality between abrasion and frictional energy dissipation [19]. In 1986, Shepherd extended Schallamach's theory advancing the concept of wear abrasion proportional to the sliding length in the operating conditions where the lateral stress is larger than a certain threshold stress, once a certain constant coefficient of friction is supposed [20]. Later, Sueoka adapted his research on wear pattern formations at the contact interface with a rotating system to automotive tires, assuming that the polygonal wear is caused by vertical force variation as a result of the first vertical natural mode of the tire belt and approximating the tire by a rigid ring model, where the wear formulation is based on Shepherd's abrasive model with a time delay concept [21]. Le Maitre proposed an empirical statistical methodology taking into account road surface, weather, driving styles, routes, vehicles, and other vehicle parameters able to affect tire wear and to classify the tire usage as mild, medium, and severe [22]. More recently, Braghin proposed an experimentally determined friction and wear model employing data acquired under controlled laboratory conditions, resulting in a friction power scaled with tire-road contact patch area and two wear constants relating to the rubber compound temperature and the road surface roughness [23]. Veen investigated irregular tire wear, coupling Shepherd's formulation, and the rigid ring vertical dynamics [24]. Huang et al. proposed a recent theory able to estimate the tread mass removal employing the instantaneous local three-dimensional pressure and sliding local distributions [25,26]. The necessity to shift the paradigm and to formulate the tire wear model from a multidisciplinary perspective was first pointed out in [27], where an impressive experimental campaign was conducted to quantify potential thermal, degradation, and material influences.

However, it must be highlighted that these references, empirically relying on restricted experimental data, generally access the goodness of the model behavior towards a specific working range, in terms of the interface temperature, road characteristics, or rubber materials adopted [28]. Indeed, a multitude of different phenomena, such as the effect of road roughness [29], vehicle running conditions [30], and the viscoelastic properties of the compound [31] all affect the wear rate. The different nature of these phenomena makes the operation of analyzing the acquired experimental data fairly complex, given the difficulty in extracting the truly relevant information directly correlated to the target variable of the activity.

First of all, all the data available, provided by an industrial partner, have been pre-processed to extract several key performance indicators (KPIs) able to synthesize the complexity of the starting database. After that, diverse statistical approaches have been applied to the reference datasets, developing different regression models [32] to understand their applicability and overall reliability. To this end, tire wear acquisitions available per each dataset collected have been employed to train the diverse algorithms in a supervised learning approach and to validate their reliability.

In this work, mainly three strategies have been analyzed and subsequently implemented, with a progressive increase in the data pre-processing phase: the implementation of feed-forward neural networks [33], the reduction in the dimensionality of the database through a feature selection carried out with the principal component analysis [34] coupled with the development of linear multiple regression, and finally, the implementation of a linear multiple regression based on the combination of different linear correlations based on the physical nature of the data. Despite the obvious simplification of the problem under examination, these approaches have been chosen with the intention of developing simple models capable of providing greater information on wear levels in scenarios not yet explored from an experimental point of view. The reliability of the different approaches proposed was evaluated through a series of statistical indicators [35] and by applying the different models to an additional testing database not used in the training phase, confirming the improved robustness of the last methodology to be presented.

The paper is structured as follows. In Section 2, all the different machine learning methods based on a statistical approach have been presented in detail, describing how

it has been possible to apply these strategies to the wear rate estimation; after that, in Section 3 an overview of the reference database is provided, indicating how it was possible to extract the different KPIs, given the complexity of the multivariate dataset in question. Subsequently, in Section 4, the different results obtained with the three approaches have been illustrated, showing the strengths and weaknesses of each procedure.

2. Methodologies

In order to develop a robust and accurate model for predicting tire wear, various approaches have been explored and applied to the reference database, which will be described in detail in the following section. This work primarily focuses on statistical methods based on machine learning as they offer valuable tools for building predictive models. Machine learning techniques allow the designing of mathematical models capable of learning and identifying patterns within a given database. They find applications in a wide range of fields, including:

- **Data Clustering:** This approach involves dividing the data into distinct subgroups based on specific features used to establish the clusters [36,37].
- **Classification Problems:** In classification, the goal is to predict outcomes among a finite number of different classes [38].
- **Regression Problems:** Regression aims to predict numerical values, making it particularly relevant for tire wear prediction [33,39].

Machine learning algorithms can be broadly categorized based on their learning process:

- **Supervised Learning:** In supervised learning, the algorithm learns from labeled examples present in the training dataset. Classification and regression problems are typical examples of supervised learning, where the labeled data provide the necessary supervision during the learning process [40].
- **Unsupervised Learning:** Unlike supervised learning, unsupervised learning does not rely on class-labeled examples in the training phase. The input examples are not labeled, and the learning process is not guided by specific targets. A common problem in this category is data clustering, where patterns and relationships are discovered without prior knowledge of the classes [41].

The tire wear prediction task, using the reference database available for this research, falls under the realm of supervised learning. This is because the database includes experimental measurements of tire wear, providing labeled data for all the analyzed conditions. Since tire wear represents the gradual thinning of the tread and does not have discrete values corresponding to specific classes, predicting tread wear assumes the form of a classical regression problem. During this research activity, three different statistical machine learning approaches have been adopted and they will be explained starting from the simplest to apply up to the one that requires the most detailed pre-processing phase of the starting database:

1. Feed Forward Neural Network;
2. Principal Component Analysis and Multiple Linear Regression;
3. Physical Correlations and Multiple Linear Regression.

2.1. Feed Forward Neural Network

The first approach has been the use of an artificial neural network (ANN); it is a set of simple units, called neurons, which communicate with each other by sending signals through connections and which are often organized in different layers according to need. In most cases, an artificial neural network is an adaptive system that changes its structure based on information flowing through the network itself during the training or learning phase [42]. They can be used to simulate complex relationships between inputs and outputs that other analytic functions cannot represent. Moreover, neural networks, through learning cycles (input–processing–output) more or less numerous, are able to generalize and provide correct outputs associated with inputs that are not part of the dataset with

which the network is trained [43]. To make a neural network work correctly, it is necessary to carry out an initial training phase; to do this, there are several training algorithms that meet specific needs and purposes. In the case of supervised learning, the network is provided with a set of inputs to which known outputs correspond. By analyzing them, the network learns the link that unites them. In this way it learns to generalize, that is, to calculate new correct input–output associations by processing inputs external to the main dataset used for the learning phase and without knowing the outputs. In this way, it is possible to predict certain results by knowing the main quantities on which they depend.

Each neuron has several inputs which can be either the input signals of the problem or signals deriving from previous neurons. The different inputs are added together in relation to the weight of their connections. In the training phase of the neural network, these weights are used to improve the learning of the network. The neuron is also characterized by a bias. It can be seen as a weight connected to a dummy input equal to unity that has the task of calibrating the neuron's optimal working point. Another fundamental feature of the elementary unit is the activation function that defines the output of the neuron itself (Figure 1). In particular, this function shapes the output of the neuron related to its internal potential, which is determined by the inputs, the weights of the connections, and the bias.

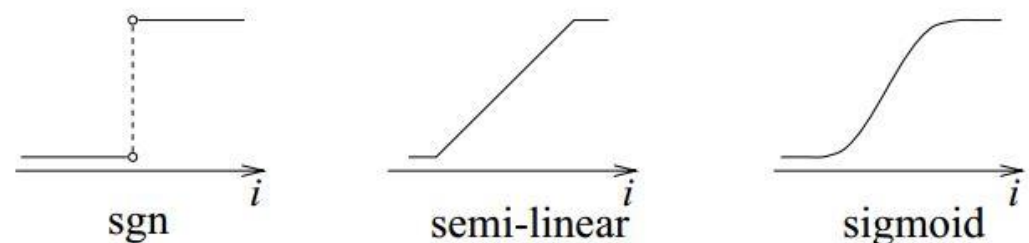


Figure 1. Most common activation functions for a unit.

A widely used activation function, employed also in the proposed approach, is the sigmoid, since its non-linearity and its continuity allow us to build more complex and effective neural networks. The output of a generic i -th neuron is evaluated as in Equation (1).

$$out_i = f(net_i) \quad (1)$$

in which f is the activation function and net_i is the internal potential of the i -th neuron, defined as in Equation (2).

$$net_i = \sum_{n=1}^d w_{ji} in_j + w_{0i} in_0 \quad (2)$$

where j indicates the number of inputs of the i -th neuron; it is possible to identify the following terms:

- in_j are the inputs;
- w_{ji} are the weights of the i -th neuron;
- in_0 is the bias of the i -th neuron, equal to 1.

The aggregation of multiple neurons designed as described and illustrated in Figure 2 gives rise, as mentioned, to complex structures called neural networks, typically organized in various layers, with each layer being formed by a certain number of neurons. Usually, in a neural network, there is an input layer, an output layer, and one or more intermediate levels called hidden layers. The choice of the number of hidden layers depends on the objective and does not follow canonical rules. Regarding the number of neurons for each level, the same consideration applies as before.

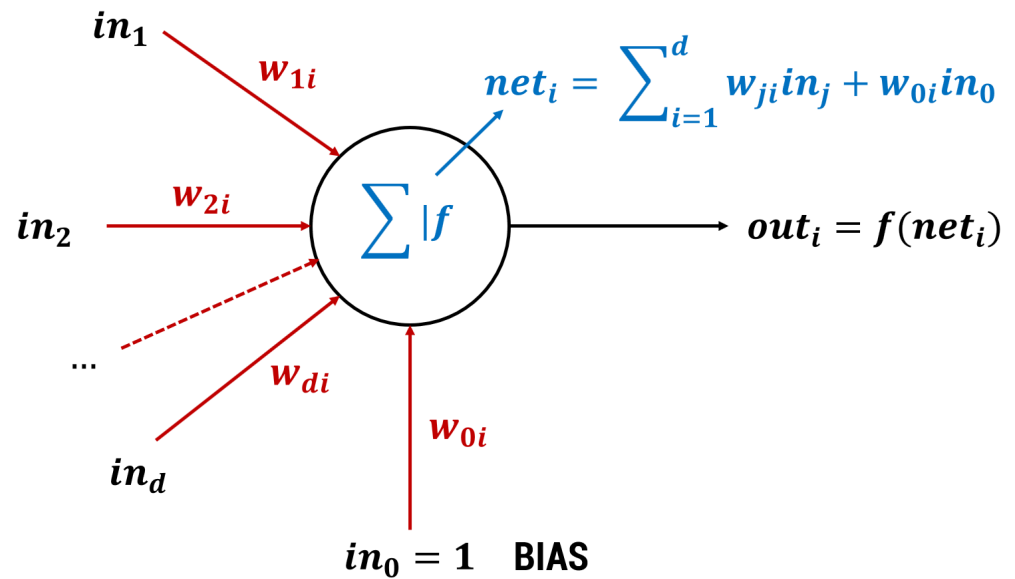


Figure 2. Neuron structure and output.

