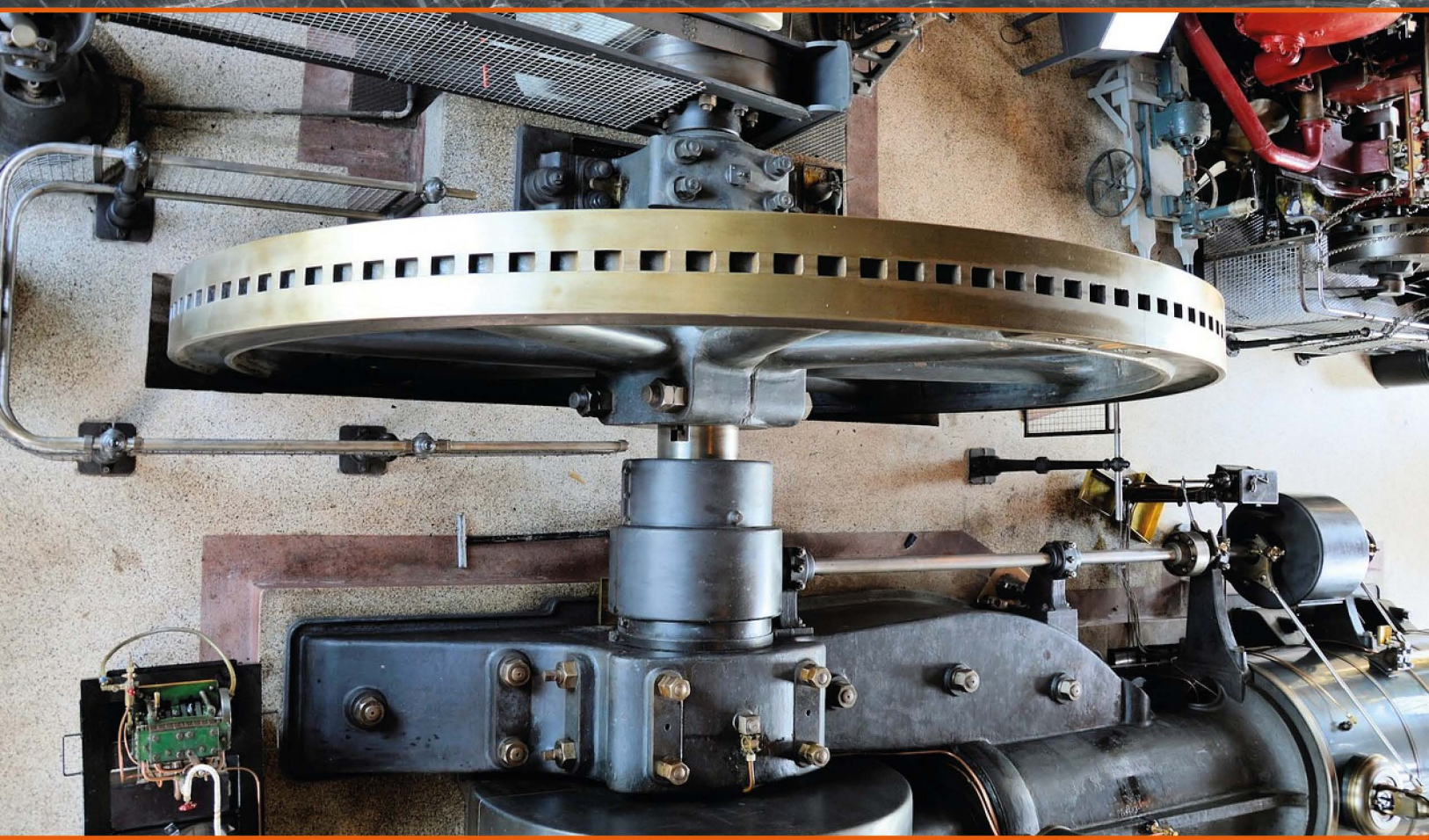
Number: 25
Redalyc 3.0
ESCI Web of science
UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
pISSN : 1390-650X
eISSN : 1390-860X
january / june 2021

INGENIUS

Revista de Ciencia y Tecnología



Ingenius

Ingenius • Issue 25 • january/june 2021. Semester Journal of Science and Tecnology of the Universidad Politécnica Salesiana of Ecuador. Publication dedicated to studies related to the Sciences of Mechanical Engineering, Electrical Engineering, Electronic Engineering, Mechatronic Engineering, Systems Engineering and Industrial Engineering.

Reviewers board

Federico Dominguez, PhD, Escuela Superior Politécnica del Litoral, Ecuador.

Enrique Carrera, PhD, Universidad de las Fuerzas Armadas, ESPE, Ecuador.

Andrés Tello, MSc, Universidad de Cuenca, Ecuador.

Cristian García Bauza, PhD, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina.

Osvaldo Añó, PhD, Universidad Nacional de San Juan, Argentina.

Thalía San Antonio, PhD, Universidad Técnica de Ambato, Ecuador.

Victor Saquicela, PhD, Universidad de Cuenca, Ecuador.

Gonzalo Olmedo, PhD, Universidad de las Fuerzas Armadas, ESPE, Ecuador.

Román Lara, PhD, Universidad de las Fuerzas Armadas, ESPE, Ecuador.

Guillermo Soriano, PhD, Escuela Superior Politécnica del Litoral, Ecuador.

María Fernanda Granda, PhD, Universidad de Cuenca, Ecuador.

Ricardo Cayssials, PhD, Universidad Tecnológica Nacional, Argentina.

Leonardo Solaque Guzman, PhD, Universidad Militar Nueva Granada, Colombia.

José Di Paolo, PhD, Universidad Nacional de Entre Ríos, Argentina.

Astrid Rubiano Fonseca, PhD, Universidad Militar Nueva Granada, Colombia.

Robinson Jiménez, PhD, Universidad Militar Nueva Granada, Colombia.

Alfonso Zozaya, PhD, Universidad de Carabobo, Venezuela.

Mauricio Mauledoux, PhD, Universidad Militar Nueva Granada, Colombia.

Luis Medina, PhD, Universidad Simón Bolivar, Venezuela.

Ernesto Cuadros-Vargas, PhD, Universidad Católica San Pablo, Perú.

Samuel Sepúlveda Cuevas, PhD, Universidad de la Frontera, Chile.

Carlos Cares, PhD, Universidad de la Frontera, Chile.

Rafael Sotelo, PhD, Universidad de Montevideo, Uruguay.

Omar Lopez, PhD, Universidad de Los Andes, Colombia.

Job Flores-Godoy, PhD, Universidad Católica del Uruguay, Uruguay.

Luis Mario Mateus, PhD, Universidad de los Andes, Colombia.

Amadeo Argüelles Cruz, PhD, Instituto Politécnico Nacional, México.

Santiago Bentancourt Parra, PhD, Universidad Pontificia Bolivariana, Colombia.

Germán Zapata, PhD, Universidad Nacional de Colombia, Colombia.

Pedro García, PhD, Universidad Autónoma de Barcelona, España.

Arturo Conde Enríquez, PhD, Universidad Autónoma de Nuevo León, México.

Alberto Cavazos González, PhD, Universidad Autónoma de Nuevo León, México.

Ernesto Vázquez Martínez, PhD, Universidad Autónoma de Nuevo León, México.

Miguel Díaz Rodriguez, PhD, Universidad de Los Andes, Venezuela.

Efraín Alcorta García, PhD, Universidad Autónoma de Nuevo León, México.

Luis Chirinos Garcia, PhD, Pontificia Universidad Católica de Perú, Perú.

Oscar Avilés, PhD, Universidad Militar Nueva Granada, Colombia.

Dora Martínez Delgado, PhD, Universidad Autónoma de Nuevo León, México.

David Ojeda, PhD, Universidad Técnica del Norte, Ecuador.

Irene Beatríz Steinmann, PhD, Universidad Tecnológica Nacional, Argentina.

Mario Serrano, Universidad Nacional de San Juan, Argentina.

Cornelio Posadas Castillo, PhD, Universidad Autónoma Nuevo León, México.

Mario Alberto Rios Mesias, PhD, Universidad de Los Andes, Colombia.

Yudith Cardinale Villarreal, PhD, Universidad Simón Bolivar, Venezuela.

Eduardo Matallanas, PhD, Universidad Politécnica de Madrid, España.

Jose Eduardo Ochoa Luna, PhD, Universidad Católica San Pablo, Perú.

Dante Angel Elias Giordano, PhD, Pontificia Universidad Católica de Perú, Perú.

Manuel Pelaez Samaniego, PhD, Universidad de Cuenca, Ecuador.

Juan Espinoza Abad, PhD, Universidad de Cuenca, Ecuador.

Pietro Codara, PhD, Universidad de Milan, Italia.

Alberto Soria, PhD, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.

José M. Aller, PhD, Universidad Politécnica Salesiana, Ecuador.

Ferney Amaya F., PhD, Universidad Pontificia Bolivariana, Medellin, Colombia.

Santiago Arango Aramburo, PhD, Universidad Nacional de Colombia, Colombia.

Diego Arcos–Avilés, PhD, Universidad de las Fuerzas Armadas, ESPE, Ecuador.

Pablo Arevalo, PhD, Universidad Politécnica Salesiana, Ecuador.

Roberto Beltran, MSc, Universidad de las Fuerzas Armadas, ESPE, Ecuador.

Leonardo Betancur, PhD, Universidad Pontificia Bolivariana, Medellin, Colombia.

Roberto Gamboa, PhD, Universidad de Lisboa, Portugal.

Paulo Lopes dos Santos, PhD, Universidad do Porto, Portugal.

Pedro André Dias Prates, PhD, Universidad de Coimbra, Portugal.

José Manuel Torres Farinha, PhD, Universidad de Coimbra, Portugal.

Celso de Almeida, PhD, Universidad Estatal de Campinas Unicamp, Brasil.

Ramon Molina Valle, PhD, Universidad Federal de Minas Gerais, Brasil.

Cristina Nader Vasconcelos, PhD, Universidad Federal Fluminense, Brasil.

João M. Ferreira Calado, PhD, Universidad de Lisboa, Portugal.

Guilherme Luz Tortorella, PhD, Universidad Federal de Santa Catarina, Brasil.

Mauro E. Benedet, PhD, Universidad Federal de Santa Catarina, Brasil.

Artemis Marti Ceschin, PhD, Universidade de Brasilia, Brasil.

Gilmar Barreto, PhD, Universidad Estatal de Campinas Unicamp, Brasil.

Ricardo Emilio F. Quevedo Nogueira, PhD, Universidad Federal de Ceará, Brasil.

Weslley Luiz da Silva Assis, PhD, Universidad Federal Fluminense, Brasil.

Ana P. Martinazzo, PhD, Universidad Federal Fluminense, Brasil.

Jorge Bernardino, PhD, Universidad de Coimbra, Portugal.

Luis Geraldo Pedroso Meloni, PhD, Universidad Estatal de Campinas Unicamp, Brasil.

Facundo Almeraya Calderón, PhD, Universidad Autónoma de Nuevo León, México.

Freddy Villao Quezada, PhD, Escuela Superior Politécnica del Litoral, Ecuador.

Jose Manrique Silupu, MSc, Universidad de Piura, Perú.

Germán Ariel Salazar, PhD, Instituto de Investigaciones en Energía no Convencional, Argentina.

José Mahomar Jananías, PhD, Universidad del BIOBIO, Chile.

Arnaldo Jélvez Caamaño, PhD, Universidad del BIOBIO, Chile.

Jorge Andrés Uribe, MSc, Centro de Ingeniería y Desarrollo Industrial, México.

Ricardo Beltran, PhD, Centro de Investigación en Materiales Avanzados, México.

Adi Corrales, MSc, Centro de Ingeniería y Desarrollo Industrial, México.

Jorge Uribe Calderón, PhD, Centro de Investigación Científica de Yucatán, México

Editorial

Editorial Abya Yala (Quito-Ecuador),
Av. 12 de octubre N422 y Wilson,
Bloque A, UPS Quito, Ecuador.
Casilla 17-12-719 Teléfonos: (593-2) 3962800 ext. 2638
email: editorial@abyayala.org

Translator

Adriana Curiel

Tiraje: 800 ejemplares

Typographic system used in the composition of this document LaTeX.

Ingenius

Journal of Science and Technology

Issue 25, january – june 2021

ISSN impreso 1390-650X / ISSN electrónico 1390-860X

The administration of the journal is done through the following parameters:

The journal uses the academic anti-plagiarism system

The articles have an identification code (Digital Object Identifier)

The editorial process is managed through the Open Journal System

The articles of the present edition can be consulted in
http://revistas.ups.edu.ec/index.php/ingenius

Universidad Politécnica Salesiana del Ecuador

INGENIUS Journal, is indexed in the following Databases and scientific information systems:

SELECTIVE DATABASES

REVIEWS EVALUATION PLATFORMS

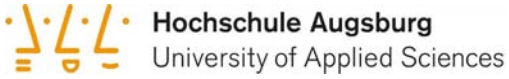
SELECTIVE DIRECTORIES

SELECTIVE SERIAL LIBRARY

SCIENTIFIC LITERATURE SEARCHERS OPEN ACCESS

OTHER BIBLIOGRAPHICAL DATABASES

CATALOG OF INTERNATIONAL UNIVERSITY LIBRARIES



Editorial

Dear readers:

The challenges left by the year 2020 make us think about the importance of research to literally keep humanity alive. The COVID-19 pandemics has disrupted all ambits of our existences; even those emporiums that believed to be indestructible have been affected and many of them have required to modify their usual activities to be able to survive. Those who did not do it succumbed to the crisis and closed their doors, generating unemployment and increasing the gap between those who have lost and those who should devise ways to survive.

Keeping active the research has not been an easy task, the restrictions generated regarding access to labs, free transit of people, importation of supplies, the risk involved by continuing processes especially those that require experimentation with groups and many other things are currently barriers that need to be overcome. Taking a quick look to the gigantic world of publications, it is evident that they have not stopped, they keep generating their periodical contents with the established quality indices. In Ecuador, for those who are in the exciting task of scientific dissemination this is palpable, and if we consider the number of papers received in INGENIUS for this number, we may understand that researchers devised processes to successfully overcome the barriers generated by the pandemics.

I consider of utmost importance to remark the innovation capability and ingenuity of researchers, and this is why the pandemics, even though it has caused many problems and human losses, for sure it has also generated great challenges. The first and the one that motivated the teamwork of thousands of researchers worldwide has been the development of a vaccine (in a record time). However, we should not forget that a continuous research will be required to provide support to the different changes undergone by the virus which will modify its interaction with human beings. It should be also considered that the well-being of people has been affected and it will be necessary a joint action to recover and improve our lifestyles. For this, it is essential: to generate proposals and developments to recover the sources of employment and improve the economy of all families, contributing to innovative ideas that enable optimizing productivity, improving communication systems, strengthening social security systems, developing cutting-edge technology, but with accessible costs that enable all people to take part of the benefits obtained. And, above all, contributing so that the less advantaged have the opportunity to integrate to a dignified life and may develop to be an active part of the society.

The teamwork must be strengthened and privileged, and thus we are very pleased to see that many of the contributions that come to INGENIUS are made by multidisciplinary groups and from different countries, which has given rise to papers of great value which for sure will contribute to the national and international scientific community in the search for products, systems, equipment, processes that enable an accelerated development to achieve the tranquility and calm that this pandemics took away.

Finally, I invite you to revise the papers chosen for this edition: very valuable works from different parts of the world which, thanks to the virtual and communication systems, show that it is possible to achieve the synergy necessary to obtain successful results.

And do not forget that if research does not continue and strengthen the world, it will be exposed to an imminent extinction.

John Calle-Sigüencia, PhD

EDITOR IN CHIEF

Índice

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.01
pISSN: 1390-650X / eISSN: 1390-860X

HFO-1234ze Analysis as an Ecological Alternative In Domestic Refrigeration

Análisis del HFO-1234ze como alternativa ecológica en la refrigeración doméstica

Cristian Andrade Terán[1,*]

Received: 03-05-2020, Reviewed: 22-07-2020, Accepted after review: 05-08-2020

Abstract

Food refrigeration is an essential process in homes, and thus a home refrigerator becomes an indispensable appliance. Being this one of the biggest consumers of electrical energy and contamination due to the refrigerant used for its operation, it is important to look for alternatives that improve this process. This study aims to implement the HFO, R12354ze as an ecological alternative in domestic refrigeration, in response to environmental demands to reduce climate change and deterioration of the ozone layer. Through a thermodynamic and heat transfer analysis, simulating the cooling cycle and the behavior of the fluid in heat exchange using specialized software and CFD, the HFO is presented as an acceptable alternative achieving cooling parameters which are between 5% and 8% different from common refrigerators currently used, with an environmental cost up to 99% lower, without altering their energy efficiency. Taking advantage of the properties of the HFO, it is possible to improve the coefficient of performance of the cooling cycle by 12%.

Keywords: CFD, HFO, Kigal amendment, Paris agreement, refrigeration, R1234ze.

Resumen

La refrigeración de alimentos es un proceso esencial en los hogares, por lo que un refrigerador doméstico se convierte en un electrodoméstico indispensable. Siendo este uno de los mayores consumidores de energía eléctrica y de contaminación por el refrigerante que ocupa para su funcionamiento, es importante buscar alternativas que mejoren este proceso. En este estudio se pretende implementar un HFO, el R12354ze, como alternativa ecológica en la refrigeración doméstica, en respuesta a las demandas ambientales para reducir el cambio climático y el deterioro de la capa de ozono. Mediante un análisis termodinámico y de transferencia de calor con simulaciones del ciclo de refrigeración y el comportamiento del fluido en el intercambio de calor utilizando *software* especializado y CFD, se presenta al HFO como una alternativa aceptable logrando parámetros de refrigeración en rango entre 5 %-8 % de diferencia con refrigeradores comunes utilizados actualmente, con un costo ambiental de hasta un 99 % más bajo, sin alterar su eficiencia energética. Aprovechando las propiedades del HFO en el ciclo de refrigeración se logra mejorar el coeficiente de desempeño del ciclo de refrigeración en un 12 %.

Palabras clave: Acuerdo de París, CFD, HFO, enmienda Kigal, refrigeración, R1234ze

[1,*]Ingeniería Mecánica, Universidad Politécnica Salesiana, Quito-Ecuador
Corresponding author ✉: candradet@est.ups.edu.ec. https://orcid.org/0000-0002- 0220-7481

Suggested citation: Andrade Terán, C. (2021). «HFO-1234ze Analysis as an Ecological Alternative In Domestic Refrigeration». Ingenius. N.° 25, (january-june). pp. 9-19. doi: https://doi.org/10.17163/ings.n25.2021.01.

1. Introduction

Historically, the mechanical refrigeration that uses fluids for heat transfer has passed through various stages to adapt to environmental needs. Starting with the Vienna Convention for protecting the ozone layer (1985), which forced to reach the Montreal Protocol (1995) regarding substances that deteriorate such layer (mainly CFC and halogens); this agreement was successful and the proposed goals were accomplished. In 2003 it was seen as the most successful protocol until then; it was frozen in 2015, and it is expected to completely eliminate the HCFCs by 2030 in the developed countries, and by 2040 in the remaining countries [1].

As a consequence, the United Nations create other treaties and agreements for protecting the ozone layer and the environment, among which it is highlighted the Paris Agreement in 2015 that created the 2030 agenda, which requests to accelerate the reduction of worldwide emissions of greenhouse gases, seeking to maintain the global temperature increase far below 2 °C in this century [2]. In this context, in 2016, with the Kigali amendment to the Montreal Protocol, an agreement is reached to gradually eliminate the hydrofluorocarbons (HFC); even though they have a null or low Ozone Depletion Potential (ODP), they possess a high Global Warming Potential (GWP), which increases the planet temperature. This amendment took effect on January 1st, 2019, to reduce the production and consumption of HFC in more than 80% during the next thirty years [3].

With the meeting in Kigali it is agreed to reduce up to placating the use of refrigerants with high GWP, promoting the switching to alternatives less harmful to the enviroment and improving the energy efficiency in refrigeration and air conditioning. The opportunity to change is very accepted, and in many cases results profitable in industrial refrigeration and refrigerated transport, and to a lesser extent in commercial refrigeration; however, the change results more costly in the cases of mobile and commercial air conditioning and domestic refrigeration and, therefore, it is more difficult to move away from HFCs. In addition, in underdeveloped countries and countries with high environmental temperature the change results more costly and difficult compared to developed countries [4]. With the regulations and taxes to the refrigerants with high GWP, it is seeked to make more complicated for manufacturers to produce household appliances with these refrigerants, making industry and researchers to focus on looking for more ecological alternatives, which do not result in relatively high costs in the adaptation or replacement in refrigeration and air conditioning equipment which has been in use.

HFOs, also called hydrofluoroolefins, might be one of the main alternatives for the field of refrigeration; they are currently the fourth generation of refrigerants. In the International Fair of Green Energies in Refrigeration and Air Conditioning, the chemical engineer Nohora Clavijo explained that the HFOs are double bond organic compounds with a shorter atmospheric life, and consequently they have less environmental impact. In addition, the Honeywell company communicated that mixtures HFO/HFC will be utilized to reduce the use of hydrofluorocarbons.

In experimental analyses conducted to compare HFO with the HFC-134a in a vapor compression system, the difference in the coefficient of performance (COP) obtained for the R1234yf is between 3 and 11% smaller to the one obtained using R134a, and for R1234ze is only between 2 and 8% below, using the same compressor for the three cases [5]. In a research involving an energy analysis in a vapor compression system, Yataganbaba *et al.* [6] found R1234yf and R1234ze as appropriate replacements for R134a, with this being the refrigerant mostly used in domestic refrigeration. The HFOs may work similarly in domestic refrigeration systems, without having the environmental penalty of the HFC.

1.1. Low levels of GWP and high safety

The GWP, or global warming potential, of a chemical substance is a consequence of the combination of its radiative forcing or climatic forcing (change in the net irradiance in the transition zone between the troposphere and the stratosphere due to a change in the atmospheric concentration of a gas) and its useful atmospheric life (time it remains in the atmosphere without being disintegrated) [7].

Since hydrofluoroolefins are fluoridated compounds without carbon, they have low levels of GWP and a null ODP, fulfilling the regulations established to protect the ozone layer and the enviroment. In addition, HFOs, as opposed to hydrocarbons (HC), do not have a high risk of flammability, with ASHRAE safety classification between A1 and A2, low or null flammability.

Another great advantage of these refrigerants is their compatibility with a great variety of lubricant oils, and it is not necessary to make important adaptations in existing systems that operate with HFC or HCFC. It does not cause wear in materials due to corrosion, and due to their properties, it does not cause high fatigue in the compressor.

1.2. Implementation of R1234ze in domestic refrigeration

Most vapor compression domestic refrigerators use HFC as working fluid in the cycle; these do not directly damage the ozone layer, but contribute to the greenhouse effect, with R134a being the mostly used. They are also being employed as ideal substitutes to hydrocarbons for domestic refrigeration, products such

as R600a, which has a very low contribution to global warming; however, they are highly flammable if they are not used properly, and it is necessary to replace the complete installation. On the contrary, the HFO-1234ze needs ten times more concentration and 250,000 times more energy than hydrocarbons to become inflamed, only above 30 °C. It is catalogued as a low toxicity and slightly flammable fluid [8–10].

The refrigerant (HFO) R1234ze has efficiency characteristics very similar to the HFC R134a and operates at similar pressures, and so it is not necessary to make important changes to existing systems. The atmospheric life of the refrigerant as a residue is only eighteen days, much smaller than the thirteen years of the R134a. A unique characteristic of the R1234ze is the absence of flammability when mixed with air at less than 30 °C of ambient temperature. For this reason, it is considered not flammable for manipulation and storage [11], [12]. In the research work by Ngoc [13] about a refrigeration cycle that uses HFO-1234ze, HFO-1234yf, R22 and R32 as alternatives to R134a, results show that the cycle that uses R1234ze as refrigerant has the highest coefficient of performance.

Due to these compatibilities and characteristics, a refrigeration system of medium or low pressure|, in this case a domestic conventional refrigerator, may be coupled to operate with R1234ze without requiring important changes to the system.

2. Materials and methods

In the study, the refrigeration cycle that will operate with R1234ze was designed according to the data obtained from a functional domestic refrigerator that operates with R134a. The Genetron Properties software provided by Honeywell and the database of the Buffalo – NY-USA Research Lab are used to model this thermodynamic cycle. In addition, for analyzing the heat transfer of the refrigerant in the refrigeration cycle, a CFD study is conducted using Fluent from ANSYS, in which tests of evaporation in the corresponding device and of temperature distributions in the household appliance are carried out, simulating the real operating conditions.

2.1. Characteristics of the refrigerant

The HFO 1234ze is a pure fluid with low toxicity and flammability, constituted by fluor, hydrogen and oxygen molecules. Table 1 presents the properties of this fluid and compares them with refrigerant 134a.

Table 1. Properties of R134a and R1234ze

Property	R134a	R1234ze
Molecular weight (kg/kmol)	102	114
Critical temp. (°C)	101,1	109,4
Density (kg/m^3)	511,9	489,23
Critical pressure (kPa)	4059	3636
ODP	0	0
GWP	1300	1
ASHRAE classification	A1	A2L

Data taken from the Honeywell catalog [12] and REFPROP [14]

2.2. Refrigeration cycle

For studying the refrigeration cycle, cycle conditions in a refrigerator with R134a for subtropical climate, nominal power of 0.104 kW, and operating with a 1/5 HP Embraco EM3U50HLP compressor, are taken. Table 2 presents the properties of this reference refrigerator.

Table 2. Characteristics of the reference refrigerator

Characteristic	Value
Name	No-frost domestic refrigerator
Type of refrigerant	R134a
Mass of refrigerant	95 kg
Nominal power	104 W
Nominal volume	250 l
Freezer volumen	55 l
Refrigerator volumen	184 l

The refrigerator operates with evaporation and condensation temperatures of –23.3 °C and 54.4 °C, respectively, and the thermodynamic cycle for the R134a is developed according to these conditions. The parameters of the cycle are defined by means of the thermodynamic properties of the refrigerant at the operating conditions in the refrigerator. Figure 1 shows the P-h diagram of the refrigeration cycle with R134a which was made using EES [15] and the data obtained experimentally.

In the calculation of the fundamental performance parameters of the refrigeration cycle, the coefficient of performance (COP) is calculated using Equation 1, and this parameter represents the ratio between the cooling capacity and the energy cost caused by the compression cycle.

$$COP = \frac{Q_e}{W_c} \tag{1}$$

$$Q_e = (h_1 - h_4)\dot{m} \tag{2}$$

Where:

Q_e: Cooling capacity , [W]
W_c: Power consumed, [W]
h_1: Evaporator output enthalpy, [kJ/kg]
h_4: Evaporator input enthalpy, [kJ/kg]
$\dot{m}$: Mass flow of refrigerant, [kg/s]

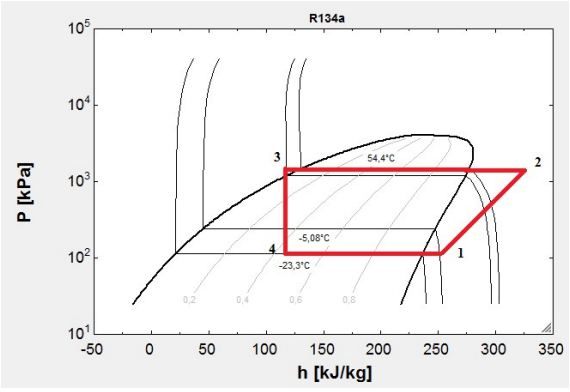


Figure 1. P-h diagram, real refrigeration cycle with R134a [15]

The isentropic efficiency of the compressor is the entropy generated in the real compression compared to the one generated ideally. This parameter is defined by Equation 3.

$$\eta = \frac{h_{2s} - h_1}{h_{2a} - h_1} \tag{3}$$

Where:

η: isentropic efficiency
h_{2s}: Cooling capacity, [kJ/kg]
h_{2a}: Power consumed, [kJ/kg]

2.3. Refrigeration cycle adapted to R1234ze

For comparing the performances of R1234ze and R134a, the cycle of the reference domestic refrigerator is theoretically adapted to operate with HFO, without altering the operating conditions. For this purpose the refrigeration cycle is designed, using Genetron Properties, maintaining the original parameters of the refrigerator under study and entering the power consumed, condensation temperature, evaporation temperature and compressor efficiency. This simulation will be called case A. Figure 2 and Figure 3 present the P-h and T-s diagrams, respectively, of the refrigeration cycle with R1234ze.

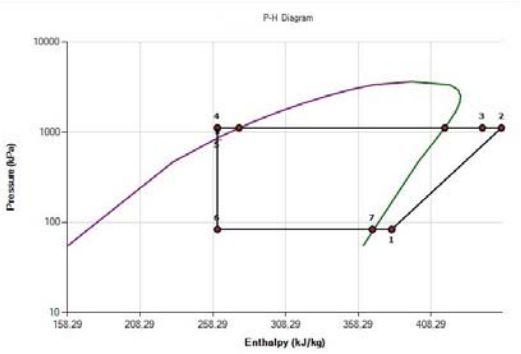


Figure 2. P-h diagram, real refrigeration cycle with R1234ze

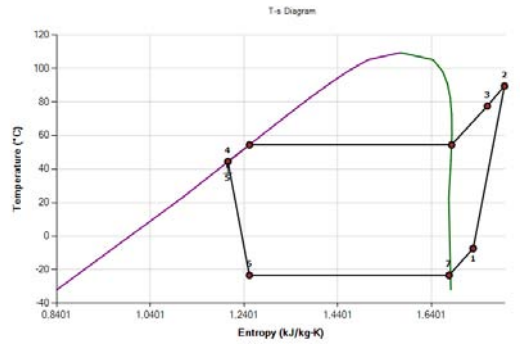


Figure 3. T-s diagram, real refrigeration cycle with R1234ze

2.3.1. Improvements to the thermodynamic cycle

According to the indications of the R1234ze suppliers, the COP of the cycle could be improved compensating the cooling capacity of the system with particular adjustments in it, and besides, improving the performance of the compressor would increase the efficiency of the refrigerator with HFO. The quality of the mixture at the outlet of the evaporator may be better with R1234ze than with R134a, if the heat flow is sufficiently low or if the mass velocity is sufficiently high [16] besides, it has been experimentally seen that the COP improves for cycles with R1234ze at greater values of evaporation temperature [5]. On the other hand, Sánchez *et al.* [17] indicate that a compressor with larger displacement should be used, to achieve the same cooling capacity than with a system with R134a.

An improved cycle is proposed, where some changes are carried to improve the performance of the refrigeration cycle. The operating parameters of the compressor are modified, the condensation and evaporation temperatures are readjusted adapting the operating pressures, the physical characteristics of the system are maintained, as well as the power supplied. It should

be remarked that the modifications are theoretical, and for experimental operation physical changes to the installation of the domestic refrigerator would be required. However, these repairs are minor compared to the required to implement hydrocarbons [10].

Figure 4 presents the thermodynamic diagram of the refrigeration cycle for the cases studied. In case B the performance of the compressor is improved, increasing the volumetric displacement and its isentropic efficiency. In case C the operating temperatures are modified, raising the evaporation temperature by 5.3 °C and reducing the condensation temperature by 4.4 °C. All these modifications seek to improve the performance of the refrigerant in the thermodynamic refrigeration cycle.

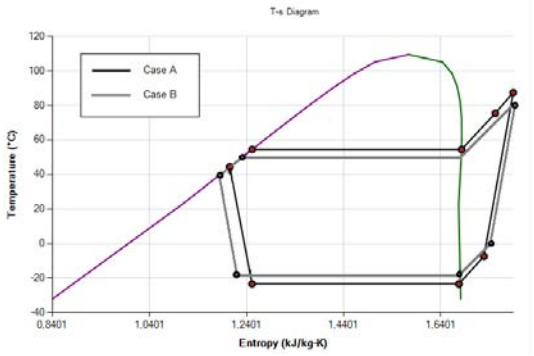


Figure 4. Improvements to the real refrigeration cycle with R1234ze, Case A and Case B, T-s Diagram

2.4. CFD study of heat transfer

In the CFD study it is analyzed what is occurring with the refrigerant in the heat exchanger that carries out the heat absorption, simulating the pass of the refrigerant through the evaporator to obtain data about phase change, temperature and behavior of the R1234ze refrigerant. In addition, it is simulated the air flow in the refrigerator with the cooling capacity supplied by the HFO, to compare with the temperatures achieved in the compartments with the R134a, which are –13° C for the freezer and 7 °C for the refrigerator, with an average operation at an ambient temperature of 18 °C.

The study is carried out in a geometric model with the dimensions of the reference refrigerator that operates with R134a. The model has a total volume of 239 l of space of the service container, between freezer and refrigerator. The evaporator consists of a pipe of copper type K with coil shape and a nominal diameter of 3/8 inches, located in the area of the freezer. The geometry of the control volume for the refrigerant in the evaporator and the control volume for the air within the refrigerator are shown in Figure 5.

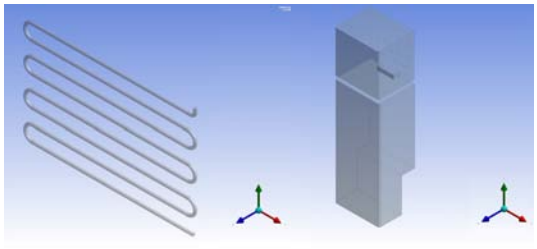


Figure 5. Geomtry of the control volume for simulation. a) Refrigerant in the evaporator. b) Air within the refrigerator

2.4.1. Mathematical models for simulation

For the surrounding air within the refrigerator, the volume of fluid model is used to simulate air circulation and temperature change. The Euler-Euler approach is utilized for the multiphase model in the evaporator, which solves the momentum and continuity equations. Two different phases that interact in the phase change are modeled, the first as liquid or liquid-vapor mixture and the second as vapor. Using the concept of volume fraction with regime equations for each phase, for the concentration of mass in the interaction of the phases it is utilized the model established by Equation 4.

$$A_i = \frac{6\alpha_p(1-\alpha_p)}{d_p} \tag{4}$$

Where:

A_i: Area of the interface
α_p: Volume fraction of phase q
d_p: Diameter of the bubble

The effects of viscosity and turbulence in the fluid are simulated with the $k-\varepsilon$ RNG model, which includes the eddy effect in the turbulence and improves the effect of the Reynolds number for the effective viscosity [18]. The model comprises expressions for each variable, Equation 5 and Equation 6.

For k:

$$\frac{\partial}{\partial r}(pk)+\frac{\partial}{\partial_{xi}}(pku_i) = \frac{\partial}{\partial xj}\left(\alpha_k\mu_{eff}\frac{\partial}{\partial xj}\right)+G_k+G_b-\rho\varepsilon-Y_M+S_k \tag{5}$$

For ε:

$$\frac{\partial}{\partial r}(p\varepsilon)+\frac{\partial}{\partial_{xi}}(p\varepsilon u_i) = \frac{\partial}{\partial xj}\left(\alpha_\varepsilon\mu_{eff}\frac{\partial}{\partial xj}\right)+C_{1\varepsilon}\frac{\varepsilon}{k}(G_k+C_{3s}G_b)- -C_{2\varepsilon}\rho\frac{\varepsilon^2}{k}-R_\varepsilon+S_\varepsilon \tag{6}$$

Where:

$G_{k:}$ Generation of turbulence kinetic energy due to average velocity
G_b: Generation of turbulence kinetic energy due to buoyancy
Y_M: Contribution of the fluctuating expansion on the compressible turbulence at the general dissipation rate
S and C: Constants and terms defined for each variable
$R\varepsilon$: Term added in the RNG method to increase the precision
αk and $\alpha\varepsilon$: Inverse of the effective Prandtl numbers for k and ε, respectively
μ_{eff}: Effective viscosity

The conservation of energy and mass are established by means of balancing the amount of mass that gets in and the amount that gets out; for Euler model the laws that govern these parameters are defined by Equation 7 and Equation 8, respectively.

Conservation of energy:

$$\frac{\partial}{\partial_t}\left(\alpha_q\rho_q h_q\right) + \nabla\cdot\left(\alpha_q\rho_q\vec{u}_q h_q\right) = -\alpha_q\frac{\partial p_q}{\partial t} + \bar{\bar{t}}_q : \nabla\vec{u}_q - \nabla\cdot\vec{q}_q + S_q + \sum_{p=1}^{n}\left(Q_{pq} + \dot{m}_{pq}h_{pq} - \dot{m}_{qp}h_{qp}\right) \quad (7)$$

Conservation of mass:

$$\frac{\partial}{\partial_t}\left(\alpha_q\rho_q\right) + \nabla\cdot\left(\alpha_q\rho_q\vec{u}_q\right) = \sum_{p=1}^{n}\left(\dot{m}_{pq} + \dot{m}_{qp}\right) + S_q \quad (8)$$

Where:

h_q: Enthalpy specified for the secondary phase
Q_{pq}: Intensity of heat exchange between the primary and secondary phases
h_{pq}: Enthalpy of phase change
$\vec{v}_q$: Velocity of the secondary phase q
$\dot{m}_{pq}$: Mass transferred from the primary to the secondary phase
$\dot{m}_{qp}$: Mass transferred from the secondary to the primary phase

Lee model defines the mass transfer by condensation or evaporation of one phase to the other, the liquid phase is defined as the primary phase and the vapor phase as secondary depending on the conditions of the refrigerant when entering the exchanger, this model is described by Equation 9.

$$\frac{\partial}{\partial_t}\left(\alpha_p\rho_p\right) + \nabla\cdot\left(\alpha_p\rho_p\vec{V}_p\right)\dot{m}_{qp} - \dot{m}_{pq} \quad (9)$$

Where:

ρ_p: Vapor phase
$\vec{V}_p$: Velocity of vapor
$\dot{m}_{pq}$: Rate of mass transferred in evaporation
$\dot{m}_{qp}$: Rate of mass transferred in condensation

In addition, Moraga correlation is integrated in the interaction between phases to define the Lift coefficient to simulate the effects caused by the eddy produced as the fluid passes as liquid, the drag coefficient is defined by Schiller-Naumann correlation to simulate the drag in the liquid surface between phases [18].

3. Results and discussion

The thermodynamic parameters obtained in the cycle of the domestic refrigerator with R134a, used as base for the cycle with R1234ze, are shown in Table 3.

Table 3. Parameters of the refrigeration cycle with R134a

Parameter	R134a
Evaporation temp.	–23,3 °C
Overheating	18,3 °C
Condensation temp.	54,4 °C
Subcooling	8 °C
Mass flow	0,0012 kg/s
Isentropic efficiency	0,688
Power consumed	0,104 kW

With the thermodynamic tests conducted in the simulations for R1234ze and the data obtained for R134a, it may be seen that the difference in the coefficient of performance (COP) of the cycle with R1234ze is reduced compared to the cycle with R134a; Figure 6 presents these results.

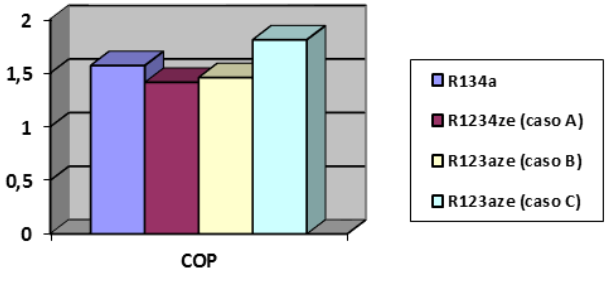


Figure 6. Coefficient of performance (COP)

The most important thing in a refrigeration system is the capacity to produce cold or to eliminate heat; for this purpose, it is calculated the cooling capacity and the heating capacity of the refrigeration cycle. Figure

7 compares these parameters for the cycle with R134a, and cases A, B and C for R1234ze.

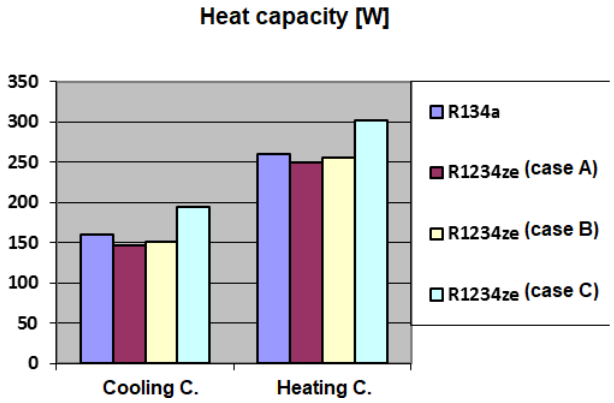


Figure 7. Cooling and heating capacity in thermodynamic cycles

It may be observed that case C has the highest heat capacity due to the modifications carried out at the operating temperatures; however, to carry out this modification it is necessary a greater energy consumption by the compressor, which would reduce the energy efficiency of the refrigerator. On the other hand, cases A and B have capacities very similar to the cycle with R134a.

Another parameter which exhibited important changes in the thermodynamic study, is the discharge temperature of the compressor. This information is useful because a large discharge temperature produces high fatigue in the compressor, and besides it is unfavorable in the thermodynamic cycle since it increases the work consumed by the compressor. Figure 8 shows the data obtained regarding temperature variation in the compressor for both refrigerants.

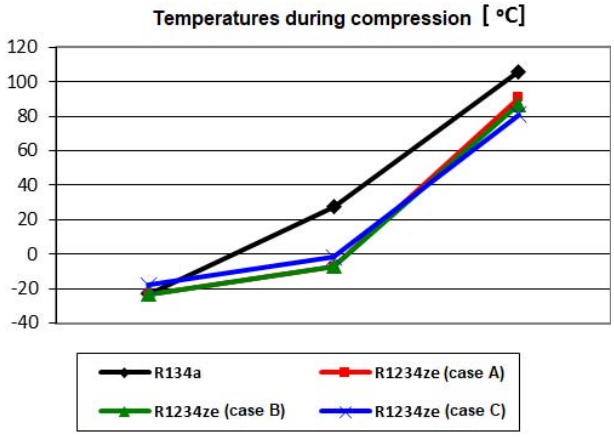


Figure 8. Temperature variation in the compression process

Due to the properties of the HFO, it generates less fatigue in the compressor because of a smaller discharge temperature, and besides it is highly compatible with various materials as opposed to hydrocarbons; this is a great advantage for HFOs in vapor compression cycles. Another important factor for the performance of the compressor is the velocity at which the refrigerant should flow. Table 4 presents the mass flow required for each cycle studied.

Table 4. Mass flow required for each cycle [kg/s]

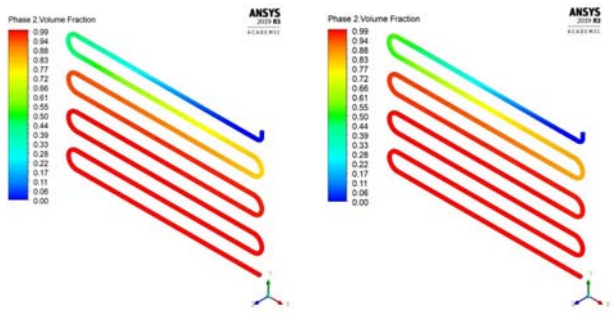
R134a	R1234ze (A)	R1234ze (B)	R1234ze (C)
0,0012	0,0013	0,0014	0,0016

3.1. Results of the CFD study

3.1.1. Refrigerant in the evaporator

In a vapor compression refrigeration cycle, the evaporator is responsible for absorbing heat, transporting it to the compressor, and then expelling it to the environment in the condenser. In this absorption process, the refrigerant experiences a phase change from liquid-vapor mixture to vapor, and the performance of the refrigerant in the process is one of the main factors that determines the cooling capacity of the cycle.

Figure 9 shows the phase change along the evaporator for the R1234ze refrigerant, to verify that the HFO accomplishes the complete evaporation in the exchanger and does not deliver wet vapor to the compressor, which may deteriorate it or produce malfunctioning during compression. The performance is compared with the refrigerant R134a, having a scale from 0 to 1 for the quality of the mixture, with 0 corresponding to the entering state (0.46 for R1234ze and 0.44 for R134a) to the evaporator as liquid-vapor mixture and 1 for saturated vapor.

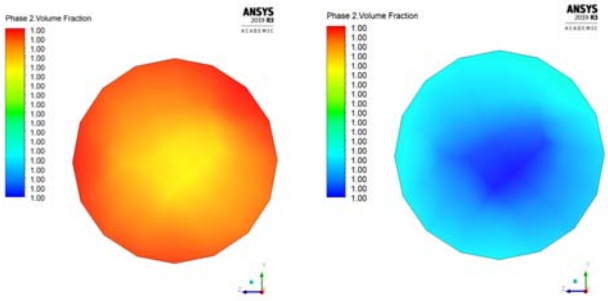


(a) (b)

Figure 9. State of the refrigerant during the phase change in the evaporator. a) R1234ze, b) R134a

The characteristics of the refrigerant at the inlet of the evaporator, such as operating pressure, quality and mass flow, are given for the thermodynamic cycle previously described for each refrigerant. It may be observed that in these conditions the evaporation

is similar for the two refrigerants under comparison, the R134a requires a shorter pipe length to complete the phase change; however, the HFO delivers a larger percentage of superheated vapor. Figure 10 shows a contour of average area measured at the outlet of the evaporator with the total quality of the mixture.



(a) (b)

Figure 10. Quality of the refrigerant at the outlet of the evaporator. a) R1234ze, b) R134a

Values of refrigerant temperature at the outlet of the evaporator are also obtained, Figure 11. It is observed in the results that there is overheating during heat absorption, the temperature is greater at the fluid surface in contact with the wall of the evaporator, reaching –6.83 °C in average, and remaining at the saturation temperature of –23.3 °C within the fluid.

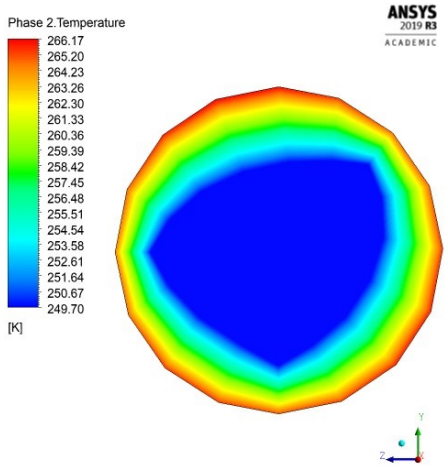


Figure 11. Refrigerant (R1234ze) temperature at the outlet of the evaporator

Compared to the HFC-134a, the HFO-1234ze reaches a higher overheating temperature at the end of the evaporation process, this variation is very small and thus it will no produce an important effect at the inlet of the evaporator. Figure 12 shows a comparison of the temperature at different points of the fluid after it leaves the evaporator.

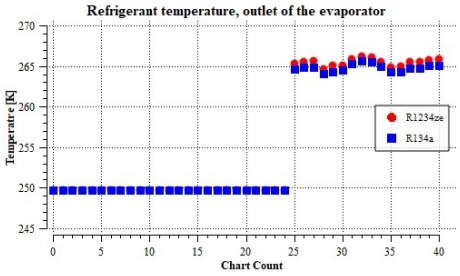


Figure 12. Comparison of overheating temperatures

The results obtained with the CFD study of heat transfer are in accordance with the results found by the thermodynamic analysis conducted in the evaporator with R1234ze; Table 5 shows the data for each case.

Table 5. Results of the studies for the evaporator

Parameter	CFD	Termodynamic study
Evaporation temp [°C]	–23,3	–23,3
Overheating temp. [°C]	–6,83	–7,3
Mass flow [kg/s]	0,0013	0,0013
Quality of the fluid	1	1
Cooling capacity [W]	146,3	146,4

3.1.2. Air flow within the refrigerator

For the air flow in the total control volume of 239 l in the compartments, 55 l for the freezer and 184 l for the refrigerator, the air circulation is simulated at an average velocity of 1.103 m/s generated by the fan of a no-frost domestic refrigerator, which operates at 2070 RPM [18]. Figure 13 schematizes the air flow within the refrigerator.

