Multi-objective analysis of performance in mileage based on the minimization of several vehicle polluting features

Gloria Vanegas[1], *Franklin* Samaniego[2,*], *Marlon* Basantes[3], and *Annabelle* Lizarzaburu[2]

[1]Instituto Superior Tecnológico Bolívar, Ambato 180102, Ecuador
[2]Universidad de Guayaquil, Guayaquil 090514, Ecuador
[3]Universidad Nacional de Chimborazo, Riobamba 060108, Ecuador

Abstract. This paper analyses the fuel consumption and emissions of and from various kinds of automobiles. The aim is to identify the set of automobiles models that produce the least pollution and provide the higher mileage. To complete the analysis, a multi-objective optimization problem (MOP) has been proposed with a visual representation methodology of the Pareto front (Level Diagram); in this way, it has been determined that the highest compromise values corresponding to the utopian point determine a mileage performance of 16.30 $[km/l]$. Finally, it is important to highlight that the MOP has facilitated the analysis process, which helps the Decision Maker (DM) in the adequate selection of the final solution, based on the available knowledge of the set of optimal solutions.

1 Introduction

The transportation industry is constantly evolving, which has a strong impact on the global economy, on trade and on people's mobility, as well as being a significant environmental and public health issue. An example has been described in [1], where, it details different emissions and fuel consumption of gasoline and hybrid light-duty vehicles in Toronto, Canada.

In specific, the transport sector contributes significantly to global pollution, being responsible for almost 37% of the carbon dioxide emitted worldwide during the year 2021 (even when pandemic restrictions [2], [3]). On the other hand, atmospheric pollution caused by nitrogen oxide (NOx) from automobiles in the atmosphere contributes to the formation of photochemical ozone (smog) and leads to health altering consequences, as well as contributing to global warming and can cause acid rain. (see Fig. 1). Thus, several strict regulations have been generated around the world to reduce polluting emissions from vehicles, and for this reason automobile manufacturers make constant efforts to optimize the performance of their internal combustion engines in order to comply with these regulations [4,5].

A brief survey of the current literature describes several works that attempt to optimize NOx emissions, a problem that remains a challenge today [6, 7]. In that sense, works focused

* Corresponding author : franklin.samaniegor@ug.edu.ec

on the design of a machine learning algorithms have been presented [8, 9] and predict the NOx levels at the engine output. On the other hand, as different photocatalytic systems have different catalysts, independent of scale and geometry, an experimental correlation of data has been widely used as a reference for the description of the photocatalytic reduction of CO_2, developing heterogeneous photocatalytic kinetic models [10] for catalysts with different energetic sites [11-13].

Fig. 1. Example of vehicular pollution in the city.

A wide variety of real problems, such as pollution, require the search for solutions that simultaneously satisfy multiple performance criteria. Thereby, the aim of this work focuses on collecting a set of different air pollution data from vehicle pollution between 2015 and 2018 (electric and/or hybrid vehicles have been left aside), analysing through a Multi-Objective Optimization (MOP) methodology and determining which vehicles, its models and characteristics have contributed the least amount of pollution and also provide the best gas performance in km/l. A set of solutions that represent the best compromises among the different criteria has been determined by means of the Pareto-optimal set and the visualization of the objective space through the Pareto Front [14], a process summarized in Fig. 2.

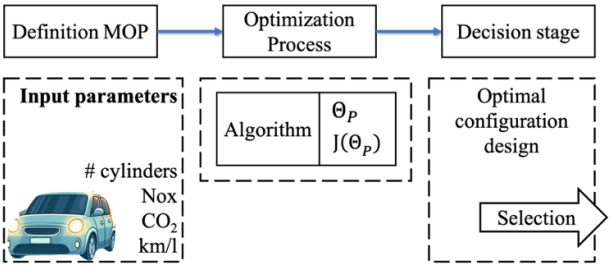

Fig. 2. Block diagram of the MOP.

The organization of this paper is described below. In section 2 a brief mathematical and procedural description of the stages completed. In section 3, the results achieved from the resolution of the problem, are shown. Finally, section 4 presents the final conclusions.

2 Methodology

This section is split into two parts, where the first part describes the data exploited throughout the study, which have been modelled through a brief identification of mathematical models, while the second part describes the MOP.