In the literature, it is possible to find different configurations of neural networks [44]. Having analyzed the nature of the problem under examination, in which the value to be predicted is not temporally correlated to the others (each series of KPIs identifies a different tire that corresponds to a different wear level), the NN with the simplest architecture, the feed-forward, has been chosen. Feed-forward Neural Networks are structures in which connections link neurons of a certain level only with neurons of a subsequent level. So, backward connections or connections between neurons of the same level are not allowed in these networks (Figure 3). Feed-forward networks do not have memory about what occurred at previous times, so the output is determined only by the current inputs. These kinds of networks are the foundation for computer vision, natural language processing, and other work, such as making predictions concerning the future course of a quantity.

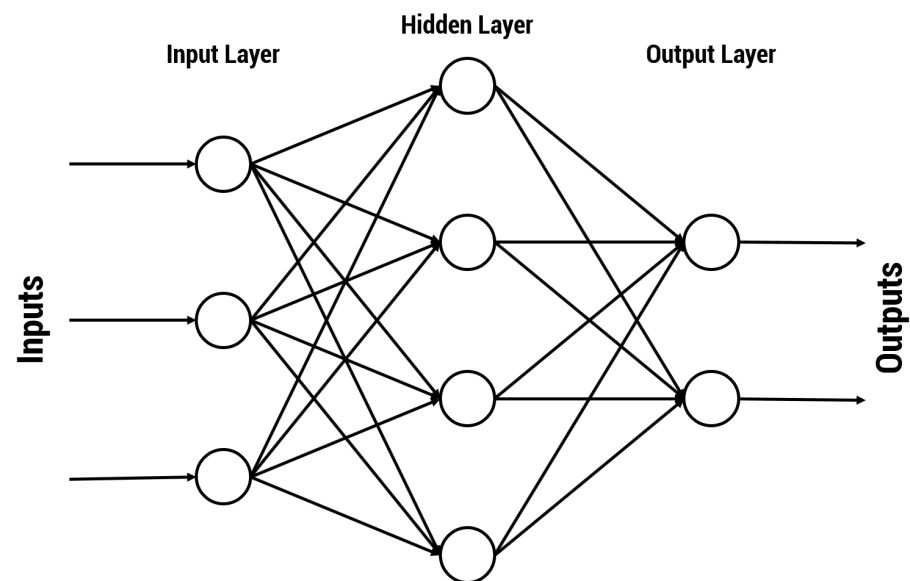


Figure 3. Feed-forward Neural Network graph.

Once the network topology has been set, that is, once the type of network, the number of layers, and the number of neurons for each layer have been chosen, it is necessary to train the network. Training the network means solving the problem of optimizing the weights and biases related to the neurons referring to a certain subset of the total dataset, called the training set, with the aim of gradually reducing the prediction error. To evaluate the error and train the network correctly, an error or loss function and an algorithm must be introduced. Often, the backpropagation algorithm is used to implement the optimization routine for the weights and biases in order to reduce this error. The backpropagation algorithm is based on the gradient descent method or on methods similar to this and has the aim of reducing the error between the result of the network and the target up to a value suitable for the type of application. In simple terms, after each forward pass through a network, backpropagation performs a backward pass while adjusting the model's parameters (weights and biases). So, the mechanism of backpropagation repeatedly adjusts the weights and biases of the connections in the network in order to minimize a measure of the difference between the actual output vector of the net and the desired output vector. More specifically, it happens that, at the beginning, the network initializes the weights and biases randomly and produces results. These results are compared with the target results and from this comparison the loss error function is evaluated. With the gradient descent method, it is possible to find the weights and biases that will yield a smaller loss in the next iteration by calculating the partial derivatives of the error function with respect to the single weights and biases of the neurons. So, the objective is to find out which node is responsible for most of the loss in every layer in order to penalize it by giving it a smaller weight value and thus lessening the total loss of the model.

In order to minimize the difference between the neural network's output and the target output, we have to understand how the model performance changes with respect to each parameter in our model. To do that, it is necessary to calculate the partial derivatives between our loss function and each weight. For each epoch, that is, for each work cycle of the neural network, the backpropagation algorithm is used to subtract the corresponding derivatives from the weights multiplied by a "learning rate" (which avoids abrupt variations in the weights) in order to create the optimization routine. This procedure continues until the error function reaches a sufficiently low value in relation to our needs. It is important to observe that the minimization of an error function in the manner described requires, as mentioned, target values of the output, and so it is a situation of supervised learning, since for each input there is a desired output defined.

After the training phase, there is the validation phase, which exploits another subset of the total dataset, called the validation set. This phase serves to avoid the error reaching a definitive and specific minimum for the training set causing overfitting. Overfitting is the phenomenon for which the network fits the training data too well and adapts to it, invalidating the work of the net with another dataset and, therefore, obstructing the generalization process. So, the validation process is necessary to stop the training phase and iterations of this before overfitting occurs. In short, during this phase, the error relating to the validation set is evaluated and when this error starts to increase instead of decreasing, the training stops. Finally, there is a last phase, called the test phase, which uses an additional set of data, known as the test set, which aims to evaluate the performance and the quality of the network. Regarding the division of the dataset into the three subsets mentioned, typically an equal division is not used, as a larger subset is assigned to the computationally more expensive phase, which is that of training.

2.2. Principal Component Analysis—PCA

One issue encountered when dealing with a set of multivariate data is the sheer abundance of variables, making it challenging to employ simple techniques for obtaining an informative initial assessment of the data. Specifically, multivariate data analysis pertains to a statistical analysis category that involves more than two dependent variables, ultimately yielding a single output, as seen in the presented case study.

Principal Components Analysis (PCA) is a multivariate method designed to address this problem by aiming to reduce the dimensionality of a multivariate dataset while retaining as much of the original variation as possible. This objective is accomplished by transforming the original variables into a new set of variables known as principal components. These components are linear combinations of the original variables, and they are uncorrelated and ordered in a manner that prioritizes the first few components accounting for the majority of the variation observed across all the original variables. As a result, conducting a principal components analysis generates a limited number of new variables that can effectively serve as substitutes for the initial large number of variables [34].

In other words, the main goal of principal components analysis is to describe variation in a set of correlated variables, $x^T = (x_1, \dots, x_q)$, in terms of a new set of uncorrelated variables, $y^T = (y_1, \dots, y_n)$, each of which is a linear combination of the x variables. The new variables are derived in decreasing order of “importance” in the sense that y_1 accounts for as much of the variation in the original data as possible amongst all linear combinations of x . Then y_2 is chosen to account for as much of the remaining variation as possible, subject to being uncorrelated with y_1 , and so on. The new variables defined by this process, y_1, \dots, y_n , are the principal components (Figure 4).

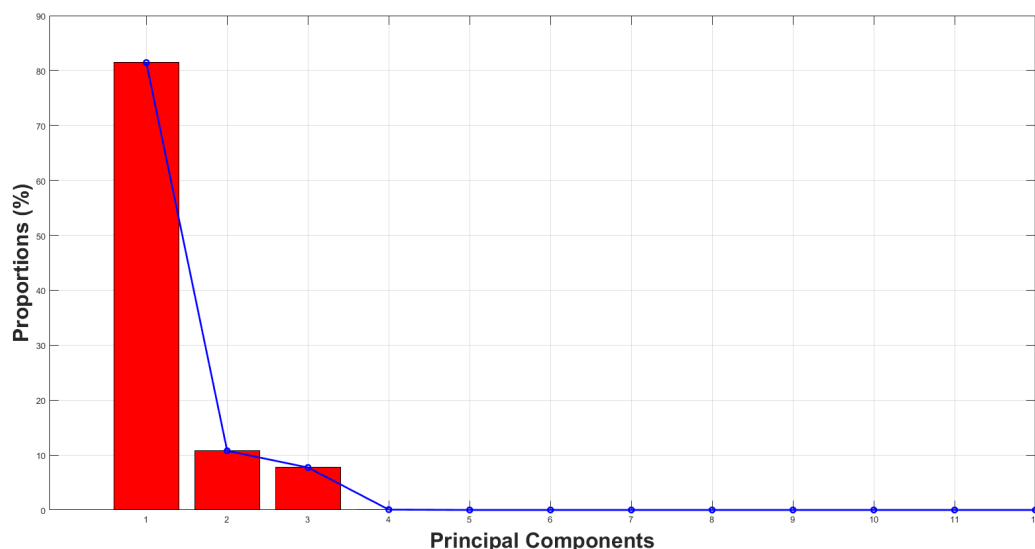


Figure 4. Principal components’ proportions.

The general hope of principal components analysis is that the first few components will account for a substantial proportion of the variation in the original variables and can, consequently, be used to provide a convenient lower-dimensional summary of these variables that might prove useful for a variety of reasons. The first principal component of the observations is that linear combination of the original variables whose sample variance is greatest amongst all possible such linear combinations (Equation (3)). The second principal component is defined as that linear combination of the original variables that accounts for a maximal proportion of the remaining variance subject to being uncorrelated with the first principal component. Subsequent components are defined similarly.

$$y_i = \sum_{j=1}^q a_{ij}x_j \quad (3)$$

In order to perform a PCA, the following steps have to be carried out:

1. calculation of the covariance matrix of the normalized experimental data;
2. evaluation of the coefficients a_{ij} of the principal components as eigenvectors of the covariance matrix;

3. calculus of the Principal Components y_i as a linear combination of the eigenvectors with normalized experimental data;
4. evaluation of the proportions, as in Equation (4):

$$Prop_i\% = \frac{\lambda_i}{\sum \lambda} \quad (4)$$

where λ_i is the eigenvalue of the correlation matrix.

The PCA made it possible to understand which variables are able to explain the variance of the studied dataset. In the second approach proposed in this work, Principal Components will be used as starting variables to evaluate a multiple linear regression.

2.3. Multiple Linear Regression

Linear Regression is probably one of the most powerful and useful tools available to the applied statistician. This method uses one or more variables to explain the values of another. Statistics alone cannot prove a cause-and-effect relationship but it can show how changes in one set of measurements are associated with changes in the average values in another. With this approach, the data analyst specifies which of the variables are to be considered explanatory and which are the responses to these [45].

This process requires a good understanding of the data and a preliminary study of them. A regression model does not imply a cause-and-effect relationship between the variables. Even though a strong empirical relationship may exist between two or more variables, this cannot be considered evidence that the regressor variables and the response are related in a cause-and-effect manner. To establish causality, the relationship between the regressors and the response must have a basis outside the sample data; for example, the relationship may be suggested by theoretical considerations. The regression equation is only an approximation to the true functional relationship between the variables of interest. These functional relationships are often based on physical or other engineering or scientific theory, that is, knowledge of the underlying mechanism [32].

A regression model that involves more than one regressor variable is called a multiple regression model. The term "linear" is used because the general equation of this kind of model (Equation (5)) is a linear function of k regressors $\beta_0, \beta_1, \dots, \beta_k$ and not because it is a linear function of the x s.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (5)$$

This model describes a hyperplane in the k -dimensional space of the regressor variables x_j [32]. Models that include interaction effects may also be analyzed by multiple linear regression methods. For example, suppose that the model is expressed as in Equation (6).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e \quad (6)$$

If $x_3 = x_1 x_2$ and $\beta_3 = \beta_{12}$, then Equation (6) can be written as Equation (7).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \quad (7)$$

which is a linear regression model.