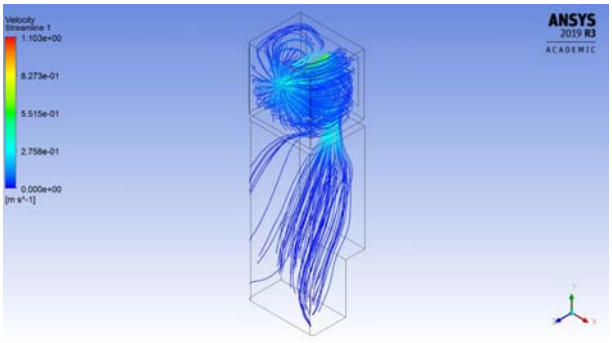


Figure 13. Air flow within the refrigerator

Tests were conducted with this flow for the cooling capacities obtained in each refrigeration cycle with R1234ze in the cases A, B and C under study. Figure 14 presents the results in a volume rendering of the air temperature, obtained with the cooling capacities for each case. Heat flows of –146.3 W for case A, –151.4 W for case B and –193.7 for case C were simulated.

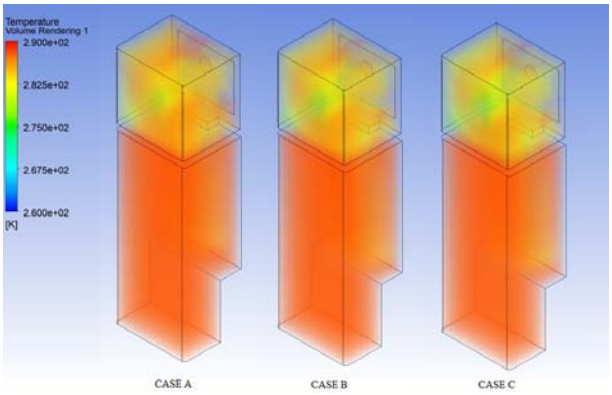


Figure 14. Volume rendering of temperature in the control volume for cases A, B and C

The temperatures reached in the freezer and in the refrigerator do not show a large variation for the three cases under study. Figure 15 schematizes the results of the temperatures at various points of the freezer, and Figure 16 the results for particles in the refrigerator.

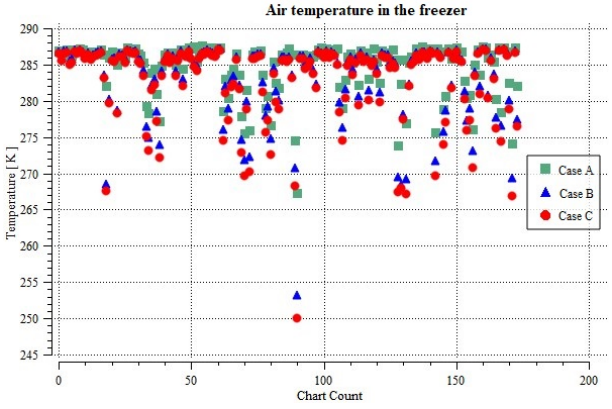


Figure 15. Comparison of temperatures in the freezer for cases A, B and C

The temperature in the freezer reaches an average value of –6 °C for case A, with peaks up to –8 °C; for case B it reaches an average value of –8 °C, with peaks up to –18 °C; and for case C with system redesign it reaches an average value of –12 °C, with peaks up to –22 °C.

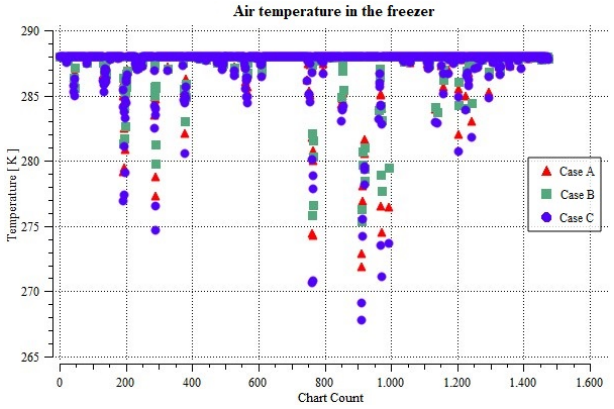


Figure 16. Comparison of temperatures in the refrigerator for cases A, B and C

The temperature in the refrigerator reaches an average value of 6 °C for case A with peaks up to 2 °C; for case B it reaches an average value of 4 °C, with peaks up to 0 °C; and for case C it reaches an average value of 2 °C, with peaks up to –4 °C. These results do not take into account losses due to air that enters when the doors are opened.

4. Conclusions

The substances of the HFO family are considered as the fourth generation of fluoridated refrigerants, due to their physical and environmental properties, and they are an excellent alternative for designing refrigeration or air conditioning systems. The HFO-1234ze is one of these alternatives, which due to its characteristics is a highly likely replacement of R134a in domestic refrigeration systems. Its contribution to global warming is 99.9% smaller than R134a and 75% smaller than R600a, the refrigerants mostly used in domestic refrigeration, and due to their characteristics of null toxicity and slight flammability, they fulfill environmental and safety regulations currently established.

With the results obtained in the thermodynamic and heat transfer tests conducted with Genetron Properties and ANSYS Fluent, besides the mathematical analysis, it may be concluded that refrigerant HFO-1234ze is suitable for working in a domestic refrigerator, showing necessary refrigeration parameters without having the environmental cost of conventional HFCs and the high flammability of hydrocarbons.

With the same operating conditions in a domestic refrigerator, the HFO has a cooling capacity only 8% lower than HFC-134a; a cooling capacity 5% lower is achieved with adaptations in the compression process, and redesigning the system for different working temperatures enables yielding a cooling capacity 20% larger; however, with this redesign the energy efficiency diminishes 2%. The coefficient of performance (COP) of the cycle is 1.57 with R134a, 1.41 with R1234ze

without adjustments, 1.46 for R1234ze with improvements to the compressor and 1.8 for R1234ze with redesign.

The discharge temperature in the compressor is substantially smaller with refrigerant R1234ze compared to R134a, being 16 ºC smaller in average, which significantly favors the useful life and the energy consumption of the compressor. It could be also observed that the intake temperature does not show important variations between the two fluids, and that, in all cases, not wet vapor is delivered to the compressor. The temperatures reached, both within the freezer and within the refrigerator, are very similar to the experimental temperatures of the reference refrigerator, and thus it is expected to achieve a correct refrigeration when using refrigerant R1234ze.

When performing physical adaptations to a refrigeration system to operate with R1234ze, data about cooling capacity and operating temperatures and pressures should be considered. It is necessary to take advantage of the thermodynamic characteristics of the refrigerant in the cycle, because the refrigerant works better with higher evaporation temperatures and its properties make it better for the compression process. In future research works, it is recommended to study mixtures of HFO and HFC to totally replace the latter, without having to replace the devices that constitute the refrigeration system.

References

[1] ONU. (2019) Día Internacional de la Preservación de la capa de ozono, 16 de septiembre. [Online]. Available: https://bit.ly/31DBnQf

[2] D. Fahey, P. A. Newman, J. A. Pyle, B. Safari, M. P. Chipperfield, D. Karoly, D. E. Kinnison, M. Ko, M. Santee, and S. J. Doherty, *Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project-Report No. 58*. World Meteorological Organization, 01 2018. [Online]. Available: https://bit.ly/3oil19H

[3] L. Höglund-Isaksson, P. Purohit, M. Amann, I. Bertok, P. Rafaj, W. Schöpp, and J. Borken-Kleefeld, "Cost estimates of the kigali amendment to phase-down hydrofluorocarbons," *Environmental Science & Policy*, vol. 75, pp. 138–147, 2017. [Online]. Available: https://doi.org/10.1016/j.envsci.2017.05.006

[4] A. Mota-Babiloni, J. Navarro-Esbrí, A. Barragán, F. Molés, and B. Peris, "Drop-in energy performance evaluation of r1234yf and r1234ze(e) in a vapor compression system as r134a replacements," *Applied Thermal Engineering*, vol. 71, no. 1, pp. 259–265, 2014. [Online]. Available: https://doi.org/10.1016/j.applthermaleng.2014.06.056

[5] A. Yataganbaba, A. Kilicarslan, and I. Kurtbaş, "Exergy analysis of r1234yf and r1234ze as r134a replacements in a two evaporator vapour compression refrigeration system," *International Journal of Refrigeration*, vol. 60, pp. 26–37, 2015. [Online]. Available: https://doi.org/10.1016/j.ijrefrig.2015.08.010

[6] M. O. McLinden, A. F. Kazakov, J. Steven Brown, and P. A. Domanski, "A thermodynamic analysis of refrigerants: Possibilities and tradeoffs for Low-GWP refrigerants," *International Journal of Refrigeration*, vol. 38, pp. 80–92, 2014. [Online]. Available: https://doi.org/10.1016/j.ijrefrig.2013.09.032

[7] D. Sánchez, I. Arauzo, J. Catalán Gil, R. Cabello, R. Doménech, and E. Torrella, "Evaluación energética de una instalación frigorífica empleando refrigerantes de bajo GWP," in *CYTEF 2016 – VIII Congreso Ibérico / VI Congreso Iberoamericano de las Ciencias y Técnicas del Frío Coimbra-Portugal, 3-6 mayo, 2016*, 05 2016. [Online]. Available: https://bit.ly/31EdI29

[8] W. C. Whitman and W. M. Jonhson, *Tecnología de la refrigeración y aire acondicionado Tomo II*. Editorial Paraninfo, 2000. [Online]. Available: https://bit.ly/34u3xiN

[9] N. Jara and C. Isaza-Roldan, "Análisis comparativo de sistemas de refrigeración doméstica utilizando refrigerantes R600a y R134a," *Revista I+T+C: Investigación, Tecnología y Ciencia*, vol. 1, pp. 7–15, 08 2015. [Online]. Available: https://bit.ly/2TnGIXo

[10] GASSERVEI, "Ficha técnica R-1234ze," GASSERVEI, Tech. Rep., 2020. [Online]. Available: https://bit.ly/3dUDJ2e

[11] Honeywell, "The environmental alternative to traditional refrigerants. solstice ze refrigerant (HFO-1234ze (E))," Honeywell, Tech. Rep., 2018. [Online]. Available: https://bit.ly/3mhCY6d

[12] N. A. Lai, "Equations of state for HFO-1234ze(E) and their application in the study on refrigeration cycle," *International Journal of Refrigeration*, vol. 43, pp. 194–202, 2014. [Online]. Available: https://doi.org/10.1016/j.ijrefrig.2013.11.011

[13] E. W. Lemmon, I. H. Bell, M. L. Huber, and M. O. McLinden, *REFPROP Documentation Release 10.0*. National Institute of Standards and Technology (NIST), 2018. [Online]. Available: https://bit.ly/3omeSct

[14] K. A. Klein and F. L. Alvarado, *EES-Engineering Equation Solver*. Version 6.648 ND, F-Chart Software, Middleton, 2004.

[15] G. A. Longo, S. Mancin, G. Righetti, and C. Zilio, "R1234yf and R1234ze(e) as environmentally friendly replacements of R134a: Assessing flow boiling on an experimental basis," *International Journal of Refrigeration*, vol. 108, pp. 336–346, 2019. [Online]. Available: https://doi.org/10.1016/j.ijrefrig.2019.09.008

[16] D. Sánchez, R. Cabello, R. Llopis, I. Arauzo, J. Catalán-Gil, and E. Torrella, "Energy performance evaluation of R1234yf, R1234ze(e), R600a, R290 and R152a as low-GWP R134a alternatives," *International Journal of Refrigeration*, vol. 74, pp. 269–282, 2017. [Online]. Available: https://doi.org/10.1016/j.ijrefrig.2016.09.020

[17] ANSYS, *ANSYS Fluent Theory Guide*. SAS IP, Inc., 2013. [Online]. Available: https://bit.ly/3jk7GtC

[18] H. Benjumea, C. Isaza-Roldan, S. Rio, N. Jara, and J. Ospina, "Simulación del flujo de aire al interior de un refrigerador doméstico no-frost," in *VII Congreso Ibérico de Ciencias y Técnicas del Frío*, 06 2014. [Online]. Available: https://bit.ly/2TnHwvE

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.02
pISSN: 1390-650X / eISSN: 1390-860X

Impact of oversampling algorithms in the classification of Guillain-Barré syndrome main subtypes

Impacto de los algoritmos de sobremuestreo en la clasificación de subtipos principales del síndrome de Guillain-Barré

Manuel Torres-Vásquez[1,2], José Hernández-Torruco[1],
Betania Hernández-Ocaña[1], Oscar Chávez-Bosquez[1*]

Received: 15-05-2020, Reviewed: 22-07-2020, Accepted after review: 25-09-2020

Resumen

Guillain-Barré Syndrome (GBS) is a neurological disorder where the body's immune system attacks the peripheral nervous system. This disease evolves rapidly and is the most frequent cause of paralysis of the body. There are four variants of GBS: Acute Inflammatory Demyelinating Polyneuropathy, Acute Motor Axonal Neuropathy, Acute Sensory Axial Neuropathy, and Miller-Fisher Syndrome. Identifying the GBS subtype that the patient has is decisive because the treatment is different for each subtype. The objective of this study was to determine which oversampling algorithm improves classifier performance. In addition, to determine whether balancing the data improves the performance of the predictive models. Three oversampling methods (ROS, SMOTE, and ADASYN) were applied to the minority class. Three classifiers (C4.5, SVM and JRip) were used.

Abstract

El síndrome de Guillain-Barré es un trastorno neurológico donde el sistema inmune del cuerpo ataca al sistema nervioso periférico. Esta enfermedad es de rápida evolución y es la causa más frecuente de parálisis del cuerpo. Existen cuatro variantes de SGB: polineuropatía desmielinizante inflamatoria aguda, neuropatía axonal motora aguda, neuropatía axonal sensorial aguda y síndrome de Miller-Fisher. Identificar el subtipo de SGB que el paciente contrajo es determinante debido a que el tratamiento es diferente para cada subtipo. El objetivo de este estudio fue determinar cuál algoritmo de sobremuestreo mejora el rendimiento de los clasificadores. Además, determinar si balancear los datos mejoran el rendimiento de los modelos predictivos. Aplicamos tres métodos de sobremuestro (ROS, SMOTE y ADASYN) a la clase minoritaria, utilizamos tres clasificadores (C4.5, SVM y JRip).

[1,*]División Académica de Ciencias y Tecnologías de la Información, Universidad Juárez Autónoma de Tabasco, Cunduacán, Tabasco, México. Corresponding author ✉: oscar.chavez@ujat.mx.
https://orcid.org/0000-0001-8475-0914 https://orcid.org/0000-0003-3146-9349
https://orcid.org/0000-0001-5700-7615 https://orcid.org/0000-0002-0324-9886
[2]Tecnológico Nacional de México campus Centla, División Sistemas Computacionales, Frontera, Centla, Tabasco, México.

Suggested citation: Torres-Vásquez, M.; Hernández-Torruco, J.; Hernández-Ocaña, B. and Chávez-Bosquez, O. (2021). «Impact of oversampling algorithms in the classification of Guillain-Barré syndrome main subtypes». Ingenius. N.° 25, (january-june). pp. 20-31. doi: https://doi.org/10.17163/ings.n25.2021.02.

The performance of the models was obtained using the ROC curve. Results show that balancing the dataset improves the performance of the predictive models. The SMOTE Algorithm was the best balancing method, in combination with the classifier JRip for OVO and the classifier C4.5 for OVA.

Keywords: ADASYN, Classifiers, Unbalance, ROS, SMOTE, Wilcoxon.

El rendimiento de los modelos se obtuvo mediante la curva ROC. Los resultados muestran que balancear el *dataset* mejora el rendimiento de los modelos predictivos. El algoritmo SMOTE fue el mejor método de balanceo en combinación con el clasificador JRip para OVO y el clasificador C4.5 para OVA.

Palabras clave: ADASYN, clasificadores, desbalanceo, ROS, SMOTE, Wilcoxon.

1. Introduction

The Guillain-Barré Syndrome (GBS) is defined as an autoimmune polyradiculoneuropathy and is the most frequent cause of acute generalized paralysis [1]. The GBS occurs when the immune system attacks part of the peripheral nervous system. This disease evolves rapidly and is characterized by weakness of the legs which further advances to the arms, "ascending paralysis". The initial symptoms are muscle weakness and tingling in the extremities. The severe cases require mechanical ventilation. The cause is unknown, but two thirds of the cases precede to a respiratory infection or acute gastroenteritis. It has been recently associated to the Zika virus. The GBS affects between 0.4 and 2.4 cases per 100,000 inhabitants/year. It appears at any age, but it often shows a higher frequency in people between 50 and 80 years old. It is slightly more frequent in men than in women. It has a mortality rate between 2% and 8%. Most people eventually recover completely when the disease is mild or moderate, and in other cases there may remain harms in the nervous system for long time or even permanently [2]. Electrophysiological and nerve conduction studies determine the tests for diagnosing GBS. There are four main subtypes of GBS:

- Acute Inflammatory Demyelinating Polyneuropathy **(AIDP)**.
- Acute Motor Axonal Neuropathy **(AMAN)**.
- Acute Sensory Axial Neuropathy a **(AMSAN)**.
- Miller-Fisher Syndrome **(MF)**.

The recovery of the patient largely depends on the prompt identification of the subtype of GBS. Each subtype should be treated in a different manner, and the treatment and costs vary according to the subtype developed by the patient. In severe cases that generate temporary or permanent immobility, the rehabilitation therapies are often long and costly generating psychological and economic implications to the sick person and to the relatives.

Machine Learning is a branch of Artificial Intelligence that uses different mathematical, statistical and optimization techniques, with the purpose of developing information analysis tools so that computers «learn» through examples [3]. At present, disciplines such as finance, oil, marketing, sales and health utilize automatic learning as technological tool to make predictions. Specifically, in the health area, an increasing number of models are being developed for diagnosing diseases such as cancer [4], [5], diabetes [6], [7], Parkinson [8] and Alzheimer [9], with excellent results.

Classification algorithms are in charge of analyzing the data provided and determining the patients that are healthy and the ones that are sick. However, one of the most common problems in medical diagnosis is the disproportionality of cases. In real life there are more healthy patients than sick patients. For example, if it is desired to diagnose patients with diabetes, it will be found that a larger number of people are healthy and a smaller number are sick with diabetes. This disproportionality in the data is known as data unbalancing. There are two types of unbalancing: binary and multiclass unbalancing. Binary unbalancing occurs when in a dataset of two classes, one of the classes has a larger number of data (majority class) with respect to the other class (minority class). On the other hand, multiclass unbalancing occurs when the dataset comprises more than two classes, and the data distribution is unequal for each of the classes [10].

Data unbalancing may affect the result of the classifiers since it tends to bias the results towards the majority class (healthy patients). The standard classification algorithms are built for balanced data, i.e., the same number of healthy and sick cases. For example, for the case of patients with diabetes, the classifier will ignore the patients with diabetes and will only take into account healthy patients. The problem is that it is desired to determine sick patients and not the healthy ones. For this reason, it is necessary to use techniques that help balancing the data.

In the specialized literature there are three techniques most commonly used to overcome the problem of data unbalancing [11].

- **At the data level.** This technique adds or eliminates data to the class, until balancing the dataset. This technique is also known as sampling and is divided in three groups:
  - Oversampling: consists in adding data to the minority class until reaching balance with the majority class.
  - Downsampling: consists in eliminating data from the majority class until reaching equilibrium with the minority class.
  - Hybrid: this technique combines oversampling and downsampling simultaneously, to reach a better balance between classes.
- **At the algorithm level.** They adapt or create classification algorithms to reinforce the prediction of the class.
- **Sensitive cost.** considers the costs associated with the incorrect classification of the samples. It uses different cost matrices that describe the costs of incorrectly classifying any particular data example.

The technique at the data level is one of the most popular because it is independent of the classifier used, and besides the data are treated before being used

by the classifier. The oversampling technique is the most commonly used since it adds data to the minority class. There are different oversampling techniques that generate data, yielding good results with respect to downsampling which may eliminate important data and affect the result of the classifier [12].

On the other hand, besides data unbalancing, the distribution of the instances affects the results of the classifiers [13]. There are techniques that add synthetic data to the minority class and locate them in strategic places to resolve the unbalancing problem and the position of the instances.

The objective of this study was twofold. The first was identifying which of the three oversampling algorithms used to balance the original GBS dataset improves the results of the classification algorithm. The second objective was to establish if balancing the data improves the performance of the predictive models created with balanced data, with respect to models created with unbalanced data. For this purpose, Wilcoxon statistical test is utilized to know if there is a statistically significant difference between such models. At present, there are no studies in the specialized literature to identify the main subtypes of the GBS using Automatic Learning algorithms. In previous studies [14], [15], predictive models were created using the original unbalanced dataset. In this experimental study, the training subsets are balanced using three oversampling techniques (ROS, SMOTE and ADASYN). Results demonstrate that balancing the data improves the performance of the predictive models. A performance of 90% was achieved in some cases.

For this study, two binarization techniques (OVO and OVA) were first used to create 10 binary subsets. Then, the subsets were divided in training sets with 66% of the data, and test sets with 33% of the data. Once the training data were obtained, three balancing methods (ROS, SMOTE and ADASYN) were applied to oversample the minority class and balance it with the majority class. Once the data were balanced, three classification algorithms were applied with different approaches: C4.5 (decision tree), SVM (Support Vector Machine), JRip (Ripper). The performance of the predictive models was determined using the Area Under the Curve (AUC) of the ROC Curve. The results of the predictive models are the average of the AUC for 60 runs. At last, Wilcoxon test is applied to the models created with balanced data that outperformed the models created with unbalanced data, to know if there is a statistically significant difference between such models.

2. Materials and Methods

2.1. Dataset

The dataset used in this study is a collection of 129 patients diagnosed with GBS. One of the 4 main subtypes of GBS was identified to each of these patients. Table 1 shows the main features of the dataset.

Table 1. Features of the dataset

Characteristic	Value
Number of classes	4
Number of instances	129
Number of attributes	16
Class 1 Instances (AIDP)	20
Class 2 Instances (AMAN)	37
Class 3 Instances (AMSAN)	59
Class 4 Instances (MF)	13

This information was obtained through the National Institute of Neurology and Neurosurgery of the City of Mexico (Instituto Nacional de Neurología y Neurocirugía de la Ciudad de México). The original dataset comprises 356 variables. In a previous paper 16 variables were identified as the most relevant ones [16]. The first 4 variables were clinical, and the following 14 belong to the nerve conduction test. The variables used in the experiments are shown in the following:

v22: Symmetry (in weakness)
v29: Affectation of extraocular muscles
v30: Ptosis
v31: Cerebellar implication
v63: Amplitude of the left median motor nerve
v106: Area under the curve of the left ulnar motor nerve
v120: Area under the curve of the right ulnar motor nerve
v130: Amplitude of the left tibial motor nerve
v141: Amplitude of the right tibial motor nerve
v161: Area under the curve of the right peroneal motor nerve
v172: Amplitude of the left median sensory nerve
v177: Amplitude of the right median sensory nerve
v178: Area under the curve of the right median sensory nerve
v186: Latency of the right ulnar sensory nerve
v187: Amplitude of the right ulnar sensory nerve
v198: Area under the curve of the right sural sensory nerve

2.2. Automatic learning algorithms

2.2.1. Oversampling algorithms

Oversampling algorithms are a technique at the data level that add data to the minority class, with the purpose of balancing the unbalanced dataset. There are diverse algorithms to oversample the classes. Three

techniques that generate instances with different approaches were used for this study:

1. The Random Oversampling (ROS) Algorithm obtains a random sample from instances of the minority class and makes a copy of them. The duplicated instances are placed randomly in the dataset. ROS is a non-heuristic method whose objective is to balance the minority class with the majority class [17].

2. The Synthetic Minority Oversampling Technique (SMOTE) oversamples the minority class generating synthetic instances with the purpose of balancing the minority class with the majority class [18]. The new synthetic instances are generated through interpolation between various instances of minority classes, based on the nearest neighbor rule. SMOTE performs this procedure in the «feature space». The procedure to generate synthetic data is the following: (a) Determine the oversampling percentage necessary to be generated. (b) In order to generate the synthetic objects, carry out the following procedure: (b1) Randomly select an instance of a minority class. (b2) Randomly choose its k-nearest neighbors according to the Euclidean distance. (b3) Take the difference between the feature vector and each of the selected neighbors. (b4) Multiply the difference times a random number between 0 and 1. (b5) Add this last value to the original value of the sample. (b6) Return the synthetic sample. (c) The new synthetic sample will be placed between the instance originally selected and each of the k-nearest neighbors.

   The main difference between SMOTE and ROS is that ROS duplicates data from the minority class and adds them randomly. SMOTE generates synthetic data and places them in a neighborhood of the minority class.

3. The Adaptive Synthetic Sampling Approach for Imbalanced Learning (ADASYN) is an extension of SMOTE. ADASYN has two objectives: the first is creating synthetic instances through linear interpolation between the instances of the minority class, to reduce the imbalance of the minority class with the majority class of the dataset. The second objective that makes ADASYN different with respect to SMOTE is that the data generated adaptively changes the decision boundary adding data in the zone of the minority class difficult to learn compared to the data of the minority class which are easy to learn, through a density distribution. ADASYN seeks to give more weight to the data of the minority class which are difficult to learn [19].

2.2.2. Classification algorithms

Three classification algorithms that determine their results through different approaches were utilized. The objective is contrasting the results of each of them:

1. Decision tree (C4.5): Is a supervised learning algorithm in which each branch node represents a choice among various options, and each leaf node represents a decision. The classification technique is performed by means of division criteria, with a structure of inverted tree, similar to a flowchart. It handles continuous and discrete characteristics. It has high precision, stability, is fast, easy to interpret and robust in the presence of noise. C4.5 bases its results in a hierarchical and inductive learning manner, i.e., in the discovery of patterns from examples [20].

2. Support Vector Machine (SVM): Is a supervised learning algorithm which is employed for binary classification. It belongs to the family of linear classifiers, i.e., the original data are resized by means of a mathematical function to search for a linear separability between them. SVM is based on the concept of constructing an optimal hyperplane, i.e., it creates a straight line that separates the classes. The objective is to find the best hyperplane that divides the dataset and maximizes the margin between the classes [21].

3. Ripper (JRip): Is one of the most popular algorithms for classification problems, with a rule-based approach. The classes are examined in increasing size, and an initial set of rules is generated for the class using the reduced incremental error JRip (RIPPER). It proceeds treating all examples of a particular sense in the training data as one class, and finding a set of rules to cover all members in that class. Afterwards, it passes to the following class and does the same, repeating this procedure until all classes are covered [22].

2.3. Performance Measure

The performance of the classification algorithms is evaluated using the graph or curve of the Receiver Operating Characteristics (ROC), and the Area Under the Curve (AUC). The ROC curve measures how well are the predictions classified, as well as the quality of the model predictions [23]. The ROC curve is defined as the sensitivity, which is the rate of true positives shown in Equation 1. The 1-specificity is the rate of false positives, shown in Equation 2. For this experiment, it is used to identify among one of the GBS subtypes.

$$sensitivity = \frac{VP}{VP + FN} \tag{1}$$

$$1 - specificity = \frac{FP}{VN + FP} \tag{2}$$

The Area Under the Curve (AUC) enables identifying a class. For example, recognizing if a patient suffers a particular disease or is healthy. In this performance measure, the values $\geq .900$ are considered excellent models. The values $\geq .700$ indicate that they are good models. However, values $\leq .500$ are considered bad models.

2.4. Binarization techniques

In classification problems it is common to find datasets that are constituted by more than two classes, which are known as multiclass datasets. Some classification algorithms are only capable of discriminating between two classes. For this reason, it is common to transform a multiclass problem in binary subproblems. Two binarization techniques are found in the literature; One-Vs-One (OVO) and One-Vs-All (OVA) [24].

The OVO technique divides a problem of n classes into $n(n-1)/2$ binary subproblems, forming all possible pairs of classes. The OVA technique takes a class as the minority class, and the remaining classes are combined to form the majority class. This procedure is performed n times according to the number of classes that constitute the dataset. The OVO and OVA binarization techniques are used to discriminate one class from the others. In medical diagnosis problems, they are used to identify a sick patient from other healthy patients. Figures 1 and 2 show the 4 subsets obtained with the OVA technique and the 6 subsets obtained with the OVO technique, from the original GBS dataset.

2.5. Validation

For each classification, the model is validated using the train-test evaluation. The dataset is divided in two subsets of data. The first is the training data, which are used to build the model. The second are called test data, which are maintained apart and are used to evaluate the model. It is employed $\frac{2}{3}$ of the dataset for training and $\frac{1}{3}$ for testing the model.

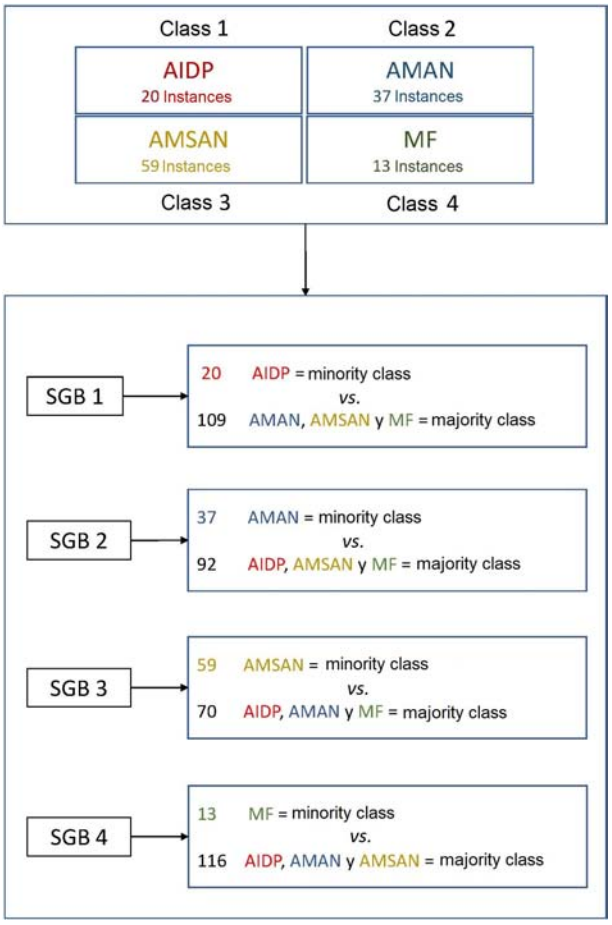


Figure 1. Binarization One-Vs-All (OVA)

3. Experimental procedure

As first step, the original unbalanced multiclass dataset was converted in two binary subproblems. Two different binarization techniques (OVO and OVA) are utilized. The difference between the two binarization techniques is the following: the OVO technique creates all possible combinations that can be formed with the n classes that constitute a dataset; on the other hand, OVA takes one class and converts it in minority class and the remaining classes are combined to form the majority class. OVA creates subsets depending on the total number of classes in the original dataset. The objective of creating binary subsets is that the balancing methods used in this study identify only two classes, the minority class which is oversampled until it is balanced with the majority class.

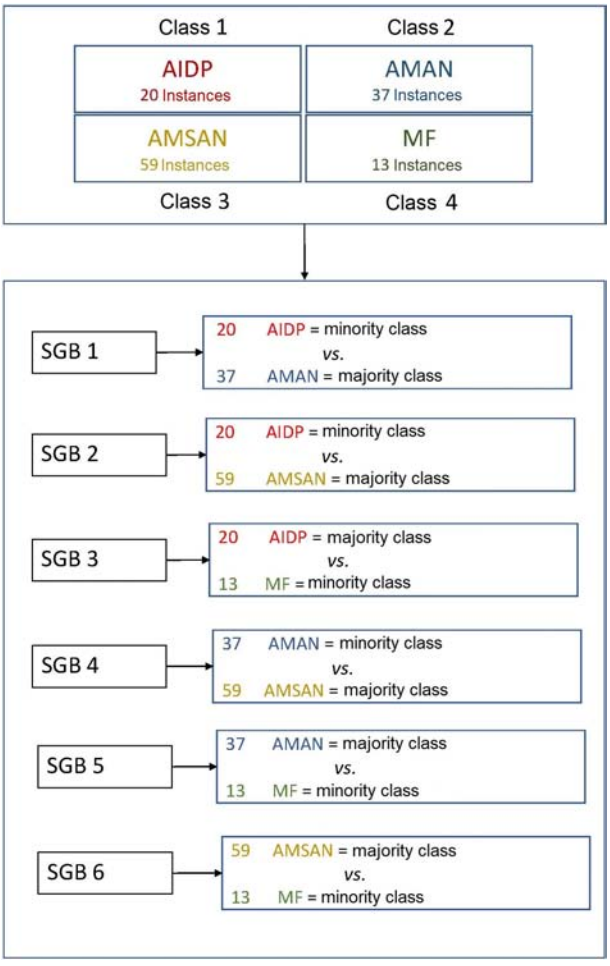


Figure 2. Binarization One-Vs-One (OVO)

Table 2. Subsets obtained with the OVA technique

Subset	Minority class	Majority class
SGB1	20	109
SGB2	37	92
SGB3	59	70
SGB4	13	116

A total of 10 binary datasets were obtained applying the two binarization techniques. Table 2 shows the 4 datasets created with the OVA technique, and Table Tabla 3 shows the 6 binary datasets created with the OVO technique. The first column shows the subsets obtained with binarization technique. The second column contains the number of instances that constitute the minority class. The third column shows the number of instances in the majority class. It may be observed that the OVA technique has a greater data unbalance between the minority class and the majority class, with respect to the OVO technique.

Table 3. Subsets obtained with the OVO technique

Subset	Minority class	Majority class
SGB1	20	37
SGB2	20	59
SGB3	13	20
SGB4	37	59
SGB5	13	37
SGB6	13	59

As a second step, each of the 10 subsets were divided. The data were split in 2/3 for training and the remaining 1/3 for testing. In the following, three oversampling algorithms (ROS, SMOTE and ADASYN) were applied to the minority class in the training data, until balancing it with the majority class. The testing data were used to measure the performance of the models obtained.

Table 4. Results of the balanced subsets applying oversampling methods to the minority class for OVA

Subset	Data A	Data B	Class a	Class b
SGB1	14	59	73	73
SGB2	25	37	61	62
SGB3	40	7	47	47
SGB4	9	69	78	78

Data A: Unbalanced training data.
Data B: Data generated with SMOTE, ROS and ADASYN.
Clase a: Balanced minority class.
Clase b: Original majority class.

Table 4 shows the 4 balanced subsets for the OVA technique. Table 5 shows the 6 balanced subsets for the OVO technique. The first column shows the subsets of the binarization technique. The second column shows the minority class with the number of instances that constitute it. The third column shows the number of instances that were generated for each oversampling algorithm. Columns 4 and 5 show the balanced minority class and majority class, respectively.

The next step was obtaining the predictive models applying three classification algorithms (C4.5, SVM and JRip) to the 10 balanced subsets. It were conducted 60 independent runs calculating the Area Under the Curve (AUC) for the 10 subsets. The predictive models are the result of the average of the AUCs for the 60 runs. On the other hand, the same procedure was carried out using the unbalanced subsets to obtain predictive models with unbalanced data.

Table 5. Results of the balanced subsets applying oversampling methods to the minority class for OVO

Subset	Data A	Data B	Class a	Class b
SGB1	14	9	23	23
SGB2	14	26	40	40
SGB3	9	5	14	14
SGB4	25	15	40	40
SGB5	9	16	25	25
SGB6	9	31	40	40

Datos A: Unbalanced training data.
Datos B: Data generated with SMOTE, ROS and ADASYN.
Clase a: Balanced minority class.
Clase b: Original majority class.

The last step was to compare the performance of the models obtained with balanced data, with the models obtained with unbalanced data. The Wilcoxon statistical test was used to know if there is a statistically significant difference between the models, provided that balanced models have outperformed unbalanced models. A significance value 0.05 was utilized.

The experiments were conducted in the R software, designed for statistical analysis. The R Studio version 1.2.1335 was utilized as integrated development environment. The packages used for balancing the data were: unbalanced for the ROS algorithm [25], DMwR for the SMOTE algorithm [26] and UBL for the ADASYN algorithm [27]. RWeka 0.4-39 [28] was used for the classification algorithms C4.5 and JRip, and 071 1.7-0 [29] was used for the SVM classifier.

The linear SVM classifier was optimized through the tune function, assigning values of 0.001, 0.01, 0.1, 1, 10, 50, 80, 100 for the parameter C. The JRip and C45 classifiers do not require optimization of hyperparameters.

4. Results and Discussion

Tables 6 and 9 show the results of the predictive models obtained. Three balancing methods (ROS, SMOTE and ADASYN) were applied. Six unbalanced subsets obtained were oversampled with the OVO binarization technique, and four subsets with the OVA binarization technique. Each value is the average of the results obtained through 60 runs. The classifiers C4.5, SVM and JRip were applied once the training set was balanced. The models were evaluated using the ROC metrics. The Wilcoxon statistical test was applied to the balanced models against the unbalanced models, when the balanced models outperformed the unbalanced models, with the objective of knowing if the performance of the balanced models obtained a statistically significant difference.

The structure of the Tables is the following: the first column shows the subsets obtained by means of the OVO and OVA binarization techniques, the GBS subtypes that constitute it, as well as the number of instances for each subtype. The second column shows the three classifiers used to obtain the predictive models for each subset. The third column shows the results of the predictive models using unbalanced data. Columns 4, 5 and 6 show the models obtained using balanced data applying three oversampling techniques (ROS, SMOTE and ADASYN). It is also observed that the values in bold letter are the predictive models which, besides outperforming unbalanced models, obtained a statistically significant difference. Table 6 shows the results of the 72 predictive models obtained using the OVO binarization technique.

Of these models, 18 were created with unbalanced data and 54 were obtained using balanced data applying the three oversampling methods. It was found that 32 balanced models could not outperform the unbalanced models. Other 15 balanced models outperformed the unbalanced models, but no statistically significant difference was found. On the other hand, 7 balanced models outperformed the unbalanced models, and in addition they had a statistically significant difference.

The best results were achieved with the subset GBS6, obtaining 3 models with statistically significant difference. On the other hand, the subsets GBS2 and GBS4 had 2 models each with statistically significant difference. The subsets GBS1, GBS3 and GBS5 exhibited the worst performance with respect to the unbalanced models, since a statistically significant difference was not found in any of them.

With respect to the balancing methods, Table 7 shows the results of the ranking obtained for each method. These results were obtained assigning a position to each method depending on its performance with each subset. For every row, a value is assigned to each oversampling method. In the first row, a value of 1 is assigned to SMOTE, since it obtained the best performance. A value of 2 was assigned to ROS since it obtained the second-best performance, and finally the value of 3 is assigned to ADASYN because it was the method with the worst performance. This operation is performed for every row. Subsequently, all the values for each method are added and divided by the number of rows to obtain the average. For example, SMOTE obtained the first place 5 times, the second place 6 times, the third place 5 times and the fourth place 2 times. The sum of these values is 40, which is divided by the number of rows in the table, 18 for this case. The result is 2.222, which holds number 1 in the ranking [30] because it is the lowest average.

For OVO, the SMOTE algorithm was the balancing method with the best performance, with an average score of 2.2222. The algorithms ADASYN and ROS held the second place, because both obtained the same average score of 2.7222. With respect to the classifiers, Table 8 shows that the JRip classifier obtained the best performance with an average score of 1.6667. The C4.5

Table 6. Table of results of the predictive models applying ROS, SMOTE and ADASYN to oversample the minority class.

Subset	Classifier	Unbalanced data	Balancing applying ROS	Balancing applying SMOTE	Balancing applying ADASYN
GBS1	C4.5	0.9604	0.9514	0.9576	0.9292
AIDP-AMAN	SVM	0.9576	0.9465	0.9618	0.9486
20-37	JRip	0.9563	0.9507	0.9403	0.9396
GBS2	C4.5	0.8585	0.8160	0.8551	0.8529
AIDP-AMSAN	SVM	0.8472	0.8306	0.8333	0.8484
20-59	JRip	0.8260	0.8178	**0.8549***	**0.8545***
GBS3	C4.5	0.8132	0.8111	0.7965	0.7854
AIDP-MF	SVM	0.6556	0.6340	0.6535	0.6792
20-13	JRip	0.8556	0.8493	0.7382	0.8396
GBS4	C4.5	0.9258	0.9093	0.9093	0.8897
AMAN-AMSAN	SVM	0.8760	0.8692	0.8827	0.8845
37-59	JRip	0.8782	**0.9059***	**0.9065***	0.8877
GBS5	C4.5	0.8736	0.8826	0.8868	0.8486
AMAN-MF	SVM	0.8806	0.8729	0.8847	0.8910
37-13	JRip	0.8854	0.8958	0.8889	0.8833
GBS6	C4.5	0.8007	**0.8411***	0.7839	0.8209
AMSAN-MF	SVM	0.7089	**0.7600***	0.7534	**0.7746***
59-13	JRip	0.8580	0.8561	0.8720	0.8264

The values are the average of 60 runs of the ROC curves using OVO.

classifier obtained the second place with an average score of 1.8333. At last, the SVM classifier obtained the worst performance with an average score of 2.500.

Table 7. Results of the ranking by balancing method for OVO

Method	Ranking	Average score
SMOTE	1	2.2222
ADASYN	2	2.7222
ROS	2	2.7222

Table 9 shows the results of 48 predictive models, obtained using the OVA binarization technique. Among these, 12 models were created with unbalanced data and 36 were obtained using balanced data applying three oversampling methods. It was found that 15 balanced models could not outperform the unbalanced models; 9 balanced models outperformed the unbalanced models, but no statistically significant difference was found. On the other hand, 12 balanced models outperformed the unbalanced ones, and besides had a statistically significant difference.

Table 8. Results of the ranking by classifier for OVO

Classifier	*Ranking*	Average score
JRip	1	1.6667
C4.5	2	1.8333
SVM	3	2.5000

The best performances were obtained with the subsets GBS1 and GBS4. In the subset GBS1, 8 balanced models improved the unbalanced models, of which 5 models obtained a statistically significant difference. In the GBS4 subset, 6 balanced models outperformed the unbalanced models, of which 5 models obtained a statistically significant difference. With the subset GBS2, 5 balanced models outperformed the unbalanced models, but only 2 models obtained a statistically significant difference. In the SGB3 subset it performed the worst. Only 3 balanced models exceeded the unbalanced data, without finding a statistically significant difference.

Table 10 shows the results of the ranking for the balancing methods applying the OVA binarization technique. The SMOTE algorithm obtained the best performance with an average score of 1.9167, holding the first place. The ADASYN algorithm obtained the second place, with an average score of 2.1667. At last, ROS was the balancing algorithm with the worst performance, holding the third place with an average

Table 9. Table of results of the predictive models applying ROS, SMOTE and ADASYN to oversample the minority class

Subset	Classifier	Unbalanced data	Balancing applying ROS	Balancing applying SMOTE	Balancing applying ADASYN
GBS1	C4.5	0.7894	0.7873	0.8042	**0.8162***
AIDP-ALL	SVM	0.7162	0.7262	**0.7750***	**0.7722***
20-109	JRip	0.7826	0.7921	**0.8102***	**0.8215***
GBS2	C4.5	0.8729	0.8653	0.8900	0.8949
AMAN-ALL	SVM	0.8564	0.8489	0.8490	**0.8871***
37-92	JRip	0.8608	0.8513	0.8699	**0.8949***
GBS3	C4.5	0.8723	0.8455	0.8795	0.8493
AMSAN-ALL	SVM	0.7948	0.7982	0.7881	0.7827
59-70	JRip	0.8470	0.8358	0.8442	0.8536
GBS4	C4.5	0.7808	0.7806	**0.8951***	0.7331
MF-ALL	SVM	0.6464	**0.7590***	**0.7516***	**0.6991***
13-116	JRip	0.8319	0.8440	**0.8826***	0.7882

The values are the average of 60 runs of the ROC curves using OVA.

score of 3.0833.

With respect to the classifiers, the results of the ranking are observed in Table 11. The C4.5 classifier obtained the first place with an average score of 1.2500. The JRip classifier ended up in the second place, with an average score of 1.500. The third place was obtained by the SVM classifier, with an average score of 2.7500.

Table 10. Results of the ranking by balancing method for OVA

Method	*Ranking*	Average score
SMOTE	1	1.9167
ADASYN	2	2.1667
ROS	3	3.0833

The OVA binarization technique obtained the best results. A total of 36 predictive models were obtained with balanced data. Among these, 12 predictive models obtained a statistically significant difference. The SMOTE algorithm was the balancing method with the best results. The JRip classifier was the best algorithm, according to the ranking.

Table 11. Results of the ranking by classifier for OVA

Method	*Ranking*	Average score
C4.5	1	1.2500
JRip	2	1.5000
SVM	3	2.7500

The OVA binarization technique obtained the worst performance. A total of 54 predictive models were obtained with balanced data, of which 7 predictive models achieved a statistically significant difference. The ADASYN algorithm obtained the best performance as oversampling method. The C4.5 classifier obtained the best performance, since it obtained the lowest average score.

5. Conclusions

In this research work, three oversampling algorithms (ROS, SMOTE and ADASYN) were explored, with the objective of knowing which obtains the best performance; in addition, to know if balancing the original dataset improves the performance of the predictive models obtained with unbalanced data. These experiments were conducted with a real dataset of patients diagnosed with some subtype of GBS. Initially, binary subsets were created applying two techniques (OVO and OVA) to the original dataset. It was obtained 10 subsets divided in: 6 subsets with the OVO technique, and 4 subsets with the OVA technique. Each subset was split into 66% of the data for training and 34% of the data for testing. The minority classes of the training subsets were oversampled applying ROS, SMOTE and ADASYN, with the purpose of balancing the minority class with the majority class. Once the subsets were balanced, three classifiers were applied: C4.5, JRip and SVM. The results are the average of 60 runs of the ROC curve. Wilcoxon test was applied to the predictive models obtained with balanced data that outperformed the models with unbalanced data, to know if there is a statistically significant difference between them.

The OVA binarization technique obtained the best result compared to the OVO technique. Applying the OVA technique, 36 predictive models were obtained

with balanced data, of which 12 had a statistically significant difference. The best algorithm for balancing the data was SMOTE with respect to ROS and ADASYN. The SMOTE algorithm improved the performance of the predictive models according to their oversampling features. SMOTE adds instances to the minority class, extrapolating new instances instead of duplicating them, as it is done by the ROS algorithm. The ROS algorithm copies instances of the minority class and adds them randomly, duplicating information that may confuse the classifiers. On the other hand, ADASYN is a variant of SMOTE which adds instances to the minority class that are difficult to learn, specially the ones located in the decision boundary; this approach may not be enough information for the classifier to identify the classes and improve the result. The C4.5 classifier obtained the best performance according to the average score for OVO.

The results demonstrate that balancing the data improves the performance of the predictive models obtained with unbalanced data. On the other hand, using automatic learning algorithms in disease diagnosis problems is feasible, and may contribute to the identification of the GBS subtype that a patient gets. As future works, hybrid oversampling and downsampling techniques will be explored, in addition to the use of other classifiers.