In Mexico City, there has been a constant record of information about fuel efficiency, in car models from 2011 to 2018 year [15]. However, the set of cars and their characteristics taken for this study covers the 2011-2014 models. Specifically, the vehicle emission data collected and related to this work, which have also been defined as the target functions are:

the engine characteristic (Number of Cylinders), city performance $[km/l]$ and rating of their CO_2 gas emissions $[g/km]$ and Nox$[g/1000km]$. It is important to note that the units of measurement for each objective are different, therefore, data standardization is necessary [16].

2.1 System identification

To provide a quick definition of the equation systems for the described data, a simple polynomial regression has been performed on the set of the different data described. Thus, starting from the analysis of a single variable $\boldsymbol{Y}$ in terms of the independent variable vector $\boldsymbol{X}$, the simple linear regression is defined as:

$$\boldsymbol{Y} = \beta_0 + \beta_{1\boldsymbol{X}} + \varepsilon \tag{1}$$

where, ε denotes an unknown random error with mean zero and conditional on a scalar variable $\boldsymbol{X}$. In this model, for each unit increase in the $\boldsymbol{X}$ value, the conditional expectation of $\boldsymbol{Y}$ increases by β_1 units.

Whereas the polynomial regression model is defined as:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_m x_i^m + \varepsilon_i \quad (\mathrm{i} = 1, 2, \ldots, \mathrm{n}) \tag{2}$$

Therefore, it is possible to express the Regression in matrix terms, from the terms of the design matrix $\boldsymbol{X}$, a response vector $\boldsymbol{Y}$, and a vector of random errors ε. Hence, the polynomial regression model is defined as a system of linear equations, such that:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & & x_2^m \\ 1 & x_3 & x_3^2 & \cdots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3}$$

being its matrix notation represented by:

$$\boldsymbol{Y} = X\,\beta + \varepsilon \tag{4}$$

Therefore, the vector of estimated polynomial regression coefficients is describe as:

$$\hat{\beta} = (X^t X)^{-1} X^t \boldsymbol{Y} \tag{5}$$

assuming $m < n$ for the matrix be invertible, whether all x_i values are distinct.

Thus, considering a change of variable $\theta = x$, the 4 objective functions used for this study have been defined, where, $f_1(\theta)$ represents CO_2, $f_2(\theta)$ represents Nox, $f_3(\theta)$ represents fuel efficiency and $f_4(\theta)$ represents the number of cylinders of the vehicles. Table 1 shows the coefficients defined for each function.

Table 1. Polynomial coefficients.

	$\boldsymbol{\theta_i^5}$	$\boldsymbol{\theta_i^4}$	$\boldsymbol{\theta_i^3}$	$\boldsymbol{\theta_i^2}$	$\boldsymbol{\theta_i^1}$	$\boldsymbol{\theta_i^0}$
$\boldsymbol{f_1(\theta_1)}$	7.685	-19.429	19.221	-9.060	2.579	-0.034
$\boldsymbol{f_2(\theta_2)}$	0.662	-3.200	6.534	-7.329	4.911	-2.000
$\boldsymbol{f_3(\theta_3)}$	17.000	-40.184	33.702	-11.296	1.504	0.052
$\boldsymbol{f_4(\theta_4)}$	11.234	-25.394	20.827	-7.344	1.372	0.029

It is important to note that each data set has a total of 4,422 samples, which have been normalized to the interval $[0, 1]$. In the same way, the maximum values of each function depend on their characteristics, thus, have been normalized between $[0, 1]$, to generate the polynomial models.

The results of the systems can be seen in Fig. 3, which describes the data collection of the vehicle models and its characteristics (blue continuous lines), while the red lines describe the curves of each system, obtained through the polynomial regression model.

2.2 Description of the Multi-objective Optimization Problem.

In contrast to single-objective optimization, where an n-dimensional decision vector that optimizes a scalar function is sought, in multi-objective optimization it is attempted to optimize a vector function which elements represent the different objective functions.

It is important to highlight that any value of θ_i (as described in Table 1), defines a spatial point among the interval $[0,1]$. Therefore, within the boundaries of θ_i, there exists an infinite number of spatial points with an infinite number of combinations with their corresponding infinite number of functions, with which its corresponding objective functions can be generated.