It is more convenient to deal with multiple regression models if they are expressed in matrix notation (Equation (8)). This allows a very compact display of the model, data, and results. y_i is a random variable (response variable) with a distribution that depends on a vector of known explanatory values x_i (predictor or regressor).

$$Y = X\beta + e \quad (8)$$

where the independent, normally distributed errors e_i have zero means and constant variances σ^2 . Additionally, it usually assumes that the errors are uncorrelated. This means

that the value of one error does not depend on the value of any other error. The different terms presented in Equation (8) can be described as follows:

- Y is an $n \times 1$ vector:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- X is an $n \times (k + 1)$ matrix:

$$\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

- β is an $(k + 1) \times 1$ vector:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

- e is an $n \times 1$ vector:

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

The regression coefficients β are parameters that need to be estimated from the observed data (y_i, x_i) . This process is also called fitting the model to the data or “training”. Thus, regression analysis is an iterative procedure in which data lead to a model and a fit of the model to the data is produced. The quality of the fit is then investigated, leading either to modification of the model or the fit or to the adoption of the model.

2.4. Model Performance Indicators

In order to assess the goodness of a regression model, it is necessary to define some specific indices able to examine how well the model represents the data from which it is derived and to what extent it is possible to use the model for predictive purposes. This type of analysis is known as model validation and may be carried out with different types of statistical tools. In particular, in this research activity, three different indicators have been considered with the aim to compare the results obtained for the approaches implemented.

2.4.1. Coefficients of Determination

The coefficient of determination, R^2 , represents the proportion of the variation in the dependent variable “explained” by variation in the independent variables [46]. The sum of the squared deviations of the dependent variable about its mean (Equation (9)—the “total” variation in the dependent variable) can be broken into two parts, called the “explained” variation (Equation (10)—the sum of squared deviations of the estimated values of the dependent variable around their mean) and the “unexplained” variation (the sum of squared residuals). R^2 is measured either as the ratio of the “explained” variation to the “total” variation or, equivalently, as minus the ratio of the “unexplained” variation to the “total” variation, and thus represents the percentage of variation in the dependent variable “explained” by variation in the independent variables.

The total variation in the dependent variable y about its mean is called SST (the Total Sum of Squares) and it is evaluable as in Equation (9).

$$SST = \sum_n (y_i - \bar{y})^2 \quad (9)$$

The “explained” variation, the sum of squared deviations of the estimated values of the dependent variable about their mean, is called SSR (the Sum of Squares due to Regression), and its formulation is shown in Equation (10).

$$SSR = \sum_n (\hat{y}_i - \bar{y})^2 \quad (10)$$

The “unexplained” variation, the sum of squared residuals, is called SSE (the Sum of Squares Error). After these definitions, is it possible to determine the coefficient of determination R^2 , expressed by Equation (11):

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (11)$$

Because $0 \leq SSR \leq SST$, it follows that $0 \leq R^2 \leq 1$. Values of R^2 that are close to 1 imply that most of the variability in y is explained by the regression model. The statistic R^2 should be used with caution since it is always possible to make R^2 large by adding enough terms to the model.

The R^2 statistic adjusted to account for degrees of freedom is called the “adjusted R^2 ” or R_{adj}^2 . It is derived from Equation (11). Estimation of these variances involves corrections for degrees of freedom, yielding Equation (12) [46].

$$R_{adj}^2 = R^2 - \left[\frac{k-1}{n-k} \right] (1 - R^2) = 1 - \left[\frac{n-1}{n-k} \right] (1 - R^2) \quad (12)$$

where k is the number of independent variables and n is the number of observations.

2.4.2. Root-Mean-Squared Residual Error

Root Mean Square Error ($RMSE$) is a standard way to measure the error of a model in predicting quantitative data [35]. Formally it is defined as in Equation (13).

$$RMSE = \sqrt{\sum_n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (13)$$

$RMSE$ is always non-negative, and a value of 0 (almost never achieved in practice) would indicate a perfect fit to the data. The quadrature formulation in $RMSE$ (squaring the differences and then taking the square-root) emphasizes the contribution of the larger differences. This is the same functional form as standard deviation σ , so it is directly comparable to that quantity.

3. Database

The data used for the construction of the database come from the acquisitions provided by an industrial motorsport partner and collected during a race event; indeed, all the data have been normalized for confidentiality reasons. The reference vehicle is a single-seater vehicle, equipped with high-performance tires, available with several different tread compounds. All the data have been acquired in dry conditions since the reference tires are all slick tires. For each event, road roughness profiles, vehicle telemetry data, and tire compound properties were provided and the wear levels of the individual sets of tires used were disclosed. This enormous amount of data has made it possible to obtain a large database that can be manipulated according to the model to be built. It was decided to process the data in order to extrapolate KPIs relating to individual runs, performing a series

of different operations on the starting acquired data. All the data have been acquired with high-precision sensors, among the best performing on the market, with an adequate level of accuracy for the purpose of this research.

Subsequently, efforts were made to refine the data by removing outliers, which refer to observations that deviate from the established pattern exhibited by other data points. The motivation behind identifying outliers is the recognition that they can significantly impact the resulting estimates, potentially in an undesirable manner. It is crucial to thoroughly examine outliers in order to determine if any plausible explanation exists for their atypical behavior. Possible factors contributing to outliers include measurement or classification errors by sensors, as well as erroneous data entry into the computer system.

Most multivariate datasets can be represented in the same way, namely in a rectangular format known from spreadsheets, in which the elements of each row correspond to the variable values of a particular observation of a single wheel and the elements of the columns correspond to the values taken by a particular variable. Data can be written in such a rectangular format as in Table 1.

Table 1. Multivariate dataset.

Observation	Variable 1	...	Variable q
1	$x_{1,1}$...	$x_{1,q}$
\vdots	\vdots	\vdots	\vdots
n	$x_{n,1}$...	$x_{n,q}$

n is the number of units, q is the number of variables recorded on each observation, and $x_{i,j}$ denotes the value of the j th variable for the i th unit. The observation part of the table above is generally represented by an $n \times q$ data matrix.

The global amount of data, organized by runs, has been divided into two different datasets; the largest one has been used for the training and validation phase of the models, whereas the second one has been adopted to assess the models' performance during the testing phase. The number of observations for these two databases is reported in Table 2.

Table 2. Training and testing database—number of observations.

Dataset	N° Observations
Run Training Database	200
Run Testing Database	80

Taking into account the different areas of analysis and the information necessary to characterize the single observation, the organization of the dataset was modeled to improve its use, highlighting the three categories of KPIs:

- Road Roughness Indices;
- Telemetry Indices;
- Tire Viscoelastic Indices.

3.1. Road Roughness Indices

Surface roughness has a big impact on several physical phenomena and plays a fundamental role in the wear of sliding surfaces. For this reason, its characterization is also a fundamental step in this research activity. Nowadays, available optical surface profiling instruments allow for a detailed measurement of surface roughness covering several length scales. This enables the validation of a mathematical statistical description of pavement texture within the framework of self-affine surfaces and hence provides a holistic characterization of surface roughness covering several length scales within a few characteristic parameters. Generally, when talking about road texture, it is possible to describe its complexity by distinguishing the two most relevant scales: macro-roughness and micro-roughness. The first one represents the surface irregularities of longer wavelengths. The micro

one, instead, is produced by fluctuations of short wavelengths characterized by asperities (local maxima) and valleys (local minima) of varying amplitudes and spacing [47].

The roughness measurements contained in the reference databases of this study have been collected on several locations on a series of different tracks. The roughness KPIs considered for a specific circuit are obtained as mean values of those evaluated on the various locations of the same track. In order to extract this relevant information, different techniques commonly adopted in the literature have been applied. First of all, each profile is studied with a statistical analysis, from which it is possible to obtain three relevant indicators [48]. Considering a profile, $z(x)$, profile heights are measured from a reference line. The center line, or mean line, is defined such that the area between the profile and the mean line above the line is equal to that below the mean line. The first KPI is the arithmetic mean of the absolute values of vertical deviation from the mean line through the profile, also known as R_a , evaluable as expressed in Equation (14).

$$R_a = \frac{1}{L} \int_0^L |z - m| dx \quad (14)$$

where $m = \frac{1}{L} \int_0^L z dx$ and L is the sampling length of the profile.

By studying the shape of the probability density function, it is also possible to define two other statistical height descriptors: skewness and kurtosis [49,50]. The first one represents the degree of symmetry of the distribution function, while the latter refers to the peakedness of the distribution and it is a measure of the degree of pointedness or bluntness of a distribution function. They are defined as in Equations (15) and (16), respectively.

$$Sk = \frac{1}{\sigma^3 L} \int_0^L (z - m)^3 dx \quad (15)$$

$$Ku = \frac{1}{\sigma^4 L} \int_0^L (z - m)^4 dx \quad (16)$$

in which σ is the square root of the arithmetic mean of the square of the vertical deviation from the mean line. For the complete characterization of a profile or a surface, the parameters just discussed are not sufficient [29]. These parameters are seen to be primarily concerned with the relative departure of the profile in the vertical direction only; they do not provide any information about the slopes, shapes, or sizes of the asperities or about the frequency and regularity of their occurrence. To complete the profile characterization it is possible to evaluate some spatial parameters, such as the wavelength. Although the profile contains a wide, continuous range of wavelengths, the macro wavelength of the profile can be defined as in Equation (17).

$$\lambda = 2\pi \frac{\int_0^L |z(x) - m| dx}{\int_0^L |z(x)| dx} \quad (17)$$

To properly parameterize the profile from a spatial point of view, exploring also the micro-scale, two further techniques, based on the concept of self-affine surfaces [51,52], have been adopted: the height different correlation function (HDC) [53,54] and the power spectral density (PSD) [55,56]. Thanks to the application of these two methods, it is also possible to evaluate the correlation length, ζ_{\perp} , also known as peakyness, normal to the surface:

$$\sigma^2 = \frac{\zeta_{\perp}^2}{2} \quad (18)$$

To summarize, applying different strategies to the acquired road profile, it is possible to obtain a series of KPIs specific for each circuit, as indicated in Table 3.

Table 3. Road roughness indices.