References

[1] P. A. van Doorn, "Guillain-Barré syndrome," in *Dysimmune Neuropathies*. Elsevier, 2020, pp. 5–29. [Online]. Available: https://doi.org/10.1016/B978-0-12-814572-2.00002-9

[2] A. Tellería-Díaz and D. Calzada-Sierra, "Síndrome de Guillain-Barré," *Revista de Neurología*, vol. 34, no. 10, pp. 966–976, 2002. [Online]. Available: https://doi.org/10.33588/rn.3410.2001280

[3] E. Alpaydin, *Introduction to Machine Learning*. MIT press, 2020. [Online]. Available: https://bit.ly/2HvdROG

[4] J. A. Cruz and D. S. Wishart, "Applications of Machine Learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, p. 117693510600200, jan 2006. [Online]. Available: https://doi.org/10.1177/117693510600200030

[5] A. R. Vaka, B. Soni, and S. Reddy K., "Breast cancer detection by leveraging Machine Learning," *ICT Express*, may 2020. [Online]. Available: https://doi.org/10.1016/j.icte.2020.04.009

[6] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Applied Computing and Informatics*, dec 2018. [Online]. Available: https://doi.org/10.1016/j.aci.2018.12.004

[7] N. P. Tigga and S. Garg, "Prediction of Type 2 Diabetes using Machine Learning classification methods," *Procedia Computer Science*, vol. 167, pp. 706–716, 2020. [Online]. Available: https://doi.org/10.1016/j.procs.2020.03.336

[8] Z. K. Senturk, "Early diagnosis of parkinson's disease using machine learning algorithms," *Medical Hypotheses*, vol. 138, p. 109603, may 2020. [Online]. Available: https://doi.org/10.1016/j.mehy.2020.109603

[9] A. Khan and S. Zubair, "An improved multimodal based Machine Learning approach for the prognosis of Alzheimer's disease," *Journal of King Saud University - Computer and Information Sciences*, apr 2020. [Online]. Available: https://doi.org/10.1016/j.jksuci.2020.04.004

[10] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. Springer International Publishing, 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-98074-4

[11] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, may 2017. [Online]. Available: https://doi.org/10.1016/j.eswa.2016.12.035

[12] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, apr 2018. [Online]. Available: https://doi.org/10.1613/jair.1.11192

[13] K. Napierala and J. Stefanowski, "Types of minority class examples and their influence on learning classifiers from imbalanced data," *Journal of Intelligent Information Systems*, vol. 46, no. 3, pp. 563–597, jul 2015. [Online]. Available: https://doi.org/10.1007/s10844-015-0368-1

[14] J. Canul-Reich, J. Frausto-Solís, and J. Hernández-Torruco, "A predictive model for Guillain-Barré syndrome based on single learning algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1–9, 2017. [Online]. Available: https://doi.org/10.1155/2017/8424198

[15] J. Canul-Reich, J. Hernández-Torruco, O. Chávez-Bosquez, and B. Hernández-Ocaña, "A predictive

model for Guillain-Barré syndrome based on ensemble methods," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–10, 2018. [Online]. Available: https://doi.org/10.1155/2018/1576927

[16] J. Hernández-Torruco, J. Canul-Reich, J. Frausto-Solís, and J. J. Méndez-Castillo, "Feature selection for better identification of subtypes of Guillain-Barré syndrome," *Computational and Mathematical Methods in Medicine*, vol. 2014, pp. 1–9, 2014. [Online]. Available: https://doi.org/10.1155/2014/432109

[17] A. Fernández, S. del Río, N. V. Chawla, and F. Herrera, "An insight into imbalanced big data classification: Outcomes and challenges," *Complex & Intelligent Systems*, vol. 3, no. 2, pp. 105–120, 2017. [Online]. Available: https://doi.org/10.1007/s40747-017-0037-9

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, jun 2002. [Online]. Available: https://doi.org/10.1613/jair.953

[19] H. He, Y. Bai, E. A. García, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, jun 2008. [Online]. Available: https://doi.org/10.1109/IJCNN.2008.4633969

[20] S. Ruggieri, "Efficient C4.5 [classification algorithm]," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 438–444, 2002. [Online]. Available: https://doi.org/10.1109/69.991727

[21] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support Vector Machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000. [Online]. Available: https://doi.org/10.1093/bioinformatics/16.10.906

[22] A. Rajput, R. P. Aharwal, M. Dubey, S. Saxena, and M. Raghuvanshi, "J48 and JRip rules for e-governance data," *International Journal of Computer Science and Security (IJCSS)*, vol. 5, no. 2, p. 201, 2011. [Online]. Available: https://bit.ly/3jt2jrY

[23] R. Kannan and V. Vasanthi, "Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease," in *Soft Computing and Medical Bioinformatics*. Springer Singapore, jun 2018, pp. 63–72. [Online]. Available: https://doi.org/10.1007/978-981-13-0059-2_8

[24] A. Fernández, V. López, M. Galar, M. J. del Jesús, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowledge-Based Systems*, vol. 42, pp. 97–110, apr 2013. [Online]. Available: https://doi.org/10.1016/j.knosys.2013.01.018

[25] A. D. Pozzolo, O. Caelen, and G. Bontempi, *unbalanced: Racing for Unbalanced Methods Selection*, 2015, R package version 2.0. [Online]. Available: https://doi.org/10.1007/978-3-642-41278-3_4

[26] L. Torgo, *Data Mining with R, learning with case studies*. Chapman and Hall/CRC, 2010. [Online]. Available: https://bit.ly/3jtkeyV

[27] P. Branco, R. P. Ribeiro, and L. Torgo, "UBL: an R package for utility-based learning," *CoRR*, vol. abs/1604.08079, 2016. [Online]. Available: https://bit.ly/35yeFtU

[28] I. H. Witten, E. Frank, M. A. Hall, and C. Pañ, *Data Mining, Practical Machine Learning Tools and Techniques*, Elsevier, Ed. Morgan Kaufmann, 2017. [Online]. Available: https://doi.org/10.1145/507338.507355

[29] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2018, R package version 1.7-0. [Online]. Available: https://bit.ly/3mm1d3s

[30] A. S. Hussein, T. Li, W. Y. Chubato, and K. Bashir, "A-SMOTE: A new preprocessing approach for highly imbalanced datasets by improving SMOTE," *International Journal of Computational Intelligence Systems*, 2019. [Online]. Available: https://bit.ly/3mhotiT

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.03
pISSN: 1390-650X / eISSN: 1390-860X

Design and construction of a batch reactor with external recirculation to obtain biodisel from residual oil frying under subcritical conditions

Diseño y construcción de un reactor discontinuo con recirculación externa para obtener biodiésel a partir de aceite de fritura en condiciones subcríticas

Cristian Fabián Pérez-Salinas[1,*], Diego Fernando Núnez-Núñez[1],
Herminia del Rosario Sanaguano-Salguero[2], Luis Fernando Sánchez-Quinchuela[3]

Received: 23-05-2020, Reviewed: 06-08-2020, Accepted after review: 25-09-2020

Abstract

A batch reactor was designed and built to obtain biodiesel from frying oil under sub-critical conditions, with the purpose of reducing the reaction time to the minimum possible. The design process is focused on the selection of the material and the verification of its resistance by means of a FEM analysis from a Design of Experiments (DOE). Three levels of pressure, temperature and wall thickness, respectively, and a material categorical factor at two levels were considered. The results obtained were that the appropriate material for manufacturing the reactor is 304 stainless steel with a design safety factor of 1. For constructing the system it was also necessary to select all the complementary components. The final operation tests showed that it is possible to safely obtain the biofuel in the batch reactor with a degree of conversion 88%, in a range of 5 to 8 minutes.

Keywords: Biodiesel, Discontinuous reactor, subcritical conditions, DOE, Finite elements.

Resumen

Se diseña y construye un reactor discontinuo para obtener biodiésel a partir de aceite de fritura en condiciones subcríticas con la intención de reducir el tiempo de reacción al mínimo posible. El proceso de diseño se centra en la selección del material y la verificación de su resistencia mediante un análisis FEM a partir de un diseño experimental DOE. Se consideran tres niveles de presión, temperatura y espesor de pared, respectivamente, y un factor categórico material a dos niveles. Los resultados obtenidos permiten determinar que el material apropiado para la manufactura del reactor es acero inoxidable 304 con un factor de seguridad de diseño de 1. Para el proceso de construcción del sistema es necesario también la selección de todos los componentes complementarios. Las pruebas finales de funcionamiento muestran que es posible obtener el biocombustible en el reactor discontinuo con un grado de conversión del 88 % de manera segura en un rango de 5 a 8 minutos.

Palabras clave: biodiésel, reactor discontinuo, condiciones subcríticas, DOE, elementos finitos

[1,*]Docente investigador, Ingeniería Mecánica, Universidad Técnica de Ambato, Ecuador.
Corresponding author ✉: cf.perez@uta.edu.ec. http://orcid.org/0000-0003-4031-5464
http://orcid.org/0000-0001-5248-4084
[2]Docente investigador, Ciencias Ambientales, Universidad Estatal de Bolívar, Ecuador.
http://orcid.org/0000-0002-2885-1515
[3]Profesor, Mecánica Industrial, Instituto Superior Tecnológico Guayaquil, Ecuador.
http://orcid.org/0000-0002-4238-7376

Suggested citation: Pérez-Salinas, C. F.; Núnez-Núñez, D. F.; Sanaguano-Salguero, H, R. and Sánchez-Quinchuela, L. F. (2021). «Design and construction of a batch reactor with external recirculation to obtain biodisel from residual oil frying under subcritical conditions». Ingenius. N.° 25, (january-june). pp. 32-40. doi: https://doi.org/10.17163/ings.n25.2021.03.

1. Introduction

Most countries worldwide are concerned about reducing the emission of greenhouse gases, incorporating new sources of energy as an alternative to fossil fuels and recovering deforested territory with arable and non-arable vegetation [1]. For this purpose, the production of biofuels is a tangible alternative [2–4]. It is expected that increasing the production of biofuels will not only contribute to the conservation of the environment, but also to the economical and social development of the producing countries [5].

Conventional techniques for producing biodiesels use reactors. A chemical reactor is a complex device in which heat and mass transfer, diffusion and friction may occur together with the chemical reaction under control and safety devices. There are different types of reactors according to the form of operation, the type of internal flow and the phases they house [6–9]. In general, it is sought to determine the size and type of reactor, as well as the operation method necessary to obtain the final product [10].

One of the important parameters intended to be improved is the time for obtaining the biodiesel. Times between 20 and 180 minutes have been reported in previous studies [11]; in particular, in residual cooking oils this time is 90 minutes.

In recent years, attention has been focused on seeking raw materials different than vegetable oils such as soy, palm, etc. This is mainly due to the cost of obtaining the raw material, which is approximately 70 % of the total cost of obtaining biodiesel [6–8]. The frying oil is another source of raw material, whose advantage over other raw materials is that they are in the category of residuals. This feature gives it a great viability because it contributes to the reduction of the environmental pollution.

In Latin America, the residual frying oil has the potential to be used as a primary source for producing biodiesel [12]. There is a large industry of fast and traditional food that uses frying oil, which becomes a waste after being used. For example, in Colombia 35 % of the yearly production of 162 million liters of cooking oil becomes a residual that is discarded in sewers [13], and Spain produces 150 million liters of used vegetable oil yearly [14].

On the other hand, in the industrial, educational and residential sectors, there are countless products and systems developed to accomplish a specific task. All of them employ engineering materials chosen to achieve an optimal performance. The process for obtaining a product comprises three macro-processes: design, manufacturing and operation tests. The design stage starts determining the requirements and constraints to further carry out a material selection process [15].

In recent years, two methods for material selection have been used in engineering development: traditional and graphical. The former is based on the knowledge and experience of the engineer, the latter on graphical maps of engineering materials organized according to their physical and mechanical properties [16].

The graphical method was developed as a very important support in the stage of conception and development of the product [17, 18]. The graphical method from Michael Ashby has been the most useful one in recent years. It is a methodology that acts as a guide for material selection, considering the attributes that are related with each other by means of graphical selection tables [18, 19]. Performance indices are utilized, which refer to groups of material properties with the purpose of maximizing or minimizing them according to the specific requirement. They are derived from the objective function of the system, and are expressed by means of mathematical equations.

Various authors have reported the use of the Finite Element Analysis (FEA) technique for designing and verifying chemical reactors subject to multiphase modeling [20–23]. Studies, such as [24–26], mention that parametric design provides mechanical systems with the capability of synthesizing, simplifying and economizing the design process, which enables amplifying and exploring possibilities of solution.

In the current study, it was carried out the design and construction of a batch reactor with external recirculation to obtain biodiesel from residual frying oil under subcritical conditions, with the purpose of reducing the production time within a controlled and safe framework that fulfills engineering requirements. To this effect, it was proceeded to select the materials and verify the resistance under the stress regime using the finite elements technique.

2. Materials and methods

The following methodology was applied: selection of materials, design by means of finite element analysis (FEA), construction of the system and operation tests. It is remarked the fact that the design of the reactor was initiated with the observation of previous designs, to come up with a completely new design which enables reducing the time to obtain the biodiesel.

2.1. Selection of materials

It was utilized the graphical method from Ashby, with the aid of the CES EDUPACK educational software. It was defined the required functionality, the objective and the constraints of the reactor (Table 1). From Hooke´s law which gives the tensile stress in the elastic zone (σ) as the product of Young´s modulus (E) times the strain (ϵ), and the formula of mass (m) as a function of volume and density (ρ), it was determined the performance index (Equation 1).

Table 1. Definition of the design requirements

Definition	Detail
Functionality	House the thermo-chemical process
Objective	Minimize mass of the reactor (ρ) and maximize mechanical resistance (σ)
Constraints	Withstand minimum and maximum temperatures of 160 °C and 400 °C, respectively
	Withstand minimum and maximum pressures of 250 psi and 3000 psi, respectively
	Resistance to oxidative and corrosive processe
	Low cost
	Good machinability and solderable

$$m = \frac{F}{\epsilon \cdot E} \cdot L \cdot \rho = F \cdot L \cdot \left[\frac{\rho}{\sigma}\right] \Rightarrow Performance\_indexño = \frac{\rho}{\sigma} \quad (1)$$

Taking all the functionalities and initial constraints for the reactor, it was obtained the possibility of 600 alternatives of material among a total of 3900 options. The adjustment of the selection was carried out according to the specific conditions of service (high mechanical resistance, good performance at high temperatures and resistant to chemical elements), local availability of the material at the smallest cost possible and a high density which enables the moving stability of the system. (Figure 1).

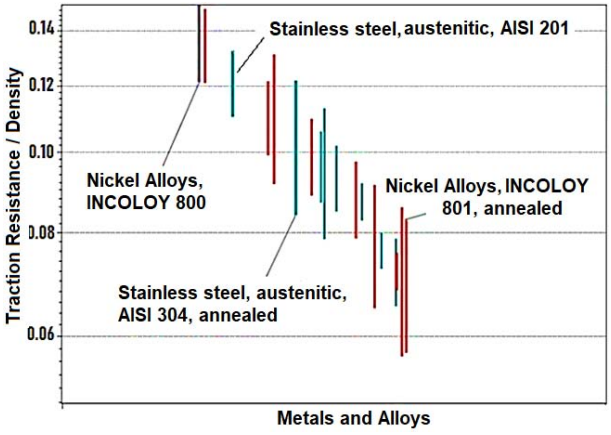


Figure 1. Plot stress-density for selection of materials [18]

The input variables were: a maximum price of 10 USD, density between 5000 and 8000 kg/m^3, service temperature of 500 °C, resistance to alkaline substances and substances with pH greater than 7, molding and machinability capabilities that result in 18 possibilities of material framed in three groups: austenitic stainless steels, annealed austenitic stainless steels and nickel alloys.

2.2. Finite elements analysis (FEA)

Once the selection of materials for the reactor has been performed, it was chosen the AISI 304 steel and the nickel alloy to enter the FEA experimental design. On the other hand, the dimensions of the batch reactor were stated such that it yields a continuous recirculation capability of 1 liter.

A complete factorial design was stated where it was taken into account the three main factors (constraints/requirements) of the reactor; pressure, temperature and wall thickness in three levels, respectively. A categorical variable (material) in two levels was also considered. Table 2 shows the experimental design. The ranges of pressure and temperature values were based on studies regarding obtaining biodiesel in subcritical and supercritical conditions.

Table 2. Design of experiment

Factors	Levels		
	Bajo	Medio	Alto
Pressure (Mpa)	1,72	4,83	20,7
Temperature (°C)	160	200	400
Wall thickness (mm)	1,5	2	3
Material	AISI 504 steel	Allow Nickel	-

It was performed a thermomechanical analysis by means of finite elements using a multi-parametric symmetric 2D model created in Ansys APDL, to evaluate the mechanical resistance of the reactor to the internal conditions of pressure and temperature specified in the design of the experiments. It was used a quarter of the cross section due to the two planes of symmetry existing in the model, Figure 2. For the two materials used in this study it was defined an elastic linear isotropic model of the material, whose properties are shown in Table 3.

Table 3. Mechanical and thermal properties for the linear elastic isotropic model of the material

Mechanical and thermal properties	Material	
	AISI 304	Ni 800
Elasticity modulus (GPa)	193	196
Poisson coefficient	0,3	0,34
Yield stress (MPa)	220	335
Density (kg/m^3)	7850	7940
Service temperature (°C)	850	816
Conductivity (W / kg-K)	16,2	11,5
Specific heat (J/kg-°C)	500	460
Thermal expansion coefficient (x 10^6 °C^{-1})	17	12

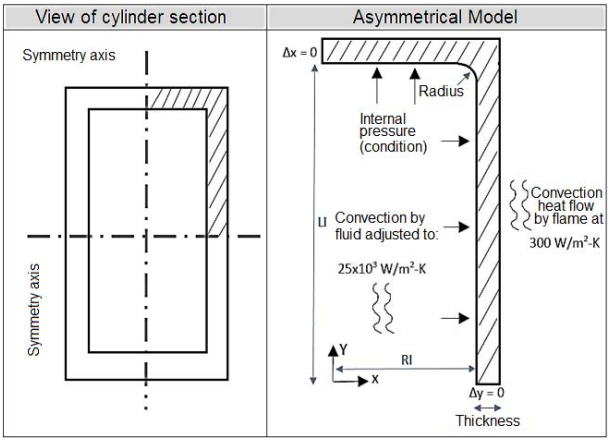


Figure 2. Geometry and limiting conditions of a model of asymmetrical reactor of quarter of axis for a thermomechanical analysis

The model was discretized with PLANE182 quadrilateral elements in their axisymmetric version; for all cases, the size of the element was established at 1/5 of the cylinder wall thickness, because this value assured a convergence smaller than 5%. The model has 1830 nodes and 1520 elements. The limiting conditions for the stationary thermomechanical analysis are shown in Figure 2.

Two types of finite elements analysis were carried out: one thermal stationary with the temperature and convection conditions specified to obtain displacements and thermal strains, and then a static structural analysis with the boundary conditions indicated such that both studies were coupled in a multi-physical analysis. The validation of the finite elements model of the first type of analysis was carried out by means of the Thin Wall Pressure Vessel theory (TWPVt), and the circumferential or tangential (SZ) and longitudinal or axial (SY) stresses were evaluated, because this theory does not take into account the concentration of stresses, such as covers, holes or abrupt changes in the cross section. In the second finite element analysis, the influence of pressure and temperature was evaluated, and the results of the equivalent Von Mises stress were compared to the yield stress of each material to determine the safety factor of the reactor.

$$\sigma_e = \sqrt{(\sigma_x + \sigma_y) - 3(\sigma_x \sigma_y - \tau_{xy}^2)} \qquad (2)$$

where:

σ_e=Von Mises stress
σ_x=Stress normal to the X axis
σ_y=Stress normal to the Y axis
τ_{xy}^2=Shear stress in the XY plane

According to the theory of thin wall pressure vessels (internal diameter/thickness ratio greater than or equal to 10), there are two main stresses which are a function of the internal pressure (p), internal radius (r) and wall thickness (t). These are longitudinal or axial and circumferential or tangential stresses, whose calculation formulas are (3) and (4), respectively:

Longitudinal or axial stress (SY):

$$\sigma_1 = \frac{pr}{2t} \qquad (3)$$

Circumferential or tangential stress (SZ):

$$\sigma_t = \frac{pr}{t} \qquad (4)$$

2.3. Construction of the system

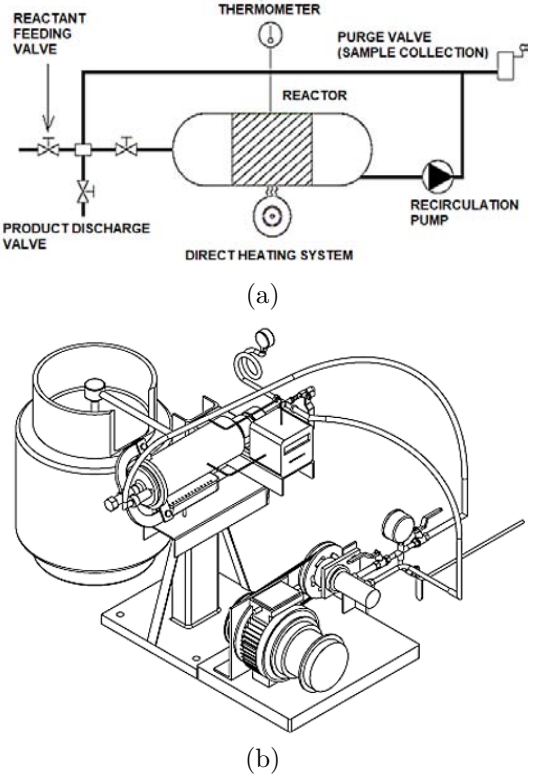


Figure 3. Batch reactor a) System scheme, b) 3D configuration

The batch reactor (Figure 3) is constituted by a tank, a circulation tube, and a recirculation pump to stir the reagents. The tank consists of a cylindrical tube with a length of 280 mm, an internal diameter of 71 mm and a wall thickness according to the minimum thickness resulting from the design (FEM analysis). The tank is coupled to a tube of length 1000 mm and diameter of 12 mm and a high-pressure pump that rotates at 250 rpm for recirculation and steering. As the tube, the reactor will be manufactured using the appropriate material studied in section 2.2 that is available at the lowest cost. The loading flow (3 l/min) is driven by a Motovario TXF005 motor. A thermocouple connected to the temperature controller and

a pressure gauge to measure the internal pressure of the vessel, were place on the reactor cover. The cover comprises a flange with a teflon gasket. Two Problock pressure gauges, from 0 to 150 psi, were placed on the upper part of the reactor. The temperature of the reactor is recorded by a Syscon AKC CB100 temperature controller.

2.4. Operational tests

The tests were carried out at temperatures of 160, 180 and 200 °C, with methanol-oil molar ratios of 6:1 and 9:1, catalyst percentage (NaOH) of 0.5 and 1 % and for a reaction time of 5 to 10 minutes to verify the functionality of the batch reactor with continuous recirculation capacity of 1 liter, reaching pressures of 250 psi (subcritical condition). For obtaining crude biodiesel, it will be subject to distillation and wash with atomized water until the wash water comes out clean. After eliminating the secondary products, the biodiesel will be heated at 105 °C for ten minutes to dry or eliminate the water and the methanol residues. After cooling, a volume of refined biodiesel was obtained.

3. Results and discussion

3.1. Numerical and analytical results

The results of the static structural analysis are shown in Table 4. The AISI 304 stainless steel and the Ni 800 nickel alloy have very similar elasticity module and Poisson ratio, which indicates that the results of the longitudinal (SY) and circumferential (SZ) stresses practically do not vary, it only changes the safety factor, defined as the ratio between the yield stress of the material and the equivalent maximum stress achieved. It is important to take into account that, in this analysis, the safety factor was only obtained with the circumferential and longitudinal stresses, which does not necessarily represent a complete stress condition of the point under analysis. According to the results of the finite elements model, it satisfactorily agrees with the theory of the thin wall pressure vessels. The maximum error was approximately 4.2%.

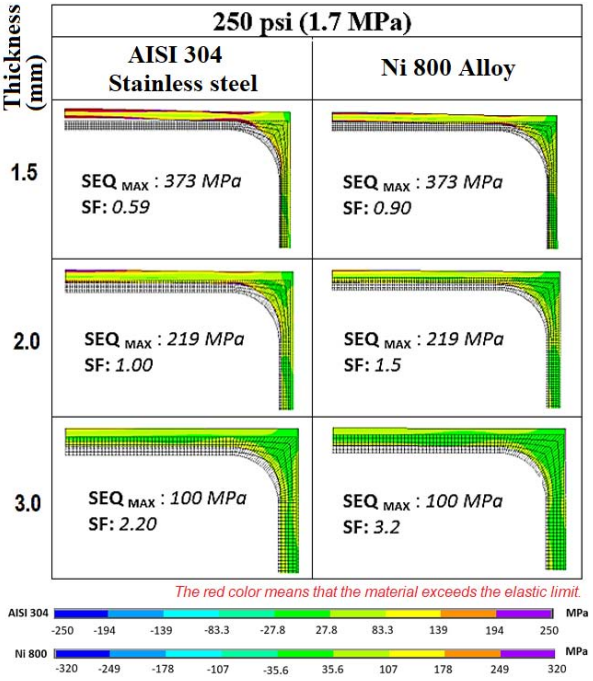
3.2. Influence of the internal pressure and temperature of the reactor

The study of the thermomechanical coupling of a quarter of the cross section of the reactor was utilized to determine the influence of the internal pressure and the temperature. Tables 5, 6 and 7 show the influence of the internal pressure for a condition of constant temperature at 160 °C. The color legend was separated for the AISI 304 stainless steel and for the Ni 800 nickel alloy, since they have a different elastic limit. The red color in Table 4 means that the material exceeds the elastic limit and there is the possibility of permanent deformations that lead to the failure of the bioreactor. The deformation scale with respect to the not deformed condition is the same for each case.

Table 4. Comparison of the longitudinal σ_l (SY) and circumferential σ_t (SZ) stresses from the finite element analysis (FEA) and the thin wall pressure vessels theory (TWPVt)

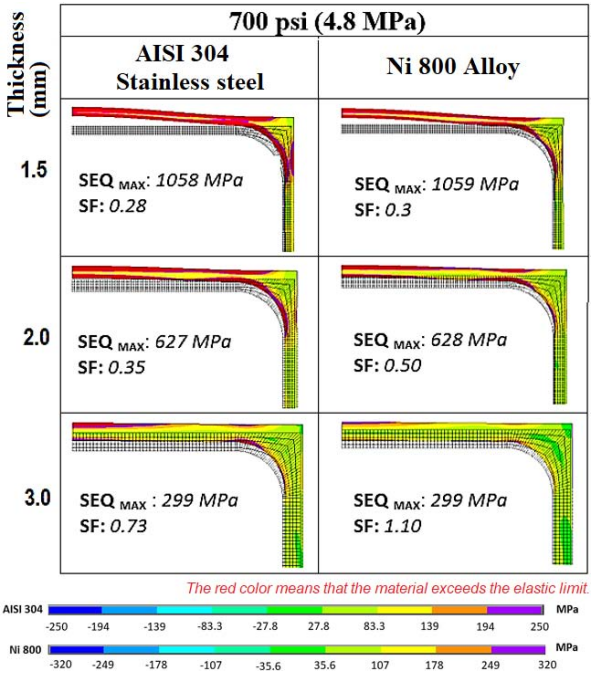
Exp	Pressure (psi)	DOE Thickness (mm)	Material	FEA SY	TWPVt σ_l	Error %	Safety factor	FEA SZ	TWPVt σ_t	Error %	Safety factor
1	250	1,5	AISI 304 / Ni 800	2857	2917	2,1	10,5/15,8	5960	5834	2,1	5,0/7,6
2	250	2	AISI 304 / Ni 800	2127	2188	2,8	14,1/21,2	4502	4375	2,8	6,7/10,0
3	250	3	AISI 304 / Ni 800	1399	1458	4,1	21,4/32,2	3045	2917	4,2	9,9/14,8
4	700	1,5	AISI 304 / Ni 800	8066	8236	2,1	3,7/5,6	16827	16472	2,1	1,8/2,7
5	700	2	AISI 304 / Ni 800	6003	6177	2,8	5,0/7,5	12711	12354	2,8	2,4/3,5
6	700	3	AISI 304 / Ni 800	3951	4118	4	7,6/11,4	8597	8236	4,2	3,5/5,2
7	3000	1,5	AISI 304 / Ni 800	34614	35346	2,1	0,9/1,3	72216	70692	2,1	0,4/0,6
8	3000	2	AISI 304 / Ni 800	25784	26510	2,7	1,2/1,7	54552	53019	2,8	0,5/0,8
9	3000	3	AISI 304 / Ni 800	16956	17673	4,1	1,8/2,7	36898	35346	4,2	0,8/1,2

Table 5. Numerical analysis of the influence of the internal pressure of the reactor



Constant temperature = 160 °C and internal pressure 250 psi (1.7 MPa)

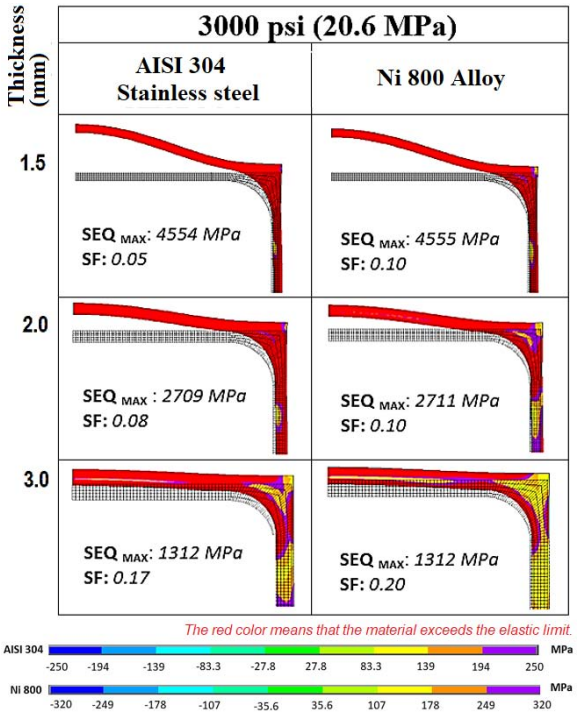
Table 6. Numerical analysis of the influence of the internal pressure of the reactor



Constant temperature = 160 °C and internal pressure 700 psi (4.8 MPa).

The finite element analysis shows similar results of equivalent Von Mises stress for both materials, because the elasticity modules and Poisson ratios are very similar; however, the safety factors are different because there is a difference of about 100 MPa between their elasticity limits. The achieved maximum Von Mises stress was larger for thicknesses of 1.5 mm of the reactor wall, a minimum for thicknesses of 3 mm. For an internal pressure of 250, 700 and 3000 psi, the maximum Von Mises stress was 373, 1058 and 4555 MPa, respectively.

Table 7. Numerical analysis of the influence of the internal pressure of the reactor



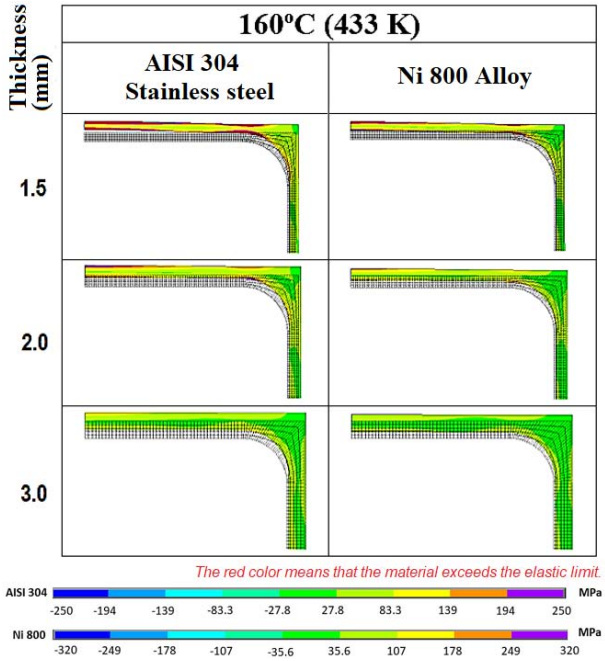
Constant temperature = 160 °C and internal pressure 3000 psi (20.6 MPa)

With the conditions of the axisymmetric finite elements modeling developed in this study, it is verified that for the case of the AISI 304 stainless steel the design is safe only for subcritical conditions of pressure up to 250 psi and a minimum thickness of 2 mm, for which a safety factor of 1.00 was obtained. The safety factor is calculated as the ratio between ratio between the elastic stress of the material and the equivalent maximum stress achieved.

For the Ni 800 nickel alloy, the internal pressure may be increased up to 700 psi with a thickness of 3 mm, where the safety factor is 1.10. For conditions of supercritical internal pressure (3000 psi), the safety factors are below 1.00 in all cases, which indicates that the material has a very high probability of failure, For this condition, the thickness of the reactor wall must be drastically increased.

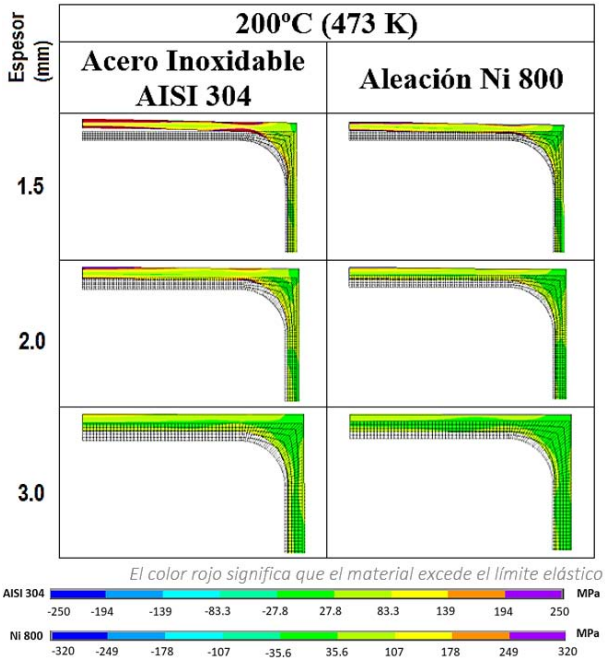
Tables 8, 9 and 10 show the influence of the temperature for a condition of constant internal pressure at 250 psi.

Table 8. Analysis of the influence of the temperature of the reactor



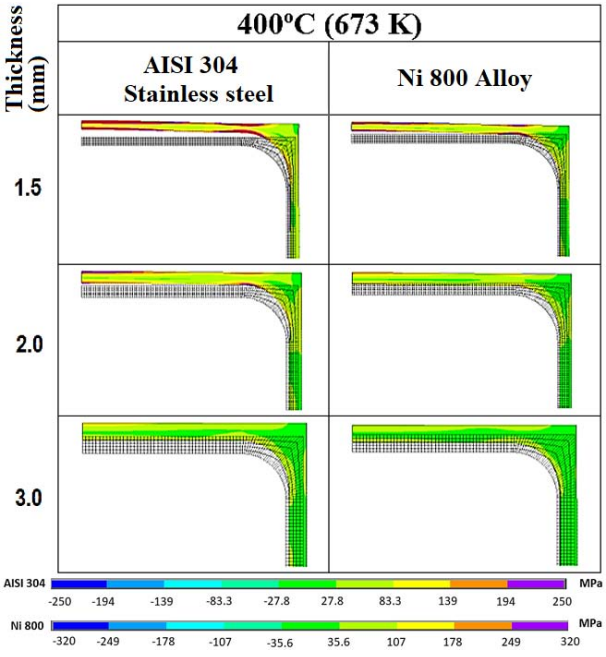
Constant pressure = 250 psi and temperature of 160 °C

Table 9. Analysis of the influence of the temperature of the reactor



Constant pressure = 250 psi and temperature of 200 °C

Table 10. Analysis of the influence of the temperature of the reactor



Constant pressure = 250 psi and temperature of 400 °C

For both materials, it may be verified that the equivalent maximum stresses have a minimum variation as the temperature increases to values of 160, 200 and 400 °C; therefore, in this model the results are only influenced by the change in internal pressure. It is assumed that this is due to two factors: the temperature gradient is minimum at the reactor wall due to the stable condition that is reached with the internal thermal fluid (biodiesel) and the external convective flow (fire). On the other hand, none of the temperatures reached exceeds the service temperatures of the materials, 850 °C for the AISI 304 stainless steel and 816 °C for the Ni 800 nickel alloy. Therefore, in a further research it is important to develop a fatigue analysis to estimate the useful life of the reactor.

3.3. Funcionality test

With the prototype developed according to section 2.3 and the considerations of the numerical analysis (FEM), it was proceeded to conduct the operational tests in the lab.

After executing the procedure stated in the methodology, it was obtained a volume of refined biodiesel of transparent yellow color. The densities of biodiesel obtained were analyzed (smaller density is better), taking as reference temperatures of 160 °C, 180 °C and 200 °C used for the transesterification (Figure 4). When comparing the means of the biodiesel results, statistically significant differences were found; the largest density occurred for 200 °C (0.89 g/ml), while at 160

°C and 180 °C the density was 0.88 g/ml. The time for obtaining the biodiesel starts from 5 to 10 minutes, being stable to obtain it after 8 minutes.

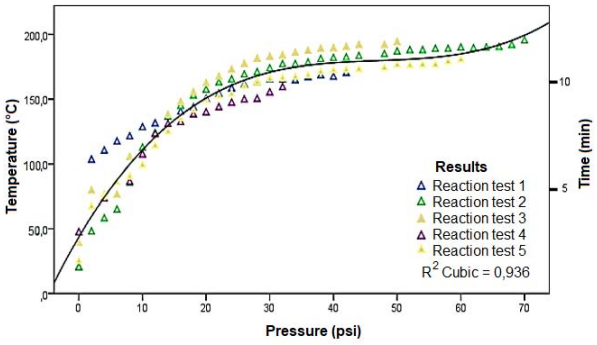


Figure 4. Behavior of the internal reaction of the reactor as a function of the temperature, pressure and time

4. Conclusions

The graphical method of material selection enables reducing the time for selecting the material. The great number of options for manufacturing a reactor makes essential the need for a refinement in the selection when determining the service conditions, such as temperature gradients, pressure ranges, chemical influences, pH, and considerations such as cost and availability of the material.

It was performed a symmetric finite element multiparametric analysis in 2D axis using the Ansys APDL software, to evaluate the mechanical resistance of the reactor. Such analysis enables determining the influence of the variables in the thermomechanical behavior of the reactor. Results show that the design is safe only for subcritical conditions of pressure up to 250 psi and a minimum thickness of 2 mm for the case of the AISI 304 stainless steel, where a safety factor of 1.00 was obtained. In the case of the Ni 800 nickel alloy, the internal pressure may be increased up to 700 psi with a minimum thickness of 3 mm, where the safety factor is 1.10. At the limiting conditions of this model, the temperature does not have influence on the Von Mises equivalent stress.

A volume of 38.48 cm^3 was obtained as a function of the diameter and length of the tubular section, which represents the 3,47% of the total volume of the reactor, thus, it was considered insignificant. On the other hand, the recirculation flow managed by the pump during the operation of the reactor was 3 l/min; as a consequence, the fluid that passes through the pipe has a residence time (cylindrical section / pump flow) equal to 0.026 min (1.6 seconds), which makes evident that the effect of the behavior of the piston flow in the tubular section is not significant with respect to the reaction process that occurred in the tank of the reactor.

Through the reaction process, it was observed that, as temperature increases, more soap is formed, which increases the biodiesel density. However, with respect to the biodiesel density obtained in subcritical conditions, these values are within the European standard that establishes a minimum range of 0.86 g/ml and a maximum of 0.90 g/ml. This biodiesel quality was achieved in an approximate time of 5-8 minutes.

References

[1] Y. Ma and Y. Liu, "Chapter 21 - biodiesel production: Status and perspectives," in *Biofuels: Alternative Feedstocks and Conversion Processes for the Production of Liquid and Gaseous Biofuels (Second Edition)*, second edition ed., ser. Biomass, Biofuels, Biochemicals, A. Pandey, C. Larroche, C.-G. Dussap, E. Gnansounou, S. K. Khanal, and S. Ricke, Eds. Academic Press, 2019, pp. 503–522. [Online]. Available: https://doi.org/10.1016/B978-0-12-816856-1.00021-X

[2] L. Faba, E. Díaz, and S. Ordóñez, "Recent developments on the catalytic technologies for the transformation of biomass into biofuels: A patent survey," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 273–287, 2015. [Online]. Available: https://doi.org/10.1016/j.rser.2015.06.020

[3] J. Xu, G. Xiao, Y. Zhou, and J. Jiang, "Production of biofuels from high-acid-value waste oils," *Energy & Fuels*, vol. 25, no. 10, pp. 4638–4642, 2011. [Online]. Available: https://doi.org/10.1021/ef2006723

[4] M. Mohammadi, G. D. Najafpour, H. Younesi, P. Lahijani, M. H. Uzir, and A. R. Mohamed, "Bioconversion of synthesis gas to second generation biofuels: A review," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4255–4273, 2011. [Online]. Available: https://doi.org/10.1016/j.rser.2011.07.124

[5] F. A. Avellaneda Vargas, "Producción y caracterización de biodiésel de palma y de aceite reclicado mediante un proceso batch y un proceso continuo con un reactor helicoidal," 2010. [Online]. Available: https://bit.ly/3jqIgdX

[6] Y. Wang, P. L. Shiyi Ou, and Z. Zhang, "Preparation of biodiesel from waste cooking oil via two-step catalyzed process," *Energy Conversion and Management*, vol. 48, no. 1, pp. 184–188, 2007. [Online]. Available: https://doi.org/10.1016/j.enconman.2006.04.016

[7] Y. Zhang, M. A. Dubé, D. D. McLean, and M. Kates, "Biodiesel production from waste

cooking oil: 2. economic assessment and sensitivity analysis," *Bioresource Technology*, vol. 90, no. 3, pp. 229–240, 2003. [Online]. Available: https://doi.org/10.1016/S0960-8524(03)00150-0

[8] S. Zheng, M. Kates, M. Dubé, and D. McLean, "Acid-catalyzed production of biodiesel from waste frying oil," *Biomass and Bioenergy*, vol. 30, no. 3, pp. 267–272, 2006. [Online]. Available: https://doi.org/10.1016/j.biombioe.2005.10.004

[9] A. Srivastava and R. Prasad, "Triglycerides-based diesel fuels," *Renewable and Sustainable Energy Reviews*, vol. 4, no. 2, pp. 111–133, 2000. [Online]. Available: https://doi.org/10.1016/S1364-0321(99)00013-1

[10] B. H. Pedro, A. S. John, and G. Cano, "Estudio experimental de las variables que afectan la reacción de transesterificación del aceite crudo de palma para la producción de biodiesel," *Scientia et Technica*, vol. 1, no. 24, ene. 2004. [Online]. Available: https://doi.org/10.22517/23447214.7323

[11] B. R. Moser, *Biodiesel Production, Properties, and Feedstocks*. New York, NY: Springer New York, 2011, pp. 285–347. [Online]. Available: https://doi.org/10.1007/978-1-4419-7145-6_15

[12] A. da Silva César, D. E. Werderits, G. L. de Oliveira Saraiva, and R. C. da Silva Guabiroba, "The potential of waste cooking oil as supply for the brazilian biodiesel chain," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 246–253, 2017. [Online]. Available: https://doi.org/10.1016/j.rser.2016.11.240

[13] A. Villabona Ortiz, R. Iriarte Pico, and C. Tejada Tovar, "Alternativas para el aprovechamiento integral de residuos grasos de procesos de fritura," *Teknos revista científica*, vol. 17, no. 1, pp. 21–29, jul. 2017. [Online]. Available: https://doi.org/10.25044/25392190.890

[14] Gobierno de España. (2020) Ministerio de Transición Ecológica y el Reto Demográfico. [Online]. Available: https://bit.ly/37zY23X

[15] W. D. Callister Jr. and D. G. Rethwisch, *Materials science and engineering: an introduction*. John Wiley & Sons, Inc., 2018. [Online]. Available: https://bit.ly/3oo2Krl

[16] H. A. González and D. H. Mesa, "La importancia del método en la selección de materiales," *Scientia Et Technica*, vol. X, pp. 175–180, 2004. [Online]. Available: https://bit.ly/37EHrMs

[17] R. V. Rao, "A material selection model using graph theory and matrix approach," *Materials Science and Engineering: A*, vol. 431, no. 1, pp. 248–255, 2006. [Online]. Available: https://doi.org/10.1016/j.msea.2006.06.006

[18] M. F. Ashby and K. Johnson, *Materials Selection in Mechanical Design*. Elsevier Ltd., 2016. [Online]. Available: https://doi.org/10.1016/C2009-0-25539-5

[19] M. F. Ashby, H. Shercliff, and D. Cebon, *Materials: engineering, science, processing and design*. Butterworth-Heinemann, 2018. [Online]. Available: https://bit.ly/3dSpRW4

[20] T. E. Tezduyar and Y. J. Park, "Discontinuity-capturing finite element formulations for nonlinear convection-diffusion-reaction equations," *Computer Methods in Applied Mechanics and Engineering*, vol. 59, no. 3, pp. 307–325, 1986. [Online]. Available: https://doi.org/10.1016/0045-7825(86)90003-4

[21] W. F. Ramírez, *Computational methods for process simulation*. Butterworth-Heinemann, 1997. [Online]. Available: https://bit.ly/3mhsfc3

[22] A. Chakrabarty, S. Mannan, and T. Cagin, *Multiscale Modeling for Process Safety Applications*. Butterworth-Heinemann, 2015. [Online]. Available: https://bit.ly/2HujBbc

[23] W. B. J. Zimmerman, *Process modelling and simulation with finite element methods*. World Scientific, 2004. [Online]. Available: https://bit.ly/2G0a9vM

[24] R. Woodbury, *Elements of parametric design*. Routledge, 2010. [Online]. Available: https://bit.ly/2Hu2c2s

[25] R. de Luca, P. Fanelli, S. Mingozzi, G. Calabró, F. Vivio, F. Maviglia, and J. You, "Parametric design study of a substrate material for a demo sacrificial limiter," *Fusion Engineering and Design*, vol. 158, p. 111721, 2020. [Online]. Available: https://doi.org/10.1016/j.fusengdes.2020.111721

[26] J. Monedero, "Parametric design: a review and some experiences," *Automation in Construction*, vol. 9, no. 4, pp. 369–377, 2000. [Online]. Available: https://doi.org/10.1016/S0926-5805(99)00020-5

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.04
pISSN: 1390-650X / eISSN: 1390-860X

Prediction algorithm of fuel consumption for anhydrous ethanol mixture in high-altitude cities

Algoritmo de predicción del consumo de combustible para mezcla de etanol anhídrido en ciudades de altura

Fabricio Espinoza[1], Fredy Tacuri[1], Wilmer Contreras[1,*], Javier Vázquez[1]

Received: 15-05-2020, Reviewed: 05-08-2020, Accepted after review: 30-09-2020

Abstract

In the present research work, a mathematical model is obtained for predicting specific fuel consumption in a 1.4-liter Otto cycle engine with electronic injection without making modifications, when using as fuel gasoline mixtures with concentrations in volume of 0%, 25%, 50%, 75% and 100% of anhydrous ethanol. For the analysis of results, a simplex lattice reticular mixture experiment design was carried out, which was subject to an urban driving cycle in the city of Cuenca at 2558 m above sea level in a roller power bank. The data acquisition and the development of the algorithm were carried out through an analysis of descriptive statistical methods. The validation of the algorithm was performed through residual analysis. As a main result, there is a mathematical model that enables predicting the engine fuel consumption, for ranges of ethanol concentration from 0% to 100% in the gasoline without needing to conduct real tests.

Keywords: Ethanol anhydride; explanatory model; fuel consumption

Resumen

En la presente investigación se obtiene un modelo matemático de predicción del consumo específico de combustible en un motor ciclo Otto de 1,4 litros con inyección electrónica sin hacer modificaciones, al usar como combustible mezclas de gasolina con concentraciones a partir de 0 %, 25 %, 50 %, 75 % y 100 % en volumen de etanol anhidro. Para el análisis de los resultados se realizó un diseño de experimento de mezcla reticular *simplex lattice*, el cual se sometió a un ciclo de conducción urbano de la ciudad de Cuenca a 2558 m s. n. m., en un banco de potencia de rodillos. La adquisición de datos y la obtención del algoritmo fueron a través de un análisis de métodos estadísticos descriptivos. La validación del algoritmo se realizó por medio del análisis de residuos. Como resultado principal se cuenta con un modelo matemático, el que permite predecir el consumo de combustible del motor, para rangos de concentración de etanol del 0 % al 100 % en la gasolina sin la necesidad de realizar pruebas reales.