Hence, to obtain an optimal solution, a multi-objective problem (MOP) is solved using evolutionary algorithms based on the concept of e-dominance [17]. Thus, it is necessary to define the decision variables, the initial conditions of the process, the MOP constraints and the index vector to be optimized to represent the Pareto front. Assuming each function $f_i(\theta)$, then $J^{ideal}(\theta) = [J_1(\theta), J_2(\theta), \dots, J_s(\theta)]$ is the vector of objectives, where J_i denotes the i-th objective.

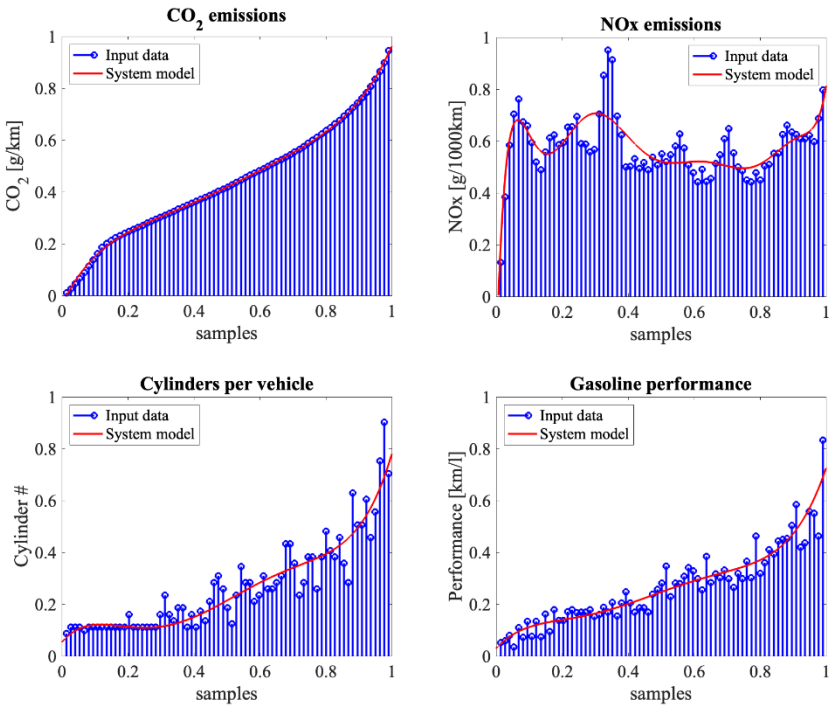

Fig. 3. Polynomial regression models from the collected data.

Consequently, $J_i = \min(f_i(\theta)), \in G_i: [i = 1, \dots, m]$, where J_i depend on the vector of decision variables θ. Assuming that: D represents the decision space within a subset D, where θ is the vector of decision variables composed of a set of θ_i, with θ_i within $[0,1]^D$. Hence, the MOP problem has been described as:

$$min\, \boldsymbol{J}(\theta) = \min_{\theta \in D}[J_1(\theta), J_2(\theta), \dots, J_m(\theta)] \tag{6}$$

subject to:

$$\begin{gathered} g_q(\theta) \leq 0 \\ h_q(\theta) = 0 \\ \theta_{li} \leq \theta_i \leq \theta_{ui}, i = [0, \dots, n] \end{gathered} \tag{7}$$

where, $\theta \in \mathbb{R}^n$ is the decision vector, D is the decision space, $\boldsymbol{J}(\theta) \in \mathbb{R}^m$ is the objective vector, $g(\theta)$ and $h(\theta)$ are the constraint vectors, and the upper and lower boundaries of the decision space, represented by θ_{li} and θ_{ui} respectively. Consequently, there is no single optimal solution for the model; in fact, there is a set of optimal solutions with different trade-offs among the objectives, where none is better than the others. Hence, using the definition of dominance, the Pareto set Θ_p is the set of each non-dominated solution.

As such, the definition of Pareto dominance is expressed in each case in which one solution θ^1 dominates another solution θ^2, such that, $(\theta^1 \prec \theta^2)$, if

$$\forall i \in B, J_i(\theta^1) \leq J_i(\theta^2) \wedge \exists k \in B: J_k(\theta^1) < J_k(\theta^2) \tag{8}$$

where, $J_i(\theta), i = B := [1 \cdots m]$ are the objectives to be optimized. Therefore, the Pareto optimum is defined as.

$$\begin{gathered} \Theta_P = \theta \in D \backslash \nexists \tilde{\theta} \in D: \tilde{\theta} \prec \theta \\ J(\Theta_P) = \{J(\theta) | \theta \in \Theta_P\} \end{gathered} \tag{9}$$

where, Θ_p and $J(\Theta_P)$ are solutions to the MOP.