Symbol	Name	Unit of Measure
λ_{Macro}	Macro Wavelength	mm
λ_{micro}	micro Wavelength	mm
H	Hurst Coefficient	(–)
Ra_{Macro}	Center Line Average Macro	mm
Ra_{micro}	Center Line Average Micro	mm
ξ_{\perp}	Peakyness	mm
Sk	Skewness	(–)
Ku	Kurtosis	(–)

3.2. Telemetry Indices

Motorsport cars are equipped with hundreds of sensors, able to monitor several aspects related to vehicle performance. Combining the raw acquisitions with other relevant parameters of the vehicle or among them, it is possible to obtain more and more information about the real operating conditions of the vehicle during races. The cars can be seen, in this way, as rolling sensor networks, constantly gathering and transmitting information about the car and driver to the racing team. In this context, for the presented activity, telemetry data have been analyzed and pre-processed in order to extract relevant information to be used in the predictive models. In particular, vehicle and tire velocities, tire temperatures and inflation pressure, external air and track temperatures, and wheel camber variation have been considered starting from the acquisitions of dedicated sensors. In addition, tire interaction forces, contact patch extensions, and slip indices have been evaluated thanks to mathematical models starting from the acquired channels. For what concerns the slip indices, it is possible to define a longitudinal slip ratio and a lateral slip angle, as described in Equations (19) and (20), respectively.

$$s_x = \frac{v_x - \Omega R}{v_x} \quad (19)$$

$$s_y = \frac{v_y}{v_x} = \tan(\alpha) \approx \alpha \quad (20)$$

in which v_x is the longitudinal component of the wheel center velocity, v_y is the lateral component of the wheel center velocity, R is the pure rolling radius, and Ω is the angular velocity of the wheel. The lateral slip angle is often also indicated with α . Considering only the numerators of Equations (19) and (20), it is possible to isolate the so-called sliding velocities along the longitudinal and lateral direction, reported in Equations (21) and (22).

$$v_{s_x} = v_x - \Omega R = s_x v_x \quad (21)$$

$$v_{s_y} = s_y v_x \quad (22)$$

By combining the information relating to the interaction forces and the sliding velocities, it is possible to introduce the *friction power* calculated along the longitudinal and lateral equation, as shown in Equations (23) and (24).

$$FP_x = F_x v_{s_x} \quad (23)$$

$$FP_y = F_y v_{s_y} \quad (24)$$

These quantities instantly define the thermal power due to the pneumatic–soil interaction or better, to the tangential stresses that, in the contact area in sliding with respect to the

ground, perform work that is dissipated in heat. This power will hereinafter be referred to as friction power and it will be indicated with FP (Equation (25)).

$$FP_{tot} = FP_x + FP_y \quad (25)$$

To build the database and synthesize the amount of data collected during a run in a single representative value, two different methods have been adopted:

- For the quantities representing energies or powers, the sum of which still has physical meaning, the sum of these quantities was evaluated during the entire run.
- All the other quantities, instead, have been reduced by considering their average value and the standard deviation during the entire run.

Finally, after these operations, the different KPIs related to telemetry data have been defined, as reported in Table 4.

Table 4. Telemetry indices.

Symbol	Name	Unit of Measure
$T_{Surf\ Avg}$	Average Tire Surface Temperature	°C
$T_{Surf\ Std}$	Tire Surface Temperature Standard Deviation	°C
$T_{Inner\ Avg}$	Average Tire Inner Liner Temperature	°C
$T_{Inner\ Std}$	Tire Inner Liner Temperature Standard Deviation	°C
$T_{Road\ Avg}$	Average Track Temperature	°C
$T_{Ambient\ Avg}$	Average External Air Temperature	°C
$F_x\ Avg$	Average Longitudinal Force	N
$F_x\ Std$	Longitudinal Force Standard Deviation	N
$F_y\ Avg$	Average Lateral Force	N
$F_y\ Std$	Lateral Force Standard Deviation	N
$F_z\ Avg$	Average Vertical Force	N
$F_z\ Std$	Lateral Force Standard Deviation	N
$V_{s_x\ Avg}$	Average Longitudinal Sliding Velocity	m/s
$V_{s_x\ Std}$	Longitudinal Velocity Standard Deviation	m/s
$V_{s_y\ Avg}$	Average Lateral Sliding Velocity	m/s
$V_{s_y\ Std}$	Lateral Velocity Standard Deviation	m/s
FP_x	Longitudinal Friction Power	W
FP_y	Lateral Friction Power	W
FP_{TOT}	Total Friction Power	W
$SlidDist$	Sliding Distance	m
γ_{Avg}	Average Dynamic Camber	Deg
p_{Avg}	Average Tire Inflation Pressure	Pa
CP_{Area}	Contact Patch Area	mm ²
$Dist_{TOT}$	Total Run Distance	m

3.3. Tire Viscoelastic Indices

As it is well-known, tires are a composite material made mainly of rubbery material. The external component of the tire, the tread, is subjected to violent stresses to maintain contact with the road. Its wear is one of the main sources of decay of vehicle performance and, for

this reason, it is essential to understand how its properties vary while the vehicle is moving. Like all rubbery materials, the tire tread also has to be studied as a viscoelastic material.

A viscoelastic material is a deformable material with a behavior that lies between a viscous liquid and an elastic solid. This kind of solid does not show a linear relationship between stress and applied strain. Indeed, their behavior deviates from Hooke's law and exhibits elastic and viscous characteristics at the same time. The typical response of viscoelastic materials is characterized by a strong dependence on the rate of straining $d\epsilon/dt$; the faster the stretching, the larger the stress required. The effect of stretching shows that the viscoelastic materials depend on time; the most generic equation that describes this feature is Newton's Law [31]:

$$\sigma = \eta \frac{d\epsilon}{dt} \quad (26)$$

Newton's Law shows the connection between the stress and the strain rate through the viscosity coefficient η ; every material which satisfies Equation (26) can be classified as a *viscoelastic* material. It is fundamental to underline that viscoelastic materials, after removing any deforming force, return to their original shape after a certain time period; contrariwise, when a perfectly elastic solid is subjected to a force, it distorts instantaneously in proportion to the applied load.

For viscoelastic materials, it can be assumed that sinusoidal stresses or strains of constant frequency are applied to a sample until a steady sinusoidal strain or stress results, with a fixed phase angle between the input and the output (Figure 5). For example, for a sinusoidal shear strain expressed in Equation (27), where ϵ_0 is the strain amplitude and ω is the angular frequency, the stress σ will oscillate sinusoidally, as shown in Equation (28).

$$\epsilon(t) = \epsilon_0 \sin(\omega t) \quad (27)$$

$$\sigma(t) = \sigma_0 \sin(\omega t + \delta) \quad (28)$$

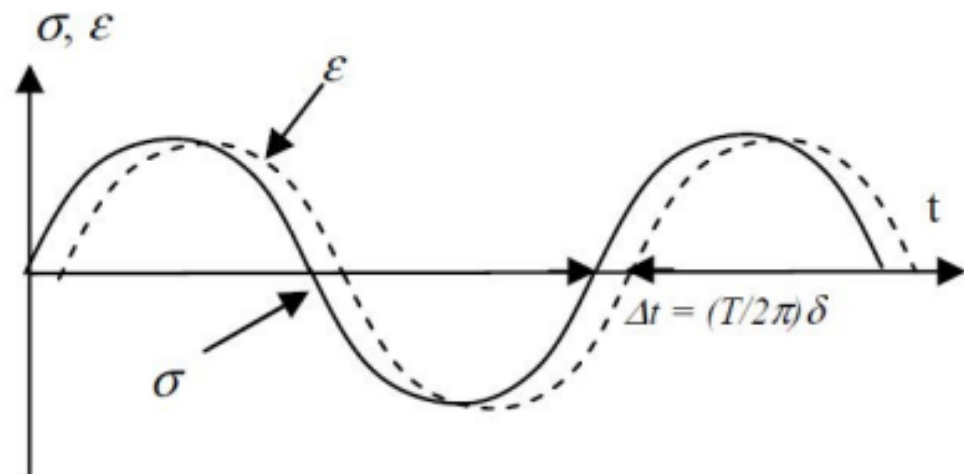


Figure 5. Strain–stress phase.

Using the trigonometric formula, Equation (28) can be rewritten as in Equation (29).

$$\sigma(t) = \epsilon_0 [E'(\omega) \sin(\omega t) + E''(\omega) \cos(\omega t)] \quad (29)$$

In the last formula, $E'(\omega) = (\sigma_0/\epsilon_0)\cos(\delta)$ is the storage modulus, while $E''(\omega) = (\sigma_0/\epsilon_0)\sin(\delta)$ is the loss modulus. The first is a measure of the elastic energy stored and recovered, while the second is a measure of the dissipated energy as heat in cyclic deformation. The ratio $E''(\omega)/E'(\omega)$ is the $\tan(\delta)$, also called the loss factor, where δ is the phase angle by which the strain lags behind the applied stress.

The modulus, the energy loss, and hysteresis of a viscoelastic material change in response to two parameters: the frequency with which the force is applied and the temperature, which produce opposite effects on the rubber. In particular, whenever the stress frequency is increased at a fixed temperature, the polymer appears in a glassy state; conversely, if the material heats up at a given stress frequency, it becomes softer and the compound exhibits rubbery behavior.

To evaluate the properties of a generic viscoelastic material in different conditions of temperature and frequency, it is possible to apply the Time-Temperature Superposition Principle (TTSP) theory [57] combined with the William–Landel–Ferry (WLF) equation [58]. In order to represent the storage modulus (E') curves at a higher or lower frequency than the reference frequency ω_0 , the shift factor, a_T , is evaluated through the empirical WLF Equation (30).

$$\log(a_T) = \log\left(\frac{\omega_0}{\omega}\right) = \frac{-C_1(T - T_0)}{C_2 + (T - T_0)} \quad (30)$$

In Equation (30), the following parameters appear:

- T_0 is the reference temperature for the storage modulus master curve;
- T is an arbitrary temperature;
- ω_0 is the reference frequency associated with the reference temperature T_0 ;
- ω is the new frequency;
- C_1 and C_2 are empirical constants adjusted to fit the values of the shift factor (a_T).

On the other hand, if the purpose is to evaluate the storage modulus at different temperatures than the reference ones, Equation (31) can be used.

$$T = T_0 - \frac{C_2 \log\left(\frac{\omega_0}{\omega}\right)}{C_1 + \log\left(\frac{\omega_0}{\omega}\right)} \quad (31)$$

In this activity, the viscoelastic properties of the different compounds available have been evaluated by adopting an innovative non-destructive technique, called VESevo [59–61]. Thanks to this innovative device, it has been possible to determine the master curve of the storage modulus and of the loss factor for the various typologies of tires. In this way, knowing the solicitation frequency and subsequently, the shifted temperature according to WLF equations, it was possible to obtain the values of the storage modulus and loss factor in specific real working conditions representative of the entire run. The viscoelastic indices reported in Table 5 have been calculated definitively for each compound.

Table 5. Viscoelastic indices.