Palabras clave: etanol anhídrido, modelo explicativo, consumo combustible

[1,*]Universidad Politécnica Salesiana, Cuenca-Ecuador. Corresponding author ✉: rcontreras@ups.edu.ec.
http://orcid.org/0000-0003-4559-4474 http://orcid.org/0000-0002-4160-2898
http://orcid.org/0000-0003-2300-9457 http://orcid.org/0000-0001-9678-5364

Suggested citation: Espinoza, F.; Tacuri, F.; Contreras, W. and Vásquez, J. (2021). «Prediction algorithm of fuel consumption for anhydrous ethanol mixture in high-altitude cities». Ingenius. N.° 25, (january-june). pp. 41-49. doi: https://doi.org/10.17163/ings.n25.2021.04.

1. Introduction

The concern about environmental pollution due to residues of incomplete combustion and the depletion of fossil fuels, motivates the study about the reformulation of mixtures with alternative fuels. A viable option is applying combinations of ethanol and gasoline, which may reduce air pollution and at the same time improve the performance of the motor compared with the unmixed oil fuel.

When evaluating the effect of these mixtures, there is a variation in the total consumption of fuel, in this line of study, according to the work by Al-Hasan, (2003) [1]. In a Toyota Tercel vehicle with a four-cylinder engine of 1.4 l capacity, four-stroke spark ignition, compression ratio of 9:1, and a maximum power of 52 kW at 5600 rpm, results in an approximate increase of 8.3%, 9%, 7% and 5.7% in fuel consumption.

When biofuels are used in internal combustion engines, consumption of such fuels increase [2]. This is due to the fact that if the air-fuel stoichiometric ratio decreases for the same revolutions per minute, same load level and same air mass, the required fuel mass should be greater [3].

Fernández, Mosquera, and Mosquera [4] demonstrated that the use of ethanol mixed with gasoline increases the consumption linearly with the mixture used.

In the research conducted in a Lada vehicle with a 1.3 l four-stroke engine and a carburation feeding system, the fuel used is a 10, 20 and 30% mixture of anhydrous ethanol in regular gasoline, of which Melo *et al.* [5] conclude that as the percentage of ethanol in the mixture increases, fuel consumption increases for all the evaluated experimental points.

In the study conducted by Delión and Rojas [6] in a 1.4 liter Ford vehicle with respect to fuel consumption for ethanol and gasoline mixtures, it was found that the «increase of fuel consumption is greater due to the increase in load or motor torque, maintaining the rpm constant, and the polluting emissions are lesser than with pure gasoline».

On the other hand, studies conducted in 1997, 1998 and 1999 by Kortum *et al.* [7], Apace [8] and Ragazzi and Nelson [9], respectively, match the research studies carried out in 2003 by Al-Hassan [1], He *et al.* [10] and Patzek [11], 2004 by Wu *et al.* [12], 2005 by Coelho *et al.* [13], Hansen *et al.*, Niven [14] and American Coalition for Ethanol [15], 2006 by Behrentz [16], Durbin *et al.* [17], Shapiro [18] and Yucesu *et al.* [19], 2008 by Acevedo *et al.* [20], as well as the more recent works by Doe *et al.* [21], in which fuel consumption increases from 1 to 6% in engines without modification using mixtures with 0-25% of ethanol, since the consumption depends on the electronic control system of the engine.

Research works conducted about fuel consumption for mixtures of anhydrous ethanol and gasoline, conclude that consumption increases as the concentration degree of ethanol in the gasoline increases.

These research works do not consider in their methodology, the development of an experimental design by mixtures together with the application of a driving cycle typical of high-altitude cities.

This work was conducted with the purpose of obtaining a mathematical algorithm that enables calculating the specific fuel consumption of a Hyundai Getz 1.4-liter vehicle, for different mixtures of anhydrous ethanol and gasoline at a height of about 2558 meters above sea level.

2. Materials and methods

The methodology applied consists of a simplex reticular (q, m) experimental design by mixtures, which considers q components and enables fitting a statistical model of order (m); this consists of all possible combinations of components or mixtures that may be formed considering that proportions may take the (m+1) values between zero and one, given by Equation 1 [22].

$$\boldsymbol{xi} = \frac{0.1}{m}, \frac{0.2}{m}, \ldots, \frac{m}{m} \tag{1}$$

The anhydrous ethanol-gasoline mixture is identified with the nomenclature (E) followed by a number, the letter represents the mixture, and the number indicates the percentage of ethanol added to the gasoline. This mixture is characterized by its density, which is obtained using the method of the pycnometer, the octane rating by means of an octane rating meter that meets the ASTM 2699 – 86 standard, and the high and low calorific value ($\boldsymbol{H}_{cs\ Ex}$), ($\boldsymbol{H}_{ci\ Ex}$) according to Equations 2 and 3, that enable calculating this property [23].

$$\boldsymbol{H}_{cs\ Ex} = \%\boldsymbol{E} \times \boldsymbol{H}_{Cs\ etanol} + \%G \times \boldsymbol{H}_{Cs\ gasolina} \tag{2}$$

$$\boldsymbol{H}_{ci\ Ex} = \%\boldsymbol{E} \times \boldsymbol{H}_{Ci\ etanol} + \%G \times \boldsymbol{H}_{Ci\ gasolina} \tag{3}$$

In order to obtain the mathematical model, the data are statistically validated using the test for outliers, and a statistical model is fitted to investigate the effect of the components on the response. A first approximation may be obtained fitting a first order model (Equation 4).

$$E(y) = \sum_{i=1}^{q-1} \beta_i \chi_i \tag{4}$$

When fitting a quadratic model, it is also necessary to incorporate the constraint $x1 + x2 + \ldots + xq = 1$, because this will give a special characteristic to the model. To illustrate the idea, it is assumed that there

are three components, $x1$, $x2$, $x3$, and thus the second order polynomial is given by Equation 5.

$$E(y) = \sum_{i=1}^{q} \beta_i \chi_i + \sum_{i<j} \sum_{j=2}^{q} \beta_{ij} \chi_i \chi_j \tag{5}$$

If the quadratic model is not enough for describing the response, the special cubic model nay be used (Equation 6).

$$E(y) = \sum_{i=1}^{q} \beta_i \chi_i + \sum_{i<j} \sum_{j=2}^{q} \beta_{ij} \chi_i \chi_j + \\ + \sum_{i<j} \sum_{j=2}^{q} \delta_{ij} \chi_i \chi_j (\chi_i - \chi_j) + \\ + \sum_{i<j} \sum_{j<k} \sum_{j=3}^{fq} \beta_{ijk} \chi_i \chi_j \chi_k \tag{6}$$

For obtaining the mathematical algorithm of fuel consumption, it is established the methodology indicated in Figure 1.

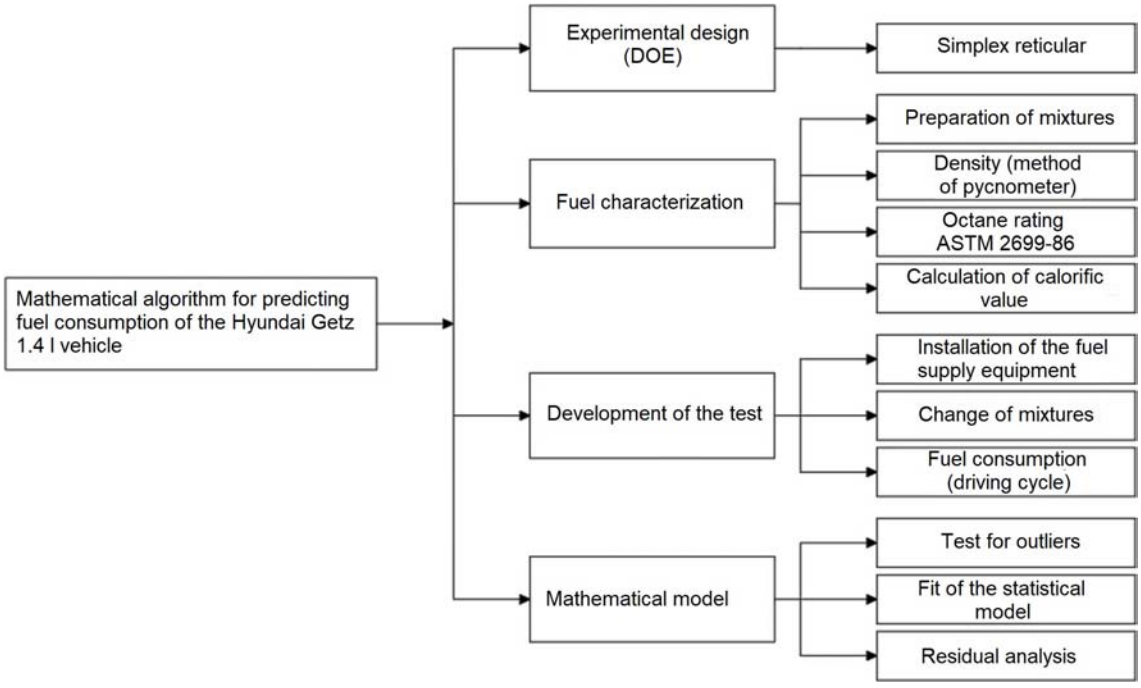


Figure 1. Methodology applied 1

2.1. Experimental design

The experimental design employed is reticular simple lattice by mixtures, using the Minitab 17 Statistical software, see Table 1, which indicates the number of components, anhydrous ethanol and gasoline for this case; design points, 21, degree of reticulum used, 4; in addition, the sequence of the experimental runs is random.

2.2. Characterization of the mixtures

The fuel used is a mixture of anhydrous ethanol with gasoline, which meet the specifications of the NTE INEN 2 478:2009 [24] and NTE INEN 935:2010 [25] standards, respectively. The used mixtures are characterized (see Table 2) indicating density, octane rating and calorific value.

Table 1. Random experimental design

Sequence established	Run sequence	Type Pt.	Blocks	ETHANOL	GASOLINE
17	1	0	1	0,5	0,5
16	2	2	1	0,25	0,75
3	3	0	1	0,5	0,5
14	4	–1	1	0,75	0,25
9	5	2	1	0,25	0,75
19	6	1	1	1	0
2	7	2	1	0,25	0,75
13	8	-1	1	0,25	0,75
10	9	0	1	0,5	0,5
1	10	1	1	0	1
12	11	1	1	1	0
21	12	–1	1	0,75	0,25
7	13	–1	1	0,75	0,25
15	14	1	1	0	1
11	15	2	1	0,75	0,25
18	16	2	1	0,75	0,25
20	17	–1	1	0,25	0,75
8	18	1	1	0	1
5	19	1	1	1	0
6	20	–1	1	0,25	0,75
4	21	2	1	0,75	0,25

Table 2. Physicochemical properties of the mixtures

Fuel	Physicochemical properties			
	Density (kg/m^3)	Octane rating (RON)	High calorific value (kJ/kg)	Low calorific value (kJ/kg)
E0	740	85,6	47 300	44 000
E25	760	90,95	42 900	39 725
E50	768,7	96,3	38 500	35 450
E75	782	101,65	34 100	31 175
E100	790,7	107	29 700	26 900

2.3. Measurement of fuel consumption

This research is developed in the city of Cuenca (Ecuador) located at 2558 meters above sea level, the fuel consumption tests are carried out in a 1.4-liter Hyundai Getz vehicle with electronic injection system and treatment of gases with a three-way catalytic converter; the compression ratio is 9,5:1, DOHC distribution system with four valves per cylinder and atmospheric type aspiration Hyundai Motor Company (2011) [26].

For performing the different tests of this research work, the original fuel feeding system of the vehicle is replaced by an alternate supply system meeting the technical specifications of the manufacturer, with the purpose of preventing alterations in the fuel mixtures.

The equipment utilized for measuring fuel consumption is the FLOW-MASTER MAHA CH-4123 flow meter. Figure 2 illustrates the installation of the equipment in the vehicle; Table 3 indicates the data obtained according to the design of experiment

Figure 2. Tests of operation 1

Table 3. Measurement of fuel consumption

Sequence established	Run sequence	ETHANOL	GASOLINE	Fuel consumption (g/km)
17	1	0,5	0,5	0,0435
16	2	0,25	0,75	0,0405
3	3	0,5	0,5	0,0405
14	4	0,75	0,25	0,037
9	5	0,25	0,75	0,0475
19	6	1	0	0,0475
2	7	0,25	0,75	0,0495
13	8	0,25	0,75	0,05
10	9	0,5	0,5	0,04
1	10	0	1	0,043
12	11	1	0	0,0515
21	12	0,75	0,25	0,045
7	13	0,75	0,25	0,0465
15	14	0	1	0,036
11	15	0,75	0,25	0,0475
18	16	0,75	0,25	0,047
20	17	0,25	0,75	0,04
8	18	0	1	0,036
5	19	1	0	0,0515
6	20	0,25	0,75	0,0355
4	21	0,75	0,25	0,049

To obtain specific fuel consumption, it was utilized the driving cycle representative of the city of Cuenca, which is identified in Figure 3; a micro-cycle is applied with the first five minutes, since they are the most distinctive ones as determined by means of a pre-experimental analysis. The test is carried out on a Maha LPS 3000 power bank.

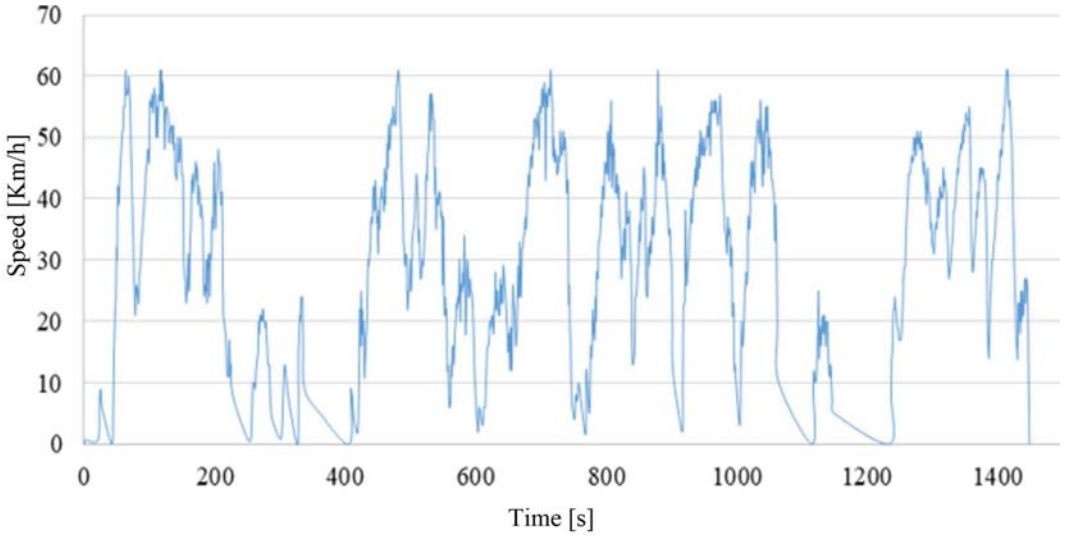


Figure 3. Driving cycle representative of the city of Cuenca

2.4. Data treatment for the response model

With the data of the experimental design as a function of the different mixtures, it is convenient to conduct validation tests of the results before proceeding to obtain the response model. The R22 Dixon test for outliers [27] was established for validating the results. After the data have been validated the model is constructed, and a residual analysis is performed with the purpose of verifying the hypotheses of normality, homoscedasticity, independence and linearity of the model.

3. Results and discussion

In order to determine the model that explains fuel consumption, a multivariable linear regression analysis is carried out with the data obtained from the DOE, and a higher order model is fitted as indicated in Table 4, where the different values of «R2» and «p-value» are observed. This information enables choosing the linear model, to meet the assumptions, and it is also observed in the «p-value» column that the higher order models are greater than the 0.05 significance level, and thus these models are not considered.

After the linear model has been selected, the estimated regression coefficients are determined (Table 5), and it is obtained the formula of the model for predicting consumption shown in Equation 7.

$$2Y = 0.05100(etanol) + 0,03455(gasolina) \quad (7)$$

Table 4. Summary of the models fitted for fuel consumption

Model	P-value	(%)	Predicted (%)	Adjusted (%)
Linear	0	94,2	93,04	93,86
Quadratic	0,274	94,6	92,66	93,94
Complete cubic	0,784	94,6	91,92	93,62
Complete cubic	0,126	95,3	91,59	94,17
Note. Value *p < 0,05				

Table 5. Regression coefficients estimated for fuel consumption

Term	Coeff.	EE of the coeff.	T	P	VIF
ETHANOL	0,051	0,000562	*	*	1,19
GASOLINE	0,03455	0,000562	*	*	1,19
Note. Value *p < 0,05					

The results indicated in Figure 4 are obtained after applying Equation 7.

In addition, in the variance analysis for the linear model, as indicated in Table 6, the «p-value» = 0,000, therefore, the model is significant and with a very good adjusted «R^2» of 93.86.

The remaining models are excluded because they do not meet the p-value assumption [28].

Once the standardized analysis of variance has been conducted, graphical results are shown in the following which confirm the fit of the model to the fuel consumption.

In order to evaluate the model of fuel consumption, a four in one residuals plot is utilized, as indicated in Figure 5. This graphical analysis with respect to the standardized residuals, enables verifying the fit of the experimental model obtained previously.

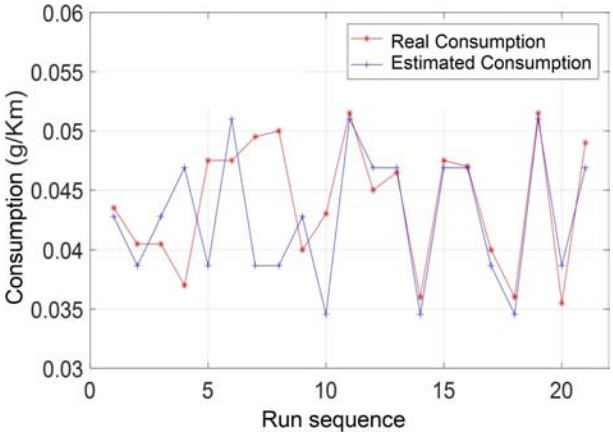


Figure 4. Result of the numerical algorithm

Table 6. Variance analysis for fuel consumption

Source	GL	SC Sec.	SC Adjust.	MC Adjust.	F	P
Regression	1	0,00061	0,000609	0,000609	306,5	0
Linear	1	0,00061	0,000609	0,000609	306,5	0
Residual error	19	3,8E-05	0,000038	0,000002		
Lack of adjustment	3	8E-06	0,000008	0,000003	1,34	0,3
Pure error	16	0,00003	0,00003	0,000002		
Total	20	0,00065				
Note. Value *p < 0,05						

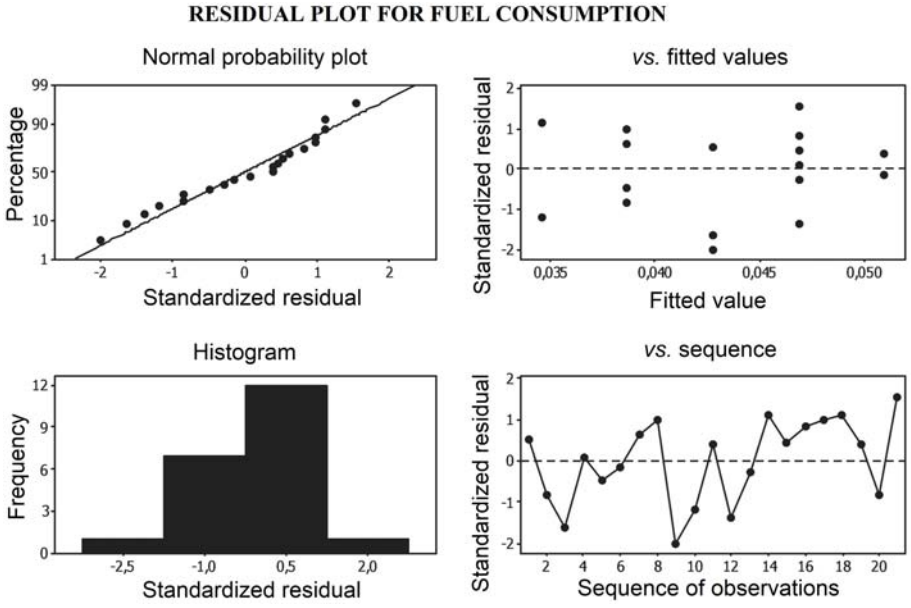


Figure 5. Residuals plots for the fuel consumption data

From the analysis of Figure 5, the following conclusions are drawn:

a. The normal probability plot shows that the residuals follow a normal distribution, since the fit to a normal trend line.

b. The Histogram plot of the residuals is bell-shaped, having a value of –2,02 for observation number nine, which is out of the allowed range (–2,0) in standard residual; however, this does not have a greater significance in the model, and the normality of the data is accepted.

c. On the other hand, in the plot of standardized residuals versus fitted values, no abnormal trend is observed that indicates a bad fit of the model, since the residuals are randomly distributed around zero, and thus it is considered that there is independence.

d. At last, in the plot of standardized residuals versus the sequence of observations, there is a random pattern around the central line, no ascending or descending trends of the observations occur that indicate a bad fit of the model.

With this analysis it is concluded that the variance is correct and that the model does not exhibit anomalies in the results, and thus it may be used to predict fuel consumption in a better manner.

Figure 6 shows a plot of the response of the mixture (ethanol-gasoline) in the fuel consumption. This enables evaluating how the components are related to the response using a fitted model.

The tracking plot of fuel consumption shown in Figure 6, provides the following information about the effects of the components:

It is observed that fuel consumption has a growing trend as the ethanol increases up to a maximum in E100, while decreases the concentration of the second component, in this case gasoline extra.

Also note that the slope of the curve is steeper in the section between 0 and 15% of ethanol, and in the section after 40%; therefore, in the range between 15

and 40% it is less step, corresponding to the zone of less fuel consumption.

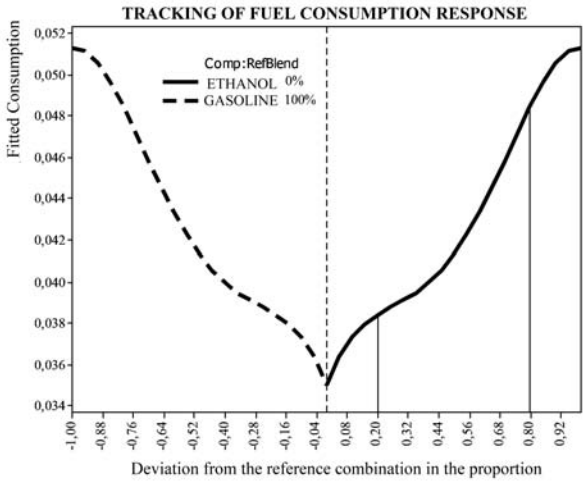


Figure 6. Tracking plot of fuel consumption response

4. Conclusions and recommendations

Specific fuel consumption is directly proportional to the percentage of ethanol in the mixture, i.e., as the percentage of ethanol in the mixtures increases, specific fuel consumption increases as well. Therefore, with an E20 mixture there is a 7% increase in fuel consumption, while in an E100 mixture there is 31.3% more consumption with respect to the gasoline of 86.5 octanes.

It was established a mathematical model that enables determining fuel consumption for different percentages of ethanol in the gasoline, which can be applied for contrasting real tests.

The increase in fuel consumption is due to the reduction of the calorific value of the mixture as the concentration of ethanol varies.

During the development of the tests, the vehicle operated correctly without showing anomalies for concentrations greater than 30%.

References

[1] M. Al-Hasan, "Effect of ethanol–unleaded gasoline blends on engine performance and exhaust emission," *Energy Conversion and Management*, vol. 44, no. 9, pp. 1547–1561, 2003. [Online]. Available: https://doi.org/10.1016/S0196-8904(02)00166-8

[2] H. L. MacLean and L. B. Lave, "Evaluating automobile fuel/propulsion system technologies," *Progress in Energy and Combustion Science*, vol. 29, no. 1, pp. 1–69, 2003. [Online]. Available: https://doi.org/10.1016/S0360-1285(02)00032-1

[3] CEPAL, "Consideraciones ambientales en torno a los biocombustibles líquidos," in *División de Desarrollo Sostenible y Asentamientos Humanos,*, 2008.

[4] S. Fernández Henao, J. Mosquera A., and J. Mosquera M., "Análisis de emisiones de CO^2 para diferentes combustibles en la población de taxis en Pereira y Dosquebradas," *Scientia et Technica*, vol. 2, no. 45, ago. 2010. [Online]. Available: https://doi.org/10.22517/23447214.385

[5] E. A. Melo Espinosa, Y. Sánchez Borroto, N. Ferrer Frontela, and N. Ferrer Frontela, "Evaluación de un motor de encendido por chispa trabajando con mezclas etanol-gasolina," *Ingeniería Energética*, vol. 33, pp. 94–102, 08 2012. [Online]. Available: https://bit.ly/38kOxFW

[6] J. Goñi Delión and M. Rojas-Delgado, "Combustibles alternativos en motores de combustión interna," *Ingeniería Industrial*, 02 2014. [Online]. Available: https://doi.org/10.26439/ing.ind2014.n032.122

[7] J. Gibbons, *Interagency Assessment of Oxygenated Fuels*. National Science and Technology Council, 1997. [Online]. Available: https://bit.ly/355W21N

[8] Apace Research Ltd, "Intensive field trial of ethanol/petrol blend in vehicles," ERDC Project No. 2511, Tech. Rep., 1998. [Online]. Available: https://bit.ly/32iZaFK

[9] R. Ragazzi and K. Nelson, "The impacts of a 10 % ethanol blended fuel on the exhaust emissions of tier 0 andtier 1 light duty gasoline vehicles al 35 f," *Colorado Department of Public Health and Environment*, 1999.

[10] B.-Q. He, Jian-Xin Wang, J.-M. Hao, X.-G. Yan, and J.-H. Xiao, "A study on emission characteristics of an efi engine with ethanol blended gasoline fuels," *Atmospheric Environment*, vol. 37, no. 7, pp. 949–957, 2003. [Online]. Available: https://doi.org/10.1016/S1352-2310(02)00973-1

[11] T. W. Patzek, S.-M. Anti, R. Campos, K. W. ha, J. Lee, B. Li, J. Padnick, and S.-A. Yee, "Ethanol from corn: Clean renewable fuel for the future, or drain on our resources and pockets?" *Environment, Development and Sustainability*, vol. 7, no. 3, pp. 319–336, Sep. 2005. [Online]. Available: https://doi.org/10.1007/s10668-004-7317-4

[12] C.-W. Wu, R.-H. Chen, J.-Y. Pu, and T.-H. Lin, "The influence of air-fuel ratio on engine performance and pollutant emission of an si engine using ethanol-gasoline-blended fuels," *Atmospheric Environment*, vol. 38, no. 40, pp.

7093–7100, 2004, 8th International Conference on Atmospheric Sciences and Applicat ions to Air Quality (ASAAQ). [Online]. Available: https://doi.org/10.1016/j.atmosenv.2004.01.058

[13] S. T. Coelho, J. Goldemberg, O. Lucon, and P. Guardabassi, "Brazilian sugarcane ethanol: lessons learned," *Energy for Sustainable Development*, vol. 10, no. 2, pp. 26–39, 2006. [Online]. Available: https://bit.ly/2JJIeBM

[14] R. K. Niven, "Ethanol in gasoline: environmental impacts and sustainability review article," *Renewable and Sustainable Energy Reviews*, vol. 9, no. 6, pp. 535–555, 2005. [Online]. Available: https://doi.org/10.1016/j.rser.2004.06.003

[15] American Coalition For Ethanol. (2005) Fuel economy study: comparing performance and costs of va-rious ethanol blends and standard unleaded gasoline. [Online]. Available: https://bit.ly/32mEmwP

[16] E. Behrentz, *Beneficios ambientales asociados con el uso de combustibles alternativos*. Centro de Investigaciones en Ingeniería Ambiental (CIIA) Universidad de los Andes, 2008. [Online]. Available: https://bit.ly/38k6cOj

[17] T. Durbin, J. W. Miller, T. Huai, D. R. Cocker III, and Y. Younglove, "Effects of ethanol and volatility parameters on exhaust emissions of light-duty vehicles." in *UC Riverside: Center for Environmental Research and Technology*. [Online]. Available: https://bit.ly/3oZFmkp

[18] E. Shapiro. (2006) Roundtable on ethanol fuel: automaker view.

[19] H. S. Yücesu, T. Topgül, C. Çinar, and M. Okur, "Effect of ethanol-gasoline blends on engine performance and exhaust emissions in different compression ratios," *Applied Thermal Engineering*, vol. 26, no. 17, 2006. [Online]. Available: https://doi.org/10.1016/j.applthermaleng.2006.03.006

[20] H. R. Acevedo G. and J. M. Mantilla G., "Viabilidad ambiental del uso de biocombustibles para motores a gasolina y diésel en colombia," *Boletin del Observatorio Colombiano de Energía, Bogota. D. C.*, pp. 3–14, 2008. [Online]. Available: https://bit.ly/38kSXg0

[21] U.S. Department of Energy, "Handbook for handling, storing, and dispensing e85 and other ethanol-gasoline blends," U.S. Department of Energy, Tech. Rep., 2013. [Online]. Available: https://bit.ly/2U1zVmF

[22] H. Gutiérrez Pulido and R. de la Vara Salazar, *Análisis y diseño de experimentos*. McGraw-Hill, 2003. [Online]. Available: https://bit.ly/36gB7rU

[23] E. A. García, "Modelización termodinámica de un motor turboalimentado y propulsado por bioetano," 2009. [Online]. Available: https://bit.ly/3k85tBO

[24] INEN, "NTE INEN 2478: Etanol anhidro. requisitos," Instituto Ecuatoriano de Normalización, Tech. Rep., 2009. [Online]. Available: https://bit.ly/354abwc

[25] ——, "NTE INEN, Gasolina. Requisitos," Instituto Ecuatoriano de Normalización, Tech. Rep., 2012. [Online]. Available: https://bit.ly/2JLGsQD

[26] Hyundai. (2011) Manual del taller. [Online]. Available: https://bit.ly/356tPIc

[27] S. P. Verma and A. Quiroz-Ruiz, "Critical values for six Dixon tests for outliers in normal samples up to sizes 100, and applications in science and engineering," *Revista Mexicana de Ciencias Geológicas*, vol. 23, pp. 133–161, 01 2006. [Online]. Available: https://bit.ly/3eBoOKG

[28] W. Contreras, J. Ortega, and R. Japa, "Aplicación de una red neuronal feed-forward backpropagation para el diagnóstico de fallas mecánicas en motores de encendido provocado," *INGENIUS*, pp. 32–40, 2019. [Online]. Available: https://doi.org/10.17163/ings.n21.2019.03

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.05
pISSN: 1390-650X / eISSN: 1390-860X

Algorithms for Table Structure Recognition

Algoritmos para el reconocimiento de estructuras de tablas

Yosveni Escalona Escalona[1,*]

Received: 03-08-2020, Reviewed: 22-09-2020, Accepted after review: 09-10-2020

Abstract

Tables are widely adopted to organize and publish data. For example, the Web has an enormous number of tables, published in HTML, embedded in PDF documents, or that can be simply downloaded from Web pages. However, tables are not always easy to interpret due to the variety of features and formats used. Indeed, a large number of methods and tools have been developed to interpreted tables. This work presents the implementation of an algorithm, based on Conditional Random Fields (CRFs), to classify the rows of a table as header rows, data rows or metadata rows. The implementation is complemented by two algorithms for table recognition in a spreadsheet document, respectively based on rules and on region detection. Finally, the work describes the results and the benefits obtained by applying the implemented algorithm to HTML tables, obtained from the Web, and to spreadsheet tables, downloaded from the Brazilian National Petroleum Agency.

Keywords: Tabular Data, HTML Tables, Spreadsheets, Conditional Random Fields, Machine Learning, Algorithm.

Resumen

Las tablas son una manera muy común de organizar y publicar datos. Por ejemplo, en el Internet se halla un enorme número de tablas publicadas en HTML integradas en documentos PDF, o que pueden ser simplemente descargadas de páginas web. Sin embargo, las tablas no siempre son fáciles de interpretar pues poseen una gran variedad de características y son organizadas en diferentes formatos. De hecho, se han desarrollado muchos métodos y herramientas para la interpretación de tablas. Este trabajo presenta la implementación de un algoritmo, basado en campos aleatorios condicionales (CRF, Conditional Random Fields), para clasificar las filas de una tabla como fila de encabezado, fila de datos y fila metadatos. La implementación se complementa con dos algoritmos para reconocer tablas en hojas de cálculo, específicamente, basados en reglas y detección de regiones. Finalmente, el trabajo describe los resultados y beneficios obtenidos por la aplicación del algoritmo para tablas HTML, obtenidas desde la web y las tablas en forma de hojas de cálculo, descargadas desde el sitio de la Agencia Nacional de Petróleo de Brasil.

Palabras clave: datos tabulados, tablas HTML, hoja de cálculo, campos aleatorios condicionales, aprendizaje automático

[1,*]Research and Development Department, SOLINTEC, Araçatuba, Brazil
Corresponding author ✉: yosveni.escalona@solinftec.com.br. http://orcid.org/0000-0003-2992-0540

Suggested citation: Excalona Escalona, Y. (2021). «Algorithms for Table Structure Recognition». Ingenius. N.° 25, (january-june). pp. 50-61. doi: https://doi.org/10.17163/ings.n25.2021.05.

1. Introduction

The volume of data available on the Web has grown in a dizzying way, which makes the Web a vast repository of data that describe our environment and our interactions. The wealth and strength of these allow the development of today's economy and society.

The data found on the Web are related to product information, articles imparting encyclopedic knowledge, presentations of state-of-the-art scientific results or reports on current financial data. A great part of those data can be found in tables, which require a particular analysis since they may be expressed in HTML, embedded in PDF documents or made available as downloadable spreadsheets, among other formats. Usually, they are organized merely and compactly as rows and columns, but they can be much more complex, with metadata and additional information.

Tables proved to be valuable sources, but their use can be very diversified, ranging from Web search to data discovery in spreadsheets and knowledge base augmentation [1]. In the literature, one finds research about methods and tools for tabular data extraction from spreadsheets, HTML table, tables embedded in PDF documents, etc. The vast majority of these methods and tools follows strategies based on heuristic rules and machine learning algorithms. The strategy for tabular data extraction and for table row classification also depends on the document format. Exploring a large set of tables has been a challenge because, in general, table semantics is not known. In [2], a corpus of over 100 million tables is presented, but the meaning of each table is rarely explicit in the table itself. Another challenge is the structure of the table. For example, the tasks described in [3–6] focus on recovering table semantics and linking table data with external sources for tables classified as genuine, with considerable data loss. These works do not consider fundamental aspects, such as the orientation of the table, and discarded those tables that are classified as non-genuine.

Another aspect to consider is based in the kind of document, for example, Correa and Zander [7] analyzed a group of methods and tools focused on extracting tabular content from PDF based on two main characteristics: ease of use and output results and the categorization of the tools based on theoretical proposals, free of cost and commercial. In [8] were developed several heuristics, which together recognize and decompose tables in PDF files and store the extracted data in a structured data format (XML) for ease of use, these heuristics are divided into two groups: table recognition and table decomposition. Other techniques were presented in [9] for data tabular extraction from PDF documents with the goal of identified table boundary where the authors describe a methodology that applies two machine learning algorithms, CRFs and Support Vector Machine (SVM). Also, have been reviewed works based in the table boundaries identification process and designed for the semantic matching and annotation of numeric and time-varying attributes in Web tables as the presented in [10–12] which annotate Web tables effectively and efficiently and identify the boundaries between attributes name rows (or columns) and its corresponding value data rows (or columns) in the table.

Furthermore, we can do special mention to the works related to the recognition of HTML Table Structure and table header detection and classification described in [13, 14] suggesting some techniques based on heuristics rules and used a learning classification algorithm for delineating kinds of tables existing into a document and detecting the structure and header types.

Finally, we make emphasis on the approach proposed in [15] that was based on machine learning techniques that cover two fundamental tasks of the table extraction process: localization of the table and identification of the row positions and types. This work is focused on the implementation of two algorithms to table recognition in Spreadsheets as well as other algorithm based on Conditional Random Fields (CRFs), to classify the kinds of rows into tables. The datasets were created with HTML tables downloaded from the Web and spreadsheet tables were downloaded from the Web site of the National Oil Agency of Brazil (ANP).

1.1. Background

Tables are frequently found in printed documents, such as books or newspapers, as well as digital documents, such as Web pages or presentation slides. But they also represent an essential concept in relational databases and spreadsheets. They may be distinguished according to their structure and orientation. A relational or horizontal [8], as illustrated in Table 1, has rows, which provide data about specific objects, called entities, and columns, which represent attributes that describe the entities.

Table 1. Example of a relational table

ID	Name	Age	Country	Job
1	Bob Smith	35	USA	Programmer
2	Jane Smith	31	USA	Teacher
3	Robert White	24	UK	Engineer

More complex types of tables exist, such as those where the attributes that describe the entities are laid vertically and the entities in a horizontal way or other kinds of structures as show Table 2 and Table 3.

Table 2. Example of a non-relational table

	Obj 1	Obj 2	Obj 3
Name	V1	V1	V1
Age	V2	V2	V2
Height	V3	V3	V3

Table 3. Example of a non-relational table with additional information

Patent Applications by Residents		
Data source: worldbanc.org (show countries top in each continent)		
Country	**Residents**	**Applications**
North America		
Unites States	307,700,000	224,912
Canada	33,739,900	5,067
Asia		
Japan	127,557,958	295,315
China	1,331,380,000	229,096

De manera más precisa, una tabla se define como:

Definition 1. A table is a pair T = (H,D) consisting of an optional header H and data D, where:

- The header $H = \{h_1, h_2, \ldots, h_n\}$ is an n-tuple of header elements h_i; if the set of header elements exists, it might be represented either as a row or as a column.
- The data nodes are organized as an (n, m) matrix consisting of n rows and m columns:

$$D = \begin{bmatrix} C_{11} & \therefore \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nm} \end{bmatrix}$$

The table row classification process consists of identifying each of the elements of a table. The general idea is based on locating the header and data in the table. It is also relevant to identify the layout elements and metadata in a table. Figure 1 shows the table row classification process denoting in different colors some of the elements present in the table: red indicates the elements that represent the titles; yellow the row header; blue the data rows; and green the additional metadata.

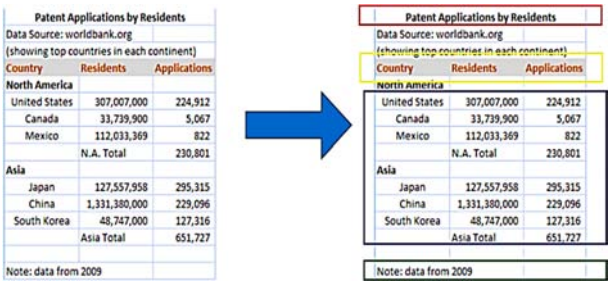


Figure 1. Tabular extraction process

The rest of the paper is organized as follows. Section 2 covers details of the implementations of the algorithms for table recognition and for table row classification. Finally, Section 3 describes experiments and results.

2. Materials and Methods

2.1. Algorithms for Table Recognition and for Table Row Classification

This section describes implementations of algorithms to recognize tables in spreadsheets and an algorithm, based on Conditional Random Fields, to classify table rows.

2.1.1. A Rule-based Algorithm to Detect Tables in Spreadsheets

Several Rule-based algorithms detect tables in spreadsheets making use of cell attributes, such as border, format, and data type. Where each cell attribute in the spreadsheet has a specific value associated with that cell. In turn, the cell border has the attributes direction, style, and color. The border may surround the cell in 4 different directions: top, bottom, left, and right.

A cell format is the visual formatting applied to the data of the cell, such as, number format, font style name, font name, font size, font bold, font italic, and font color. The detection of multiple tables in the same spreadsheet is performed by finding a separator between two tables (usually a set of empty rows), as explained in what follows [16].

Given a table T, with rn rows and cn columns, the following layout features are computed:

- Average number of columns, computed as the average number of cells per row.

$$c = \frac{1}{rn} \sum_{i=1}^{rn} c_i \quad (1)$$

where c_i is the number of cells in row i, $i = 1, \ldots rn$.

- Average number of rows, computed as the average number of cells per columns.

$$r = \frac{1}{cn} \sum_{i=1}^{cn} r_i \tag{2}$$

where r_i is the number of cells in column i, $i = 1, \ldots, cn$.

Figure 2 shows the algorithm that identifies the number of tables into a document and captures the range of the rows that represent the tables.

Algorithm 1 Table Detection and Recognition

Input: A spreadsheet document
Output: Number of tables into the document (ct)

$U \leftarrow$ The threshold to empty rows
$H \leftarrow$ Set header cells
$D \leftarrow$ Set data cells
$T \leftarrow$ Set title cells
$E \leftarrow$ Set empty rows
$R_i \leftarrow$ Numbers of cells tagged as header in row i
$X \leftarrow$ Numbers of Empty cells between a header row and title row
for $i \leftarrow 0$ to rn **do** ▷ Number of rows into the spreadsheet document
  for $j \leftarrow 1$ to cn **do** ▷ Number of columns into the spreadsheet
    if $C[i-1,j]$ **is** $Header$ **and** $C[i-1,j].type = String$ **then**
      if $C[i,j].format = C[i,j-1].format = C[i,j+1].format$ **and** $C[i-1,j].format = C[i-1,j].format = C[i,j].font = C[i-1,j].font$ **then**
        if $R_i > 0$ **then**
          $C[i,j] \in Header$
          if $C[i-1,j] \neq Empty$ **and** $C[i,j].format \neq C[i,j].format$ **then**
            $H \leftarrow H \cup \{c\}$ ▷ Adds c to the set of header cells
          end if
        end if
      end if
    else if $C[i,j]$ **and** $C[i-1,j]$ **and** $C[i+1,j] \in \{type, format\}$ **then**
      $D \leftarrow D \cup \{c\}$ ▷ Adds c to the set of data cells
    else if $C[i,j].type = String$ **and** $E \geq X$ **then**
      if $C[i,j-1]$ **and** $C[i,j+1] \in E$ **and** j *is first column* **then**
        $T \leftarrow T \cup \{c\}$ ▷ Adds c to the set of title cells
      else
        $E \leftarrow E \cup \{c\}$ ▷ Adds c to the set of empty cells
      end if
    end if
  end for
  if $len(E) = U$ **then** ▷ if the count of empty cells equals to threshold
    $ct \leftarrow ct + 1$
  end if
end for
return ct

Figure 2. Table Detection and Recognition Algorithm

Region Detection

Region Detection is computed through a graphp-based algorithm that detects tables in spreadsheets, called Remove and Conquer [17], it uses a comprehensive set of rules and heuristics based on a graph representation of a spreadsheet. The spreadsheet files contain one or more sheets and where each sheet is comprised of a collection of cells organized in rows and columns are defined as some useful definitions that help in the region detection process.

Definition 2. Let W denote the set containing all the cells of a worksheet.

Region detection consists in scanning the spreadsheet from the first cell in the left top corner until the last non-empty cell in the right bottom corner to check cells with similar formatting and to detect separators, such as empty rows, different cells formatting or different kinds of borders, such as different cell value type. A region is defined as follows. More precisely, a region is defined as follows.

Definition 3. A region is a maximal collection $R \subseteq W$ of cells from a rectagular area of the worksheet.

It also is inferred the layout role of non-empty cells in the worksheet where each non-empty cell is assigned one of the following roles: Header(H), Data(D), Title(T), Metadata or non-relational (N). This cell role is defined as following.

Definition 4. Let label: $W \rightarrow$ Labels, where Labels = {Header, Data, Title, Metadata}, be a function that maps cells to their assigned layout role. For empty cells label is undefined. We identify them using empty: $W \rightarrow \{0, 1\}$. It returns 1 for empty cells, otherwise 0.

Let the label be: Labels, where Labels = Header, Data, Title, Metadata, a function that relates to the cells their assigned design role. For empty cells, the label is not defined; these cells are identified using empty:, which returns 1 for empty cells and 0 otherwise.

The cells of a spreadsheet are grouped together so that adjacent cells have the same layout role (label) or form larger structures. These groups are called label regions, as shown in [17], as shown in Figure 3

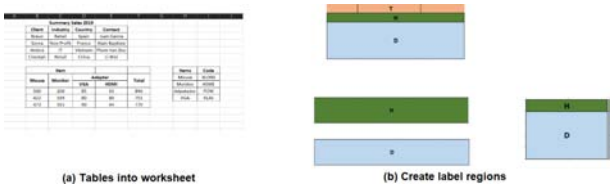


Figure 3. Creation label Regions Process

Formally, a label region is defined as follows:

Definition 5. A label region is a region LR of a spreadsheet such that, for any two cells c and c' in LR, $label(c) = label(c')$ and $empty(c) \neq 1$ and $empty(c') \neq 1$.

Figure 4(a) shows tables in a spreadsheet and Figure 4(b) indicates the regions corresponding to the table structures. The label region detection process clusters cells based on their label. It is iterated over each row to create sequences of cells having the same label. These form the base LRs. Subsequently, it merges LRs from consecutive rows, if their labels, minimum column, and maximum column match.

Table Representation through Graphs

Regions allow constructing graphs that captures the spatial interrelations of label regions. Figure 4 shows the representation of tables as a graph.

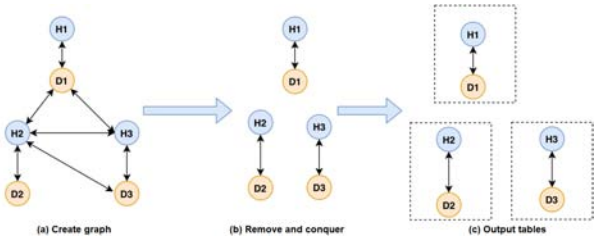


Figure 4. Table Representation through Graphs

The graph construction process consists of identifying spatial relations, such as top, bottom, left, and right, based on locating the nearest neighboring regions for each direction and identifying all vertices whose maximum row is less than the minimum row of another vertex. For each direction, a distance function is defined where all the nearest vertices are identified:

$$ND_v = \{n \in D_v \vee ddist(v,n) = \underset{u \in DV}{minddist}(u,v)\} \quad (3)$$

where D_v is the direction for the vertex v; directed edges (v,n) are created for every $n \in ND_v$.

2.1.2. The Remove and Conquer Algorithm

Remove and Conquer (RAC) is a rule-based algorithm whose objective is to separate the edges that are farther to the left and to the right direction of the graph that was created from each worksheet in a spreadsheet as shows Figure 5. The algorithm processes the strongly connected components of the graph to pair all groups formed and to detect valid tables.

The vertices are sorted in descending order of their maximum row, followed by the ascending order of their minimum row, thus the tables are searched in order inverse, from the bottom to top. Each header h is individually processed to identify vertices with minimum row greater or equal to h.

The algorithm that verifies the valid header is shown in Figure 6. All valid headers are stored in Q that represent the vertices set, including h; these vertices set is called potential tables.