3 Results

To provide a graphical and numerical representation of the set of equations described in Table 1, it is necessary to address the Decision Maker (DM) stage that selects one of the solutions. This solution implies a trade-off of minimizing pollution and maximizing the mileage performance. In this way, for this work the selection criterion has been based on the shortest distance to the ideal point. Fig. 4 and Fig. 5, depicts the results achieved in the definitions of Pareto front and Pareto set, expressed through the selected points (red color) from $J(\Theta_p^*)$ and (Θ_p^*), which has been selected from a standard norm-∞ and Level Diagram[18].

From the Pareto front, which is shown in Fig. 4, it is possible to establish different zones (A, B) in that the DM can select an optimal solution. Thus, in zone B, the points of the first model achieve individual improvements, however, the other models move away. Thereby, points within zone B (magenta box filled with red dots) in $J_1(\theta)$ on the left of the diagram produce points on the right in $J_2(\theta)$, while in $J_3(\theta)$ and $J_4(\theta)$ are shown scattered and erratic, which means that the values of $J_1(\theta)$ are relevant, however, the differences in $J_2(\theta)$, $J_3(\theta)$ and $J_4(\theta)$ are worse.

On the other hand, within zone A (dashed green box) through the DM a suitable model can be selected, provided that preferences are based on models nearest to the ideal point in zone A. Being, the values of $J(\Theta)$ and Θ_p selected using the norm-∞.

Finally, to complete the paper, an identification and validation of the parameters resulting from the MOP process has been carried out, through the values of the optimization criteria, shown in Table 2.

Table 2. Nearest points to the ideal of the objective functions.

norm	J_1	J_2	J_3	J_4
∞	0.118	0.642	0.111	0.029

Therefore, by replacing the highest compromise values corresponding to the utopian point, it has been reached the vehicle with gas emission characteristics $CO_2 = 170.00[g/km]$, $NOx = 16.00[g/100km]$, 4-cylinder engine and a performance in mileage of $16.30[km/l]$.

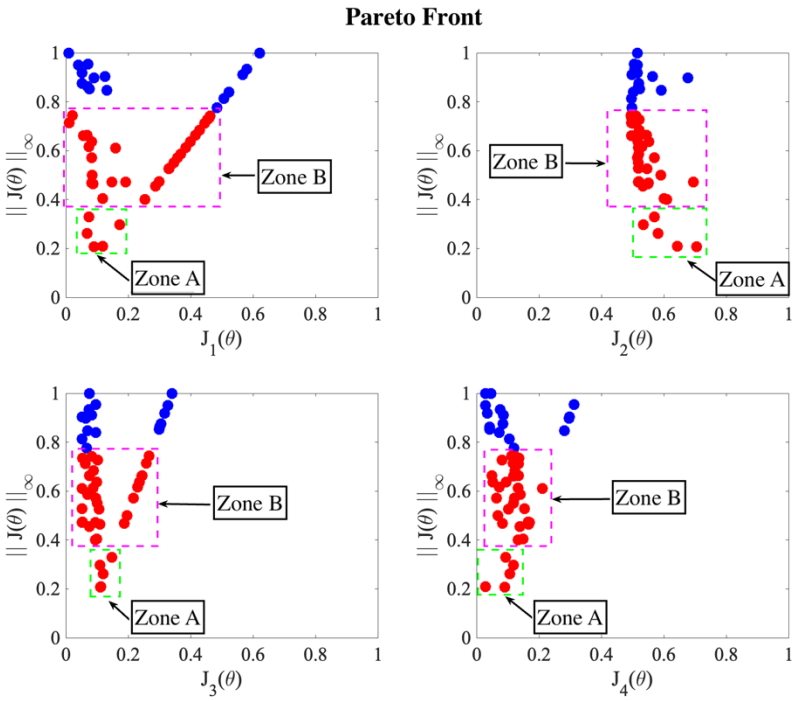

Fig. 4. Representation of the Pareto front using norm-mean, the subindices J represent the values of the functions within the defined boundaries. Nearby targets J^{ideal} are represented by the red circles.