Symbol	Name	Unit of Measure
f_{Hz}	Solicitation Frequency	Hz
T^*	Shifted Temperature	°C
$E'(T^*)$	Storage Modulus at the Shifted Temperature	MPa
$\tan(\delta)(T^*)$	Loss Factor at the Shifted Temperature	(–)

3.4. Wear Measurements

After the definition of the KPIs contained in the databases, it is necessary to spend a few words about the wear measurements and their pre-processing. The wear data were provided in the form of depth detected by reading inspection holes on the tread at the end of each run. For each tire used, the wear measurements come out reading the tread depth of each available spot thanks to a professional tire tread depth tool, widely used in the motorsport context, averaging several measures for each spot in order to obtain reliable wear evaluation. Since these spots are not equally spaced in the tread width, the acquired values have been interpolated. In this way, it has been possible to obtain a “virtual” value of the wear for each rib (Figure 6). This operation allows us to obtain wear values on the “virtual” ribs that can be interpreted as volumetric loss of material, since in this case, the

spots are equally spaced. The values thus obtained have been processed in order to obtain an average wear value along the entire length of the tread ($Wear_{Mean}$); this quantity is expressed in percentage terms with respect to the original thickness of the tread.

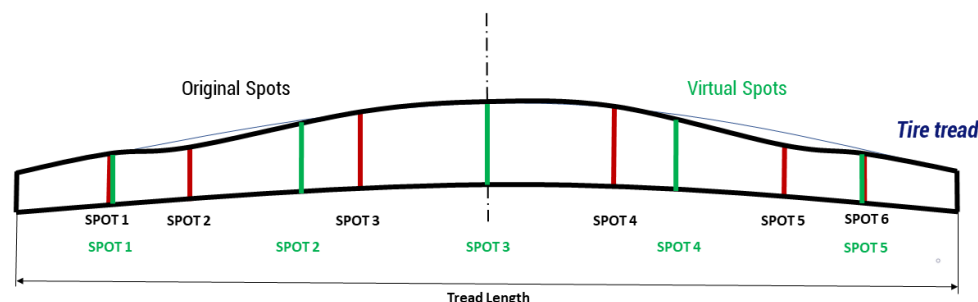


Figure 6. Wear measurements—schematic tread spots.

4. Results

In this section, the different methods, already presented in Section 2, have been applied to the available databases, in order to build proper tire tread wear prediction models. For each approach, the performance of the regression model has been assessed for the different construction phases: training, validation, and testing.

4.1. Feed Forward Neural Network

The Feed-Forward Neural Network has been designed with two hidden layers, one input layer and one output layer. The choice of two hidden layers was guided by various factors including the desire to not overcomplicate the neural network, both due to the small number of observations available and to avoid long training times; moreover, using additional hidden layers would certainly have caused overfitting problems. The number of neurons in the first hidden layer has been chosen as equal to the number of inputs. Regarding the number of neurons in the second hidden layer, a looping mechanism has been inserted in the NN to find the best combination among the neurons of the two hidden layers and to optimize the work of the network thanks to subsequent trainings triggered by the looping mechanism. Downstream of the looping mechanism, the network was taken with the pair of neurons that minimize the error function, in our case, the mean square error (MSE), with which the performance of the network is evaluated during the training. This network with the optimal pair of neurons was trained again to see if the performance was able to increase with another training. In the case of a positive response, the last trained net was preferred; otherwise, it was returned to the previous one. Finally, the best network was used to make the prediction. In addition, it was noted that networks with a high number of neurons always presented an overfitting problem, showing excellent results on the training dataset, but a poor ability to predict and generalize the problem.

To train the network, the settings reported in Table 6 have been used to optimize the generalization and speed of the net as well as the results.

Table 6. Feed-forward Neural Network settings.

Neural Network Settings	
Training Algorithms	Levenberg–Marquadrat [62]
Training Dataset	80%
Validation Dataset	20%
Number of Epochs	100
Validation Checks	50

The larger database was randomly divided into two parts: 80% was used for the training phase, while the remainder was used for validation. In addition, the testing database was used to evaluate the predictive performance of the network in the testing

phase. The number of epochs is the number of presentations to the network of all patterns of the training set, and it was chosen to find the best trade-off between the quality of the results and the NN processing time. The validation checks are used to stop the training when the validation subset error rate increases continuously for more than 50 epochs in the presented case. This has the goal of limiting as much as possible the overfitting phenomena that could compromise the generalization.

Finally, in order to reduce the number of inputs for the NN, a feature selection has been performed, taking into consideration only the KPIs that show a regression coefficient with the target, the mean wear, higher than 0.15 and that are not directly correlated to each other. In this way, twelve indicators have been selected, as illustrated in Table 7.

Table 7. Neural network inputs.

Symbol	Name	Unit of Measure	Working Range
λ_{Macro}	Macro Wavelength	mm	$3 \div 6$
Ra_{Macro}	Center Line Average Macro	mm	$4 \div 7$
ζ_{\perp}	Peakyness	mm	$0.5 \div 2$
Sk	Skewness	(-)	$-1.5 \div 0$
Ku	Kurtosis	(-)	$2.5 \div 6$
$T_{Surf Avg}$	Average Tire Surface Temperature	$^{\circ}C$	$55 \div 85$
$T_{Inner Avg}$	Average Tire Inner Liner Temperature	$^{\circ}C$	$90 \div 135$
FP_{TOT}	Total Friction Power	W	$10^{10} \div 10^{11}$
p_{Avg}	Average Tire Inflation Pressure	Pa	$80,000 \div 15,000$
$Dist_{TOT}$	Total Run Distance	m	$30,000 \div 250,000$
$E'(T^*)$	Storage Modulus at the Shifted Temperature	MPa	$30 \div 350$
$\tan(\delta)(T^*)$	Loss Factor at the Shifted Temperature	(-)	$0.20 \div 0.35$

After several attempts, the best network architecture has been found, with a number of neurons for the layers of 12–12–13–1, as reported in Figure 7.

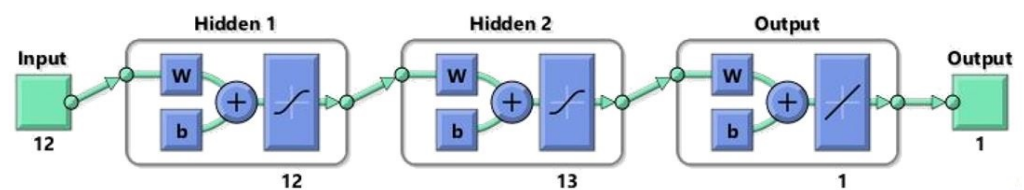


Figure 7. Neural network architecture.

Despite the good results obtained in the training phase (Figure 8), the network is not able to complete the training because it quickly reached an overfitting condition, overcoming the number of maximum validation checks acceptable. This is mainly related to the training database composition, in which the different wear levels are not explored homogeneously.

This is particularly evident for the high level of wear, demonstrated by the observations for which the network commits the bigger errors, as visible from Figure 9. The inability of the network to correctly predict high wear values makes it an ineffective tool, given the high RMSE values detected in the predictions (Table 8).

Another important limitation of this approach is that the network does not directly provide feedback on the most relevant KPIs, making structured improvement of the approach difficult. This strategy is the most easily implementable of the three proposed, but

to obtain stable results it would be necessary to have a much larger and more homogeneous dataset available.

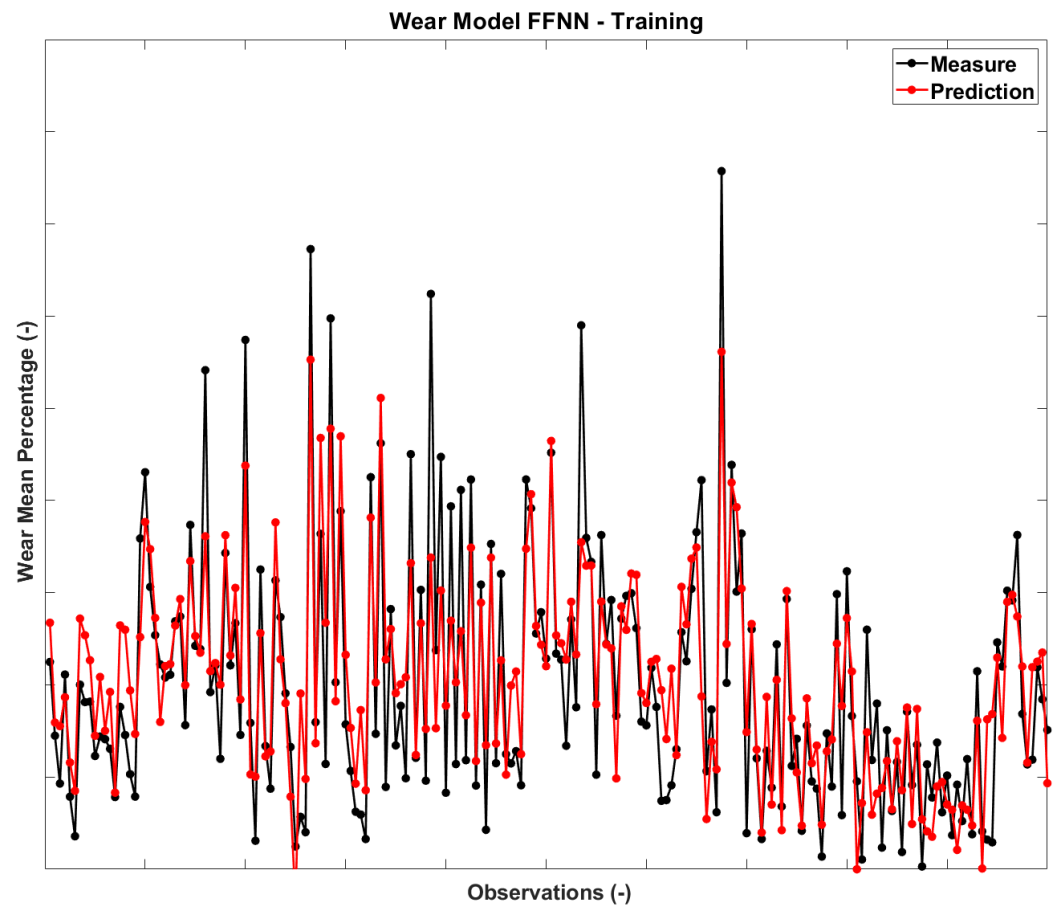
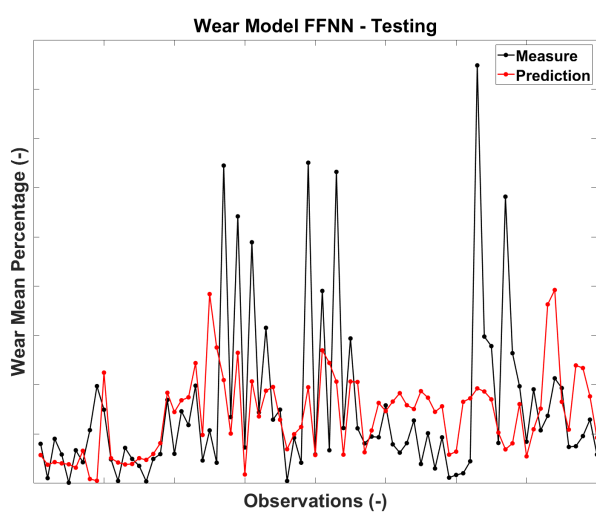
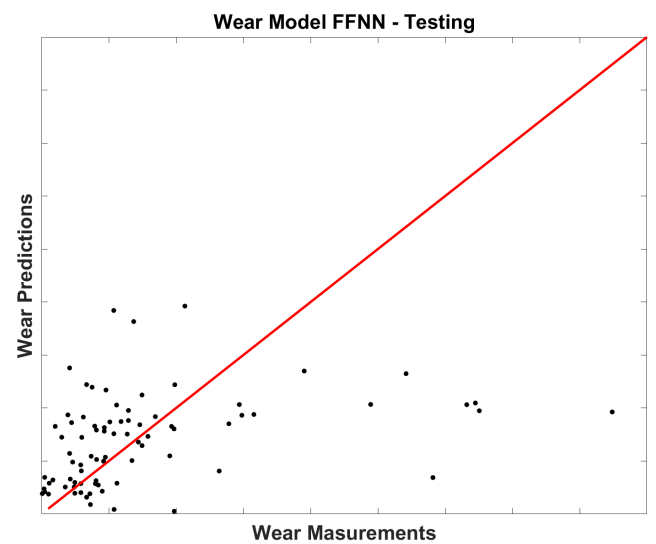


Figure 8. Neural network training.