Algorithm 2 Remove and Conquer

Input: Graph representation of a worksheet
Output: Tables into document

$P \leftarrow 0$
$E_l \leftarrow \{e \in E / dir(e) = Left$ **and** $ldist(e) > 1\}$
$E_r \leftarrow \{e \in E / dir(e) = Right$ **and** $rdist(e) > 1\}$
$E \leftarrow (E_l \cup E_r)$
for $G^s \in getSCC(G)$ **do**
  $LQ \leftarrow Null$
  $S^{'} \leftarrow \{v \leftarrow S^{'} / lbl(v) = Header\}$
  if $|S^{'}_H| > 0$ **then**
    for $h \leftarrow S^{'}_H$ **do**
      if $h \leftarrow LQ$ **then**
        $Q \leftarrow \{s \in S / rmin(s) \geq rmin(h)$ **and** $hasPath(s,h,E_s)\}$
      end if
      if $isValid(h,Q,0.5)$ **then**
        $P \leftarrow P \cup \{LQ\}$
        $S^{'} \leftarrow S^{'} / Q$
      else if $LQ = Null$ **then**
        if $|Q| = 1$ **and** $isAligned(h,LQ)$ **then**
          $LQ \leftarrow LQ \cup \{h\}$
        end if
      end if
    end for
  end if
  $P \leftarrow P \cup \{LQ\}$
end for
$U \leftarrow U \cup \{S\}$ ▷ Remaining unpaired
$P, U \leftarrow handleOverlapping(P,U)$
for $u \in U$ **do** ▷ Find nearest table left or right
  $N, dist \leftarrow getNearestVertices(u, (E_l \cup E_r))$
  $P^{'} \leftarrow \{P \leftarrow P / 0 < |N \cap P|\}$
  if $|P^{'}| = 1$ **and** $dist \leq 3$ **then**
    $P \leftarrow P \cup \{u\}, where\ P \in P^{'}$
  end if
end for
return P, U

Figure 5. Remove and Conquer Algorithm

Input: h: a Header vertex, Q: vertices to form table with, th: threshold for alignment ratio
Output: $True$ if h is valid, $False$ otherwise

if $|\{q \in Q | rmin(q) > rmax(h)\}| > 0$ **then**
  $Q_{\mathcal{H}} \leftarrow \{q \in Q | lbl(q) = Header$ **and** $rmin(q) \leq rmax(h)$ **and** $rmin(q) \geq rmin(h)\}$
  $X \leftarrow \emptyset$; $X' \leftarrow \emptyset$
  for all $u \in Q_{\mathcal{H}}$ **do**
    $X \leftarrow X \cup \{x \in \mathbb{N} | cmin(u) \leq x \leq cmax(u)\}$
  for all $v \in Q \setminus Q_{\mathcal{H}}$ **do**
    $X' \leftarrow X' \cup \{x \in \mathbb{N} | cmin(v) \leq x \leq cmax(v)\}$
  return $\dfrac{|X \cap X'|}{|X'|} \geq th$ **and** $|X| > 1$
else
  return $False$

Figure 6. Header Validity Check

The algorithm ensures that other vertices connected to h are not left isolated. Those vertices paired with a valid header are subtracted from the vertices set and then ordered to create set S'. The valid headers are appended to the set of valid headers, called LQ. The vertices that represent potential tables, called Q, are not appended directly to the tables set P because the algorithm needs to check that h is not connected to other vertices. The tables that cannot be formed are stored in U. Then, in the last step the algorithm, it attempts to pair the tables in U with the nearest table on their left or right.

2.1.3. A Machine Learning Algorithm for Table Row Classification

One important contribution in this work is the identifying and classification of the kinds of rows that compose a table through the implementation of a machine learning algorithm, this case, Conditional Random Fields(CRFs) which is based on features set, values of the cells, as well the classes that represent the table structure.

CRFs are undirected graph models, introduced by Lafferty *et al* [18], that can act as classifiers for sequence labeling tasks. They are frequently used for natural language processing, such as part-of-speech tagging. The CRF algorithm defines X as a random variable over data sequences to be labeled, and Y as a random variable over the corresponding label sequences. Figure 7 shows a structure of a linear Conditional Random Field.

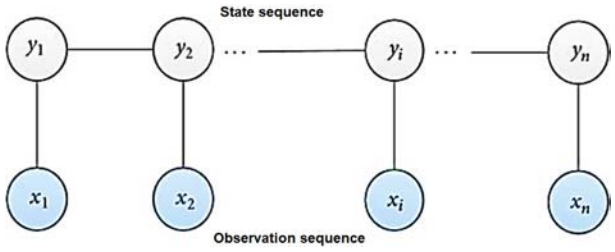


Figure 7. Structure of a linear chain Conditional Random Field

In our problem of classifying table rows, the input sequence x corresponds to a series of rows of a given table, while the label sequence y is the series of labels assigned to the observed rows. Each row x in is assigned exactly one label in y.

Formally, Conditional Random Fields are defined as follows:

Definition 6. Let $G = (V, E)$ be a graph and let $Y = (Y_v)_{v \in V}$ be a sequence of random variables indexed by the vertices of G. A conditional random field is a pair (X, Y) such that, when conditioned on X, the random variables Y_v obey the Markov property with respect to the graph.

$$P(Y_v|X, Y_w, w \neq v = P(Y_v \vee X, Y_w \approx v) \quad (4)$$

where $w \approx v$ means that w and v are neighbors in G.

The probability $P(X \vee Y)$ of a state sequence Y, given an observation sequence X, is:

$$P(X \vee Y) = \frac{1}{Z(x)} exp(\sum_j \lambda_j f_j(Y_{i-1}, Y_i, X, i) + \\ + \sum_k \mu_k g_k(Y_i, X, i)) \quad (5)$$

where $f_j(Y_{i-1}, Y_i, X, i)$ is a transition feature function of the observation sequence and of the labels at positions i and i-1 in the label sequence; $g_k(Y_i, X, i)$ is a state feature function of the label at position i and the observation sequence; and λ_j and μ_k are parameters to be estimated from training data.

In the data table scenario, X represents the list of rows in the table, and Y represents the corresponding row classes. Each relational data table has a schema, which, in the context of data tables, consists of attribute names, values, and types, where attribute names are column titles, attribute types are the types of values in the column, and attribute values correspond to data values in the column's cells. Column names are stored in a special row or rows, usually near the head of the table, called header rows, while the data is stored in rows referred to as data rows.

The data table may also contain descriptions of data, which it refers to the metadata. In corresponding to the criteria addressed above are identified each type of row according to the properties of each cell in a data table. Then, we focused our problem in assign one label each row where each row consists of constituent cells, which can exhibit different sets of attributes. The feature selection process involves the extraction of a collection of attributes for individual cells and combining attributes from all cells in the row, in order to construct a set of row features. Consider the ideas addressed above and an example of a simple table with header and data as showed in Figure 8.

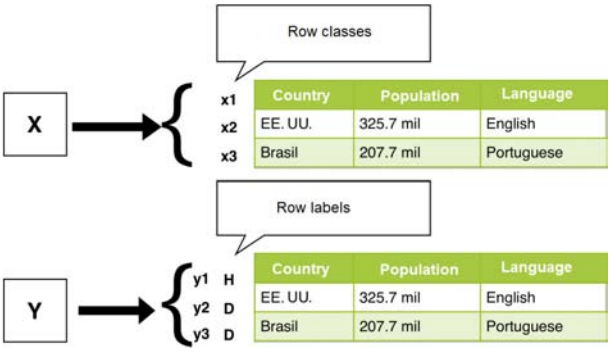


Figure 8. Example of a labelled table

The X represents a vector with the rows of the table and Y represents another vector with the tags of each row x of the table.

2.1.4. Row Classes

According to the structure of tables and the Definition 6 was defined the kinds of classes as shown 4.

Table 4. Row classes

Label	Description
H	Represents the header row in the table
D	Data rows contains data records
N	Non-relational metadata

2.1.5. Feature Set

In any machine learning algorithm, a feature is an individual measurable property or characteristic of a phenomenon being observed [19]. So, each feature was partitioned into three categories considering aspects related to the layout, styles, and values which we call layout attributes.

Layout attributes are the cells that are commonly found in header rows, which usually contain merged table cells with centered text.

Style attributes are various properties derived from stylesheets, such as font type, font color, font weight or underlined text.

Value attributes are those that represent cells where the stored information is exclusively linked to the data rows. Header rows often contain relatively short textual values, rather than numbers or dates.

Let see an example like the CRF algorithm works in our table classification problem given the transition feature $f_j(Y_{i-1}, Y_i, X, i)$ and features function $g_k(Y_i, X, i)$:

> x is a row into the data table.
>
> j is a position-row in the table (each feature is associated with a position); more than one feature associated with the same position.
>
> y_{j-1}, y_j are the tags (classes) assigned to rows j and $j-1$ de x

Then, the feature function and the state function are as follows:

$$f_1(Y_{i-1}, Y_i, X, i) = \begin{cases} 1 if x_j \in heady_j = H \\ 0, otherwise \end{cases} \quad (6)$$

$$f_2(Y_{i-1}, Y_i, X, i) = \begin{cases} 1 if x_j \in daty_j = D \\ 0, otherwise \end{cases} \quad (7)$$

$$g_1(Y_i, X, i) = \begin{cases} 1 if (x_j isacell \in x) \wedge (x_j \in rowfeatures) \wedge y_j \\ 0, otherwise \end{cases} \quad (8)$$

$$g_2(Y_i, X, i) = \begin{cases} 1 if (x_j isacell \in x) \wedge (x_j \in rowfeatures) \wedge y_j = D \\ 0, otherwise \end{cases} \quad (9)$$

The full list of individual cell attributes is given in Table 5. The features are divided according to the kind of attributes that they represent.

Table 5. Cells attributes

Layout	Style	Value	Spatial
IsMerged	IsBold	IsEmpty	RowNumber
Aligment	IsItalic	IsText	ColNumber
	IsUnderlined	IsNumber	NumNeighbor
	IsColored	IsDate	MatchStyle
	Font	IsAlpNum	MatchType
	Format	IsCapital	
	Border	TotalWord	

2.1.6. Similarity between rows

Another characteristic that was taken into account to generalize the training data was the row similarity [20] where a unique feature is assigned to each unique combination (c, r) where c is the number of cells exhibiting an attribute and r is the number of cells in the row. Then, two rows R_x and R_y are considered similar with respect to a certain cell attribute α if the logarithm of their widths are equal and the logarithm of the number of cells exhibiting or lacking attribute α. This approach is known as feature binnig and can be defined as follow.

Formally, for a row R_i of length r in which c cells exhibit a specific cell attribute α, is assigned feature $R_\alpha = (a, b)$ to $R_j(a, b)$, where a and b are the bin and computed as follows.

$$a = \begin{cases} 0, & if c = 0 \\ \lfloor log_2(c) + 1 \rfloor, & if 0 < c \leq r/2 \\ \lfloor log_2(r - c) + 1 \rfloor^-, & if\ r/2 < c < r \\ 0^- & if c = r \end{cases} \quad (10)$$

$$b = \lfloor log_2(r) \rfloor \quad (11)$$

The aim of the bins is given by:

1. Differentiate Between Table Widths.
2. Aggregate Wide Tables.
3. Highlight Uniform Rows.

3. Results and discussion

This section presents the experiments performed to test the accuracy of the implementation of the table row classifier as well as the experiments with table recognition in spreadsheet documents.

3.1. Preprocessing

The preprocessing task focuses on two table scenarios, HTML tables and spreadsheets tables, with the goal of removing irrelevant content or content that will not provide information for our table row classifier. Other aspects that were considered were the structure of the tables and the information present in both kinds of tables. This work does not cover in details whole the preprocessing of tables so we only emphasize those elements that we consider most important. In the case of the spreadsheet tables we highlight that the dataset had a predefined annotation, but with many errors related to the identification of ranges of data rows and head rows.

3.2. Main Characteristics of the Datasets used for Testing

Table 6 shows the statistics of the annotation process in both datasets. Each row of each table was annotated with the label corresponding to its class: "H" for header, "D" for data, etc.

Table 6. Annotated tables

	HTML	Spreadsheet
Annotated tables	105	252
Annotated rows	13,025	227,638
Header rows	105(<1 %)	252(<1 %)
Data rows	12,920(99 %)	227,254(98 %)
Other row classes	0(0 %)	132(<1 %)

The above table indicates that a critical aspect of both HTML and spreadsheets tables is that the percentage of header rows is very low, due to the fact that the tables obtained were simple tables with simple schemes (tables with a single header row followed by one or more data rows).

3.3. Table Classification Experiments

This section presents the experiments to evaluate the table classification solution proposed. In a first stage, the algorithm was trained with 80% of the data and tested with 20% of the data, randomly selected. The algorithm used L-BFGS as the optimization method and regularizations parameters L1 and L2 set to 0.1 and 0.01. The experiments with HTML and spreadsheet tables were performed separately to expose the differences between the two table formats. The performance metrics adopted were precision, recall, f1-score, support.

3.3.1. Results

Shows the results obtained. We observe that the precision value for spreadsheet tables was higher than that for HTML tables due to two main factors: (1) the features of the spreadsheet tables have a better definition; (2) we guarantee a correct definition for data rows. The recall was similar for both kinds of tables, as well as the F1-score. An important point in this analysis relates to the number of rows classified as non-relational in the spreadsheet dataset, due to the fact that we annotated spreadsheets tables manually, as opposed to the HTML tables, where some rows could have been identified as data rows or header rows, being in fact non- relational rows (Table 7).

Table 7. Results for HTML and Spreadsheets tables

Row class	Precision	Recall	F1-Score	Support
HTML				
D	0.966	0.982	0.970	2,496
H	0.955	0.992	0.970	17
N	0.980	0.980	0.970	92
Spreadsheets				
D	0.997	0.985	0.994	39,08
H	0.969	0.993	0.983	49
N	0.985	0.965	0.974	5

Note: row labels are as in Table 4:
D: Data rows
H: Header rows
N: Non-relational metadata (a note, clarification, etc.)

3.3.2. Cross Validation

Validation is the process of deciding whether the numerical results quantifying hypothesize between variables are acceptable as descriptions of the data this is a helpful process when there is not enough data to train a model and there is a large imbalance in the number of objects in each class. Then, was applied a k-fold strategy known as stratified k-fold, which is a slight variation in the k-fold cross-validation technique, such that fold contains approximately the same percentage of samples of each target class as the complete set.

Table 8 shows the results obtained for both datasets. We observe that for case of HTML tables the best results were achieved for k=2 and k=3 and that the average accuracy was 0.958 and for spreadsheets tables each k=1...,5 is similar and that the average accuracy was 0.997.

Table 8. Accuracy of cross-validate method for HTML and spreadsheets tables

HTML					
Stage	**K = 1**	**K = 2**	**K = 3**	**K = 4**	**K = 5**
Test	0,92	0,98	0,98	0,94	0,97
spreadsheets					
Stage	**K = 1**	**K = 2**	**K = 3**	**K = 4**	**K = 5**
Test	0,997	0,998	0,996	0,998	0,996

3.3.3. Confusion matrix

As in any classification problem, there are aspects that may be improved. In our experiments, we have to examine the rows in each class that were confused with rows in another class. We then used a confusion matrix, as shown Figure 9 and Figure 10. Each cell of the matrix shows the percentage of all classified rows that were actually of the class with the label shown in the first column, but which the classifier assigned the row label shown in the first row. The shaded cells with blue color stronger in the diagonal show correct row classifications, while the remainders show incorrect classification.

Ideally, our classifier would result in zeroes for the values off the diagonal. However, our model indeed misclassified rows. In the case of spreadsheets tables, we observed that, for both data rows and header rows, erroneous results were obtained with respect to the non-relational rows, that is, a considerable number of data rows and header rows were identified as non-relational rows. In the HTML tables, the erroneous results for non-relational rows was larger than for spreadsheet tables, being 7.9% and 6.6% for data rows and header rows, respectively.

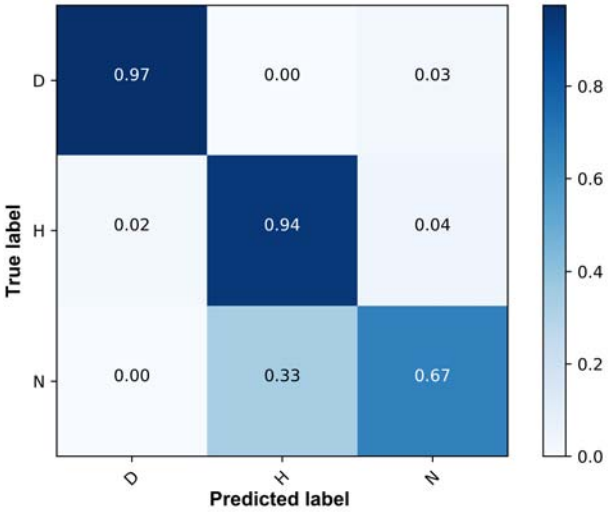


Figure 9. Confusion matrix for spreadsheets tables

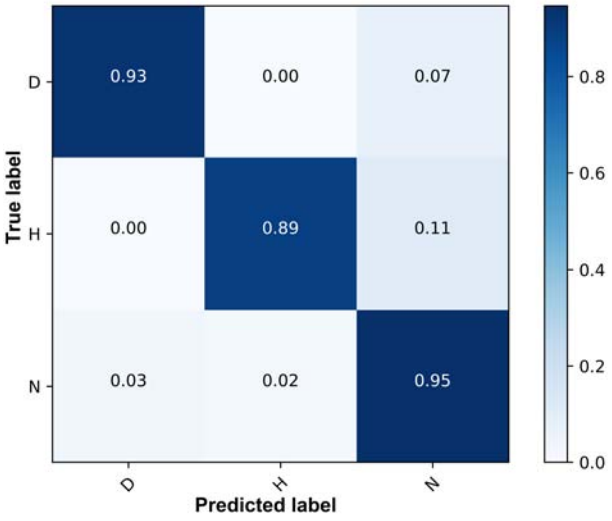


Figure 10. Confusion matrix for HTML tables

This deserves some explanation: (1) the difference between the average number of rows of the HTML tables and the average number of rows of spreadsheets tables; (2) in our classification process, a given row is classified as "non-relational metadata" when the row cannot be identified as data or header; (3) the spreadsheets tables have a better definition in term of features, for example, tables depend on properties encapsulated into CSS files.

3.4. Rule-based Table Detection Algorithm

The rule-based detection algorithm was applied to worksheets in a sample set that contained tables with different layouts and embedded charts. Table 9 summarizes the results obtained, which analyzed a total of 1,000 spreadsheet documents, detected 1,481 tables and misclassified 141.

Table 9. Results for the detection rule-based algorithm

Spreadsheet document	**1000**
Tables	1481
Tables misclassified	141
Single Table	700
Multi Table	158

The algorithm failed for multi-tables with internal separators that are less than the thresholds defined. In that case, the algorithm would consider the two tables as a single table. Also, would not recognize tables correctly when the table cells do not have attributes or separators (e.g, a table with no borders, not font formatting, no background colors, and no empty rows that separate headers and table title) and did not discover tables where the number of empty cells to the right and left is extremely large.

3.4.1. Experiments with the Remove and Conquer Table Detection Algorithm

The Remove and Conquer algorithm were applied to the same dataset. This algorithm detected tables that could not be recognized by the rule-based algorithm and maximized the match between a proposed table P and a true table T, which is equivalent to maximizing the number of cells that they have in common and minimizing the number of cells by which they differ. Table 10 shows the results if we compare with Algorithm 1, we can observe that the number of tables misclassified decreased and the number of multi-tables detected increased.

Table 10. Table recognized through of RAC

Spreadsheet document	1000
Tables	1481
Tables misclassified	141
Single Table	650
Multi Table	230

3.5. Results of the algorithms and environment description

Before entering details about the execution times of the algorithms let's explain the characteristics main of the environment: Portable Computer (PC) Lenovo 80YH model with 8 Gigabytes of Random Memory. Processor Intel(R) Core i7-7500 with 2.70 GHz. Board Graphic Intel(R) 620 with 128 of Memory. Operating System Windows 10 Home of 64 Bits. Table 11 shows the execution times each of the Algorithms.

Table 11. Execution Time to the Tables Recognition Algorithms

Algorithm	**Execution time (s)**	**CPU (%)**	**Memory(%)**
Remove and Conquer	114,28	4,3	1,3
Rule-based	69,19	3,7	1
Conditional Random Field	376,57	11,5	2,5

4. Conclusions

In this work we have described the implementation of three algorithms to classify rows of a table and recognize tables in spreadsheet documents respectively. We performed experiments to test the performance of the table row classifier using HTML and Spreadsheet tables. The experiments show that the classifier obtained excellent results for both kinds of tables. Also was applied a k-fold cross-validation where were obtained results similar to the other experiments reported in [20].

To summarize, the contributions of this work were:

- A table row classifier, applicable to both HTML and spreadsheet tables.
- Experiments to validate the classifier.
- Two datasets containing annotated HTML and Spreadsheets tables to train and validate table row classifiers.
- The implementation of two algorithms for table recognition in spreadsheet documents.

As future work, we propose to increment the number of instances and classes in our datasets and add more complex features. We expect that CRF can also be applied to other non-tabular classification tasks involving content of various formatting and layouts. In general, CRF may help constructing generic information extraction systems.

References

[1] M. Yakout, K. Ganjam, K. Chakrabarti, and S. Chaudhuri, "Infogather: Entity augmentation and attribute discovery by holistic matching with web tables," in *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 97–108. [Online]. Available: https://doi.org/10.1145/2213836.2213848

[2] M. J. Cafarella, A. Halevy, D. Z. Wang, E. Wu, and Y. Zhang, "Webtables: Exploring the power of tables on the web," *Proc. VLDB Endow.*, vol. 1, no. 1, pp. 538–549, Aug. 2008. [Online]. Available: https://doi.org/10.14778/1453856.1453916

[3] E. Koci, M. Thiele, O. Romero, and W. Lehner, "Table identification and reconstruction in spreadsheets," in *Advanced Information Systems Engineering*, E. Dubois and K. Pohl, Eds. Cham: Springer International Publishing, 2017, pp. 527–541.

[4] P. Venetis, A. Halevy, J. Madhavan, M. Paşca, W. Shen, F. Wu, G. Miao, and C. Wu, "Recovering semantics of tables on the web," *Proc. VLDB Endow.*, vol. 4, no. 9, pp. 528–538, Jun. 2011. [Online]. Available: https://doi.org/10.14778/2002938.2002939

[5] G. Limaye, S. Sarawagi, and S. Chakrabarti, "Annotating and searching web tables using entities, types and relationships," *Proc.*

VLDB Endow., vol. 3, no. 1–2, pp. 1338–1347, Sep. 2010. [Online]. Available: https://doi.org/10.14778/1920841.1921005

[6] T. F. Varish Mulwad and A. Joshi, "Generating Linked Data by Inferring the Semantics of Tables," in *Proceedings of the First International Workshop on Searching and Integrating New Web Data Sources*, September 2011, co-located with VLDB 2011. [Online]. Available: https://bit.ly/3p8s1q0

[7] A. S. Corrêa and P.-O. Zander, "Unleashing tabular content to open data: A survey on pdf table extraction methods and tools," in *Proceedings of the 18th Annual International Conference on Digital Government Research*, ser. dg.o '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 54–63. [Online]. Available: https://doi.org/10.1145/3085228.3085278

[8] B. Yildiz, K. Kaiser, and S. Miksch, "pdf2table: A method to extract table information from pdf files." [Online]. Available: https://bit.ly/3k2ejBa

[9] Y. Liu, P. Mitra, and C. L. Giles, "Identifying table boundaries in digital documents via sparse line detection," in *CIKM '08*, 2008. [Online]. Available: https://bit.ly/369nWcm

[10] T. Kieninger, "Table structure recognition based on robust block segmentation," 1998, pp. 22–32. [Online]. Available: https://bit.ly/38k4YT9

[11] M. Zhang and K. Chakrabarti, "Infogather+: Semantic matching and annotation of numeric and time-varying attributes in web tables," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 145–156. [Online]. Available: https://doi.org/10.1145/2463676.2465276

[12] Z. Zhang, "Towards efficient and effective semantic table interpretation," in *The Semantic Web – ISWC 2014*, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, and C. Goble, Eds. Cham: Springer International Publishing, 2014, pp. 487–502. [Online]. Available: https://doi.org/10.1007/978-3-319-11964-9_31

[13] H. Masuda and S. Tsukamoto, "Recognition of html table structure," 2004. [Online]. Available: https://bit.ly/3p8xL2Q

[14] J. Fang, P. Mitra, Z. Tang, and C. L. Giles, "Table header detection and classification," in *AAAI*, 2012. [Online]. Available: https://bit.ly/2IcT3vy

[15] D. Pinto, A. McCallum, X. Wei, and W. B. Croft, "Table extraction using conditional random fields," in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, ser. SIGIR '03. New York, NY, USA: Association for Computing Machinery, 2003, pp. 235–242. [Online]. Available: https://doi.org/10.1145/860435.860479

[16] I. A. Doush and E. Pontelli, "Detecting and recognizing tables in spreadsheets," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, ser. DAS '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 471–478. [Online]. Available: https://doi.org/10.1145/1815330.1815391

[17] E. Koci, M. Thiele, W. Lehner, and O. Romero, "Table recognition in spreadsheets via a graph representation," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, 2018, pp. 139–144. [Online]. Available: https://doi.org/10.1109/DAS.2018.48

[18] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289. [Online]. Available: https://bit.ly/3lbW1yE

[19] J. L. Solé, *Book review: Pattern recognition and machine learning. Cristopher M. Bishop. Information Science and Statistics.* Springer, 2007. [Online]. Available: https://bit.ly/3l7doRq

[20] M. D. Adelfio and H. Samet, "Schema extraction for tabular data on the web," *Proc. VLDB Endow.*, vol. 6, no. 6, pp. 421–432, Apr. 2013. [Online]. Available: https://doi.org/10.14778/2536336.2536343

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2020.06
pISSN: 1390-650X / eISSN: 1390-860X

Analisys of the Efficiency of a Conventional Ventilated Brake Disc Compared to a Hyperventilated Disc by Machining

Análisis de la eficiencia de un disco de freno convencional ventilado con respecto a un disco hiperventilado mediante mecanizado

Vicente Rojas[1,*], Johnny Pancha[2], Vicente Romero[3], Jorge Lema[3]

Received: 06-04-2020, Reviewed: 22-09-2020, Accepted after review: 17-10-2020

Abstract

The objective of this study is to redesign a ventilated brake disc as a hyperventilated disc to compare the differences in temperature and brake distances; for this purpose, a monitoring system was installed in the vehicle which consisted of the implementation of two temperature sensors located near each disk and a data acquisition card. With the implementation of hyperventilated discs, the temperature generated by the braking friction could be reduced. The temperature values that occur between the discs when braking was monitored. To obtain the temperature values of the brake discs, road tests were conducted with different types of discs: ventilated discs on the two front wheels, hyperventilated discs on the two front wheels and mixed discs. In the mixed discs there is a ventilated disc on the right front wheel and a hyperventilated disc on the left front wheel.

Resumen

Este estudio tiene como finalidad rediseñar un disco de freno ventilado a un disco hiperventilado para comparar las diferencias de temperatura y distancias de frenado, para ello se instaló un sistema de monitoreo en el vehículo, que consistió en la implementación de dos sensores de temperatura ubicados cerca de cada disco y una tarjeta de adquisición de datos. Con la implementación de los discos hiperventilados se pudo disminuir la temperatura generada por la fricción del frenado. Se realizó el monitoreo de valores de temperaturas que se producen entre los discos al momento de frenar. Para obtener los valores de temperaturas de los discos de frenos, se realizaron pruebas de ruta con diferentes tipos de discos: discos ventilados en las dos ruedas frontales, discos hiperventilados en las dos ruedas frontales y discos mixtos. En los discos mixtos van un disco ventilado en rueda delantera derecha y uno hiperventilado en rueda delantera izquierda.

[1,*]Universidad Politécnica Salesiana, Ecuador. Corresponding author ✉: erojas@ups.edu.ec. http://orcid.org/0000-0001-5658-3055.
[2]Escuela Superior Politécnica de Chimborazo, Ecuador. http://orcid.org/0000-0001-7320-2154.
[3]Universidad Castilla La Mancha, España. http://orcid.org/0000-0003-2317-7071
[4]Universidad Tecnológica Indoamérica, Ecuador. http://orcid.org/0000-0002-1515-4526

Suggested citation: Rojas, V.; Pancha, J.; Romero, V. and Lema, J. (2021). «Analisys of the Efficiency of a Conventional Ventilated Brake Disc Compared to a Hyperventilated Disc by Machining». Ingenius. N.° 25, (january-june). pp. 62-69. doi: https://doi.org/10.17163/ings.n25.2020.06.

By displaying time data of the brake discs, it was possible to conclude that hyperventilated discs have better heat dissipation, since they have better ventilation. From all the results obtained on the route tests, it was possible to visualize the temperature behavior in the discs at the moment of braking and it was evidenced that hyperventilated discs tend to heat up less than normal discs, thus leading to a decrease in time and stopping distance.

Keywords: Brake Disc, Ventilated Disc, Hyperventilated Disc, Disc Temperature.

Mediante visualización de datos en tiempo en los discos de freno se pudo concluir que los hiperventilados poseen una mejor disipación de calor ya que presentan una mejor ventilación. A partir de todos los resultados de las pruebas de rutas obtenidas, se pudo visualizar el comportamiento de temperatura en los discos al momento de frenar y se evidenció que los discos hiperventilados tienden a calentarse menos que los normales, llevando con esto a disminuir el tiempo y distancia de frenado.

Palabras clave: disco de freno, disco ventilado, disco hiperventilado, temperatura de disco

1. Introduction

As a vehicle circulates, its state varies continuously, i.e., it accelerates, brakes or turns. These phenomena are produced by a large number of forces, and their sum is known as vehicle dynamics. If the sum of all forces is zero, it means that the vehicle is at rest. If it is different than zero, it will be moving. Likewise, all these forces vary as a function of a physical magnitude called acceleration, responsible for modifying the speed and direction of any object. For example, accelerating the vehicle corresponds to a positive acceleration, and the case of braking corresponds to a negative acceleration [1].

The braking system is, undoubtedly, the most important component for the road safety of the automobile, since the total or partial stopping of the vehicle and, consequently, the integrity of its passengers, rely on it. In general, 70% of the kinetic energy produced during movement is absorbed by the front disc brakes, and the remaining by the rear disc brakes, which are often drums. These systems are based on the friction to stop the movement of the vehicle, having as operating principle the hydraulic pressure that pushes the brake pads against the cast iron disc. As a consequence, the behavior produced by this type of devices, through the kinetic energy, is creating a considerably high heat during braking, thus increasing the temperature by friction; this heat is quickly dissipated by the surrounding air by means of the convection phenomenon (heat transfer produced between masses at different temperature) [2].

The geometrical characteristics of the discs depend on the load capacity and on the operation, which are important factors at the initial design stage. In most cases, the design of the discs should avoid the overheating that arises between the brake and the pad due to the effect of friction, by appropriate selection of the physical, mechanical and chemical properties that are produced because in some cases the types of materials do not behave correctly, thus having negative effects on the effectiveness of the braking process. At the mechanical design stage in ventilated disc brakes, it is very important to analyze the behavior of the associated thermofluids (surrounding air), i.e. observing the characteristics and operation of the fluids on the disc surface, always guaranteeing the effectiveness of the braking process and the heat dissipation through the surface and the ventilation channels [3].

During the braking process, the heat produced by the friction between the brake pads and the disc is not dissipated quickly, which depends on the geometry and on the manufacturing material. Therefore, when a very hard brake is produced on the disc brake, large amount of heat may be accumulated in a short period of time, which causes high temperature gradients. In these conditions, the functionality and safety of the braking system may be compromised [4]. The disc brakes have been widely used in vehicles due to their correct behavior since they absorb 70% of the kinetic energy produced during movement, which has been its main advantage with respect to the drum brake. When repetitive brakes are produced in the brake discs the temperature increases due to the friction with the pad, thus generating temperature gradients. This heat is quickly dissipated by the surrounding air by means of the convection phenomenon (heat transfer produced between masses at different temperature). The high temperatures may cause vaporization of the brake fluid, brake wear, bearing failure, thermal cracks (fading) and vibrations. For this reason, on many occasions the performance of the system is reduced; as a result, it is very important to predict the behavior of the types of convection existing in the dissipation of heat to the environment, with the purpose of evaluating its efficiency taking into account its design and initial geometry [5].

The brake system is an essential safety system to avoid any accident when driving all types of vehicles. The problem existing in these systems is due to the overheating of its components, and thus it is indispensable to monitor the behavior of the temperature in the brake discs [6]. One of the indispensable factors for analysis and study are the temperature changes when activating the brakes. In recent times, with the technological growth, ventilated discs are being implemented, which help to reduce overheating and avoid traffic accidents to a large extent [7].

The use of these modern systems is only applied in high range vehicles, due to their greater cost with respect to a conventional system [8]. The objective of the implementation of ventilated discs in the present work is to improve the dissipation of the heat produced during the braking action, at a lower and accessible cost.

2. Materials and methods

2.1. Design of the disc

The software Solidworks was used for designing the hyperventilated disc [9], starting from a normal brake disc of a Chevrolet Dmax 4x4 vehicle.

The main axes were molded on the initial planar face, as can be seen in Figure 1, on which there will be holes and slots, key parts for developing the project, which were drawn taking into account the geometry and shape of the disc, on which we will work afterwards.

The dimensions of holes, slots, depths and distances, detailed in Figure 2, were chosen according to the criteria of the authors, which will be centered on the previously described axes.

The initial molding detailed in Figure 3 was performed considering the dimensions and details of the brake disc, with the purpose of visualizing its shapes and geometries in the software.

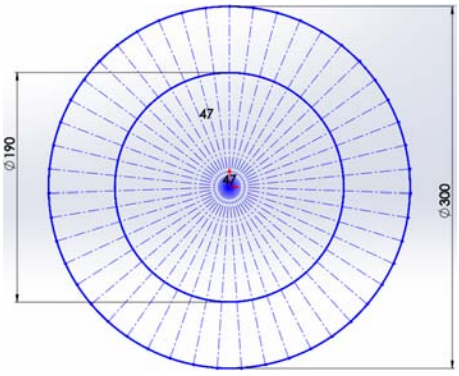


Figure 1. Delineation of the main geometrical axes

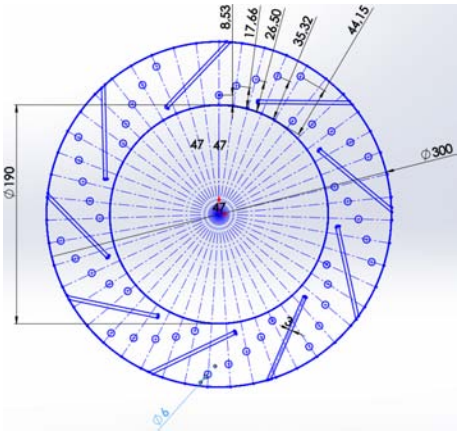


Figure 2. Drawing of holes and slots

Figure 3. Molding of the ventilated disc [3]

2.2. Machining of the brake disc

In this part, the modifications previously designed with the software were machined to the normal disc. All geometrical data were sent to a CNC machining tool, in this case the milling machine, which enabled achieving a greater precision and facilitating the work. It was necessary to export all planes previously made in Solidworks, to the software Mastercam, Figure 4, that controls the CNC milling machine, a very practical software, and besides it is compatible with Solidworks [10].

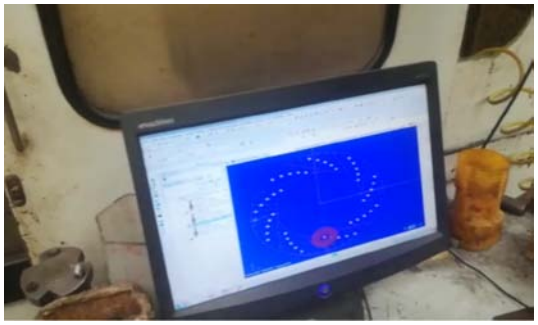


Figure 4. Allocation of the points to be drilled on the disc

After fixing the working coordinates, it was proceeded to drill the through holes that will ventilate the brake disc. Figure 5 shows the drilling and the execution of the G code.

Figure 5. Drilling of the brake disc

After drilling, it was executed the software for grinding the slots that will be employed for ventilation and for removing the chips in the disc. The depth and thickness of the slots were determined by discretion of the authors. Care was taken of not compromising the disc thickness; it was chosen a depth of 2.5 mm and thickness of 3 mm, as can be seen in Figure 6.

Figure 6. Grinding of ventilation slots

2.3. Implementation of hardware and sensors

The amount of hardware currently available in the market is significant, which enables obtaining a very important level of flexibility when searching for an appropriate configuration [11].

The Arduino card was chosen as processor for temperature monitoring and sensors (MLX90614), for acquiring the temperature data from the brake discs, as shown in Figure 7 and Figure 8, respectively.

Figure 7. Arduino card

Figure 8. Infrared temperature sensor

2.4. Programming of the data acquisition software

The brake discs temperature monitoring system consists of temperature sensors, Arduino electronic board, serial communications cable and a notebook. The software LabVIEW 2017 [12] was used for programming and for assigning the parameters.

The graphical screens were modified and renamed according to the requirement of the project temperature monitoring system, resulting in the graphical screen (Figure 9).

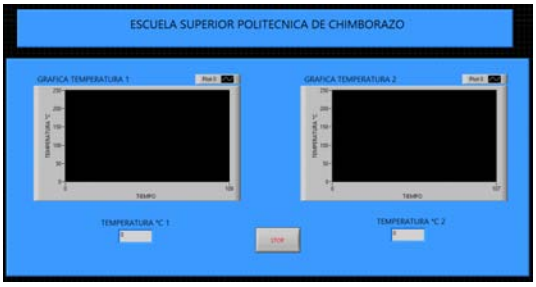


Figure 9. Graphical screen of temperature

2.5. Implementation of discs, sensors and data acquisition system

For the correct installation of the brake discs, it is required: piston retracting tool, comparator clock, torque wrench and other tools, as seen in Figure 10.

Figure 10. Implementation of the ventilated disc

The data acquisition system for monitoring the brake discs temperature was installed using the Arduino card and distance temperature sensors (MLX90614).

For easy access and manipulation, the Arduino card was located inside the cabin of the automobile (Figure 11).

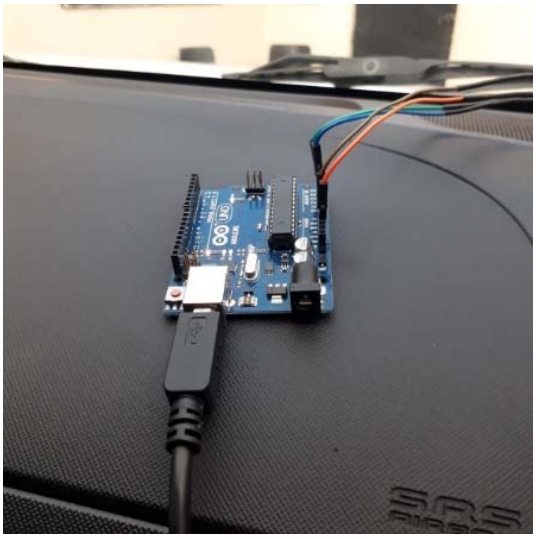


Figure 11. Location of the Arduino

The temperature sensors are installed very close to the ventilated discs, located in the protecting sheet of the brake disc which was used a support base for mounting the sensors, as seen in Figure 12.

Figure 12. Location of the temperature sensors

3. Results and discussion

Initially, the temperature of the original brake discs of the Dmax 4x4 was monitored, in an urban and interurban circuit.

Afterwards, it was carried out the acquisition of temperature data corresponding to the redesigned ventilated brake discs; the test was conducted in the same route previously selected.

At last, the temperature data were monitored, installing an original disc in the left front wheel and a hyperventilated disc in the right front wheel. In this section the temperature data with normal discs were acquired, at different vehicle speeds, in the selected route (Tambo-Cañar).

Figure 13 details the temperature data monitored with hyperventilated brake discs, obtaining a range of working temperatures between 80 °C and 100 °C.

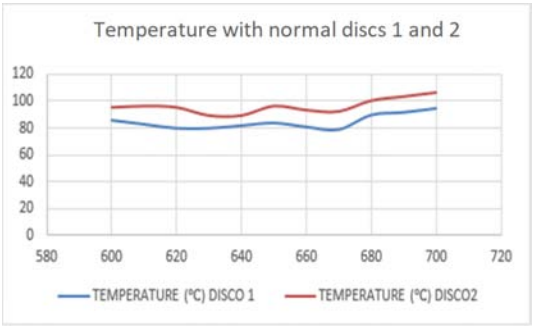


Figure 13. Temperature of the ventilated discs

For acquiring the temperature data with the original discs, the tests are conducted in the same route selected (Tambo-Cañar), and the results obtained with ventilated discs are shown in Figure 14. The obtained range of working temperature is between 90 °C and 130 °C.

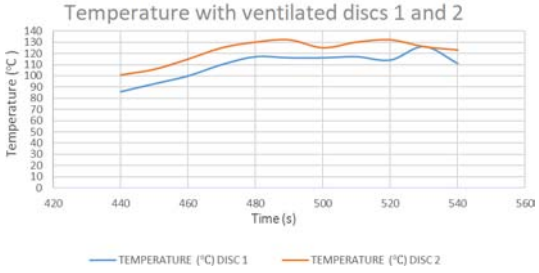


Figure 14. Temperature of the hyperventilated discs

Test of braking distance with original discs

The vehicle was tested with normal discs, under the conditions shown in Table 1.

Table 1. Test conditions and results for braking distance with conventional discs

Initial speed (km/h)	**100**
Initial time (s)	1637
Initial temperature (°C)	64
Final speed (km/h)	0
Final time (s)	1641
Final temperature (°C)	82
Braking distance (m)	37,5
Braking time (s)	4

Braking test with hyperventilated discs

Hyperventilated discs were installed in the test vehicle, and tests were conducted under the conditions shown in Table 2.

Table 2. Test conditions and results for braking distance with hyperventilated discs

Initial speedl (km/h)	100
Initial time (s)	178
Initial temperature (℃)	39
Final speed (km/h)	0
Final time (s)	181
Final temperature (℃)	55
Braking distance (m)	25
Braking time (s)	3

At last, the two previous tests of braking distance with original and hyperventilated discs were compared, and the results are shown in Figure 15.

To make the comparison plots, the initial and final times were discarded since they are associated to the period during which the program is run, and thus they are not significant values for comparing the braking distance.

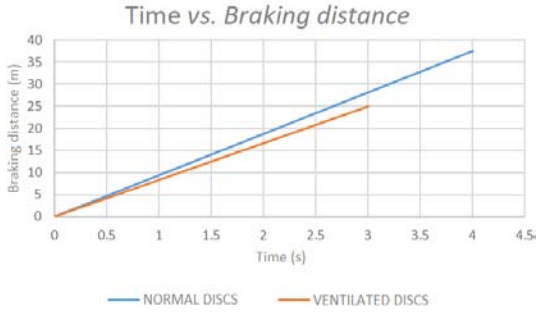


Figure 15. Comparison of braking distances of the original and hyperventilated discs

A braking distance of 37 m was obtained with the original discs, and a braking distance of 25 m with the hyperventilated discs. Considering the computational load and the operating time for mechanizing, the redesign cost with respect to the original disc represents a 30% increase of the market value.

4. Conclusions

Thanks to the Solidworks software the redesign was performed in an easy and precise manner, and thereby the disc could be further mechanized with the help of a CNC milling machine to obtain exact and reliable redesign results.

The system for monitoring the braking discs temperature enabled visualizing, in real time, the data of heat generated by the thermal loads when the vehicle is braked.

The ventilated brake discs dissipated the heat more effectively than normal brake discs, achieving better efficiency, less braking time and distance, and greater driving safety.

The redesign cost has an increment of 30% with respect to the original disc, justifying this value with the data of braking time and distance obtained in this study.

For future studies it may be analyzed the concentration of axial stresses in the modified disc, with the purpose of verifying its durability and useful life time.

References

[1] L. G. Ayala Ayala and J. P. Vallejo Orbe, "Adaptación de un sistema de frenos ABS a un vehículo Fiat, para mejorar la seguridad del frenado," 2013. [Online]. Available: https://bit.ly/3l8an3l

[2] R. García-León, M. Acosta, and E. Solano, "Análisis del comportamiento de los frenos de disco de los vehículos a partir de la aceleración del proceso de corrosión," *Tecnura*, vol. 19, 07 2015. [Online]. Available: http://doi.org/10.14483/udistrital.jour.tecnura.2015.3.a04

[3] R. A. García-León, R. D. Echavez Díaz, and E. Flórez Solano, "Análisis termodinámico de un disco de freno automotriz con pilares de ventilación tipo NACA 66-209," *INGE CUC*, vol. 14, no. 2, pp. 9–18, 2018. [Online]. Available: https://doi.org/10.17981/ingecuc.14.2.2018.01

[4] R. A. García León, "Estudio térmico en tres frenos de disco ventilados, utilizando el análisis de elementos finitos," *DYNA*, vol. 84, no. 200, pp. 15–30, 2017. [Online]. Available: http://dx.doi.org/10.15446/dyna.v84n200.55663

[5] R. A. García León and E. N. Flórez Solano, "Estudio analítico de la transferencia de calor por convección que afectan a los frenos de disco ventilados," *Tecnura*, vol. 20, pp. 15–30, mar. 2017. [Online]. Available: https://doi.org/10.14483/22487638.11676

[6] E. Águeda Casado, T. Gómez Morales, and J. Martín Navarro, *Sistemas de transmisión de fuerzas y trenes de rodaje*, ser. Ciclos formativos. Ediciones Paraninfo, 2012. [Online]. Available: https://bit.ly/2IcXcj6

[7] Aficionados a la Mecánica. (2020.) Ingeniería automotriz. [Online]. Available: https://bit.ly/3k7EzKC

[8] Y. Carranza Sánchez and R. Beltrán, "Transferencia de calor de estado inestable en forros para frenos," 07 2003. [Online]. Available: https://bit.ly/354APFf

[9] F. Kreith, M. S. Bohn, and R. M. Manglik, *Principios de Transferencia de Calor*. Cengage Learning Editores, 2012. [Online]. Available: https://bit.ly/36dOPfh

[10] F. P. Incropera and D. P. DeWitt, *Fundamentos de transferencia de calor*. Pearson Educación, 1999. [Online]. Available: https://bit.ly/3ezzLw8

[11] Kashima University. (2016) Curso ingeniería automotriz. [Online]. Available: https://bit.ly/3mXaTRW

[12] D. Wenner, *Manual práctico del cuidado y reparación del automóvil*. México Continental, 1984. [Online]. Available: https://bit.ly/32jIr4P

Scientific Paper Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.07
pISSN: 1390-650X / eISSN: 1390-860X

Evaluation of a FMCW radar as a teaching tool in the automotive and telecommunications engineering careers

Evaluación de un radar FMCW como herramienta didáctica en las carreras de Ingeniería Automotriz y Telecomunicaciones

Pablo J. Mavares F.[1,*]

Received: 10-05-2020, Reviewed: 22-09-2020, Accepted after review: 17-10-2020

Abstract

In recent decades, advanced driver-assistance systems (ADAS) have evolved to be available in much of the vehicles manufactured today; it is very im-portant to keep teaching in this area up-to-date. This paper presents a frequency modulated continuous wave radar that works in the 24 GHz ISM band. The purpose of this work is to evaluate its performance and suitability to be used as a didactic tool in teach-ing in the automotive and telecommunications engi-neering careers with an emphasis on the mathematics and telecommunications subjects, under scenarios feasible to be found in university labs. For this pur-pose, the measurement scenario is described, as well as the hardware, firmware and a generic algorithm implemented in MATLAB based on fast Fourier transforms to obtain Range-Doppler maps that allow, in conjunction with the CFAR algorithm, to improve detection of objects when compared to the detection from a fixed level. The results presented demonstrate that the accuracy and precision of the radar are within the parameters for a short-range radar for vehicles, also finding a tool with great didactic potential with which students can understand today´s applications of mathematics in the field of telecommunications, especially in radars that serve ADAS systems.