One of the benefits of working with multi-objective analysis methodologies focuses on its simplicity and speed of analysis. In this way, a good understanding analysis focuses on the definition of the Pareto front and its optimums. Therefore, from these results, the Decision Maker can select the best solution according to the design parameters.

Finally, it is important to highlight that in this work a set of design parameters have been defined. However, it is possible to increase the parameters within the optimization process by adding various features of study such as HC, CO, and SO. Which implies the definition of new equations, consequently, the new results of the MOP process extend the scopes of the analysis.

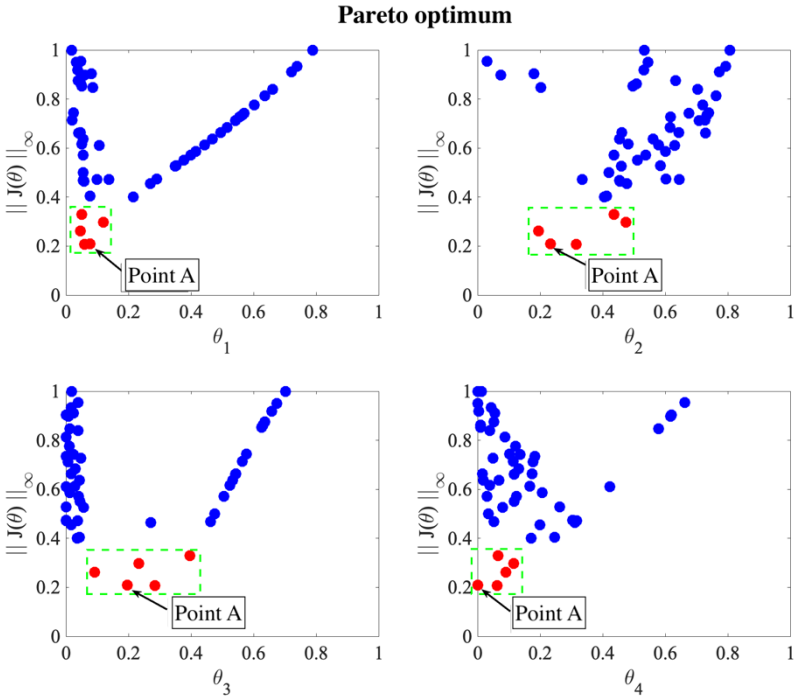

Fig. 5. Representation of the Pareto optimal parameters. Nearby targets J^{ideal} are represented by the red circles.

4 Conclusions

The work presents an identification and determination of polluting factors CO_2, NOx and Fuel Efficiency in automobiles with engines having different numbers of cylinders. It has been demonstrated that the application of MOP on a system based on real data collection and recording has produced a real result.

One of the advantages of working with multi-objective analysis methodologies focuses on its simple and fast analysis, in addition to the improvement in exploration, by modifying the probabilities of the local spatial search around the best solutions found. In this way, a good understanding analysis is focused on the definition of the Pareto front and its optimums, thus the designer is in charge of selecting the best solution according to the design parameters.

The results can be extended by adding parameters within the optimization process such as: HC, CO, and SO_2. However, increasing parameters means identifying new systems, however, with the definition of new equations, the results of the MOP process extend the scope of the analysis.

Finally, it should be emphasized that a MOP methodology allows for an adequate analysis, despite the conflict among function systems of different physical magnitudes, assumed as simultaneous objectives.

References

1. R. Tu, J. Xu, A. Wang, M. Zhang, Z. Zhai, M. Hatzopoulou, Real-world emissions and fuel consumption of gasoline and hybrid light duty vehicles under local and regulatory

drive cycles. Science of The Total Environment. **805**, 150407 (2022). doi: 10.1016/j.scitotenv.2021.150407.