(a) Neural Network Testing



(b) Measures vs. Predictions

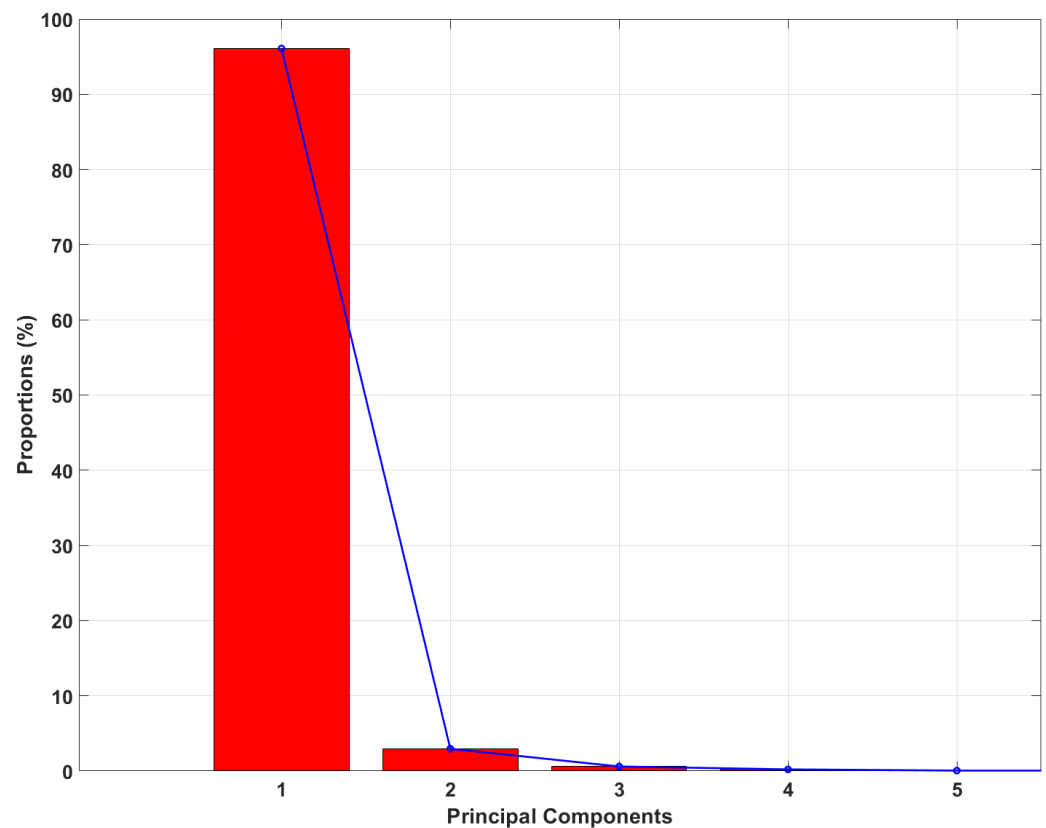
Figure 9. Neural network testing.

Table 8. Feed-Forward Neural Network performance indicators.

Model	R^2	R^2_{adj}	RMSE
FFNN—Training	0.7301	0.7144	7.1456
FFNN—Testing	0.2492	0.1278	16.062

4.2. Principal Component Analysis and Multiple Linear Regression

As already described in Section 2, Principal Component Analysis gives the opportunity to reduce the dimensionality of the database, maintaining its global variance. The preliminary operation to perform a PCA is to investigate the correlations among the input variables in order to select only the uncorrelated inputs; this operation is a bit different from the one performed with the NN, in which the target is also involved in this pre-processing stage. After this operation, it has been possible to define seventeen input variables that did not show mutual correlations. Thanks to the application of the PCA, it has been possible to reduce the dimensionality, identifying three relevant Principal components able to synthesize almost more than 99% of the variance of the starting training dataset, as shown in Figure 10.

**Figure 10.** Principal Component Analysis.

After that, the three Principal Components have been used as input to build a multiple linear regression. The predictions made with the regressors evaluated, for both the training and the testing dataset, are illustrated in Figure 11.

As is possible to see from the two last plots, this regression model is more conservative than the one identified with neural networks, always estimating intermediate levels of wear, with a high forecast error for both lightly and heavily worn tires. This is also highlighted by the low values of R^2 and R^2_{adj} reported in Table 9.

Despite the more detailed analysis of the starting database, this approach does not show acceptable results in predicting wear. Furthermore, the manipulation of the original database is complex and dependent on the knowledge of the coefficients that identify the principal components.

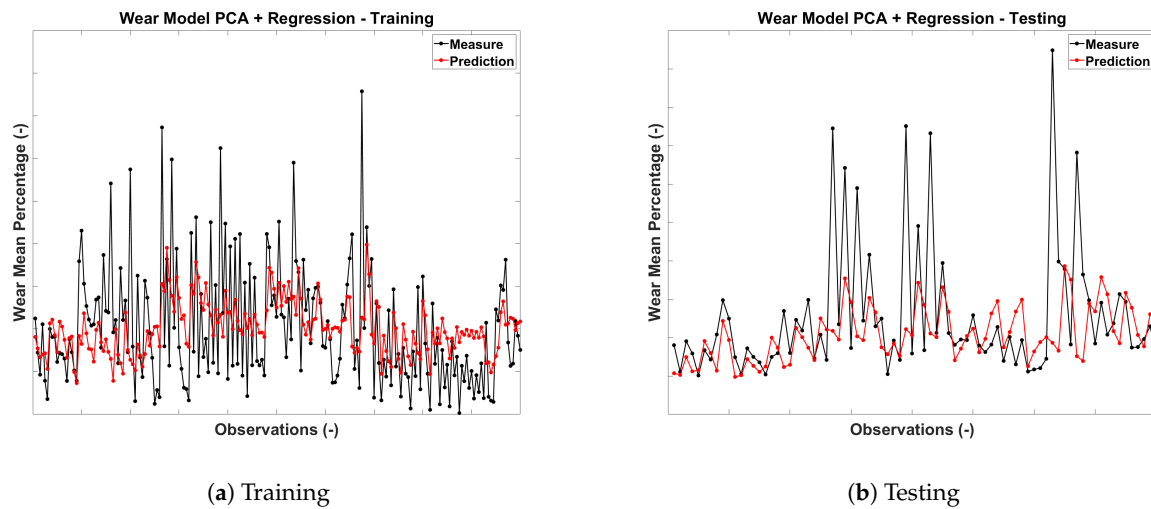


Figure 11. PCA + Regression.

Table 9. PCA + regression performance indicators.

Model	R^2	R^2_{adj}	RMSE
PCA + Regression—Training	0.1961	0.1262	12.332
PCA + Regression—Testing	0.1710	−0.0395	17.103

4.3. Physical Correlations and Multiple Linear Regression

In the last approach proposed, before developing the linear multiple regression model, a series of database investigations were carried out to find correlations between the physical variables, in order to more robustly define the set of inputs with which to define the regression.

First of all, considering the information provided, it was realized that in several cases, the front wheels show phenomena of irregular wear, which are difficult to quantify using KPIs. In addition, due to the steering dynamics, the front tires were subject to comparable friction power with each other, but not with the rear tires; for these reasons, two separate investigations were conducted, one for the tires of the front axle, the other for the rear. The dependence between mean wear and friction power was then analyzed, considering the accumulated energy dissipated in the entire run. Starting from the Mean Wear Percentage vs. Total Friction Power plot, an interpolating line has been defined for each circuit (Figure 12).

In order to find a robust correlation among wear, friction power, temperature, and road indexes, the following procedure has been carried out:

1. Two reliability indexes have been identified to obtain further information on the robustness of the wear/friction correlation:
 - R_1 —Obtained normalizing the RMSE among the experimental points on the plot Wear vs. FP_{TOT} and the linear fitting; it is used to weight fittings with other performance indicators.
 - R_2 —Obtained normalizing the number of consistent experimental points on the plot Wear vs. FP_{TOT} ; it is used to remove circuits with a lower number of acquisitions from the fittings.
2. The slopes obtained for each circuit were compared with several temperature indicators with the aim to find other correlations.
3. Finally, the residuals between the fitting line and the experimental points obtained from the combination among wear, friction power, and temperature were plotted against the road indexes already evaluated for each circuit.

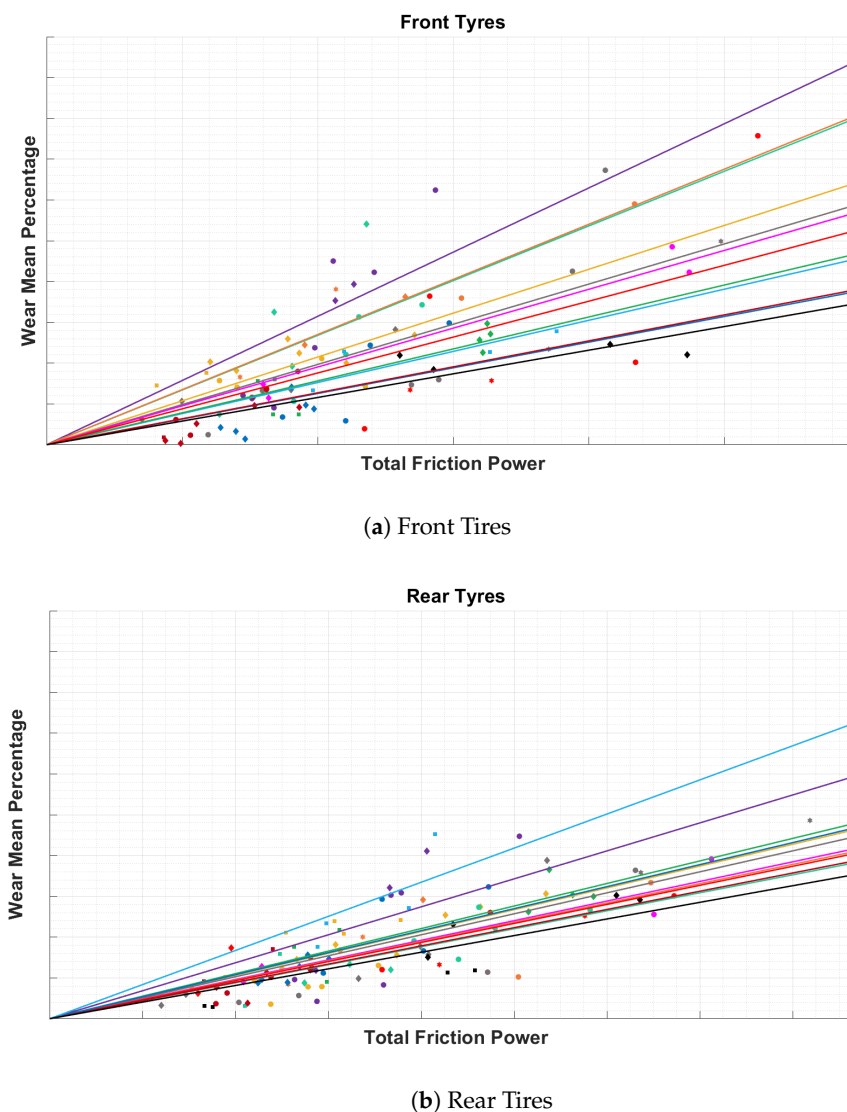


Figure 12. $Wear_{Mean}$ —Each color represents a different track vs. FP_{TOT} .