Keywords: FMCW radar, Range-Doppler map, CFAR, didactic tool.

Resumen

En las últimas décadas los sistemas avanzados de asistencia al conductor (ADAS) han evolucionado hasta estar disponibles en gran parte de los vehículos fabricados hoy en día; mantener actualizada la enseñanza en esta área es de vital importancia. Este artículo presenta un radar de onda continua modulado en frecuencia que trabaja en la banda de 24 GHz ISM. El propósito es evaluar su desempeño e idoneidad para usarse como herramienta didáctica en la enseñanza en las carreras de Ingeniería Automotriz y de Telecomunicaciones con énfasis en las asignaturas de Matemáticas y Telecomunicaciones con escenarios factibles de encontrar en los laboratorios universitarios. Para ello se describe el escenario de medición, así como el *hardware*, el *firmware* y un algoritmo genérico implementado en MATLAB basado en transformadas rápidas de Fourier para obtener mapas RangeDoppler que permiten junto con el algoritmo CFAR mejorar la detección de objetos al comparar con la detección a partir de un nivel fijo. Se presentan resultados que demuestran que la exactitud y precisión del radar se encuentran dentro de los parámetros para un radar de corto alcance para vehículos, encontrándose, además, una herramienta con gran potencial didáctico, con la cual los estudiantes pueden comprender las aplicaciones que hoy tienen las matemáticas en el ámbito de las telecomunicaciones, especialmente en radares que sirven a sistemas ADAS.

Palabras clave: radar FMCW, mapa Range-Doppler, CFAR, herramienta didáctica

[1]Departamento de Sistemas de Automatización, Instituto Tecnológico El Pacífico – Ecuador.
Corresponding author ✉: pmavares@tecnologicopacifico.edu.ec http://orcid.org/0000-0001-7342-6262

Suggested citation: Mavares, P. J. (2021). «Evaluation of a FMCW radar as a teaching tool in the automotive and telecommunications engineering careers». Ingenius. N.° 25, (january-june). pp. 70-80. doi: https://doi.org/10.17163/ings.n25.2021.07.

1. Introduction

The radar has been traditionally employed in the military and aeronautical industries for several years, mainly due to the high complexity and the high cost associated to its use. However, with the evolution of electronics and integrated circuits (IC) it is becoming more common every day to find radars with specifications appropriate for other tasks, among which it can be mentioned the automotive industry [1]. In particular, in recent years they have been employed in advanced driver-assistance systems (ADAS) [2]. For this purpose, radar systems measure directly or indirectly the position, velocity, and even acceleration of a great variety of objects such as other vehicles, pedestrians or cyclists, thanks to the execution of different algorithms [3].

Although video cameras, ultrasound and other systems are also currently used, radars have the advantage of being little affected or unaffected by environmental conditions such as temperature, illumi-nation, dust, etc. This is why they can operate individually or together with other technologies to guar-antee the high accuracy, precision, reliability and adaptability required by ADAS systems [4].

Besides objects detection, radars have other applications such as the identification of physical condi-tions, in particular the friction coefficient of a road. This can be employed to issue alerts or take actions to avoid accidents, especially under adverse conditions such as rain or ice [5], [6].

In addition, it is possible to use transponders or tags to send information about the users in a road, such as other vehicles or pedestrians, and also information about the driving conditions, in order to avoid accidents and traffic violations [5].

Frequency modulated continuous wave (FMCW) radars have been satisfactorily applied for quite long time for measuring the position and velocity of objects [7].

FMCW radars basically operate at frequencies of 24 and 77 GHz in the automotive industry. For a short range of up to tens of meters it is possible to use a radar of 24 GHz, while for a long range of about 250 m it is used a radar at 77 GHz [3], [5].

Various research works have been conducted from a technical perspective to evaluate different FMCW radar systems, both at 24 and 77 GHZ, for their use in the automotive industry [6–12]. How-ever, in this work it is considered the use of these radars from a different perspective, which besides evaluating their performance also focuses on their potential use as a didactic tool that enables enhancing the understanding of concepts which on many occasions may be quite complex to understand, such as the Fast Fourier Transform (FFT), the radars and their applications in telecommunications and in the automotive industry. For this purpose, it is proposed to use radar evaluation boards considering the benefits in the development of specific and transversal competencies in students.

1.1. Important parameters in an FMCW radar

In order to process and obtain position, velocity and angle it is necessary to consider different equa-tions, whose derivation may be found in [4], [13]. These equations are derived from the following gen-eral idea about the operation of an FMCW radar. The radar transmits a signal with a frequency that var-ies linearly (TX) and with a particular bandwidth, this signal is reflected by a body located in the radi-ated space of the antenna and arrives as a signal to the receiving antenna (RX) with a delay time pro-portional to the distance of the body. These signals are mixed to obtain a signal of intermediate fre-quency (IF), whose frequency is proportional to the distance of the body (Figure 1). Therefore, bodies at different distances will generate IF signals of different frequencies. On the other hand, small differ-ences in position will generate different phases of the IF signals which enables determining the velocity of the bodies [2], [4], [14].

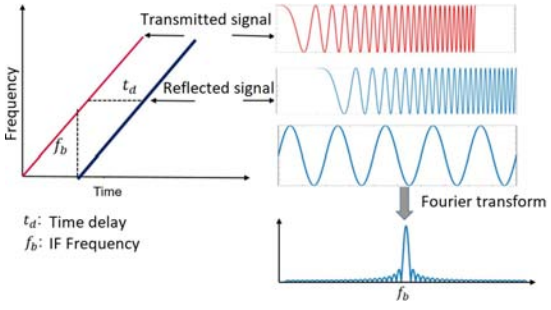


Figure 1. Basic principle of operation of an FMCW radar [15]

The value of the distance to a body is obtained from Equation 1.

$$R = \frac{cT_c f_b}{2B} \tag{1}$$

Where:

R is the distance of the object
c is the speed of light
T_c is the tiempo of the chirp
f_b is the frequency of the IF signal (beat)
B is the bandwidth

The distance resolution or capability to identify two objects close to each other is determined according to Equation 2.

$$\triangle R = \frac{c}{2B} \tag{2}$$

Where:

$\triangle R$ is the resolution
c is the speed of light
B is the bandwidth

The maximum distance may be determined as a function of the sampling frequency, according to Equation 3.

$$R_{max} = \frac{f_s c}{2S} \tag{3}$$

Where:

R_{max} is the range or maximum distance
f_s is the sampling frequency
c is the speed of light
S is the slope of the modulation $S = B/T_c$

On the other hand, the maximum velocity is given by Equation 4.

$$v_{max} = \frac{c}{4 f_c T_c} \tag{4}$$

Where:

v_{max} is the maximum velocity
c is the speed of light
f_c is the frequency of the chirp
T_c is the time of the chirp

Now, in case of having different objects different frequencies will appear and, thus, instead of applying Equation 1, one of the most common techniques is to perform a spectral analysis from the fast Fourier transform (FFT) according to Figure 2, to obtain a Range-Doppler map.

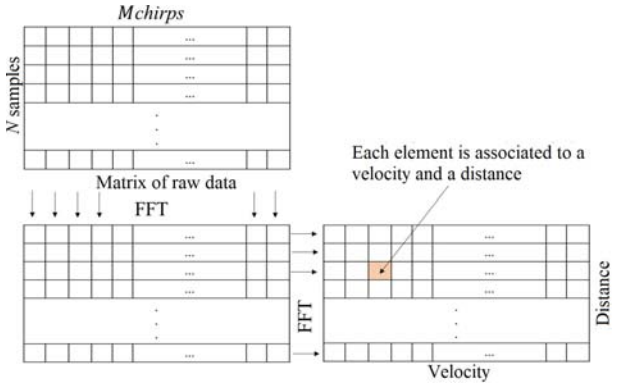


Figure 2. 2D Range-Doppler FFT

The angle of arrival (AoA) may be obtained from a 3D fast Fourier transform (3D FFT), such that the peaks of the frequency spectrum correspond to a particular angle, which is normally carried out when there are four receiving antennas and one or more transmitting antennas to obtain a multiple-input multiple-output (MIMO) array [8]. When the circuit has only two receiving antennas it is preferable to obtain the AoA from Equation 5.

$$\alpha = sin^{-1} \frac{\lambda \triangle \varphi}{2\pi d} \tag{5}$$

Where:

α is the angle of arrival
λ is the wavelength
$\triangle \varphi$ is the phase difference between the signal of the antennas
d is the distance between the antennas

1.2. Applications of radars in the automotive industry

In recent years radars have been employed in vehicles mainly for safety reasons, such as ADAS sys-tems, anticipating needs and taking the lead when necessary [16]. In the automotive industry, applica-tions of radars are mainly divided in short range radars (SRR), which enable blind spots detection (BSD), lane change assistance (LCA), cross traffic alert (CTA) in the front and rear areas, lateral impact alert and alert for cyclists in the lateral way (Figure 3). On the other hand, there are medium and long range radars (MRR and LRR) responsible for automatic emergency braking (AEB) in front of collisions with pedestrians and with other vehicles, as well as adaptive cruise control (ACC) [17].

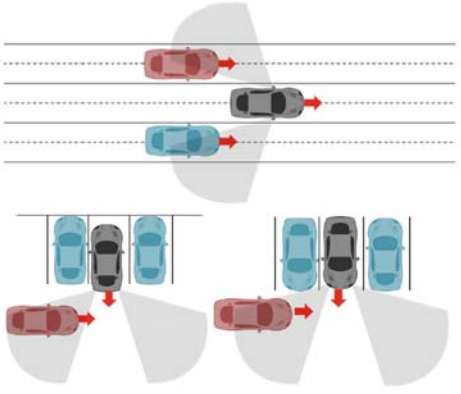


Figure 3. Example of short-range scenario. Detection of blind spots and cross traffic alert

The radar chosen in this work, whose features are detailed later, is a short range one and thus it re-sponds to the typical requirements shown in Table 1.

Table 1. Requirements of short range radar [17]

Requirements of the radar parameters	BSD Requirements	CTA Requirements
Range (m)	1-50	2-60
Range accurancy (m)	± 0,10	pm0,20
Range resolution (m)	0,75	1
Velocity (m/s)	-70 to + 70	-70 to + 70
Velocity accurancy (m/s)	± 0,1	± 0,1
Velocity resolution (m/s)	0,25	0,3
Azimuth (°)	±75	±40
Azimuth accurancy (°)	±5	±5
Azimuth resolution (°)	15	15
Elevation (°)	±6	±10
Elevation accurancy (°)	pm0,1	±5
Elevation resolution (°)	-	-

1.3. Didactic tools and competency-based learning

Among the aspects involved in engineering teaching, lab practices are of vital importance. According to [18], such practices must be capable of providing enough information that enables defining and charac-terizing activities in which students develop specific and transversal competencies.

For example, Arduino has been frequently employed in recent years thank to its ease of use and low acquisition cost [19]. This is something that has influence in countries where there are no adequate re-sources to acquire equipment which may be rather expensive, and therefore using low-cost development boards, in this case radar ones, may generate extensive benefits. Specifically, they enable improving access to devices which, as it has been mentioned, have almost been of exclusive use in the military and aeronautical areas. Similarly, it enables improving access to pioneer technology being developed in high, medium, and even low range vehicles. In this respect, it is important that future engineers and techni-cians are appropriately prepared in Latin America with the knowledge and competencies to be able to perform maintenance to vehicles and, similarly, participate in research and development in these areas, since current vehicles are equipped with a great number of electronic devices [20].

On the other hand, theory teaching benefits from and harmonizes with adequate experimental methodologies, not only in traditional labs (e.g., electronics or electric circuits), but also in those subjects that, traditionally, do not involve experimental activities, such as Mathematics.

2. Materials and methods

At present, different technological companies have brought to the market boards that in principle serve to evaluate the performance of the integrated circuits that they offer, such as the case of Analog Devic-es who offer the Demorad for different chipsets including the ADF5901 (microwave integrated circuit at 24 GHz with 2 transmission channels), ADF5904 (receiver of 4 channels at 24 GHz) and AD4159 (in charge of generating triangular chirps or saw tooth) and other circuits that provide a complete radar system integrated in a board [21]. Infineon also offers the Distance2Go [22] or Position2Go [23] boards, among others.

The Position2Go board has been chosen for this work since it is one of the simplest (from the point of view of the architecture) that can be found in the market, relatively easy to use and at an affordable cost. Its performance is evaluated for obtaining Range-Doppler maps, position and object detection by means of the CFAR (Constant False Alarm Rate) algorithm, which is one of the simplest to implement and understand from the didactic point of view.

2.1. Hardware

The Position2Go board includes all elements necessary for generating, receiving, and processing the signals. It is divided in four important sections: the radio frequency (RF) section includes the CI BGT24MTR12 that is the primary responsible for generating and receiving signals at 24 GHz [23], as well as three antennas, one for transmitting and two for receiving. It also has analog amplifiers that pro-vide an interface between the RF and the digital sections. It includes a frequency control section and a digital section, specifically it has a 32-bit XMC4700 ARM microcontroller to sample and process data, all this mounted on a board that enables access through a USB 2.0 connection that also has a CAN connection in case that a direct communication with an electronic control unit (ECU) of a vehicle is required. Therefore, it may be said that it is a complete radar integrated system that enables its direct control and programming by means of some connection protocols pointed out (USB has been used in this work). Figure 4 shows the general architecture of the board, while Figure 5 shows the board used in this work.

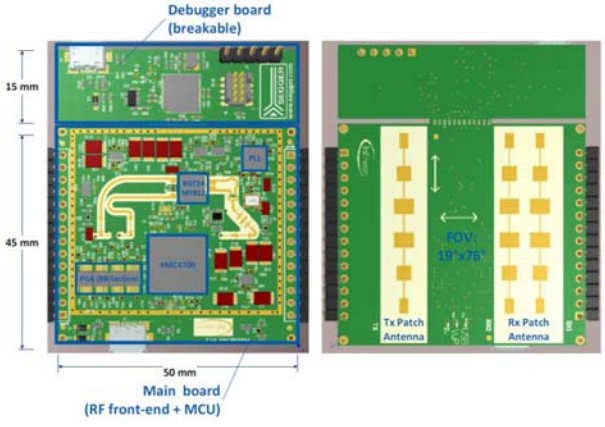


Figure 4. Position2Go Board

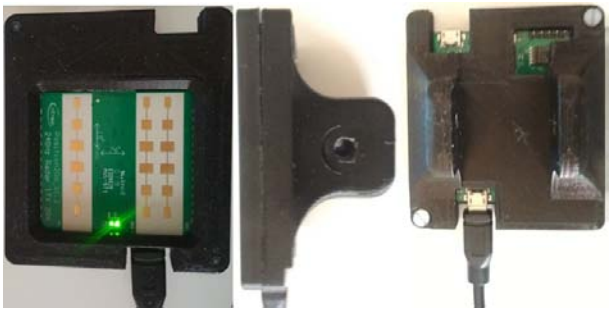


Figure 5. Position2Go board used

Table 2. System parameters

Parameters	Value
Up-chirp time: τup	301 μs
Down-chirp time:τ_{down}	100 μs
Standby time between chirps τ_{sby}	$100_{\mu s}$
Pulse repetition time: $\tau_{up} + \tau_{down} + \tau_{sby}$	$501_{\mu s}$
Bandwidth (B)	301 MHz
Sampling frequency (f_s)	850 MHz
Samples per chirp	256
Chirps per frame	16 (maximum)
Reach resolution ($\triangle R$)	75 cm^*
Minimum reach (R_{min})	0 m
Maximum reach (R_{max})	14 m^{**}

* Theoretical without windows
** Maximum distance to be evaluated

2.2. Experimental design

Position2Go includes a firmware that simplifies the control of the different ICs and peripherals through the XMC4700 microcontroller. This firmware enables performing changes in the configuration of the parameters of interest and provides the signals processed or unprocessed. There is a user interface to directly obtain the signal processed, generating the frequency spectrum and determining the position and velocity of the different objects [24]. This toolbox of Infineon (Figure 6) enables showing the calculations carried out by the microcontroller, taking into account that by default the calculations are performed with a threshold of 100 LSB, which results in a range of approximately 12 m for pedestrians and 15 m for a radar cross section (RCS) of 1 m^2 [25]. However, working with this interface limits the capability of processing the signals, and besides, it does not allow to obtain the didactic advantages sought for (although it may be a starting point). As a result, it is intended to acquire the IF signal and process it directly in MATLAB. For this purpose, an API is available which enables modifying and obtaining the values directly [26]. Table 2 shows the chosen features thanks to different lines of code implemented.

It must be emphasized that the conditions detailed are not typical for evaluation of radars. However, it has been preferred a scenario more quotidian and adaptable to the reality of a university for three reasons.

The first reason is that in the environment of a university lab it is difficult to have access to regulated conditions for a correct evaluation, such as the availability of an anechoic room (Figure 7) and square or triangular reflectors to avoid dispersion in the measurements, and in this way evaluating the accuracy and precision of the radar in an appropriate manner; just as the vehicles for tests, a static vehicle and a pedestrian are more feasible conditions in an educational environment. The second reason is that the FMCW radars at 24 GHz are used for short range and low speeds, as has been detailed previously. At last, the main didactic objective stated is the understanding and implementation of the theoretical as-pects and not of a deep evaluation of the features of the board used.

As an additional aspect it should be mentioned that due to the limitation to free circulation and the confinement during much of 2020 (especially between the months of March and July) and the closure of universities, measurements were carried out in a nonideal environment, due to the impossibility to access the working material that was in the labs.

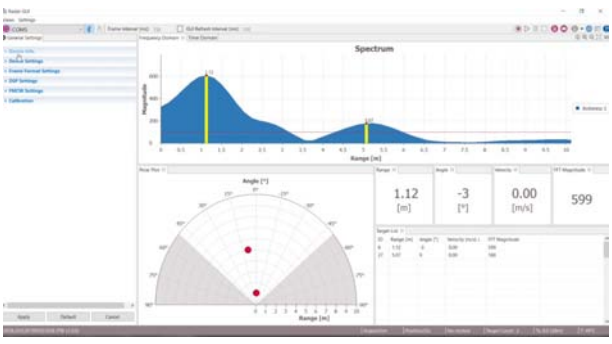


Figure 6. Radar GUI of Infineon

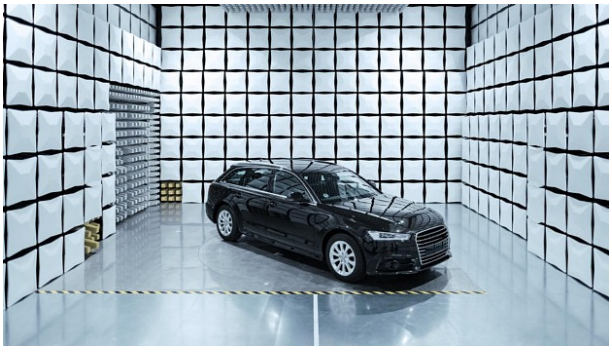


Figure 7. Anechoic room [27]

To carry out the measurements it has been considered the following scenario according to Figure 8, where it has been taken into account a static vehicle at a distance of 2.9 m, a static pedestrian at a distance of 7.5 m, a wall at a distance of 12.3 m, and besides a pedestrian who runs in a straight line from the initial position of the radar up to the wall and returns.

From the configuration detailed in Figure 8, different measurements were carried out considering the data (IF signal) provided by the firmware directly to MATLAB, where each measurement has 50 dataframes, by means of a for loop that can be modified to obtain the desired number. A dataframe consists of the raw data to be processed (IF signal) with different functions implemented in MATLAB.

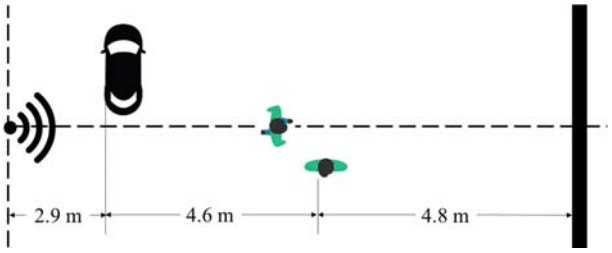


Figure 8. Measuring scenario

2.3. Processing in MATLAB

Data processing may be summarized in 4 levels. First, define the parameters detailed in Table 2 following the methods listed in *26*, calculate the measuring limits from the equations detailed, define the window to be applied, in this case it was chosen a Hamming window with the purpose of reducing the amplitude of the lobes adjacent to the main peak in the FFT, reorganize the data, since the firmware by default sends the data in a three-dimensional array. The first dimension contains all the samples per chirp for estimating the distance; the second corresponds to the different chirps per frame for estimating the velocity and the third corresponds to the antenna; however, it has been chosen to work in two three-dimensional arrays, one for each antenna and substituting the third dimension for the measurement number such that two arrays of size $N \times M \times L$ were obtained, where N is the number of samples per chirp limited in frequency corresponding to the interval of distance (*2,14* meters); M is the number of samples per frame, which corresponds to $M = 16 \times 256 = 4096$; and finally L measurements, being L=50 measurements in this case. The interval between measurements has been 0.2 s.

Then it was applied a first FFT to the first chirp in order to obtain the distance spectrum with a size of 2^{12}, therefore, $2^{12} - N$ elements are filled with zeros using zero padding, which is automatically carried out in MATLAB (see [28]); the purpose is increasing the resolution such that it is simpler to recognize two frequencies close to each other [29]. This results in the so-called Range-FFT.

Afterwards, a new FFT is applied, but this time through the different chirps, in order to obtain the Range-Doppler map.

At last, the CFAR algorithm has been applied to identify the different objects from what is detailed in [30]. This board has the disadvantage of being limited in bandwidth (200 MHz for 24 GHz) and in number of chirps/frame, therefore, the size of the cells for the application of the CFAR must be high, and in this case it was chosen 90×90 of guard and 30×30 of training, with a factor $K = 10^{8/20}$. A flow diagram with a generalization of the algorithm used is shown in Figure 9.

The programming has been developed from the demos of Infineon *23*, Analog Devices *21* and the work by Guerrero *11* applied to other development board. However, these programs have been broadly modified to include:

Capability to export recordings (.AVI) in MATLAB

Capability to create and export polar plots of the position of the bodies.

Capability to isolate and analyze desired dataframes.

Application of CFAR algorithm.

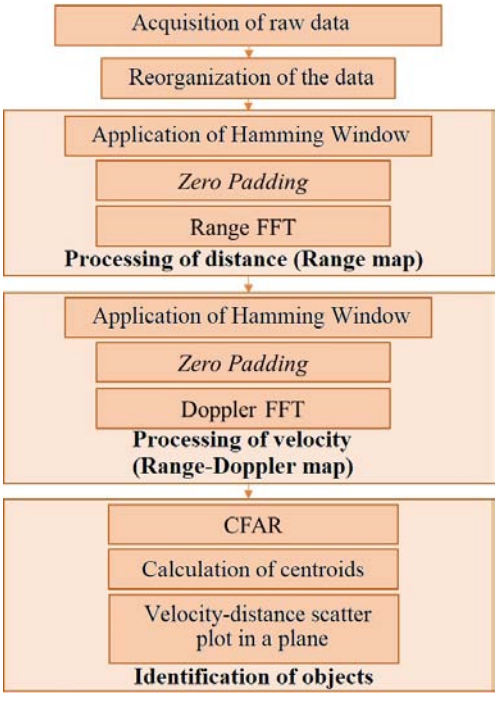


Figure 9. Signal processing diagram

3. Results and discussion

3.1. Firmware and Infineon Toolbox

From the data directly processed by the microcontroller, it was obtained the Range-Doppler map shown in Figure 10, in which it is seen various false detections or misdetections because the threshold is fixed by default as has been detailed previously.

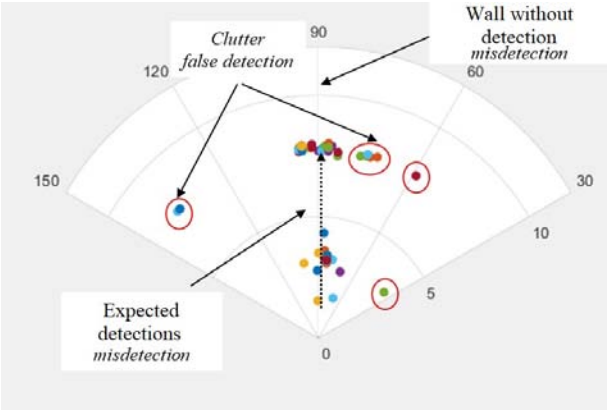


Figure 10. Measurements from the firmware

On the other hand, although this first approximation for stu-dents may seem sufficient, the fact that only processed data are shown limits the objectives sought for when using the board, i.e., it is not evaluated didactically the obtention of the results, but they are only shown without deepening in the algorithm employed for this purpose. Even though a modification of the algorithm and a direct implementation of the CFAR in the microcontroller is possible, its implementation is not simple and for this reason the rest of the results shown were obtained using the MATLAB implementation.

3.2. MATLAB

The Range-Doppler map of Figure 11 has been selected among the 50 maps generated; in this map it is possible to appreciate the bandwidth problem of the radars of 24 GHz of 200 MHz, i.e., that they do not have a good distance resolution, and in addition the chirps per frame of the Position2Go are limited to 16 by default, when in other boards of 77 GHz it may be obtained, for instance, up to 128 chirps per frame.

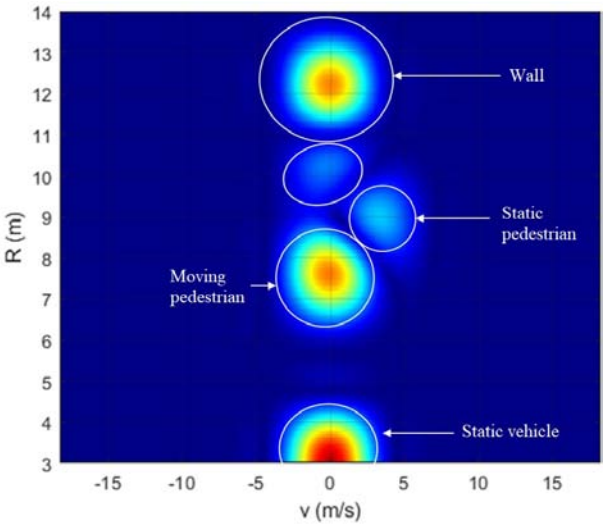


Figure 11. Range-Doppler map

Five areas may be seen in Figure 11 corresponding to relative maxima in the FFT; the identification by means of algorithms that enable automatic discrimination between the different bodies that may be in the environment is out of the scope of this work, particularly because the study has been stated as a tool for undergraduate environments, which implies a limited complexity.

At this point it is important to clarify to the student that although this map enables a fast visual identification, it is not very practical from the computational point of view because a microcontroller by itself would not be capable of identifying the bodies without an additional algorithm in charge of that (one of these algorithms is detailed in-depth in [31]). It is at this point where the application of the CFAR algorithm simplifies to dynamically obtain a detection limit, so that it is considered the signal-to-noise ratio (SNR). It is important to remark that the CFAR by itself is not capable of carrying out the discrimination about the bodies (vehicles, pedestrians, motorcycles, etc.)

Since the K factor has been configured to a relatively low value, this results in many areas when applying the CFAR (Figure 12). To overcome this issue, it has been only taken those points that have a power of at least –40 dB, value empirically obtained from various measurements (for a smaller power it generally resulted in false detections), this value may be also dynamic, considering that the power reflected depends on the radar equation which in turn depends on the distance at which the body is lo-cated; at the same time, the reflection levels are different according to the object, for example, a vehicle reflects more power than a pedestrian, considering that the electromagnetic waves (such as the ones transmitted by a radar) are reflected in the bodies according to the radar cross section which depends on different factors such as the area, the material (of the clothes or of the vehicle), the

shape, etc., and this may be used to be able to classify the bodies. Table 3 shows some of the values obtained from the centroids for each of the areas (Figure 13).

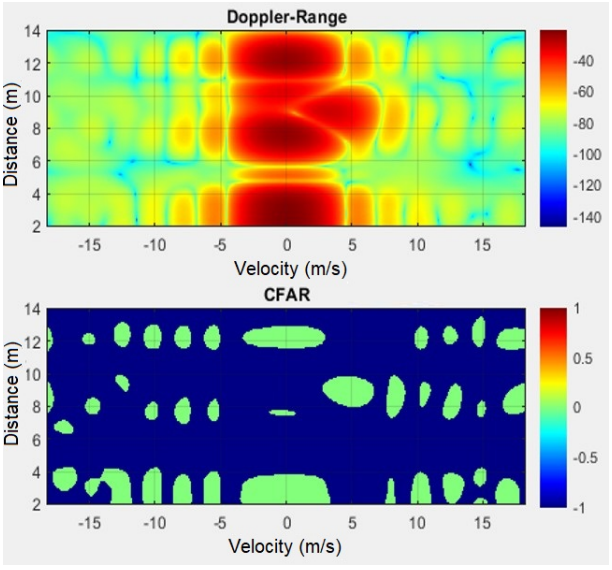


Figure 12. Range-Doppler map in dB and result of CFAR

Figure 14 shows the objects detected considering the –40 dB limit. Then, it is evident that the CFAR enables identifying objects correctly, even when the velocity of the pedestrian has been slightly higher than the real one (Table 4); in this sense, it is a contribution of this work to the detection of ob-jects with respect to the one carried out by the Infineon graphical interface. It is clarified that the results have not been optimized, and in this way, it may be shown conditions in which a student might imple-ment a non-optimized algorithm and even obtain acceptable results.

Table 3. Data sorted considering the power

Distance (m)	Velocity (m/s)	Power (dB)
2,8564	–0,0803	–21,1564
7,5806	–0,2945	–23,2399
12,2498	–0,0803	–23,3293
8,9539	4,8109	–33,7255
2,9114	–5,6320	–53,0894
⋮	⋮	⋮

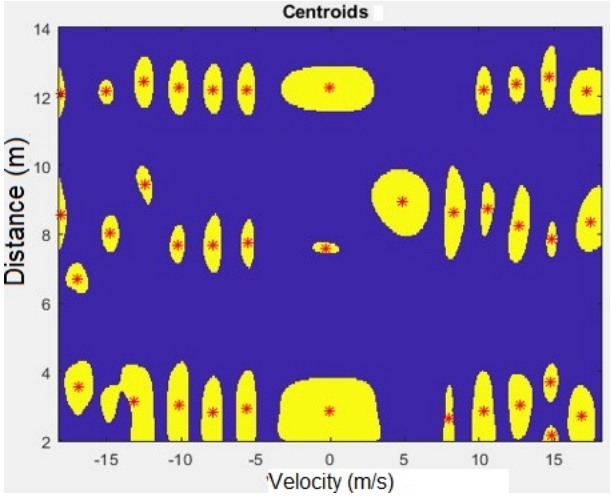


Figure 13. Centroids from the CFAR

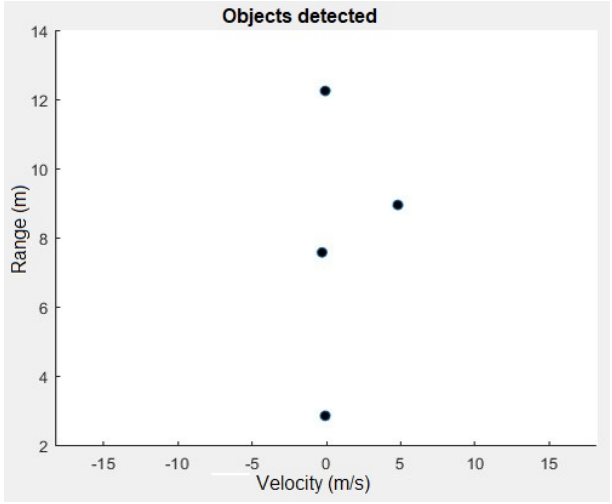


Figure 14. Identification of objects using CFAR

Table 4. Data sorted considering the power

	Measured distance (m)	Real distance (m)	Measured velocity (m/s)	Real velocity (m/s)
Vehicle	2,8564	2,9	–0,0803	0
Static pedestrian	7,5806	7,5	–0,2945	0
Wall	12,2498	12,3	–0,0803	0
Moving pedestrian	8,9539	≈ 9	4,8109	≈ 4

Figures 11 – 14 show one of the frames of one of the measurements that has been chosen; due to the structure of this document all of them cannot be shown. However, Figures 15 and 16 show the distances obtained from the application of the CFAR in each of the dataframes. Figure 15 shows the moving pedestrian, where an effective detection rate of 78.57% is achieved (without considering false detec-tions and misdetections). On the other hand, Figure 17 shows the dispersion for the static bodies. Table 5 details

some statistical values that are within the values of Table 1. However, it is important to remark that the precision and accuracy shown does not necessarily reflect the real values of the radar, due to the aforementioned constraints. The importance of these results lies on the fact that even in non-optimal conditions, the radar achieves good precision and accuracy, and thus it is expected that students yield adequate results without requiring a deep calibration and more complex algorithms.

From these results, the use of this evaluation board emerges as an appropriate tool to be employed in lab practices, because it covers specific competencies: understanding of mathematical fundamentals such as the Fourier transform, complex numbers, operations with matrices, spectral analysis, among oth-ers, as well as fundamentals of physics: kinematics, electromagnetic waves, Doppler effect, etc. On the other hand, with an adequate guide of the teacher it may be promoted transversal competencies accord-ing to [32], such as: computer concepts related to the area of study, specifically it is promoted the use of MATLAB and other computational tools. Similarly, it may be encouraged the work in teams, auton-omous learning, adaptation to new situations, capability of applying theoretical concepts in the practice, use of the Internet as an information source, among others.

All the previously mentioned in an adequate environment, which does not generate frustration on the students and an appropriate tutoring may translate in a great impact in their education.

Unfortunately, no previous case studies were found about the viability of the proposal of using ra-dars as a didactic tool and the evaluation of the impact. Therefore, this is the reason why this work ana-lyzes the viability both technical and didactic of using an evaluation board of the many available in the market.

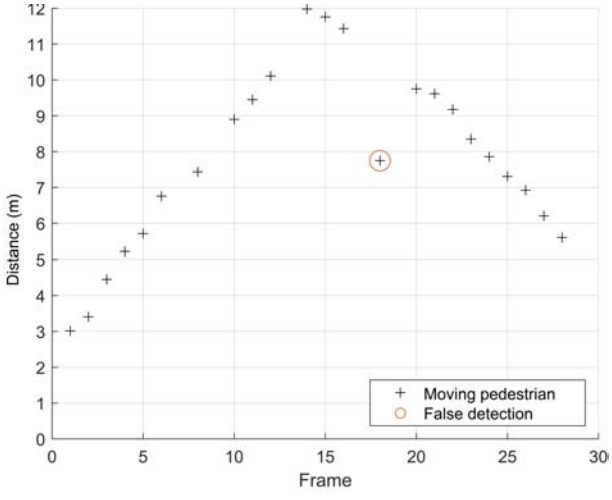


Figure 15. Identification of objects using CFAR

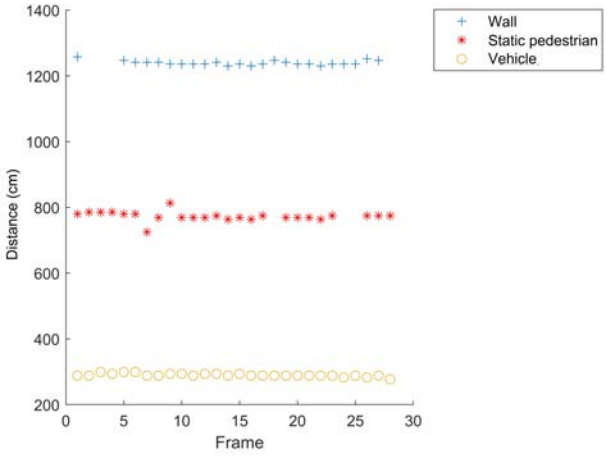


Figure 16. Identification of objects using CFAR

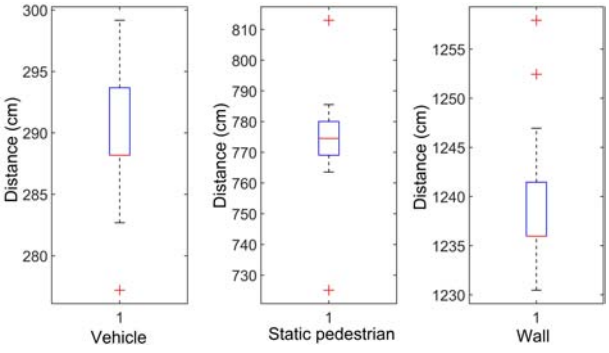


Figure 17. Identification of objects using CFAR

Table 5. Statistics for measuring the distance of the static bodies

Distance			
Real (cm)	**Mean (cm)**	**Median (cm)**	**Standard deviation (cm)**
290	289,75	288,18	4,84
750	773	774,54	14,03
1230	1239,40	1235,96	6,69

4. Conclusions

This work shows an evaluation board based on a FMCW radar which operates at 24 GHz with a bandwidth of 200 MHz. The empirical performance without considering a statistical analysis that enables determining the rate of false detections and without a precise calibration, it is more than acceptable for its use as an educational tool with the purpose of understanding the operation of short-range radars and their application in vehicles. Obtaining and processing data is relatively simple, thus helping in the de-velopment of both the researcher and the teacher who desires it. However, it is recommended to start by the application of an algorithm that involves the two-dimensional fast Fourier transform to obtain Range-Doppler maps, and then to

use a CFAR algorithm for improving the detection of objects with respect to the implementation of a fixed limit. In case it is desired to deepen in the classification of bodies, it is recommended to investigate about the different algorithms available.

Although it has not been compared with other boards, the results obtained and the low cost of the Position 2Go show that implementing its use in lab practices in Automotive and Telecommunication Engineering careers may result in many benefits in the development of specific and transversal capaci-ties of students.

References

[1] V. Jain and P. Heydari, *Automotive radar sensors in silicon technologies*. Springer Science & Business Media, 2012. [Online]. Available: https://bit.ly/3kak6ER

[2] J. Gamba, *Radar Signal Processing for Autonomous Driving*. Springer, 2020. [Online]. Available: https://bit.ly/3k8LP95

[3] C. Kumar and A. Killedar, *mmWave Radar – ADAS Applications*, 2015. [Online]. Available: https://bit.ly/3p6XeK9

[4] M. Nazir, "Automotive radar target detection using ambiguity function," Ph.D. dissertation, University of Birmingham, 2016. [Online]. Available: https://bit.ly/3n2kI1g

[5] P. Pursula and V. Viikari, "Novel automotive radar applications," in *Millimetre Wave Days Proceedings*, 2011, millimetre Wave Days 2011: 6th ESA Workshop on Millimetre-Wave Technology and Applications and 4th Global Symposium on Millimeter Waves; Conference date: 23-05-2011 Through 25-05-2011. [Online]. Available: https://bit.ly/2U5GZi0

[6] N. Bouhlel, S. Meric, C. Moullec, and C. Brousseau, "FMCW radar system for transponder identification," *Progress In Electromagnetics Research B*, vol. 81, pp. 101–122, 2018. [Online]. Available: http://dx.doi.org/10.2528/PIERB18032007

[7] F. Ali and M. Vossiek, "Detection of weak moving targets based on 2-D range-doppler FMCW radar Fourier processing," in *German Microwave Conference Digest of Papers*, 2010, pp. 214–217. [Online]. Available: https://bit.ly/3ni0lgE

[8] J. Kim, J. Chun, and S. Song, "Joint range and angle estimation for FMCW MIMO radar and its application," 2018. [Online]. Available: https://bit.ly/2Uk63Ck

[9] B. Kim, Y. Jin, S. Kim, and J. Lee, "A low-complexity FMCW surveillance radar algorithm using two random beat signals," *Sensors (Basel)*, vol. 19, no. 3, p. 608, 2019. [Online]. Available: https://dx.doi.org/10.3390%2Fs19030608

[10] J. Svensson, "High resolution frequency estimation in an FMCW radar application," 2018. [Online]. Available: https://bit.ly/2UjCXTs

[11] E. Guerrero-Menéndez, "Frequency-modulated continuous-wave radar in automotive applications," 2018. [Online]. Available: https://bit.ly/3pp6slg

[12] M. D. Fontaine, *Traffic Monitoring*, 2009. [Online]. Available: https://bit.ly/36ne6nm

[13] V. Issakov, *Microwave Circuits for 24 GHz Automotive Radar in Silicon-based Technologies*. Springer, 2010. [Online]. Available: https://bit.ly/35nfJST

[14] C. Händel, H. Konttaniemi, and M. Autioniemi, *State-of-the-Art Review on Automotive Radars and Passive Radar Reflectors Arctic Challenge research project*. Lapland University of Applied Sciences, 2018. [Online]. Available: https://bit.ly/36u2yig

[15] Infineon, "Position2go software user manual," Infineon, Tech. Rep., 2019. [Online]. Available: https://bit.ly/38ASmqE

[16] Mikusova, Miroslava, "Crash avoidance systems and collision safety devices for vehicle occupants," *MATEC Web Conf.*, vol. 107, p. 00024, 2017. [Online]. Available: https://doi.org/10.1051/matecconf/201710700024

[17] K. Ramasubramanian and K. Ramaiah, "Moving from legacy 24 ghz to state-of-the-art 77-ghz radar," *ATZelektronik worldwide*, vol. 13, no. 3, pp. 46–49, Jun. 2018. [Online]. Available: https://doi.org/10.1007/s38314-018-0029-6

[18] J. M. Canino, V. Mena, J. Alonso, A. Ravelo, and E. García, "Prácticas de laboratorio en contextos de enseñanza-aprendizaje basados en competencias: dificultades y oportunidades," in *I Jornadas Iberoamericanas de Innovación Educativa en el ámbito de las TIC Las Palmas de Gran Canaria*, 2014. [Online]. Available: https://bit.ly/2L2K01r

[19] S. Kocijancic, "Contemporary challenges in teaching electronics to stem teachers," *AIP Conference Proceedings*, vol. 2043, no. 1, p. 020002, 2018. [Online]. Available: https://doi.org/10.1063/1.5080021

[20] M. V. Noroña M. and M. F. Gómez B., "Desarrollo e innovación de los sistemas mecatrónicos en un automóvil: una revisión," *Enfoque UTE*, vol. 10, pp. 117–127, 03 2019. [Online]. Available: https://doi.org/10.29019/enfoqueute.v10n1.350

[21] Analog Devices, "Radar demonstration platform. evaluates radar chipset including the ADF5901, ADF5904 and ADF4159." Analog Devices, Tech. Rep., 2019. [Online]. Available: https://bit.ly/2YRxHK8.

[22] Infineon, "Demo distance2go," Infineon, Tech. Rep., 2020. [Online]. Available: https://bit.ly/35Rr7VI

[23] ——, "Demo position2go," Infineon, Tech. Rep., 2020. [Online]. Available: https://bit.ly/2SUt6D8

[24] ——, "Infineon toolbox," Infineon, Tech. Rep., 2020. [Online]. Available: https://bit.ly/3lrUbtJ

[25] ——, "Application notes," Infineon, Tech. Rep., 2020. [Online]. Available: https://bit.ly/3bmpDnt

[26] ——, "MATLAB radar system API," Infineon, Tech. Rep., 2020. [Online]. Available: https://bit.ly/38z2aS5

[27] CETECOM, "Testing services for various radar applications," CETECOM, Tech. Rep., 2020. [Online]. Available: https://bit.ly/2UwRWK1

[28] MATLAB. (2020) Fourier transforms. [Online]. Available: https://bit.ly/2YPQzcz

[29] S. Hilbert. (2013) FFT zero padding. [Online]. Available: https://bit.ly/36HV2R5

[30] MATLAB. (2020) Constant false alarm rate (CFAR) detection. [Online]. Available: https://bit.ly/3dCuWRf.

[31] C. Will, P. Vaishnav, A. Chakraborty, and A. Santra, "Human target detection, tracking, and classification using 24-GHz FMCW radar," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7283–7299, 2019. [Online]. Available: https://doi.org/10.1109/JSEN.2019.2914365

[32] M. Sepúlveda, "Las competencias transversales, base del aprendizaje para toda la vida," *Universidad EAN*, 2017. [Online]. Available: https://bit.ly/2GYyDGg

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.08
pISSN: 1390-650X / eISSN: 1390-860X

Numerical simulation of the under-expanded flow in the experimental conical nozzle helios-x

Simulación numérica del flujo subexpandido en la tobera cónica experimental helios-x

San Luis B. Tolentino Masgo[1,2,*], Richard Nakka[3], Simón Caraballo[4], Jorge Mírez[2,5]

Received: 11-03-2020, Reviewed: 22-09-2020, Accepted after review: 17-10-2020

Abstract

Numerical studies of the flow field for convergent-divergent nozzles with throat length, have reported fluctuations of the flow with oblique shock waves in the throat section, for the overexpanded flow condition. However, for other flow conditions, for the same type of nozzle, knowledge is limited. In the present work, the objective is to determine the behavior of the flow in the throat length and in the divergent, for an experimental conical nozzle classified as Helios-X, for the underexpanded flow condition. 2D numerical simulations of the flow field were performed with the ANSYS-Fluent version 12.1 code, applying the RANS model. The governing equations for compressible flow, conservation of mass, momentum, energy, and state were used; as well as, for turbulence, the Menter SST $k-\omega$ model and for the viscosity as a function of temperature the Sutherland equation.

Resumen

Estudios numéricos del campo de flujo para toberas convergentes-divergentes con longitud de garganta, han reportado fluctuaciones del flujo con ondas de choque oblicuo en la sección de la garganta, para la condición de flujo sobre-expandido. Sin embargo, para otras condiciones del flujo, para un mismo tipo de tobera, el conocimiento es limitado. En el presente trabajo, el objetivo es determinar el comportamiento del flujo en la longitud de garganta y en la divergente, para una tobera cónica experimental clasificada como Helios-X, para la condición de flujo sub-expandido. Se realizaron simulaciones numéricas 2D del campo de flujo con el código ANSYS-Fluent versión 12.1, aplicando el modelo RANS. Se emplearon las ecuaciones gobernantes para el flujo compresible, conservación de la masa, cantidad de movimiento, energía y de estado; así como, para la turbulencia el modelo SST $k-\omega$ de Menter y para la viscosidad en función de la temperatura la ecuación de Sutherland.

[1]Universidad Nacional Experimental Politécnica Antonio José de Sucre, Vice-Rectorado Puerto Ordaz, Venezuela.
[2,*]Grupo de Modelamiento Matemático y Simulación Numérica, Universidad Nacional de Ingeniería, Lima, Perú. Corresponding author: ✉: sanluist@gmail.com. http://orcid.org/0000-0001-6320-6864,
[3]Cohetería experimental amateur, Canadá. http://orcid.org/0000-0002-7759-7162
[4]Departamento de Ingeniería Mecánica, Universidad Nacional Experimental Politécnica Antonio José de Sucre Vice-Rectorado Puerto Ordaz, Bolívar, Venezuela. http://orcid.org/0000-0002-0170-2448
[5]Facultad de Ingeniería de Petróleo Gas y Petroquímica, Universidad Nacional de Ingeniería, Lima, Perú. http://orcid.org/0000-0002-5614-5853

Suggested citation: Tolentino Masgo, S. L. B.; Nakka, R.; Caraballo, S. and Mírez, J. (2021). «Numerical simulation of the under-expanded flow in the experimental conical nozzle helios-x». Ingenius. N.° 25, (january-june). pp. 81-93. doi: https://doi.org/10.17163/ings.n25.2021.08.

In the section of the throat adjacent to the wall the flow exhibited fluctuations; in the axial symmetry the flow showed a stepped acceleration; in the divergent section the flow slowed in a certain region; however, the flow exited the nozzle at a supersonic speed slightly greater than Mach 3. It is concluded that in the throat length section there is a flow pattern, as well as in the divergent section.