2. H. El Hafdaoui, A. Khallaayoun, K. Ouazzani, Activity and efficiency of the building sector in Morocco: A review of status and measures in Ifrane. AIMS Energy, **11**(3), 454–485 (2023). doi: 10.3934/energy.2023024.
3. V. Rievaj, J. Gaňa, F. Synák, Comparison of emissions depending on the type of vehicle engine. Logistics & Sustainable Transport. **10**(1), 45–54 (2019), doi: 10.2478/jlst-2019-0004.
4. J. Enzmann, M. Ringel, Reducing road transport emissions in Europe: Investigating a demand side driven approach. Sustainability. 7594, **12**(18) (2020). doi: 10.3390/su12187594.
5. D. C. Carslaw, N. J. Farren, A. R. Vaughan, W. S. Drysdale, S. Young, J. D. Lee, The diminishing importance of nitrogen dioxide emissions from road vehicle exhaust. Atmos Environ X**, 1** (2019). doi: 10.1016/j.aeaoa.2018.100002.
6. B. Winkler-Ebner, M. Hirsch, L. Del Re, H. Klinger, W. Mistelberger, Comparison of virtual and physical NOx-sensors for heavy duty diesel engine application. SAE International J. Engines, **3** (2010). doi: 10.4271/2010-01-1296.
7. S. Stadlbauer, D. Alberer, M. Hirsch, S. Formentin, C. Benatzky, L. Re, Evaluation of Virtual NOx Sensor Models for Off Road Heavy Duty Diesel Engines. SAE Int J Commer Veh, **5**(1) (2012). doi: 10.4271/2012-01-0358.
8. R. Fechert, B. Bäker, S. Gereke, F. Atzler, Using machine learning methods to develop virtual NOx sensors for vehicle applications, In: Bargende, M., Reuss, HC., Wagner, A. (eds) 20. Internationales Stuttgarter Symposium. Proceedings. Springer Vieweg, Wiesbaden, 265–280 (2020). doi: 10.1007/978-3-658-30995-4_27.
9. N. J. Kempema, C. Sharpe, X. Wu, M. Shahabi, D. Kubinski, Machine-Learning-Based Emission Models in Gasoline Powertrains Part 2: Virtual Carbon Monoxide, SAE Int J Engines, **16**(6), 799-807 (2022). doi: 10.4271/03-16-06-0045.
10. L. L. Tan, W. J. Ong, S. P. Chai, A. R. Mohamed, Photocatalytic reduction of CO2 with H2O over graphene oxide-supported oxygen-rich TiO2 hybrid photocatalyst under visible light irradiation: Process and kinetic studies. Chemical Engineering Journal, **308** (2017). doi: 10.1016/j.cej.2016.09.050.
11. A. khalilzadeh, A. Shariati, Fe-N-TiO2/CPO-Cu-27 nanocomposite for superior CO2 photoreduction performance under visible light irradiation, Solar Energy, **186**, 166-174 (2019). doi: 10.1016/j.solener.2019.05.009.
12. W. A. Thompson, E. Sanchez Fernandez, M. M. Maroto-Valer, Probability Langmuir-Hinshelwood based CO2 photoreduction kinetic models, Chemical Engineering Journal, **384** (2020). doi: 10.1016/j.cej.2019.123356.
13. R. Sips, On the structure of a catalyst surface, J Chem Phys, **16**(5), 490–495 (1948), doi: 10.1063/1.1746922.
14. C. A. Coello Coello Coello, A Short Tutorial on Evolutionary Multiobjective Optimization, In: Zitzler, E., Thiele, L., Deb, K., Coello Coello, C.A., Corne, D. (eds) Evolutionary Multi-Criterion Optimization. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, (1993). doi.org/10.1007/3-540-44719-9_2
15. A. Acosta, Autos - Consumo Gasolina Mexico, Accessed: Sep. 08, 2023. [Online]. Available: https://www.kaggle.com/datasets/checoalejandro/autos-consumo-gasolina-mexico?resource=download

16. N. Morrison, D. C. Hoyle, Normalization, In: Berrar, D.P., Dubitzky, W., Granzow, M. (eds) A Practical Approach to Microarray Data Analysis, Springer, Boston, MA, 76–90 (2003). doi: 10.1007/0-306-47815-3_4.
17. J. M. Herrero, X. Blasco, M. Martínez, C. Ramos, J. Sanchis, Non-linear robust identification of a greenhouse model using multi-objective evolutionary algorithms, Biosyst Eng, **98**(3) (2007). doi: 10.1016/j.biosystemseng.2007.06.004.
18. X. Blasco, J. M. Herrero, J. Sanchis, M. Martínez, A new graphical visualization of n-dimensional Pareto front for decision-making in multiobjective optimization, Inf Sci (N Y), **178**(20), 3908–3924 (2008). doi: 10.1016/j.ins.2008.06.010.