Thanks to this procedure, a good correlation has been found with the surface average temperature (Figure 13) and, subsequently, it was noticed that residuals show good correlations with road skewness, but only for rear tires (Figure 14). As already mentioned, this means that for the front tires, the wear is also influenced by further phenomena, such as graining or blistering.

Thanks to this preliminary physical correlation, it has been possible to select six inputs in order to build two regression models for front and rear tires. The six inputs are reported in Table 10 and the rear tire regression model shows interesting results for both the training and validation phase (Figure 15).

With this methodology, the rear regression model shows the best performance if compared with the other model presented (Table 11); this is mainly due to the preliminary analysis carried out on the database, which made it possible to highlight the functional relationships between the various key performance indicators. It is important to underline that within the inputs of the regressive model, there are indicators of all three relevant areas analyzed. As regards the model developed for the front wheels, the result is in line with the models previously analyzed, with slightly higher R^2 and R^2_{adj} values.

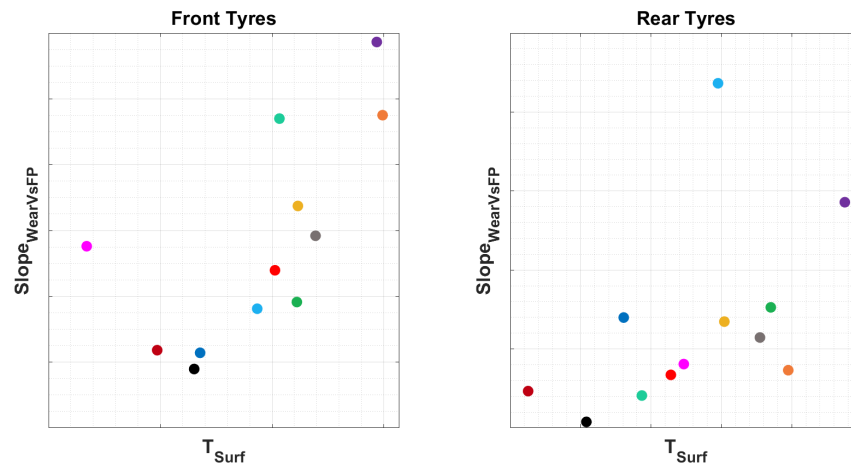


Figure 13. $Wear_{Mean} - FP_{TOT}$ Slope vs. T_{Surf} —Each color represents a different track.

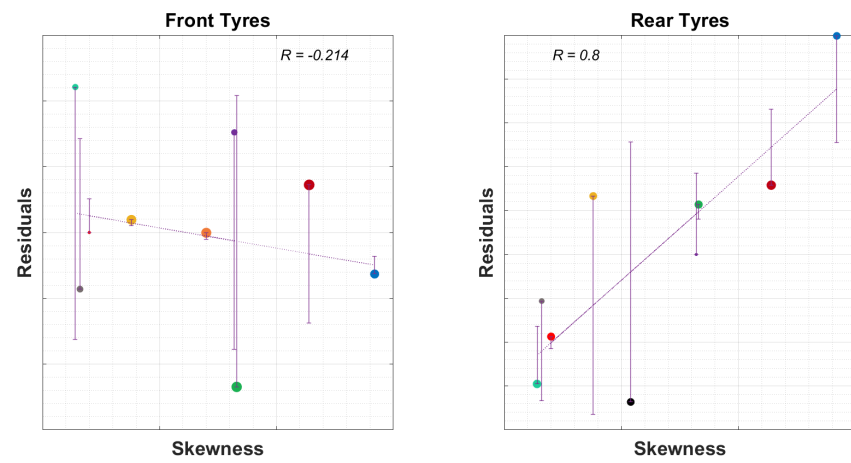
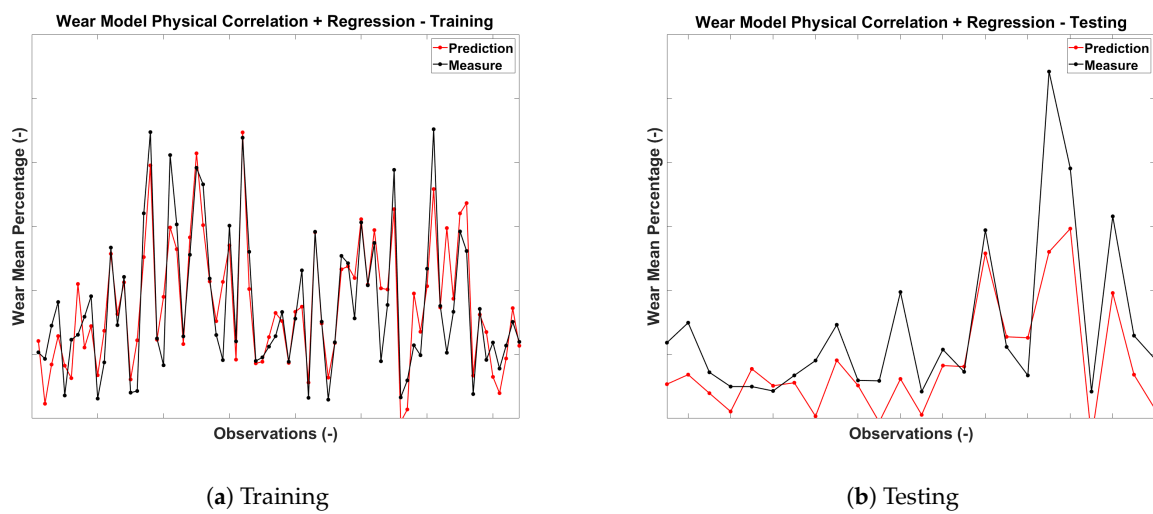


Figure 14. Residuals vs. skewness—Each color represents a different track.



(a) Training

(b) Testing

Figure 15. Physical correlations + regression.

Table 10. Physical correlations + regression inputs.

Symbol	Name	Unit of Measure
Sk	Skewness	(–)
Ku	Kurtosis	(–)
$T_{Surf\ Avg}$	Average Tire Surface Temperature	°C
FP_{TOT}	Total Friction Power	W
CP_{Area}	Contact Patch Area	mm ²
$\tan(\delta)(T^*)$	Loss Factor at the Shifted Temperature	(–)

Table 11. Physical correlations + regression performance indicators.

Model	R^2	R^2_{adj}	RMSE
Front—Correlation + Regression—Training	0.7645	0.7412	8.5303
Front—Correlation + Regression—Testing	0.3381	0.2297	17.6296
Rear—Correlation + Regression—Training	0.8126	0.7830	5.0905
Rear—Correlation + Regression—Testing	0.5242	0.4426	7.1491

For the case study presented, this third approach appears to be the most robust and able to provide indications on the expected level of wear in a predictive manner. Furthermore, this method allows a more conscious application of statistical techniques based on the knowledge of the underlying physical phenomena. The use of a multiple linear regression which directly adopts the KPIs present in the database as inputs makes the procedure easily replicable and implementable in different contexts and scenarios.

5. Conclusions

In this work, three different methodologies to build wear models based on several statistical and machine learning techniques have been presented. By exploiting different types of data, relating to telemetry, road roughness acquisitions, and the viscoelastic properties of the tire, it was possible to build a complex database made up of various performance indicators. By applying the different strategies to the reference datasets, it has been shown that the last approach presented, based on the combination of preliminary physical correlations and the identification of a multiple linear regression model, is the most effective for predicting the level of wear. This approach, despite its simplicity, can quickly provide further information on the expected level of wear in conditions not yet explored experimentally, becoming a useful tool for planning tender strategies. To further strengthen the procedure, it will be necessary to apply it to increasingly large and homogeneous databases in terms of explored wear levels and to define new innovative KPIs for the evaluation of irregular wear conditions.