Keywords: Throat, Fluctuation, Under-expanded, Simulation, Nozzle.

En la sección de la garganta, adyacente a la pared, el flujo presentó fluctuaciones, en la simetría axial el flujo presentó una aceleración escalonada; en la sección divergente, el flujo se desaceleró en cierta región, sin embargo, el flujo salió de la tobera a velocidad supersónica ligeramente mayor de Mach 3. Se concluye que en la sección de la longitud de garganta se presenta un patrón de flujo, así como, en la sección divergente.

Palabras clave: garganta, fluctuación, subexpandido, simulación, tobera

1. Introduction

The flow in nozzles in supersonic rocket motors is recurrently studied for different geometric configurations of the internal profile of the nozzle walls. The geometry of the divergent section of the nozzle may be conic, bell-shaped, parabolic, rectangular, flat, among others. Taking experimental data in a test bench of the flow within the nozzle in static conditions has limitations when the gas flow is well above the ambient temperature. For this reason, during the experimental test data are recorded with instruments sensitive to changes in pressure, temperature, vibration, in the combustion chamber, at the beginning of the nozzle, in the walls, and through the capture of images of the flow region that is discharged to the environment. When the flow is strangled in the throat and based on the pressure of the combustion chamber, the flow may be overexpanded, optimally expanded or underexpanded [1]. It should be pointed out that, for an overexpanded flow the shock occurs within the nozzle and for an underexpanded flow the shock occurs outside of it.

The experimental data obtained at the boundary of the nozzle are used to reproduce the behavior of the flow field within the nozzle, through the application of computational fluid dynamics [2], [3], which enables quantifying the thermodynamic magnitudes in a computational domain of the flow field that is not possible to obtain by experimental means in open field and labs.

CFD is a computational tool that enables to obtain approximate solutions of the reality of the physical phenomenon, if it is appropriately considered the geometry of the computational domain, the refinement of the mesh, the turbulence model, setting the errors during the iteration stage, as well as other variable control parameters.

For compressible flow there are supported and reported works about the shock waves and the expansion waves of Prandtl-Meyer [4, 5], the capturing of images of the shock waves by means of the Schlieren technique [6], the turbulence [7], the limit layer [8, 9], as well as the turbulence models [10].

Experimental studies in static and dynamic conditions for convergent-divergent nozzles with throat length have been reported with research purposes; however, the knowledge of the behavior of the flow regime present within nozzles with throat length is virtually none since it has been scarcely addressed using computational tools to determine its thermodynamic magnitudes.

A computational study of the simulation of the overexpanded flow in a conic nozzle with throat length and mean angle of 11° in the divergent section, which belongs to a motor of a sounding rocket for solid fuel, classified as ULA-1A XP [11], reported numerical results of the fluctuations of the flow velocity and of the oblique shocks that occurred in the throat section, in the range of transonic velocity, as well as the behavior of the flow fluctuation from the center to the walls of the throat. It should be noted that this type of nozzle was previously tested in static and dynamic conditions by the Group of Atmospheric and Space Sciences (GCAE, Grupo de Ciencias de la Atmósfera y del Espacio) of the Universidad de Los Andes, Venezuela [12], [13], [14].

Another study conducted for an underexpanded flow in a conic nozzle with throat length and mean angle of 15° in the divergent section, identified as Helios-X nozzle [15], which also simulated the flow field, presented its results according to the Mach number, where the flow in the throat section exhibited deceleration. Although they provide fundamental data of the flow velocity according to the Mach number, it still remains to analyze with greater detail the flow field for other thermodynamic parameters. For this reason, it is of interest to continue with the investigation for this type of nozzle, to determine the behavior of the field of density, pressure, temperature and velocity, as well as the Mach number in all the domain, and focusing with greater interest in the throat section, which might contribute and enrich the knowledge about the distribution of gradients of the thermodynamic parameters.

The Helios-X nozzle, designed by Nakka [16], was tested in a test bench in static and dynamic conditions. Figure 1 shows the record of the test in static conditions of the nozzle coupled to the Helios-X rocket motor, the rocket in the launching platform and the rocket take-off. Multiple experiments conducted by Nakka of nozzles in rocket motors for amateur solid fuel, in static and dynamic conditions, may be obtained in their web site [16].

The present work is intended to continue with the research for the underexpanded flow in the Helios-X nozzle, reported in [15], with the purpose of determining the behavior of the flow in the throat and divergent sections. In order to achieve the stated objective, the supersonic flow field was simulated to obtain the field of pressure, Mach number, velocity, temperature and density. Results show, in the wall at the inlet of the throat, a region of the flow which exhibits fluctuation, and thus in that region the flow accelerates and decelerates, in other regions of the same section, the flow velocity only slows.

Section 2 presents the equations used, the 2D computational domain and the computational solution method. Section 3 describes the results obtained and the discussion. Section 4 presents the conclusions of the analysis conducted.

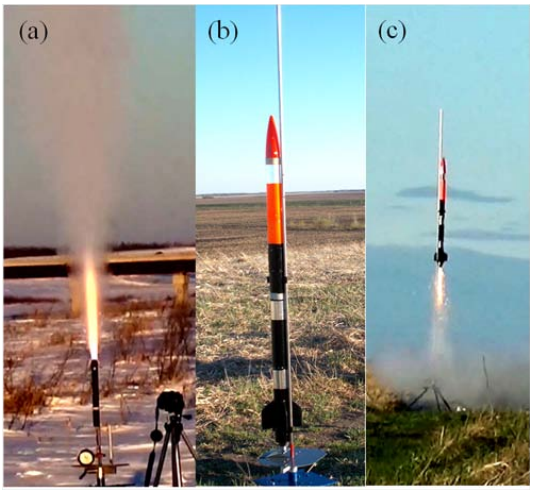


Figure 1. (a) Experimental test in static conditions of the nozzle with throat length coupled to the Helios-X rocket motor, March 2017. (b) Launch platform of the Z-30 rocket, with the Helios-X rocket motor installed. (c) Z-30 rocket take-off, May 2017 [16]

2. Materials and methods

2.1. Mathematical fundamentals

In the present work, the model of the Reynolds-averaged Navier-Stokes (RANS) equations are employed for simulating the flow field of compressible flow. The governing equations used are: the equation of conservation of mass (1), momentum (2), conservation of energy (3), and state (4). In compact form and for a flow in stationary conditions, they are expressed as:

$$\nabla \cdot (\rho u_i) = 0 \tag{1}$$

Where, ρ is the density and u the velocity.

$$\nabla \cdot (\rho u_i u_j) = -\nabla p + \nabla(\overline{\overline{\tau}}) + \nabla \cdot (-\rho \overline{u_i u_j}) \tag{2}$$

Where, p is the pressure; $\overline{\overline{\tau}}$ is the tensor of tensions and $-\rho \overline{u_i u_j}$ the Reynolds tensions. It should be noted that Equation (2) is closed, since it includes the Reynolds tensions term.

$$\nabla \cdot (u_i(\rho E + p)) = \nabla \cdot (k_{eff} \nabla T + (\overline{\overline{\tau}}_{eff} \cdot u_i)) \tag{3}$$

Where E is the total energy, T the temperature, k_{eff} the effective thermal conductivity, and $\overline{\overline{\tau}}_{eff}$ the effective tensor of tensions.

$$p = \rho R T \tag{4}$$

Where, R is the gas constant.

For compressible flow it is taken into account the ratio of pressures (5) and of temperatures (6) as a function of the Mach number, which is the dominant parameter, and are expressed as:

$$\frac{p_0}{p} = \left(1 + \frac{\gamma - 1}{2} M^2\right)^{\frac{\gamma}{\gamma - 1}} \tag{5}$$

$$\frac{T_0}{T} = 1 + \frac{\gamma - 1}{2} M^2 \tag{6}$$

Where, p_0 is the total pressure, T_0 the total temperature, γ the ratio of specific heats and M the Mach number.

The supersonic Mach number at the outlet of the nozzle for an underexpanded flow and without shock at the outlet of the divergent section is determined with Equation (7); where A/A^* is the ratio of design areas, A is the area at the outlet of the nozzle and A^* is the area of the throat:

$$\frac{A}{A^*} = \frac{1}{M} \left(\frac{1 + \frac{\gamma - 1}{2} M^2}{\frac{\gamma + 1}{2}} \right)^{\frac{\gamma + 1}{2\gamma - 2}} \tag{7}$$

The considerations of the Mach number are the following: $M < 0.3$ for incompressible flow; $0.3 < M < 0.8$ for subsonic flow; $0.8 < M < 1.2$ for transonic flow; $1.2 < M < 5$ for supersonic flow; $M > 5$ for hypersonic flow; and $M = 1$ for the flow with sonic velocity [5]. It should be noted that White [5] considers that the flow is hypersonic from a value greater than Mach 3, and Anderson [4] from Mach 5.

According to Sutherland´s law [8], the viscosity as a function of the temperature (8) is expressed as:

$$\frac{\mu}{\mu_0} = \left(\frac{T}{T_0}\right)^{\frac{3}{2}} \frac{T_0 + S}{T + S} \tag{8}$$

Where, the reference viscosity is $\mu_0 = 1,716\ kg/(m \cdot s)$, the reference temperature is $T_0 = 273,11\ K$ and the effective temperature is $S = 110,56\ K$.

For the flow turbulence, it is taken into account the Menter SST $k - \omega$ turbulence model [17], which is solved together with the equation of momentum. This turbulence model has two equations, one for the specific kinetic energy k, and the other for the specific dissipation rate ω, and thus it improves the responses in the presence of adverse pressure gradients, and flow separation.

The Menter turbulence model [17] has been compared with other turbulence models for different conditions of the compressible flow with presence of shock waves in different experimental equipment for overexpanded flow and underexpanded flow, where the numerical results for 2D domains with symmetric geometries overlap the experimental data of pressure at the walls of experimental equipment, in a nozzle [18,19] and in a transonic diffuser [20,21], with the shapes of the shock

waves of the numerical results approximately similar to the experimental ones. Applications of the Menter turbulence model support their validity for different flow regime conditions [22–25]. In 3D domains, the contribution of the physical analysis is much greater, even for asymmetric geometries and asymmetric lateral loads in the flow field. Therefore, for underexpanded flow conditions and without conditions of asymmetric lateral loads, the Menter turbulence model is appropriate for 2D simulation of the underexpanded flow of the present work.

2.2. Computational domain

The scheme of the geometry of the Helios-X experimental convergent-divergent conic nozzle with throat length [16] is shown in Figure 2, including its main dimensions and corresponding units in millimeters: throat length $L_g = 6.01\ mm$, throat diameter $D_g = 5.58\ mm$, ratio between throat length and diameter $r_{LD} = 1.07$, and the ratio of design areas $A/A^* = 5.206$.

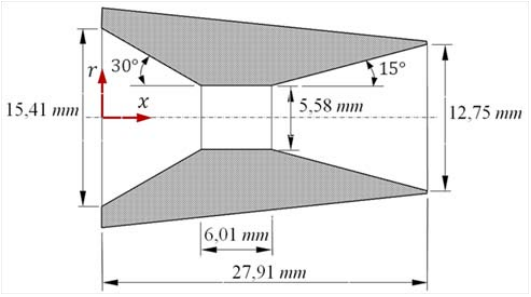


Figure 2. Geometry of the Helios-X experimental conic nozzle [16]

The 2D computational domain with axial symmetry consists of a section of the combustion chamber, the nozzle and the atmosphere, as shown in Figure 3. The purpose of taking into account a short section of 7.47 mm and not the whole section of the motor tube, is to apply the pressure load and direct the flow towards the inlet of the nozzle. The length of the section of the conic nozzle is 27.91 mm and the length of the section of the atmosphere ambient is 281 mm. The throat begins at the position x=8.51 mm and ends at the position x=14.53 mm, and its length is $L_g = 6.01\ mm$.

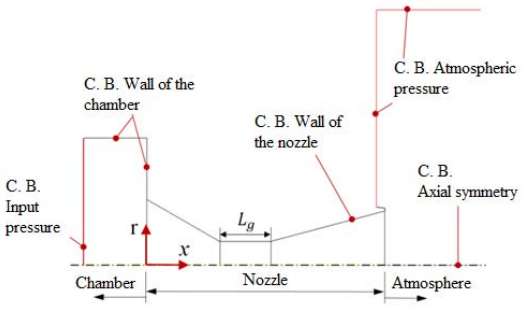


Figure 3. 2D computational domain with axial symmetry. The domain includes a section of the combustion chamber, the nozzle and a section of atmosphere

In addition, since the nozzle has a symmetric revolution geometry, a 2D domain is appropriate, which contributes to reduce the number of cells in the mesh and the iteration time when processing the computational data. In addition, the references where the boundary conditions (BC) are applied are pointed in the same figure of the domain.

The initial and boundary conditions are established as:

In the chamber of the Helios-X rocket motor, the absolute total pressure is established as 6996,11 kPa; and the total temperature as 2558 K.

In the atmosphere ambient, the pressure is established as 101.5 kPa, and the temperature as 263.15 K.

In the axis of axial symmetry, the flow velocity in the radial direction is zero. In the walls, the velocity is zero due to the no-slip condition.

The walls of the section of the combustion chamber and of the nozzle are considered adiabatic.

The effect of the gravity of the flow within the nozzle is not considered, due to the high velocity of the supersonic stream in the divergent. In the atmosphere, the flow velocity is supersonic, and thus for the length of the domain considered, the effect of gravity is considered tiny and therefore neglected.

It should be pointed out that the pressure and temperature data in the chamber of the rocket motor which are applied at the boundary conditions in the present work have been obtained by Nakka [16] through experimental means, as well as the data of the atmosphere ambient recorded with pressure and temperature measuring instruments. The experimental data of the combustion chamber and of the atmosphere ambient applied to the 2D domain, contribute to obtain the simulation of the flow field in all the domain, and enable to determine the behavior of the pressure

and temperature in the wall of the nozzle which has not been possible to measure experimentally.

For the numerical simulations in the present work, the gases burnt as a result of the combustion of the solid propellant consisting of ammonium nitrate, aluminum, sulfur and chloroprene (A24 ANCP) [16] are considered as ideal gases, and as air substance. As parameters, it is fixed the specific heat ratio $\gamma = 1,4$, the gas constant $R = 287\ J/(kg \cdot K)$, the specific heat at constant pressure $C_p = 1006,43\ J/(kg \cdot K)$ and the thermal conductivity $k = 0,0242\ W/(m \cdot K)$.

Figure 4 shows the meshed computational domain and an enlargement of the meshed section of the nozzle. The meshed was made in the ANSYS-Meshing platform and the domain was discretized through the interaction of ICEM-CFD. The mesh of the domain was refined along all the walls, due to the presence of shear stress in those regions. A meshed with triangular cells was applied in the section of the chamber and of the nozzle, and a structured mesh with quadrilateral cells was applied in the section of the atmosphere ambient, for a total of 32675 cells combined. The meshed of the domain shown in the figure is the final one, after performing a numerical convergence study in which the domain was refined three times.

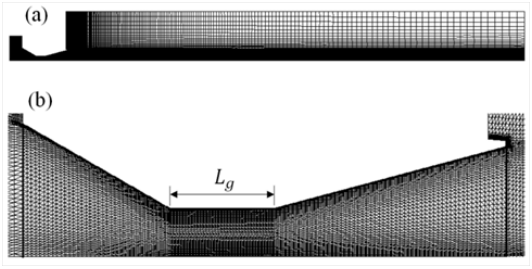


Figure 4. (a) Meshed computational domain with a total of 32675 cells combined. (b) Section of the meshed domain of the nozzle

In the numerical convergence study, the first meshed domain had 30950 cells; in the section of the nozzle the smallest cell had a dimension of $4,9 \times 10^{-5}\ mm$ and the largest one a dimension of $3,37 \times 10^{-4}\ mm$, and the largest cell dimension was 3.2 mm in the upper right corner of the atmosphere domain. The second meshed domain had 32296 cells, with a smallest cell of $4,7 \times 10^{-5}\ mm$ and a largest cell of $3,31 \times 10^{-4}\ mm$, and in the atmosphere domain the maximum cell size was 3.2 mm. The third meshed domain had 32675 cells, with minimum cell dimension of $4,67 \times 10^{-5}\ mm$ and maximum of $3,29 \times 10^{-4}\ mm$, and in the atmosphere the largest cell dimension was 3.2 mm. For the first domain, the minimum spacing of the cell in the wall of the nozzle was $y^+ = 0,98$ in average, for the second domain $y^+ = 0,95$ and for the third domain of the final meshed $y^+ = 0,94$.

For the three cases, the density of the mesh is high in the nozzle and in the region of the atmosphere where there is the supersonic stream known as plume. The end of the divergent section was taken as control point, in the axial symmetry, to numerically evaluate the Mach number, since it is a critical region due to the high flow velocity, and a percentage error of 0.04% was obtained between the third and second meshed domains; and a slightly larger error of 0.052% between the second and the first, at the outlet of the nozzle in the axial symmetry of the X axis for the third domain of the final meshed and a Mach numeric value of 3.1.

For the theory of quasi-one-dimensional flow with $\gamma = 1,4$, and $A/A^* = 5,206$ of design of the Helios-X nozzle, and without shock at the outlet of the nozzle, a Mach value of 3.217 was obtained with Equation (7) (theoretical). A magnitude difference of 0.117 was obtained between the theoretical Mach number of 3.217 and the Mach of 3.1 numerically calculated of the third meshed domain. Therefore, it is acceptable the magnitude of the result of the comparisons of Mach numbers as numerical validation. For a quasi-one-dimensional and underexpanded flow, the flow exits the nozzle uniformly and perpendicular to the outlet area, and therefore it has the same cross-sectional area, while for the numerical method, the flow has a velocity gradient of the Mach number at the outlet of the nozzle and its magnitude varies in its cross-sectional area.

The meshed domain shown in Figure 4 is appropriate and fulfills satisfactory criteria of the convergence analysis performed. Although the three domains fulfill for $y^+ < 1$, it was chosen the meshed domain with 32675 cells because it has a greater refinement in the walls, which is used in the computational simulations.

The quality of the mesh for two-dimensional cells, and the equiangular bias (Q_{EAS}) establish that it should be in $0 \geq Q_{EAS} \geq 1$, for any 2D cell [26]. For all the domain of the final mesh, it was obtained $Q_{EAS} = 0,55$. It should be stated that the numerical convergence study was performed using the Menter SST $k - \omega$ turbulence model [17] to simulate the flow turbulence.

2.3. Computational solution method

For simulating the flow in the ANSYS-Fluent 12.1 code, which applies the finite volume method (FVM), it was chosen the analysis option based on density for a compressible fluid, 2D domain with axial symmetry in the x axis. For the turbulence of the flow, it was employed the Menter SST $k - \omega$ model [17], and for the viscosity the equation of Sutherland [8].

In the solution method, it was considered the implicit formulation and Roe-FDS type of flow. For the spatial discretization, the gradient: Least Squares Cell based; for the flow, turbulent kinetic energy and the specific dissipation rate, the option: First Order Upwin.

In the residual monitor, for the absolute convergence criterion, it was established a fixed value of

0.00001, for continuity, velocity and energy. It was carried out 136400 iterations in a time close to fifty hours, to obtain the numerical convergence of the final results of the flow field, Mach number, pressure, velocity, temperature and density.

For data processing, an equipment with the following features was employed: Siragon notebook, model M54R, Intel Core 2 Duo, two processors of 1.8 GHz and RAM memory of 3 GB.

3. Results and discussion

This section presents the results of the flow field, for the input pressure load applied of 6996.11 kPa and temperature of 2558 K, ambient pressure at the outlet of the nozzle of 101.5 kPa and temperature of 263.15 K.

The variations of the magnitudes of the static pressure (Figure 5), Mach number (Figure 6), velocity (Figure 7), static temperature (Figure 8) and density (Figure 9), show in which regions of the domain are reached the maximum and minimum values. By means of the contour lines it is observed how they are distributed in different regions of the flow field, both in the section of the nozzle and in the section of the atmosphere.

The flow exiting the nozzle is underexpanded and it is shown in the atmosphere how the shock waves are constituted; and the region of the supersonic stream where these waves are present, is known in the literature as plume.

Figure 10 shows the profiles evaluated in the axial symmetry, for the region of the supersonic flow that exits from the nozzle and discharges in the atmosphere. The fluctuations due to the shock wave are shown, even in certain regions the pressure drops below atmospheric pressure, accelerating the flow in that region before the occurrence of the shock at a Mach value close to 4.75, and velocity of 2050 m/s; while the temperature drops below 500 K. After the shock, the flow is still supersonic with presence of dampen fluctuations around Mach 3, velocity of 1800 m/s and temperature of 900 K. In addition, the behavior of the pressure and density curves show an equilibrium trend.

The contour lines in the nozzle section have been considered, which provide more information regarding the distribution of the gradients of the thermodynamic parameters. The enlarged figures illustrate how the contour lines are distributed in the convergent, throat of length L_g and divergent sections: for the static pressure (Figure 11), Mach number (Figure 12), velocity (Figure 13), static temperature (Figure 14) and density (Figure 15). It is observed that at the beginning of the throat section the contour lines have a behavior different to the contour lines at the end of the same throat, as well as in the middle part.

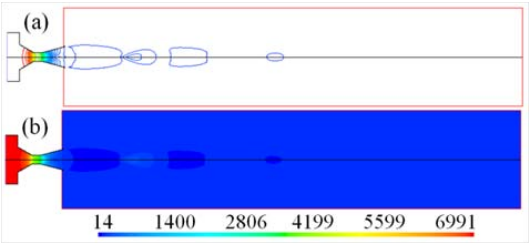


Figure 5. Field of static pressure (kPa)

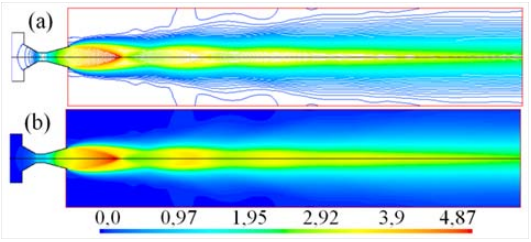


Figure 6. Field of Mach number

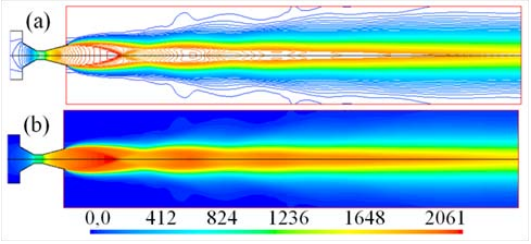


Figure 7. Field of velocity (m/s)

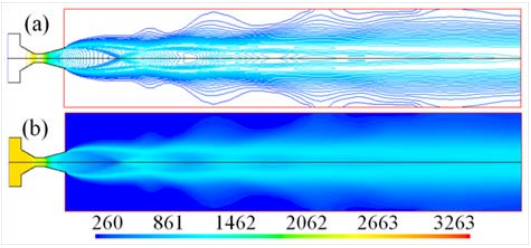


Figure 8. Field of static temperature (K)

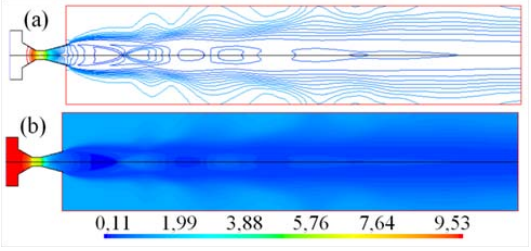


Figure 9. Field of density (kg/m^3).

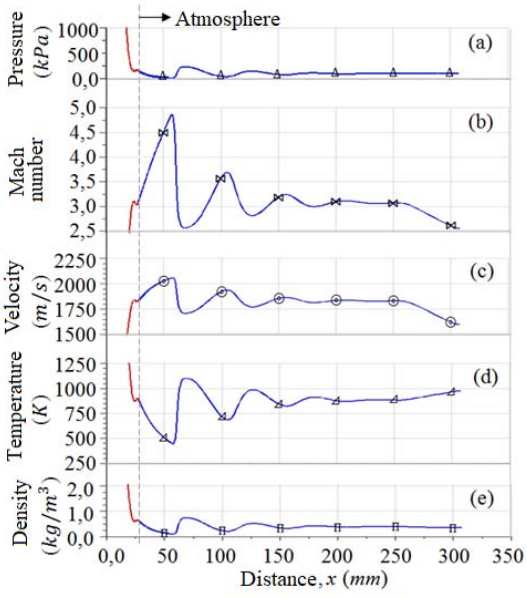


Figure 10. Profiles evaluated in the axial symmetry, in the X axis, in the region of atmosphere ambient. (a) Pressure. (b) Mach number. (c) Velocity. (d) Temperature and (e) Density

It should be pointed out that for the case of static pressure, it is observed that in the middle part of the throat length the contour lines tend to be perpendicular to the walls of the throat, and thus the pressure magnitude in the symmetry is similar to the pressure magnitude in the wall.

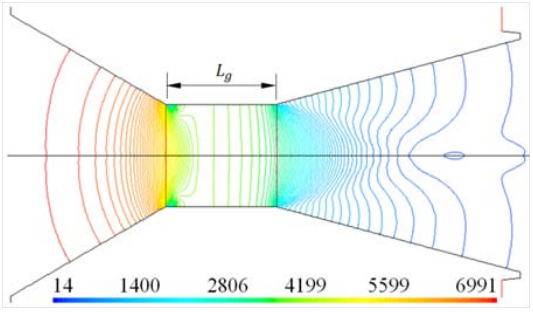


Figure 11. Contour lines of static pressure (kPa) of the flow in the nozzle section

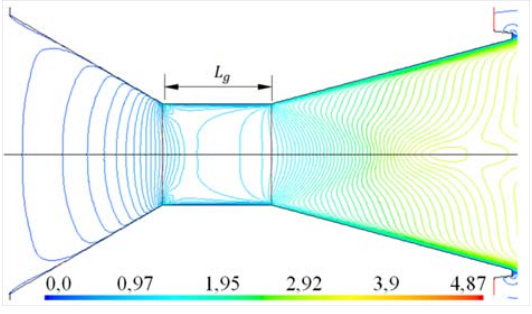


Figure 12. Contour lines of Mach number of the flow in the nozzle section

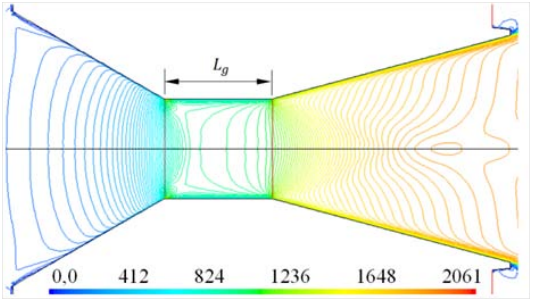


Figure 13. Contour lines of velocity (m/s) of the flow in the nozzle section

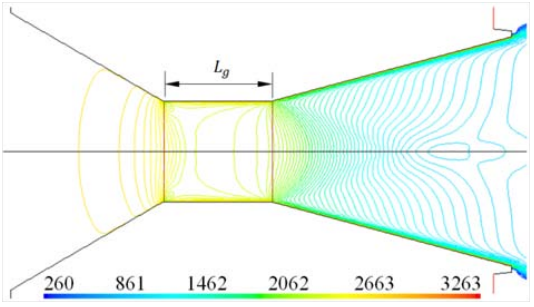


Figure 14. Contour lines of static temperature (K) of the flow in the nozzle section

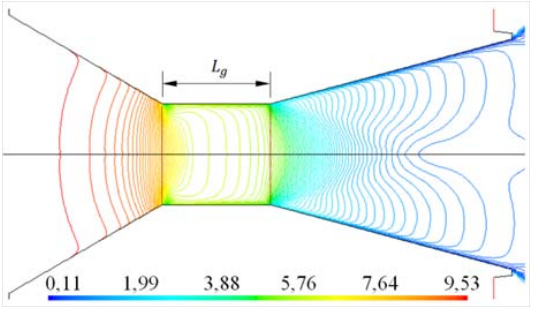


Figure 15. Contour lines of density (kg/m^3) of the flow in the nozzle section

Analyzing the behavior of the flow from another perspective, through the trajectory of the numeric curves, it is shown that the sharp change in the static pressure drop (Figure 16) occurs in the first vertex of the throat section. In the middle section of the throat, a part of the trajectories of the pressure profiles in the wall and in the axial symmetry are coincident, which indicates that a uniform flow regime is present with incidence in the acceleration of the flow. Downstream, from position x=22.5 mm, the pressure drop of the flow tends to stop, slightly increases, then decreases with smaller intensity, thus being a part of the profile trajectory with oscillatory trend; in such part, the pressure fluctuations influence the development of the

acceleration of the flow, and thus the magnitudes of the thermodynamic parameters exhibit variations.

For the case of the Mach number (Figure 17), it is shown the deceleration of the flow in the throat section, being Mach 1 in the middle part of the throat section, and greater than this value at the outlet of the throat. Close to the outlet of the divergent section, the flow velocity reaches a value greater than Mach 3, at position x=22.5 mm, and downstream a curvature is present and tends to reduce the flow up to the outlet of the nozzle, maintaining a velocity greater than Mach 3, corresponding to a supersonic flow. From position x=22.5 mm up to the outlet of the nozzle, the behavior of the trajectory of the curve shows that a flow deceleration occurs before exiting the nozzle.

In the velocity profile (Figure 18) it is observed the behavior of the velocity pattern in the axial symmetry, but it is now shown in the wall because the velocity is zero there due to the no-slip condition. In the middle part of the throat section, the flow has an estimated velocity of 920 m/s; downstream, it is shown the behavior of a part of the trajectory of the velocity profile from position x=22.5 mm, with a velocity trend slightly greater than 1800 m/s. Also, like in the previous case, for the Mach number the flow velocity decelerates before exiting the nozzle.

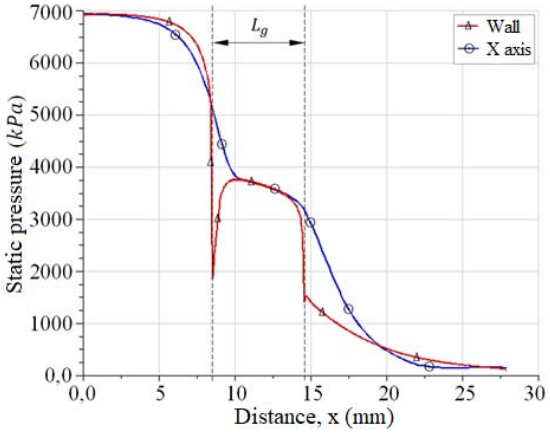


Figure 16. Static pressure profiles evaluated in the wall and in the X axis, in the nozzle section

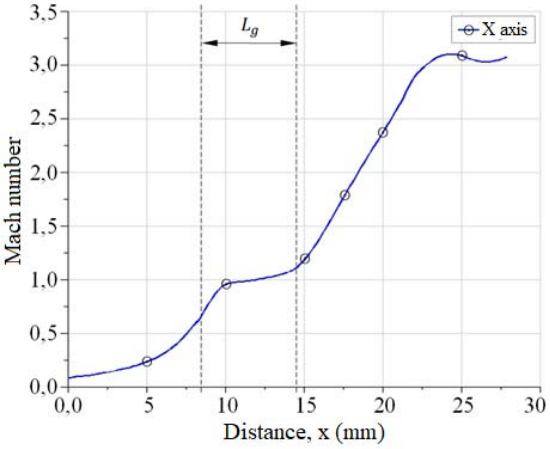


Figure 17. Mach number profile evaluated in the X axis, in the nozzle section

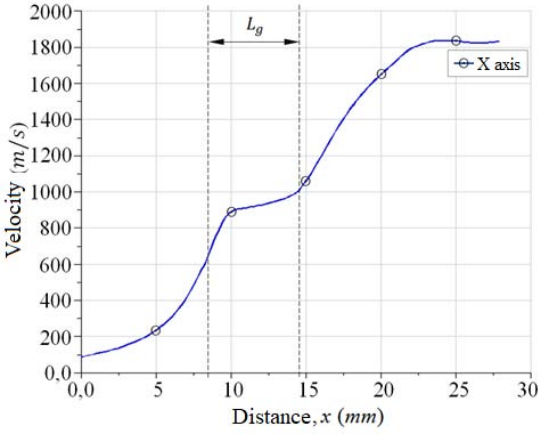


Figure 18. Velocity profile evaluated in the X axis, in the nozzle section

Likewise, for the static temperature profiles (Figure 19), it is shown the behavior of the trajectories in the wall and in the axial symmetry. It is observed in the throat section that the temperature increases and decreases, and the vertices have influence in the occurrence of sudden changes. In the case of the divergent section, the increase of the temperature in the wall to a value greater than the input temperature, is a consequence of the friction of the flow at high velocity rubbing the adiabatic wall, which drastically reduces before the flow exits the nozzle. Whereas in the axial symmetry, the temperature decreases due to the flow expansion, and it is shown the fluctuation of its magnitude from position x=22.5 mm. Regarding the density profiles (Figure 20), their behavior depends on the flow expansion and compression, in the convergent, throat and divergent sections.

The results of flow field and of the profiles that define the trajectory of pressure, Mach number, temperature, velocity and density, in the axial symmetry

and in the nozzle wall, are acceptable. This is justified because there is a magnitude difference of 0.117 between the theoretical Mach value of 3.217 obtained with Equation (7) and the numerical value of Mach 3.1 shown in Figure 17, with both values being calculated at the outlet of the nozzle with magnitudes close to each other.

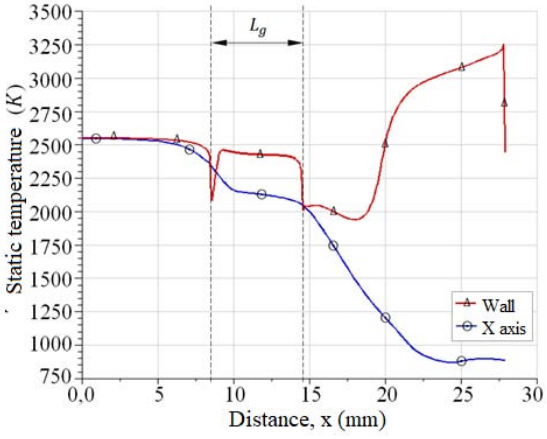


Figure 19. Static temperature profiles evaluated in the wall and in the X axis, in the nozzle section

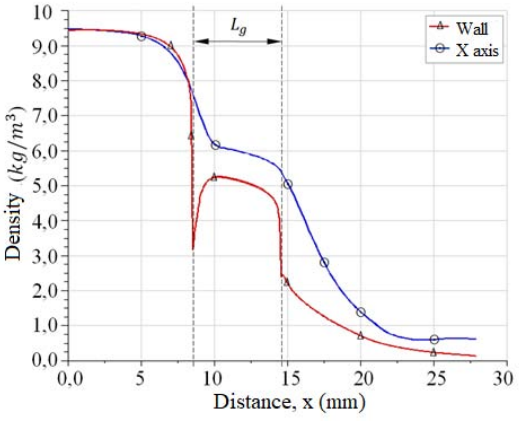


Figure 20. Density profiles evaluated in the wall and in the X axis, in the nozzle section

Enlarged details of the throat section show the behavior of the trajectories of the profiles of static pressure (Figure 21), Mach number (Figure 22), velocity (Figure 23), static temperature (Figure 24) and density (Figure 25), for different radial distances, evaluated from the axial symmetry to the wall. The length of the throat begins at position x=8.51 mm and ends at position x=14.53 mm. The profiles show the behavior of the flow region in the first vertex, at the beginning of the throat length, along the throat length, as well as when the flow enters at the beginning of the divergent section.

For the case of Mach number (Figure 22) and velocity (Figure 23), for the flow region adjacent to the wall of the throat section, the flow accelerates, decelerates and accelerates again; whereas in the axial symmetry the flow accelerates, decelerates and accelerates again when entering the divergent section; this is a consequence of the variations in static pressure that occur from the beginning of the throat section up to the end of such section (see Figure 21). Therefore, the geometry of the aerodynamic profile of the throat section of length L_g, determines the behavior of the flow pattern in such section; in this case, there is a flow pattern in the throat section, for an underexpanded flow at the outlet of the nozzle at supersonic velocity greater than Mach 3.

The variations in static temperature in the throat section are shown in Figure 24, where there is an increase of its magnitude in the wall, and variations in the vertices at the inlet and outlet of the throat (Figure 24). Similarly, for the case of the density (Figure 25), its magnitude decreases towards the wall, and in the vertex at the inlet of the throat the density decreases and increases, with a behavior similar to the trajectory of the pressure profile.

Results show that the geometry of the throat section with throat length and diameter ratio r_{LD}=1.07, begins at position x=8.51 mm and ends at position x=14.53 mm, which is a short segment of circular section, has influence on the development of the flow in the regions adjacent to the walls, in the vertices and in the symmetry. In the axial symmetry, in the X axis, the flow reached a value in the range of Mach 0.65 to 1.2, with regions of subsonic, transonic and sonic velocity, without the presence of shocks; therefore, a flow pattern is present.

In a work reported of the ULA-1A XP experimental conic nozzle [11], with mean angle of the divergent section of 11° and throat length and diameter ratio $r_{LD} = 1,10$, it showed flow deceleration with presence of oblique shock waves, in the range of Mach 0.8 to 1.4; for an overexpanded flow.

When comparing both cases for the flow in the throat region, the length of such region has influence in the development of the flow, either underexpanded or overexpanded flow.

Another flow pattern was present from position x=22.5 mm. From position x=22.5 mm up to the outlet of the nozzle, the behavior of the trajectory of the curve shows a deceleration of the flow in the divergent section, after the flow reaches the supersonic velocity Mach 3.

Similar results of the flow deceleration in the divergent section for 2D domains, was reported in [19] through a density pattern, for a mean angle of 11.01° of a flat nozzle, for overexpanded flow conditions. Another study reported flow deceleration for a conic nozzle with mean angle smaller than 5° [27], as well as for a nozzle with optimized contour and parabolic contour [28].

Therefore, the fluctuation obtained in the present

work is not an isolated result, since a region of the flow exhibits a deceleration in the divergent section, with a defined pattern.

It should be mentioned that works reported in the literature for inviscid one-dimensional flow [1,4,5,9,26], detail the increase of the flow velocity according to the Mach number in the divergent section of a nozzle, the trajectory of the curve increases and tends to curve towards the outlet of the nozzle, and exhibits no fluctuation. However, the results of the present work show that there is a fluctuation for certain region of the flow in the divergent section, even though the shock wave occurs outside of the nozzle.

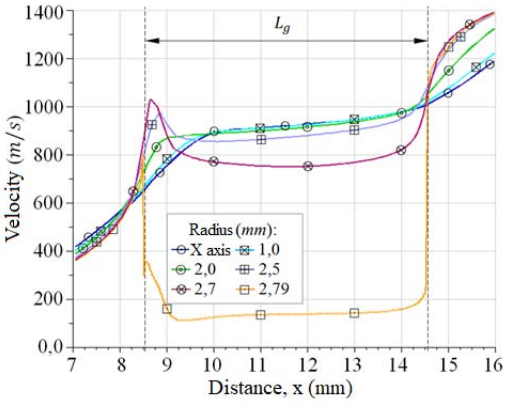


Figure 23. Velocity profiles evaluated at the throat section of the nozzle

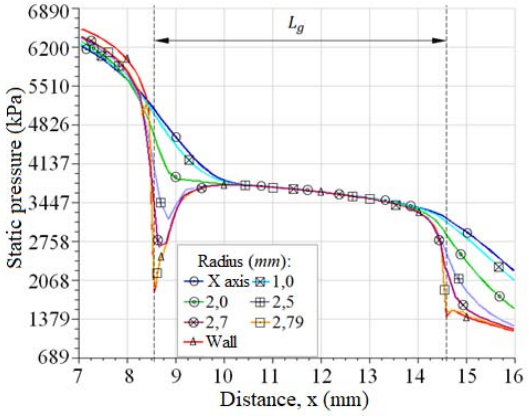


Figure 21. Static pressure profiles evaluated at the throat section of the nozzle

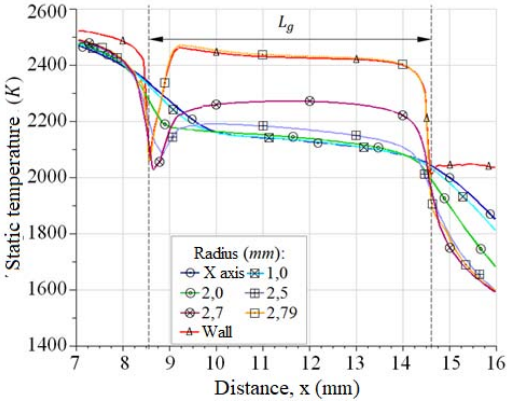


Figure 24. Static temperature profiles evaluated at the throat section of the nozzle

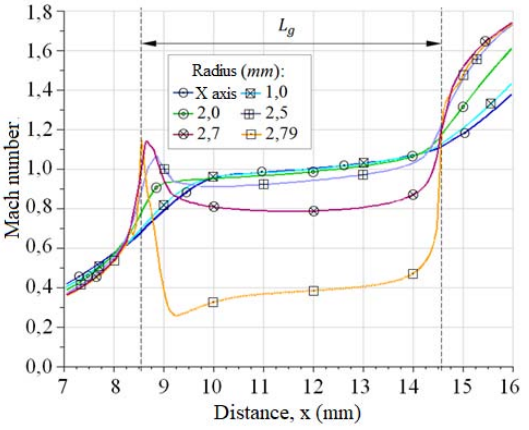


Figure 22. Mach number profiles evaluated at the throat section of the nozzle

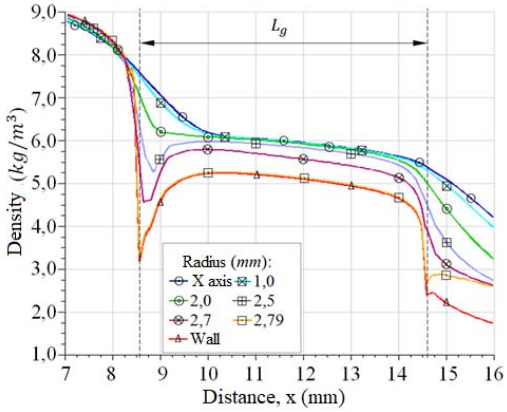


Figure 25. Density profiles evaluated at the throat section of the nozzle

4. Conclusions

According to the analyzes carried out, from the results of the numerical simulations of the underexpanded flow

it is concluded that: In the nozzle section there are two regions where the pressure drop of the flow slows, the throat length section, which is between positions x=8.51 mm and x=14.53 mm; and at the end of the divergent section, from position x=22.5 mm up to the outlet of the nozzle. Therefore, in both regions, the flow decelerates.

In the axial symmetry, at the end of the divergent section of the nozzle, the flow reaches a value slightly larger than Mach 3, and flow velocity of 1800 m/s. In the atmosphere ambient it reaches an estimated value of Mach 4.75, and flow velocity of 2050 m/s, before the occurrence of the shock. Downstream, the magnitudes of Mach number and of velocity fluctuate, up to a prolonged distance around Mach 3.

The profiles of static pressure, Mach number, velocity, static temperature and density, show how the flow develops in the throat section, where at the beginning of such section there is the vertex; a region of the flow adjacent to the wall exhibits fluctuations, causing pressure drops, and thus the flow velocity in such region adjacent to the wall accelerates and decelerates.

In the axial symmetry in the throat, at position x=8.51 mm the flow reaches an estimated value of Mach 0.68, and at the end of the throat section, at position x=14.53 mm an estimated value of Mach 1.1; thus, a flow region is subsonic, and the rest is transonic. In such section of the throat, the trajectory of the profile defines a staggered behavior.

A future work considers, for 2D and 3D domains and with different computational codes, to progressively reduce the throat length L_g, to determine if there is still influence in the flow acceleration, or if it occurs a possible fluctuation that causes oblique shock waves. Similarly, to determine if the fluctuation of thermodynamic parameters, such as Mach number, pressure and temperature, in the axial symmetry, at the end of the divergent section exhibit some significant change.

Acknowledgements

Mi agradecimiento a Jehová, mi Dios todopoderoso, mi fuente de sabiduría e inspiración. Al Centro de Estudios Energéticos de la Universidad Nacional Experimental Politécnica «AJS» Vice-Rectorado Puerto Ordaz (UNEXPO), Bolívar, Venezuela. Al Grupo de Modelamiento Matemático y Simulación Numérica (GMMNS, Group of Mathematical Modeling and Numerical Simulation) de la Universidad Nacional de Ingeniería (UNI), Lima, Perú. A Richard Nakka (Ingeniero Aeroespacial), por el aporte de datos experimentales de la tobera del motor cohete Helios-X.