Author Contributions: Conceptualization, F.F. and F.T.; methodology, A.S. and G.N.D.; software, G.N.D. and G.A.; validation, G.N.D. and F.T.; investigation, A.S. and G.A.; resources, F.F.; data curation, G.N.D.; writing—original draft preparation, G.N.D. and A.S.; writing—review and editing, G.N.D. and F.F.; visualization, G.A.; supervision, F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the project “FASTire (Foam Airless Spoked Tire): Smart Airless Tyres for Extremely-Low Rolling Resistance and Superior Passengers Comfort” funded by the Italian MIUR “Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN)”—grant no. 2017948FEN.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gunaratne, M.; Bandara, N.; Medzorian, J.; Chawla, M.; Ulrich, P. Correlation of tire wear and friction to texture of concrete pavements. *J. Mater. Civ. Eng.* **2000**, *12*, 46–54. [[CrossRef](#)]
2. Kragelsky, I.V.; Dobyichin, M.N.; Komalov, V.S. *Friction and Wear: Calculation Methods*; Elsevier: Amsterdam, The Netherlands, 2013.
3. Pacejka, H. *Tire and Vehicle Dynamics*; Elsevier: Amsterdam, The Netherlands, 2005.
4. Guiggiani, M. *The Science of Vehicle Dynamics*; Springer: Berlin/Heidelberg, Germany, 2018; p. 550.
5. Lowne, R. The effect of road surface texture on tyre wear. *Wear* **1970**, *15*, 57–70. [[CrossRef](#)]
6. Gnecco, E.; Meyer, E. *Fundamentals of Friction and Wear*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
7. Anghelache, G.; Moisescu, R.; Sorohan, Ş.; Bureţea, D. Measuring system for investigation of tri-axial stress distribution across the tyre–road contact patch. *Measurement* **2011**, *44*, 559–568. [[CrossRef](#)]
8. Singh, K.B.; Arat, M.A.; Taheri, S. Literature review and fundamental approaches for vehicle and tire state estimation. *Veh. Syst. Dyn.* **2019**, *57*, 1643–1665. [[CrossRef](#)]
9. Derafshpour, S.; Valizadeh, M.; Mardani, A.; Saray, M.T. A novel system developed based on image processing techniques for dynamical measurement of tire-surface contact area. *Measurement* **2019**, *139*, 270–276. [[CrossRef](#)]
10. De Beer, M.; Fisher, C. Stress-In-Motion (SIM) system for capturing tri-axial tyre–road interaction in the contact patch. *Measurement* **2013**, *46*, 2155–2173. [[CrossRef](#)]
11. Niskanen, A.J.; Tuononen, A.J. Three 3-axis accelerometers fixed inside the tyre for studying contact patch deformations in wet conditions. *Veh. Syst. Dyn.* **2014**, *52*, 287–298. [[CrossRef](#)]
12. Xiong, Y.; Yang, X. A review on in-tire sensor systems for tire-road interaction studies. *Sens. Rev.* **2018**, *38*, 231–238. [[CrossRef](#)]
13. Acosta, M.; Kanarachos, S.; Blundell, M. Virtual tyre force sensors: An overview of tyre model-based and tyre model-less state estimation techniques. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2018**, *232*, 1883–1930. [[CrossRef](#)]
14. Uhlar, S.; Heyder, F.; König, T. Assessment of two physical tyre models in relation to their NVH performance up to 300 Hz. *Veh. Syst. Dyn.* **2021**, *59*, 331–351. [[CrossRef](#)]
15. Menard, K.P.; Menard, N.R. Dynamic mechanical analysis in the analysis of polymers and rubbers. In *Encyclopedia of Polymer Science and Technology*; Wiley: Hoboken, NJ, USA, 2002; pp. 1–33.
16. Genovese, A.; Pastore, S.R. Development of a portable instrument for non-destructive characterization of the polymers viscoelastic properties. *Mech. Syst. Signal Process.* **2021**, *150*, 107259. [[CrossRef](#)]
17. Reye, T. Zur theorie der zapfenreibung. *Civilingenieur* **1860**, *4*, 235–255.
18. Archard, J. Contact and rubbing of flat surfaces. *J. Appl. Phys.* **1953**, *24*, 981–988. [[CrossRef](#)]
19. Schallamach, A. Recent advances in knowledge of rubber friction and tire wear. *Rubber Chem. Technol.* **1968**, *41*, 209–244. [[CrossRef](#)]
20. Shepherd, W.K. *Diagonal Wear Predicted by a Simple Wear Model*; ASTM International: West Conshohocken, PA, USA, 1986.
21. Sueoka, A.; Ryu, T.; Kondou, T.; Togashi, M.; Fujimoto, T. Polygonal wear of automobile tire. *JSME Int. J. Ser. C Mech. Syst. Mach. Elem. Manuf.* **1997**, *40*, 209–217. [[CrossRef](#)]
22. Le Maitre, O.; Süßner, M.; Zarak, C. *Evaluation of Tire Wear Performance*; Technical Report, SAE Technical Paper; SAE International: Warrendale, PA, USA, 1998.
23. Braghin, F.; Cheli, F.; Melzi, S.; Resta, F. Tyre wear model: Validation and sensitivity analysis. *Meccanica* **2006**, *41*, 143–156. [[CrossRef](#)]
24. Veen, J. *An Analytical Approach to Dynamic Irregular Tyre Wear*; Wydawnictwo Technische Universiteit Eindhoven: Eindhoven, The Netherlands, 2007; Volume 92.
25. Huang, H.; Chiu, Y.; Wang, C.; Jin, X. Three-dimensional global pattern prediction for tyre tread wear. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2015**, *229*, 197–213. [[CrossRef](#)]
26. Wang, C.; Huang, H.; Chen, X.; Liu, J. The influence of the contact features on the tyre wear in steady-state conditions. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2017**, *231*, 1326–1339. [[CrossRef](#)]
27. Grosch, K. Rubber abrasion and tire wear. *Rubber Chem. Technol.* **2008**, *81*, 470–505. [[CrossRef](#)]
28. Zhang, S.W. *Tribology of Elastomers*; Elsevier: Amsterdam, The Netherlands, 2004.
29. Gadelmawla, E.; Koura, M.; Maksoud, T.; Elewa, I.; Soliman, H. Roughness parameters. *J. Mater. Process. Technol.* **2002**, *123*, 133–145. [[CrossRef](#)]
30. Sakhnevych, A. Multiphysical MF-based tyre modelling and parametrisation for vehicle setup and control strategies optimisation. *Veh. Syst. Dyn.* **2022**, *60*, 3462–3483. [[CrossRef](#)]
31. Mark, J.E.; Erman, B.; Roland, M. *The Science and Technology of Rubber*, 4th ed.; Academic Press: Cambridge, MA, USA, 2013.
32. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013.
33. Napolitano Dell’Annunziata, G.; Arricale, V.M.; Farroni, F.; Genovese, A.; Pasquino, N.; Tranquillo, G. Estimation of Vehicle Longitudinal Velocity with Artificial Neural Network. *Sensors* **2022**, *22*, 9516. [[CrossRef](#)]
34. Everitt, B.; Hothorn, T. *An Introduction to Applied Multivariate Analysis with R*; Springer: New York, NY, USA, 2011.
35. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE). *Geosci. Model Dev. Discuss.* **2014**, *7*, 1525–1534.

36. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: An efficient data clustering method for very large databases. *ACM Sigmod Rec.* **1996**, *25*, 103–114. [[CrossRef](#)]
37. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comput. Surv. CSUR* **1999**, *31*, 264–323. [[CrossRef](#)]
38. Dietterich, T.G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.* **1998**, *10*, 1895–1923. [[CrossRef](#)] [[PubMed](#)]
39. Carputo, F.; D’Andrea, D.; Risitano, G.; Sakhnevych, A.; Santonocito, D.; Farroni, F. A neural-network-based methodology for the evaluation of the center of gravity of a motorcycle rider. *Vehicles* **2021**, *3*, 377–389. [[CrossRef](#)]
40. Karayiannis, N.; Mi, G. Growing radial basis neural networks: Merging supervised and unsupervised learning with network growth techniques. *IEEE Trans. Neural Netw.* **1997**, *8*, 1492–1506. [[CrossRef](#)] [[PubMed](#)]
41. Sathya, R.; Abraham, A. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *Int. J. Adv. Res. Artif. Intell.* **2013**, *2*, 34–38. [[CrossRef](#)]
42. Walczak, S.; Cerpa, N. Artificial Neural Networks. In *Encyclopedia of Physical Science and Technology*, 3rd ed.; Meyers, R.A., Ed.; Academic Press: Cambridge, MA, USA, 2003; pp. 631–645.
43. Mandic, D.; Chambers, J.J. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2001.
44. Da Silva, I.N.; Hernane Spatti, D.; Andrade Flauzino, R.; Liboni, L.H.B.; dos Reis Alves, S.F.; da Silva, I.N.; Hernane Spatti, D.; Andrade Flauzino, R.; Liboni, L.H.B.; dos Reis Alves, S.F. *Artificial Neural Network Architectures and Training Processes*; Springer: Berlin/Heidelberg, Germany, 2017.
45. Zelterman, D. *Applied Multivariate Statistics with R*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015.
46. Kennedy, P. *A Guide to Econometrics*, 6th ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2008.
47. Sahoo, P.; Barman, T.K. ANN modelling of fractal dimension in machining. In *Mechatronics and Manufacturing Engineering*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 159–226.
48. International Organization for Standardization. *ISO Geometrical Product Specifications; Surface Texture: Profile Method—Terms, Definitions and Surface Texture Parameters*; International Organization for Standardization: Geneva, Switzerland, 1997; Volume 4.
49. Mendenhall, W.; Beaver, R.J.; Beaver, B.M. *Introduction to Probability and Statistics*; Cengage Learning: Boston, MA, USA, 2012.
50. Sedlaček, M.; Podgornik, B.; Vižintin, J. Correlation between standard roughness parameters skewness and kurtosis and tribological behaviour of contact surfaces. *Tribol. Int.* **2012**, *48*, 102–112. [[CrossRef](#)]
51. Mandelbrot, B.B. *The Fractal Geometry of Nature*; WH Freeman: New York, NY, USA, 1977.
52. Klüppel, M.; Heinrich, G. Rubber friction on self-affine road tracks. *Rubber Chem. Technol.* **2000**, *73*, 578–606. [[CrossRef](#)]
53. Heinrich, G.; Klüppel, M.; Vilgis, T.A. Evaluation of self-affine surfaces and their implication for frictional dynamics as illustrated with a Rouse material. *Comput. Theor. Polym. Sci.* **2000**, *10*, 53–61. [[CrossRef](#)]
54. Lang, A.; Klüppel, M. Influences of temperature and load on the dry friction behaviour of tire tread compounds in contact with rough granite. *Wear* **2017**, *380*, 15–25. [[CrossRef](#)]
55. Le Gal, A.; Klüppel, M. Investigation and modelling of rubber stationary friction on rough surfaces. *J. Phys. Condens. Matter* **2007**, *20*, 015007. [[CrossRef](#)]
56. Persson, B.N. Rubber friction: Role of the flash temperature. *J. Phys. Condens. Matter* **2006**, *18*, 7789. [[CrossRef](#)] [[PubMed](#)]
57. Povofo, F.; Fontelos, M. Time–temperature superposition principle and scaling behaviour. *J. Mater. Sci.* **1987**, *22*, 1530–1534. [[CrossRef](#)]
58. Williams, M.L.; Landel, R.F.; Ferry, J.D. The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids. *J. Am. Chem. Soc.* **1955**, *77*, 3701–3707. [[CrossRef](#)]
59. Genovese, A.; Maiorano, A.; Russo, R. A Novel Methodology for Non-Destructive Characterization of Polymers’ Viscoelastic Properties. *Int. J. Appl. Mech.* **2022**, *14*, 2250017. [[CrossRef](#)]
60. Genovese, A.; Arricale, V.M.; Barbaro, M.; Marenga, L.; Sanfelice, M.; Dell’Annunziata, G.N. Use of an Innovative Methodology for the Characterization of Polymers in the Analysis of the UV-Light Radiation Effects. In *Proceedings of the International Symposium on Dynamic Response and Failure of Composite Materials*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 142–147.
61. Farroni, F.; Genovese, A.; Maiorano, A.; Sakhnevych, A.; Timpone, F. Development of an innovative instrument for non-destructive viscoelasticity characterization: VESevo. In *Proceedings of the International Conference of IFToMM Italy, Naples, Italy, 9–11 September 2020*.
62. Gavin, H. *The Levenberg-Marquardt Method for Nonlinear Least Squares Curve-Fitting Problems*; Department of Civil and Environmental Engineering, Duke University: Durham, NC, USA, 2013.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.