References

[1] G. P. Sutton and O. Biblarz, *Rocket propulsion elements*. John Wiley & Sons, 2016. [Online]. Available: https://bit.ly/35psFaR

[2] J. Blazek, *Computational fluid dynamics: principles and applications*. Butterworth-Heinemann, 2015. [Online]. Available: https://bit.ly/3pnxjhx

[3] B. Andersson, R. Andersson, L. Hakansson, M. Mortensen, R. Sudiyo, and B. van Wachem, *Computational Fluid Dynamics Engineers*. Cambridge University Press, 2011. [Online]. Available: https://bit.ly/32BPCWo

[4] J. D. Anderson, *Fundamentals of aerodynamics*. McGraw-Hill international editions. Mechanical engineering series, 1984. [Online]. Available: https://bit.ly/3eQR5ge

[5] F. M. White, *Fluid Mechanics*. McGraw-Hill series in mechanical engineering, 2011. [Online]. Available: https://bit.ly/35opmAy

[6] P. Krehl and S. Engemann, "August toepler – the first who visualized shock waves," *Shock Waves*, vol. 5, no. 1, pp. 1–18, Jun. 1995. [Online]. Available: https://doi.org/10.1007/BF02425031

[7] V. Karman, "The fundamentals of the statistical theory of turbulence," *Journal of the Aeronautical Sciences*, vol. 4, no. 4, pp. 131–138, 1937. [Online]. Available: https://doi.org/10.2514/8.350

[8] F. White, *Viscous fluid flow*. McGraw-Hill series in Aeronautical and Aerospace Engineering, 1974. [Online]. Available: https://bit.ly/3eRRCyP

[9] H. Schlichting and K. Gersten, *Boundary-Layer Theory*. Springer, 2016. [Online]. Available: https://bit.ly/36yZGAx

[10] D. C. Wilcox, *Turbulence Modeling for CFD*. DCW Industries, Incorporated, 1994. [Online]. Available: https://bit.ly/32HZnCm

[11] A. L. Tolentino, J. Ferreira, M. Parco, L. Lacruz, and V. Marcano, "Simulación numérica del flujo sobre-expandido en la tobera cónica experimental ULA-1A XP," *Unviversidad, Ciencia y Tecnología*, vol. 21, no. 84, pp. 126–133, 2017. [Online]. Available: https://bit.ly/2H4yX6k

[12] V. Marcano, P. Benitez, C. La Rosa, L. La Cruz, M. A. Parco, J. Ferreira, R. Andrenssen, A. Serra Valls, M. Peñaloza, L. Rodríguez, J. E. Cárdenas, V. Minitti, and J. J. Rojas, "Progresos alcanzados en el proyecto universitario cohete sonda ULA," *Universidad, Ciencia y Tecnología*, vol. 13, no. 53, pp. 305–316, 2009. [Online]. Available: https://bit.ly/3f73vB2

[13] L. Lacruz-Rincón, M. A. Parco-Brizuela, R. Santos-Luque, C. Torres-Monzón, J. Ferreira-Rodríguez, and P. Benítez-Díaz, "Análisis experimental de las oscilaciones de presión interna en un motor de combustible solido para cohete sonda," *Ciencia e Ingeniería*, vol. 13, no. 53, 2016. [Online]. Available: https://bit.ly/3noQfdL

[14] Universidad de los Andes. Programa espacial ULA. [Online]. Available: https://bit.ly/35tiodw

[15] S. L. Tolentino Masgo and R. Nakka, "Simulación del flujo supersónico en la tobera del motor cohete Helios-X de categoría amateur," in *Jornadas de Investigación*, 2019. [Online]. Available: https://bit.ly/3pwDY8U

[16] R. Nakka. Richard Nakka's experimental rocketry web site. [Online]. Available: https://bit.ly/2IrDZKX

[17] F. R. Menter, "Two equation eddy-viscosity turbulence models for engineering applications," *Aerospace Research Central*, vol. 32, no. 8, pp. 1598–1605, 2012. [Online]. Available: https://doi.org/10.2514/3.12149

[18] A. Balabel, A. M. Hegab, M. Nasr, and S. M. El-Behery, "Assessment of turbulence modeling for gas flow in two-dimensional convergent–divergent rocket nozzle," *Applied Mathematical Modelling*, vol. 35, no. 7, pp. 3408–3422, 2011. [Online]. Available: https://doi.org/10.1016/j.apm.2011.01.013

[19] S. L. Tolentino Masgo, "Evaluación de modelos de turbulencia para el flujo de aire en una tobera plana," *INGENIUS*, no. 22, pp. 25–37, 2019. [Online]. Available: https://doi.org/10.17163/ings.n22.2019.03

[20] Y. Liu, J. Wu, and L. Lu, "Performance of turbulence models for transonic flows in a diffuser," *Modern Physics Letters B*, vol. 30, no. 25, p. 1650326, 2016. [Online]. Available: https://doi.org/10.1142/S0217984916503267

[21] S. L. B. Tolentino Masgo, "Evaluación de modelos de turbulencia para el flujo de aire en un difusor transónico," *Revista Politécnica*, vol. 45, no. 1, pp. 25–38, abr. 2020. [Online]. Available: https://doi.org/10.33333/rp.vol45n1.03

[22] Y. Zhang, H. Chen, M. Zhang, M. Zhang, Z. Li, and S. Fu, "Performance prediction of conical nozzle using navier–stokes computation," *Journal of Propulsion and Power*, vol. 31, no. 1, pp. 192–203, 2015. [Online]. Available: https://doi.org/10.2514/1.B35164

[23] R. Jia, Z. Jiang, and W. Zhang, "Numerical analysis of flow separation and side loads of a conical nozzle during staging," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 230, no. 5, pp. 845–855, 2016. [Online]. Available: https://doi.org/10.1177/0954410015599798

[24] H. Ding, C. Wang, and G. Wang, "Transient conjugate heat transfer in critical flow nozzles," *International Journal of Heat and Mass Transfer*, vol. 104, pp. 930–942, 2017. [Online]. Available: https://doi.org/10.1016/j.ijheatmasstransfer.2016.09.021

[25] A. K. Mubarak and P. S. Tide, "Design of a double parabolic supersonic nozzle and performance evaluation by experimental and numerical methods," *Aircraft Engineering and Aerospace Technology*, vol. 91, no. 1, pp. 145–156, Dec. 2020. [Online]. Available: https://doi.org/10.1108/AEAT-12-2017-0275

[26] R. H. Pletcher, J. C. Tannehill, and D. Anderson, *Computational Fluid Mechanics and Heat Transfer*. CRC Press, 2012. [Online]. Available: https://bit.ly/3psYKX9

[27] D. Munday, E. Gutmark, J. Liu, and K. Kailasanath, *Flow Structure of Supersonic Jets from Conical C-D Nozzles*. [Online]. Available: https://doi.org/10.2514/6.2009-4005

[28] J. Östlund and B. Muhammad-Klingmann, "Supersonic Flow Separation with Application to Rocket Engine Nozzles ," *Applied Mechanics Reviews*, vol. 58, no. 3, pp. 143–177, 05 2005. [Online]. Available: https://doi.org/10.1115/1.1894402

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.09
pISSN: 1390-650X / eISSN: 1390-860X

Reducing the IoT security breach with a microservice architecture based on TLS and OAuth2

Reduciendo la brecha de seguridad del IoT con una arquitectura de microservicios basada en TLS y OAuth2

Diego Ordóñez-Camacho[1,*]

Received: 15-09-2020, Reviewed: 19-10-2020, Accepted after review: 14-11-2020

Abstract

The Internet of Things has emerged as one of the most promising trends today. The speed of its adoption, however, has caused certain gaps. Amongst the most critical there is the one related with the security of the systems involved. This project addressed the security problem in a broad way but focusing on smart-home environments, where the use of devices with widely heterogeneous technologies and multiple services, generates problems with authentication and with the confidentiality of the data, if the network is compromised. To tackle these problems, state-of-the-art technologies such as OAuth2 and TLS, among others, were put together, along with an architectural methodology of lightly coupled microservices. As a result, a secure and broad range IoT architecture was built, backed up and validated by a reference implementation. The division into functional layers enables both fixed and mobile devices and sensors, to get connected into the system transparently and fluently. The security scheme structured in three incremental levels enables a better device integration, at the level that best adapts to its computing resources and the type of information it shares. The results show the flexibility of the solution and the robustness and novelty of the security scheme presented.

Keywords: IoT, microservices, *software* architecture, systems security, TLS, OAuth.

Resumen

El Internet de las cosas es una de las tendencias más prometedoras en la actualidad. La rapidez de su adopción, sin embargo, ha provocado ciertas brechas críticas en la seguridad de los sistemas involucrados. Este proyecto analizó el problema de seguridad de una manera amplia, pero enfocándose en entornos de tipo hogar inteligente, donde el uso de dispositivos con tecnologías ampliamente heterogéneas genera problemas en la autenticación con múltiples servicios, y en la confidencialidad de los datos, si la red llegara a verse comprometida. Para atacar estos problemas, se juntaron tecnologías de última generación como OAuth2 y TLS, entre otras, junto a una metodología arquitectural de microservicios de acoplamiento ligero, para generar una arquitectura IoT segura y de amplio alcance, respaldada y validada por una implementación de referencia. La división en capas funcionales permite que tanto los dispositivos y sensores fijos como aquellos móviles, puedan acoplarse al sistema de manera transparente y fluida. El esquema de seguridad estructurado en tres niveles incrementales permite que cada equipo pueda integrarse al que mejor se adapte tanto a sus recursos computacionales como al tipo de información que debe entregar o consumir. Los resultados muestran la flexibilidad de la solución y la solidez del esquema de seguridad presentado.

Palabras clave: IoT, microservicios, arquitectura de *software*, seguridad de sistemas, TLS, OAuth

[1,*]Grupo de Investigación en Informática (GrIInf), Universidad UTE – Ecuador.
Corresponding author ✉: dordonez@ute.edu.ec. http://orcid.org/0000-0001-8390-634X

Suggested citation: Ordóñez-Camacho, D. (2021). «Reducing the IoT security breach with a microservice architecture based on TLS and OAuth2». Ingenius. N.° 25, (january-june). pp. 94-103. doi: https://doi.org/10.17163/ings.n25.2021.09.

1. Introduction

The Internet of Things (IoT) is a technology that is strongly getting in real life of people. All environments are involved, namely urban, industrial, office or home. The interest generated and the speed of adoption of the technology have produced certain disorder and informality in the process. Consequently, important elements were put aside, with the security being one of the most relevant [1].

In principle, the security of IoT does not have to be very distant from the security of a typical computer network. However, in practice there are specific difficulties associated to the environment that further complicate the security problem. Many devices for IoT are computationally limited, which hinders using various known robust security mechanisms. The great number of devices that may be involved in an IoT network, as well as the exponential increment in the number of the interactions, aggravates the problem. The diversity of equipment that are utilized, both in hardware and software, complicates the possibility of generalizing the proposed solutions [2]. There is a great variety of methods and tools that may be utilized to do the job [3]. In this work, the intention was to utilize those techniques that, as IoT, set trends and are successfully utilized in other related fields. Among these techniques, the most promising ones were the microservices [4] and OAuth2 [5], which, together with techniques and technologies perhaps a little more traditional but equally successful such as TLS [6] and MQTT [7], provided an appropriate structured environment.

The general objective was to provide the world of IoT with an alternative architecture that is secure and adapted to the new technological trends, which may be utilized in a generic way in multiple situations. More specifically, special relevance was first given to generating an alternative for smart homes, and for this reason the methodology promotes the clear division of functionalities but caring for the fluid integration and the availability of tools for creating new utilities. Second, among the multiple existing security problems, those related with the authentication of clients when calling multiple services and with the confidentiality during the transmission of information were tackled, especially when the network is compromised. At last, aware of the hardware restrictions of many devices, especially sensors, work was carried out to propose an architecture that considers a hierarchy of various levels of security, enabling an interconnection adjusted to the capabilities of various types of equipment, particularly in environments with heterogeneous technologies.

1.1. Related work

In the ecosystem of IoT devices, these are largely insecure, since they are equipment of small size and with low energy consumption; thus, they also have limited computational resources. This latter condition greatly affects when attempting to incorporate in them complex security systems [8]. At the industrial level, there are various entrepreneurships applying IoT, for example, in intelligent transportation [9, 10] and in agriculture [11]; however, one of the main current objectives related to home or office automation is the connected living. This objective requires important advances in the field of IoT, where it is necessary to provide answers to problems related with the enormous increment of devices that should interact [12]. A particular case of IoT is the one corresponding to smart homes, since the implemented solutions are frequently ad-hoc, by the users themselves, who often attempt to minimize costs and efforts, which in general results in a minimum and probably inexistent security scheme [13].

The devices involved require interconnecting in a many-to-many scheme. In order to ensure the information exchange, it is necessary to implement a system for administration of identities that scales appropriately. In this sense, Ayed, Boujezza and Riabi [14] propose a home system which combines EAP, OAuth and DTLS. Also concerned about the administration of identities and the access control, Fernández, Alonso, Marco and Salvachúa [15] confirm the possibilities of OAuth and propose an architecture compatible with services.

Bugeja, Jacobsson and Davidsson [16] analyze the IoT security problem for home systems and remark the importance of avoiding that sensors capture and distribute, indiscriminately, home data. As an example, they present the case of private conversations, which should not be published. Among the possible approaches, they mention a service-oriented approach as a viable alternative to balance between centralization and distribution of the control; furthermore, the trend in distributed software and applications seems generally directed to the use of microservices [17–19]. Delving deeper in this line, Díaz-Sánchez, Marín-López, Almenarez, Arias and Sherrat [20] highlight the benefits that an architecture based on microservices, additionally based on TLS/PKI, may have on IoT, by making lighter the development and maintenance tasks, which is beneficial for both providers and distributors and for users, while simultaneously reinforcing the interconnection security. Case studies such as the one by Urien [21] confirm in a practical manner the possibilities of these techniques.

Although SSL/TLS is the preferred security and encryption mechanism in most of the related works, it is very interesting the proposal of Hoz *et al.* [22] when analyzing the complications that there may be in IoT with TLS at a practical level. Then, such work analyzes the possibilities of using SSH and highlights the advantages provided by the data compression included in such protocol, which shows to be especially

beneficial when working on HTTP.

Khan, Anwar, Azam, Samea and Shinwari [23] contribute to the IoT environment with an approach administered by models, and precisely propose an OAuth oriented model with a strong UML inclination. Through transformations, this proposal might be adapted to a specific architecture, offering the possibility of being customized to the required environment. Another interesting proposal of architecture and security is the one presented by Kim, Wasicek, Mehne and Lee [24], where rather than a traditional mechanism such as the one presented by SSL certification authorities, it would be used an approach to local certification authorities who more frequently, but also in a lighter process, would authenticate the IoT equipment. The proposal by Pahl and Donini [25] may be considered a middle point between the two proposals just reviewed, which uses traditional certificates, but with authentication nearby, rather at the nodes level; they highlight that this mechanism might be complemented with any one of authorization, such as OAuth or similar.

Sciancalepore, Piro, Caldarola, Boggia and Bianchi [26] emphasize on OAuth, but especially with the feature of concentrating their security architecture in the gateway equipment, which houses the base station or sink node, which is in charge of the heavy procedure of authenticating, authorizing and establishing links between clients and resources. This approach is very relevant when it is considered the susceptibility of the edge computing equipment [27], through which a whole IoT system may be compromised. One of the security high points in edge IoT systems is often the one related to the use of MQTT (or similar protocols), and consequently various works such as the one by Singh, Rajan, Shivraj and Balamuralidhar [28] propose improvements on such protocol.

2. Materials and methods

For this work it was mainly considered that within the IoT ecosystem it is necessary to segment the location, range and access of the equipment involved in two layers: local or edge and centralized.

In the local layer, schematically represented in Figure 1, there are those elements that will be invariably installed in the smart household or office, as it is the case of sensors, actuators, edge processors such as gateways and brokers, and user mobile devices.

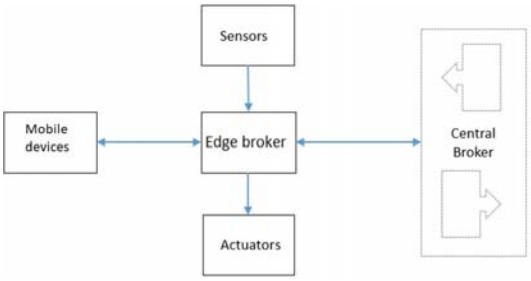


Figure 1. Components in the local or edge layer of the IoT system

The sensors are the equipment capable of capturing physical phenomena, virtual events or periodic signals. Those sensors with the necessary capability may communicate directly with the central broker; otherwise they will interact with equipment in the edge processing subsystem. The sensors will be static, when emitting a constant signal that in general would be used by mobile equipment moving around the environment as Bluetooth beacons for positioning. The dynamic sensors will capture measurements from the environment, which will vary depending on the environmental conditions, such as luminescence, temperature, humidity, among others.

The actuators will enable the interaction with hardware or software, generating events or actions. They will receive instructions either directly from the broker, or from a preprocessor when necessary. They may be local, located in the smart environment, such as light or temperature controllers; they may also be remote such as those capable of sending instructions, probably through the network, for a distant equipment, but controlled from the smart household or office, such as when it is necessary to send an SMS, mail or tweet.

At last, this layer considers the preprocessing equipment, which may be also called processors or edge brokers. They capture raw information coming from the sensors, to resend it to the central broker or to the actuator when the sensor is not capable. The information could be sent as received by the sensor or could be preprocessed and send the result. This equipment may be Raspberry Pi or small computers such as tablets.

In the centralized layer, shown in Figure 2, it is considered the equipment in charge of the general coordination of all the components, and it is considered that three subsystems are necessary: administration, processing and persistence. These subsystems are linked with each other and also with the edge layer through a centralized broker.

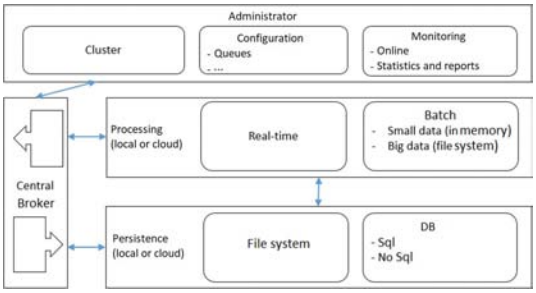


Figure 2. Components in the centralized layer of the IoT system

The administration subsystem defines the parameters, and the configuration of the system presents the web interfaces so that administrative users interact with the whole system. In the case that the central processing is performed with various equipment, it also manages the resulting cluster. Then, a main part in this subsystem is the monitoring, which will enable all type of users to review relevant information, preferably by means of dashboards and statistical charts.

For all heavy information processing, the corresponding subsystem takes the data collected from the broker and process them as defined by the applications or specific requirements of the IoT system. In general, the processing will be divided depending, mainly, on the urgency of the processing, in: real-time, which processes the data in a continuous flow, as they arrive from the sensors; in memory, which collects the information in the cluster memory, depending on the needs, and process it in small batches; and batch, which generally interacts with the storage system, for those processings where the amount of data is larger than the capacity of the living memory of the system.

The information generated by the IoT system is directed to the persistence module, which safeguards the data for further use either in model development or in the generation of reports. Various alternatives should be considered, depending on the size of the information and the way it will be accessed. At least it is necessary to consider an HDFS support, one of database both SQL and NoSQL, and in case that the system evolution requires it, a support in the cloud.

At last, the whole system, and more specifically the edge and centralized layers, should connect to each other and exchange information, which is achieved by means of a central broker, in charge of handling all the message queues thus reducing the complexity of the interactions.

3. Results

The resulting architecture design mainly considered the need of ensuring the exchange of information of all system components, always seeking to guarantee the speed of calculation. These elements require to reconcile characteristics that are often incompatible. For example, more robust cryptography systems may require more computational power than the one provided by many light devices, such as sensors.

The final architecture designed, implemented and tested, is the one sketched in Figure 3, which will be described in detail in the following. First, the components involved are pointed out, and then the security functionality is presented in a general way.

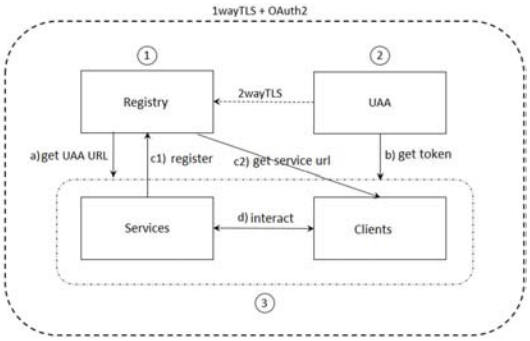


Figure 3. Secure architecture

3.1. Components

This section will attempt to define in a general way the types of components or equipment involved in the architecture proposed in Figure 3, which basically consists of the elements in charge of the registry services (Registry), the equipment providing authentication services for clients and users (UAA) and then, in a broad manner, all equipment providing general services, as well as all client equipment (Services and Clients). Although for simplicity reference will only be made to the components in singular, the architecture considers that, for each category or type of component, especially the services, these can work in clusters.

3.1.1. Registry services (REG)

The main task of the REG is to enable the services to register by means of IP and alias (service name), and make them available to the clients, which will connect to the REG to request information with which they will finally connect to the services of interest. The REG also provides a load balancing service, when detecting that a service is registered in cluster (various equipment with the same service).

The first contact point for all remaining components of the system, whether they are services or clients, is the REG, which requires a static IP; however, all the remaining components of the system may work with dynamic IPs, by means of a DNS.

3.1.2. Authentication services (UAA)

The initials UAA stand for User Authentication and Authorization. Within the proposed architecture, it basically provides the authentication service, which works under OAuth2. The UAA stores the data of all clients in the system, including their roles; with this information the UAA user services may also authorize or not the use of certain elements. Any component may connect with the UAA for requesting an access token, by means of client credentials (user and password). In a similar manner, any component of the system may request to the UAA the validation of the token received by a third party.

3.1.3. Generic services (SRV) and client equipment (CLI)

The last category of components gathers all other services and all clients. In general, these components will interact with each other after being registered/authenticated in the system with the help of the REG and the UAA. The services may be very diverse, and it is in the hands of the system administrator to decide which will be required. However, in the proposed IoT architecture, some are fundamental, and for this reason they have been implemented in the test system and will be mentioned in the following.

To enable interconnectivity and, at the same time, reduce its complexity, it was implemented a messaging service, which in the methodology is represented by the central broker (Figures 1 and 2). The broker is capable of receiving and distributing all messages circulating in the system, and it basically enables that in general all services and clients may establish a unique connection with the broker, to deposit messages and retrieve them from one or more queues. This broker may work with any communication protocol, or a combination of them. However, since in the IoT world, at least at present, the most widely used protocol is the Message Queuing Telemetry Transport (MQTT), it is the one that is used in the implementation presented here.

Other service implemented for testing the concept of the architecture is the one related to the persistence, as a necessary support to further implement batch processing. For this it was implemented a transit service that takes the information from the broker and transfers it to a Hadoop cluster [29], where different types of tools from such ecosystem may be used for information treatment. One of the cases that was worked here, given the nature of the IoT information, especially the one coming from the sensors, was the one of time series. For this purpose, two data series services were built, thus providing graphing and trend analysis, among others.

With respect to the clients, all sensors are considered here, which deliver information to the system, the actuators, which react with the environment thanks to the information of the system, and all those devices, mobile or desktop, that enable the user to enter to configure the system, collect processed information, or even also acting as sensors and actuators. For example, a mobile device may deliver information about the activity of the user or automatically send tweets.

3.2. Security schemes

The system security architecture consists of three fundamental scenarios: basic scheme, to which all elements should stick to in their transactions, unless it is specified otherwise; light scheme, generally used only when starting a work process in the system; strengthened scheme, for trusting relationships between services.

3.2.1. Basic scheme

This is the scheme that the components of the system will use in their transactions by default. This scheme is represented in Figure 3 by the dotted line that encompasses the system and uses a combination of one-way TLS plus OAuth2. All services should provide its public key infrastructure (PKI) certificate to the clients, who could thus validate it with the certification authority (CA). Similarly, all clients should provide an OAuth access token to the services so that they can validate it with the authentication service.

In the proposed scheme, the use of the TLS is especially necessary to be able to encrypt the content of the information being transmitted. It is only utilized on the side of the services to limit as much as possible the overload that may occur, especially, at the administration level (but also at the resources and processing levels), use it in all components. The security gap that appears is compensated with the use of OAuth2, through which the clients are, in turn, validated by the services.

3.2.2. Light scheme

This is a scheme that may be considered insecure, and for this reason it is only provided for those cases where an access token may not be yet obtained, or when it is considered redundant to request it.

Now, two cases exist in the working environment that implement this scheme. When the components enter the system to be able to start their transactions, it is in general necessary to have the OAuth token, but since the IP address of the UAA server may have changed, the first step is to contact the REG to request the updated IP address. For this connection the client does not yet have the token, and for this reason it is not possible to work with the basic scheme. The second implementation of this scheme was applied to avoid unnecessary redundant connections and occurs when the service receives the token and should validate it

with the UAA. Since the service was already initially validated with the UAA, is it prevented to duplicate this step.

It is for this type of cases that the light scheme comes into play. The service provides its PKI with which the communication is encrypted, but the client is not obliged to validate it (although it is recommended to do so), the service provides an access point «insecure» for the client, who does not require the token. This is the scheme provided by REG exclusively to be able to deliver the data of the UAA.

3.2.3. Strengthened scheme

Similar to the problem worked with in the light scheme, sometimes two services require to interconnect, but at least one of them (who acts as client) is not capable of obtaining its access token. Since this is about services, it is not convenient to open an insecure channel as it is done in the light scheme. Then, to maintain the security standard it was decided to implement a two-way TLS scheme, which is possible without incurring in much overload, since as they are services, they anyway have their PKI. In addition, in general the services will be run in equipment with larger processing capacity.

This implementation also requires a dedicated channel to be able to execute this type of validation, and the example is given by the communication between the REG and the UAA. The UAA is the one that provides the access tokens, and therefore it should validate itself which would generate a security hole. Then, the REG opens a dedicated channel so that at any moment a UAA service may register that way. As the UAA connects with the REG, they exchange their corresponding PKIs, validating each other by TLS, without reducing the security standard of the system.

3.3. Functionality

Referring again to Figure 3, the dotted line represents the scope of the basic security scheme, which encompasses the whole system. Numbers 1, 2 and 3 may be seen inside, enclosed in circles, which indicate the recommended starting sequence to guarantee the fluency of the service. In practice, at least the services have in their base library the functionality to retry the connection when this starting sequence is not followed. However, this may be susceptible to unnecessary delays.

First the registry server REG is started, which will provide a central access point for acquiring the contact information for the other services: all services will register in the REG their corresponding IP and alias (name of service) and all clients will look for here, by alias, the IP of the required service to be able to connect to it. REG offers three access points, each of which should handle a different security mode: the first, light, enables any client to obtain the UAA IP without additional security; the second mode, strengthened, enables the connection of the UAA by means of a two-way TLS; the last, basic, which requires OAuth2, enables clients to request information about services, and services to register their contact information.

Second, an authentication server (UAA) is started, which will provide OAuth2 credentials to the clients. A reinforced security mechanism is used between the UAA and the REG, with two-way TLS validation. The UAA connects with the REG, as well as with any other service, to deliver its IP and alias and thus being available for the whole system. Once these two services, REG and UAA, are online, all the remaining components, services and clients may start their work.

Then, at last, as point 3, any other component, either service or client, will proceed as follows: first, using the light security scheme, they will connect with the REG to request the IP address of the UAA; they establish the connection with the UAA and request the access token, using their client credentials. With the access token at hand the basic security scheme may be already utilized and, in the case of services, they will register with the REG, delivering IP and alias, and be ready to wait for the requests of the clients, or act as a client of another service, as necessary. In the case of a client, the next step is using a service, where the basic security scheme will be applied; it is connected to the REG and by means of an alias it requests the IP of the service of interest, to further connect with such service. The last action becomes that of the services that receive a request from a client, since in this case they should, by means of the light scheme, connect to the UAA and request the validation of the access token, with which they can serve the request of the client.

3.4. Implementation and tests

The implementation decision was mainly that each of the components may be run in a variety of equipment with the lowest interdependency, for which it was proceeded to work in a microservices architecture that enables their deployment either as independent processes, or within a container structure, such as Docker [30]. For desktop services and clients, it was used Java with Spring Boot in general. For the central messaging service, it was decided to work using the MQTT protocol, and for the development of the broker the Moquette library [31] was taken as base, which was modified mainly for adding support for OAuth2. For the persistence services an HDFS cluster [32] was used as base in the local network, and time series services were built on it which, according to the current interests of the project, were the more appropriate to process data coming from the sensors. It was interacted with OpenTSDB [33] and Prometheus [34].

Regarding clients, it was decided to construct generic libraries for different types of systems, which ease the process of developing specific applications. A Java client library was developed, one for Android mobile devices, one for Arduino MKR [35] and another for ESP32 [36], the last three based on Eclipse Paho [37]. The Arduino libraries were exclusively oriented for their use in controllers of sensors and actuators.

The tests of the system were carried out in a controlled office environment. The main equipment was an RPi 3B+ which acted as a Wi-Fi Gateway, providing the DNS and NTP services, among others. Controllers MKR 1010 and ESP32 were simultaneously used, which permanently received information from temperature, humidity and noise sensors, such as the DHT11 and the KY038 [38]. The RPi3B+ also housed the REG, and UAA services and the MQTT messaging broker. The persistence services on HDFS, as well as those of the TSDB were installed in Linux on an i5-4210U with 8 GB of RAM memory. In this last equipment generic services were also installed for sending and receiving messages, with which it was verified the fluency of the interaction. A N9005 was used as mobile client, which had two functionalities: bobo sensor, sending a great quantity of random numbers to the system, and light actuator, notifying every time that the measurements of the real sensors exceeded particular levels parameterized in the app. Figure 4 presents one of the setups of the test system.

In the experimentation relative to the security tests, it was decided to suppose that a possible attacker was already connected to the local network and was potentially in the capacity of making any request and capturing all the traffic. In the first test, Zap was used to perform a scanning both active and passive, and it was verified that the security was maintained (encrypted traffic), except in the case (documented in the architecture) of the light scheme.

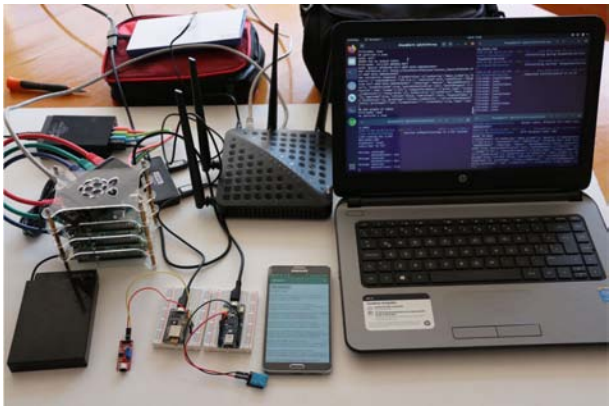


Figure 4. Setup for tests: from left to right it is observed a cluster of 3 Raspberry Pi running all services on Docker; an ESP8266 monitoring the level of noise with an KY038; an MKR1010 monitoring ambient temperature with an DHT11; an N9005 injecting random messages, and a laptop monitoring all services.

Then, Wireshark was used for a manual verification, analyzing the packets by type of connection or scheme, and it was again validated, as expected, that, except for the light scheme, all traffic was kept encrypted, as shown in Figure 5.

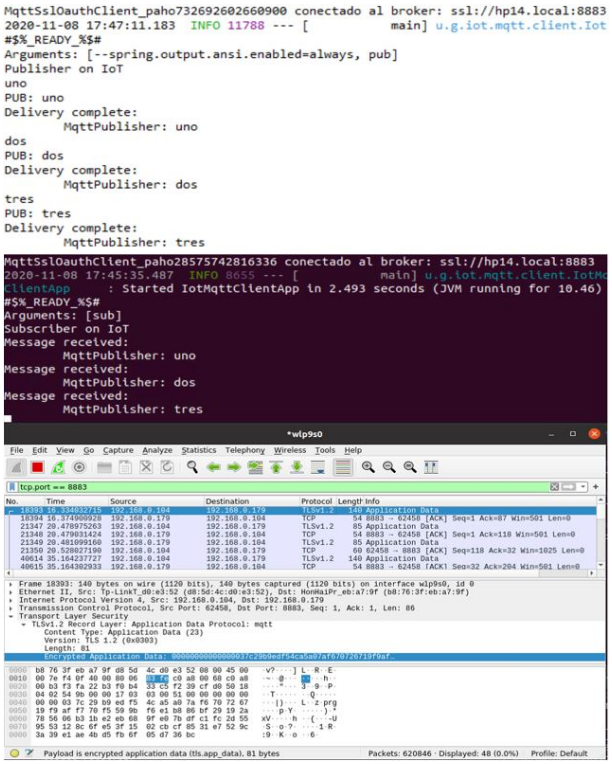


Figure 5. Run and traces: from top to bottom, first, a client sends three simple text messages; then the receiver obtains the messages and, at last, the Wireshark trace captures the packages and verifies that, during transit, they are encrypted.

4. Discussion

The process of including complex security schemes in light IoT devices, is not trivial as stated by Khan and Salah [8], and this work coincides with this statement. Two remarkable cases were that managing the TLS with self-generated certificates for MKR 1010 require regenerating all the firmware to include the CA, and it was not possible to carry this out in ESP8266 controllers due to the unavailability, in practice, of open-source libraries sufficiently complete to guarantee the expected security level.

In general, the adopted approach retakes the warnings made by Lin and Bergmann [13], and attempts to provide a sufficiently complete system so that with minimum technical support it could be implemented at home, and then directly administered by the user.

Clearly, this exhibits limitations given by the great variety of types of users that may want to be included in the IoT. However, the tests carried out indicate that the heart of the prototype would enable providing this, if it is possible to include some features in the user interface, equipment and installers.

This work confirms what was presented by Ayed *et al.*, [14] and by Fernández *et al.*, [15] regarding the problem of scaling identities and the benefit that may be obtained both with OAuth and with a services approach. Delegating the authentication to a single point, further trusting in temporary credentials, as enabled by OAuth2, maintains a light infrastructure of identities, while the distribution in services provides a great easiness when replicating and redounding services in a cluster when the work load of the system requires it. It is just this services/microservices approach which enables that working with TLS is not too demanding at the level of maintenance, as highlighted by Díaz-Sánchez *et al.* [20].

This proposal acknowledges the importance that should be given to the edge equipment at the security level, and following the line of Shapsough *et al.* [27] and of Sciancalepore *et al.* [26], it especially reinforces the gateway equipment to further enable implementing improvements in the MQTT protocol, similar to what was made by Singh *et al.* [28] but mainly including the use of OAuth2 and TLS.

5. Conclusions

This work was motivated by a current urgent need: to secure the IoT systems, especially those linked to a home environment, and making it without undermining the freedom of access of the user to collect information and to modify the configuration of the system. As it was seen, this need comes from the relative informality of the IoT, among other things, especially in the environment of a smart home.

It started with a methodological conception that stratifies the environment in layers: the one more closely related with the interaction with the space through the collection of information and execution of actions to modify the microenvironment, and the layer of centralized processing and analysis of information. Both layers were interrelated with a centralized connection that unifies them also maintaining their light coupling.

This methodology was oriented to enable an implementation based on microservices, where, besides avoiding to the extent possible any type of monolithic structure, it was provided a group of services and libraries for clients which ease to a large extent the generation of new utilities and components. The tests carried out with the services, clients and sensors generated under this infrastructure enabled confirming both the solidity and the relative easiness of use of the components.

The main point, the security, although it was not the easiest to implement, after debugging the three schemes defined for the different types of connection, it demonstrated being a robust choice, which withstood the illicit access tests. However, it is worth indicating that a methodical testing scheme specific for the field is still pending to be designed and applied, which will be considered in future works. Nevertheless, it may be stated that combining TLS, OAuth2 and MQTT produced the expected results to a large extent.

As main contributions it could be mentioned the implementation of a solid secure IoT architecture for households; the stable and fluent combination of at least three high level technologies for handling security and an implementation of functional reference that can be made publicly available for free use.

References

[1] Y. Lu and L. D. Xu, "Internet of things (IoT) cybersecurity research: A review of current research topics," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2103–2115, 2019. [Online]. Available: https://doi.org/10.1109/JIOT.2018.2869847

[2] A. Riahi Sfar, E. Natalizio, Y. Challal, and Z. Chtourou, "A roadmap for security challenges in the internet of things," *Digital Communications and Networks*, vol. 4, no. 2, pp. 118–137, 2018. [Online]. Available: https://doi.org/10.1016/j.dcan.2017.04.003

[3] P. Lea, *Internet of Things for Architects: Architecting IoT solutions by implementing sensors, communication infrastructure, edge computing, analytics, and security.* Packt Publishing Ltd, 2018. [Online]. Available: https://bit.ly/3oJ1XRl

[4] P. Jamshidi, C. Pahl, N. C. Mendonça, J. Lewis, and S. Tilkov, "Microservices: The journey so far and challenges ahead," *IEEE Software*, vol. 35, no. 3, pp. 24–35, 2018. [Online]. Available: https://doi.org/10.1109/MS.2018.2141039

[5] J. Khan, J. p. Li, I. Ali, S. Parveen, G. a. Khan, M. Khalil, A. Khan, A. U. Haq, and M. Shahid, "An authentication technique based on oauth 2.0 protocol for internet of things (IoT) network," in *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2018, pp. 160–165. [Online]. Available: https://doi.org/10.1109/ICCWAMTIP.2018.8632587

[6] C. Chan, R. Fontugne, K. Cho, and S. Goto, "Monitoring tls adoption using backbone and edge traffic," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2018, pp. 208–213. [Online]. Available: https://doi.org/10.1109/INFCOMW.2018.8406957

[7] F. Izquierdo, M. Ciurana, F. Barcelo, J. Paradells, and E. Zola, "Performance evaluation of a TOA-based trilateration method to locate terminals in WLAN," in *2006 1st International Symposium on Wireless Pervasive Computing*, 2006, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ISWPC.2006.1613598

[8] M. A. Khan and K. Salah, "IoT security: Review, blockchain solutions, and open challenges," *Future Generation Computer Systems*, vol. 82, pp. 395–411, 2018. [Online]. Available: https://doi.org/10.1016/j.future.2017.11.022

[9] J. P. Rojas, J. C. Bustos, and D. Ordóñez Camacho, "Smart public transportation at your fingertips," *Enfoque UTE*, vol. 8, no. 1, pp. 122–134, Feb. 2017. [Online]. Available: https://doi.org/10.29019/enfoqueute.v8n1.143

[10] J. P. Rojas, J. C. Bustos, and D. Ordóñez-Camacho, "Qbus: Movilidad inteligente para el usuario de transporte público," in *Proceedings of the International Conference on Information Systems and Computer Science, INCISCOS 2016*, 2016. [Online]. Available: https://bit.ly/3jZlBpE

[11] E. A. Q. Montoya, S. F. J. Colorado, W. Y. C. Muñoz, and G. E. C. Golondrino, "Propuesta de una arquitectura para agricultura de precisión soportada en IoT," *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, pp. 39–56, 2017. [Online]. Available: http://dx.doi.org/10.17013/risti.24.39-56

[12] M. Agiwal, N. Saxena, and A. Roy, "Towards connected living: 5g enabled internet of things (IoT)," *IETE Technical Review*, vol. 36, no. 2, pp. 190–202, 2019. [Online]. Available: https://doi.org/10.1080/02564602.2018.1444516

[13] H. Lin and N. Bergmann, "IoT privacy and security challenges for smart home environments," *Information*, vol. 7, no. 3, p. 44, Jul 2016. [Online]. Available: http://dx.doi.org/10.3390/info7030044

[14] H. Kaffel-Ben Ayed, H. Boujezza, and I. Riabi, "An idms approach towards privacy and new requirements in IoT," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017, pp. 429–434. [Online]. Available: https://doi.org/10.1109/IWCMC.2017.7986324

[15] F. Fernández, A. Alonso, L. Marco, and J. Salvachúa, "A model to enable application-scoped access control as a service for IoT using OAuth 2.0," in *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, 2017, pp. 322–324. [Online]. Available: https://doi.org/10.1109/ICIN.2017.7899433

[16] J. Bugeja, A. Jacobsson, and P. Davidsson, "On privacy and security challenges in smart connected homes," in *2016 European Intelligence and Security Informatics Conference (EISIC)*, 2016, pp. 172–175. [Online]. Available: https://doi.org/10.1109/EISIC.2016.044

[17] L. Sun, Y. Li, and R. A. Memon, "An open IoT framework based on microservices architecture," *China Communications*, vol. 14, no. 2, pp. 154–162, 2017. [Online]. Available: https://doi.org/10.1109/CC.2017.7868163

[18] T. Vresk and I. Çavrak, "Architecture of an interoperable IoT platform based on microservices," in *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2016, pp. 1196–1201. [Online]. Available: https://doi.org/10.1109/MIPRO.2016.7522321

[19] R. Yu, V. T. Kilari, G. Xue, and D. Yang, "Load balancing for interdependent IoT microservices," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 298–306. [Online]. Available: https://doi.org/10.1109/INFOCOM.2019.8737450

[20] D. Díaz-Sánchez, A. Marín-Lopez, F. A. Mendoza, P. A. Cabarcos, and R. S. Sherratt, "TLS/PKI challenges and certificate pinning techniques for IoT and M2M secure communications," *IEEE Communications Surveys Tutorials*, vol. 21,

no. 4, pp. 3502–3531, 2019. [Online]. Available: https://doi.org/10.1109/COMST.2019.2914453

[21] P. Urien, "Securing the IoT with TLS/DTLS server stacks embedded in secure elements: An ePlug usecase," in *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*, 2017, pp. 569–570. [Online]. Available: https://doi.org/10.1109/CCNC.2017.7983170

[22] J. D. Hoz, J. Saldana, J. Fernández-Navajas, J. Ruiz-Mas, R. G. Rodríguez, and F. d. J. M. Luna, "SSH as an alternative to TLS in IoT environments using HTTP," in *2018 Global Internet of Things Summit (GIoTS)*, 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/GIOTS.2018.8534545

[23] M. Khan, M. W. Anwar, F. Azam, F. Samea, and M. F. Shinwari, *A Model-Driven Approach for Access Control in Internet of Things (IoT) Applications – An Introduction to UMLOA.* Communications in Computer and Information Science, Springer, 2018, vol. 920. [Online]. Available: https://doi.org/10.1007/978-3-319-99972-2_16

[24] H. Kim, A. Wasicek, B. Mehne, and E. A. Lee, "A secure network architecture for the internet of things based on local authorization entities," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, 2016, pp. 114–122. [Online]. Available: https://doi.org/10.1109/FiCloud.2016.24

[25] M. Pahl and L. Donini, "Securing IoT microservices with certificates," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, 2018, pp. 1–5. [Online]. Available: https://doi.org/10.1109/NOMS.2018.8406189

[26] S. Sciancalepore, G. Piro, D. Caldarola, G. Boggia, and G. Bianchi, "Oauth-iot: An access control framework for the internet of things based on open standards," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, 2017, pp. 676–681. [Online]. Available: https://doi.org/10.1109/ISCC.2017.8024606

[27] S. Shapsough, F. Aloul, and I. A. Zualkernan, "Securing low-resource edge devices for IoT systems," in *2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI)*, 2018, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ISSI.2018.8538135

[28] M. Singh, M. A. Rajan, V. L. Shivraj, and P. Balamuralidhar, "Secure mqtt for internet of things (IoT)," in *2015 Fifth International Conference on Communication Systems and Network Technologies*, 2015, pp. 746–751. [Online]. Available: https://doi.org/10.1109/CSNT.2015.16

[29] C. Singh and M. Kumar, *Mastering Hadoop 3: Big data processing at scale to unlock unique business insights.* Packt Publishing, 2019. [Online]. Available: https://bit.ly/37Qi2O9

[30] J. Turnbull, *The Docker Book: Containerization is the new virtualization*, 2014. [Online]. Available: https://bit.ly/3m7nqRY

[31] A. Selva. (2014) Java MQTT lightweight broker. moquette. [Online]. Available: https://bit.ly/3gB82Mw

[32] M. Bhushan, *Big Data and Hadoop: Learn by Example.* BPB Publications, 2018. [Online]. Available: https://bit.ly/2W0AmP1

[33] T. Dunning and E. Friedman, *Time Series Databases: New Ways to Store and Access Data, Edition: 1. Sebastopol.* O'Reilly Media, Inc, 2014. [Online]. Available: https://bit.ly/2W1VnsU

[34] B. Brazil, *Prometheus: Up & Running: Infrastructure and Application Performance Monitoring.* O'Reilly Media, 2018. [Online]. Available: https://bit.ly/39V80xX

[35] A. Kurniawan, *Arduino MKR WIFI 1010 Development Workshop.* PE Press, 2018. [Online]. Available: https://bit.ly/37OEnvD

[36] I. Dogan and I. Ahmet, *The Official ESP32 Book.* Elektor International Media, 2017. [Online]. Available: https://bit.ly/2IzEW3G

[37] G. C. Hillar, *Hands-On MQTT Programming with Python: Work with the lightweight IoT protocol in Python.* Packt Publishing, 2018. [Online]. Available: https://bit.ly/33YpdTg

[38] B. Charles, *Beginning Sensor Networks with Arduino and Raspberry Pi.* Apress, 2013. [Online]. Available: https://bit.ly/3m5syGj

Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n25.2021.10
pISSN: 1390-650X / eISSN: 1390-860X

Vulnerability analysis with SQLMAP applied to APEX5 context

Análisis de vulnerabilidades con SQLMAP aplicada a entornos APEX 5

Esteban Crespo-Martínez[1,*]

Received: 14-09-2020, Reviewed: 01-10-2020, Accepted after review: 30-11-2020

Abstract

Databases are usually the main targets of an attack, specifically for the information that they store, since, according to Druker, information is power. In this work vulnerability tests are performed of the database of an ERP software developed in APEX 5. For this purpose, FOSS tools are used to test and analyze vulnerabilities of databases, identifying that sessions used by ERP based on Oracle APEX are carried out randomly, and besides are generated again at particular times. It is therefore concluded that, with the tests applied and the updates of SQLMAP to the date of the experiment, it has not been possible to vulnerate the ERP software with SQL injection techniques.

Keywords: APEX, Data protection, Information systems evaluation, SQL Injection.

Resumen

Las bases de datos son usualmente los principales objetivos de un ataque, específicamente por la información que en ella reside, ya que, de acuerdo con Druker, la información es poder. En este trabajo se realizan las pruebas de vulnerabilidad de la base de datos de un *software* ERP desarrollado en APEX 5. Para ello, se utilizan herramientas FOSS de prueba y análisis de vulnerabilidades de bases de datos, identificando que las sesiones que utiliza ERP basada en Oracle APEX son realizadas de manera aleatoria y que, además, son nuevamente generadas en determinados momentos. Se concluye que, con las pruebas aplicadas y las actualizaciones de SQLMAP a la fecha del experimento, no se ha conseguido vulnerar el *software* ERP con técnicas de inyección SQL.

Palabras clave: APEX, evaluación a sistemas de información, inyección SQL, protección de datos

[1,*]Universidad del Azuay, Ecuador. Corresponding author ✉: ecrespo@uazuay.edu.ec.
https://orcid.org/0000-0002-3061-9045

Suggested citation: Crespo-Martínez (2021). «Vulnerability analysis with SQLMAP applied to APEX5 context». Ingenius. N.° 25, (january-june). pp. 104-113. doi: https://doi.org/10.17163/ings.n25.2021.10.

1. Introduction

Various experts in information security agree that cyber-attacks are increasingly recurrent, and usually target web systems, thus altering or putting on risk personal information [1], especially to web applications [2] , due to their complexity, extension, high personalization and because they are usually developed by programmers with little experience in security [3].

It cannot be denied that, in this society of information and knowledge, databases contain the gold mine, which becomes one of the most important strategic elements of the organization, because from its analysis and interpretation at the right time it is possible to project strategies to stay ahead of opportunities and foresee threats according to the role of the organization in the society.

The protection of the information started to make sense when the first computer viruses appeared: they altered or erased user information, often with the purpose of demonstrating the destructing and creative capacity of the designer of the malware used to undertake the attack. The computing context was simpler, there was no intercommunication between organizations and systems were limited to operate in a centralized manner [4]. However, as a result of the boom of opportunities generated by the appearance of the Internet, attackers now see data in a different manner. Damaging or eliminating them does not make sense, data theft, copy or seizure are aspects that become the new targets.

Ojagbule *et al.* [5] mention that, today, there are more than one billion websites, and that many of them are developed by content managers such as Drupal, Joomla or WordPress, and that, according to Mohammadi y Namadchian [6], they contain important data.

According to Ojagbule *et al.* [5] and Kruegel *et al.* [6], due to the existence of a large number of sites, there is also a large number of databases subject to vulnerabilities and risks. Thereof it appears a technique known as SQL injection (SQLIA, Structured Query Language Injection Attack), which according to Santin, Oliveira and Lago [7] citing [8] and [9], is a technique where an attacker explores vulnerabilities that enable altering the SQL commands in an application, which is known as one of the vulnerabilities that generate greater impact in the organization.

Nofal and Amber [9] add that this technique usually does not have predictable or specific patterns, which becomes an important problem for researchers and developers. Badaruddin [10] concludes that the SQL injection technique is the second most common error found in web servers in Internet, with around 44.11 %.

With the purpose of discovering security failures regarding SQL injection vulnerabilities, in this work it is carried out an evaluation using SQLMAP of an Oracle database that stores information of the UDA-ERP system developed by the Universidad del Azuay on APEX 5. This paper is divided in the following sections: i) state of the art, where some concepts are established, as well as related works; ii) the methodology applied for obtaining the results, detailing the configurations made in the laboratory test equipment; iii) the results obtained after executing the tool; iv) the discussion about the results obtained and v) the conclusions and future works.

1.1. The SQL injection

According to OWASP, the SQL injection is one of the ten most dangerous and popular vulnerabilities that may appear in web environments [11], which in general are difficult to protect due to their high personalization, complexity, scale [3], technology and development by programmers with little experience in security [3] [12], causing serious damages to the businesses of the victims [13]. In addition to this, Setiawan and Setiyadi [14] state that, in a computer networks context, any existing data in a computer connected to another computer becomes insecure.

Authors Santin, Oliveira and Lago [7] citing [15] state that there are no solutions that guarantee or resolve all vulnerabilities which occur at the hardware and software levels, statement also supported by Setiawan [14]. They also add that, since many elements are not constantly updated, they are more prone to cyber-attacks. On the other hand, Kals *et al.* [12] state that there are multiple vulnerabilities to the security of web applications, as a result of generic problems of input validation. In addition, vulnerabilities may be kept secret or reported by manufacturers, either publicly or privately [16].

An SQL injection attack may be basically represented as indicated in Figure 1.

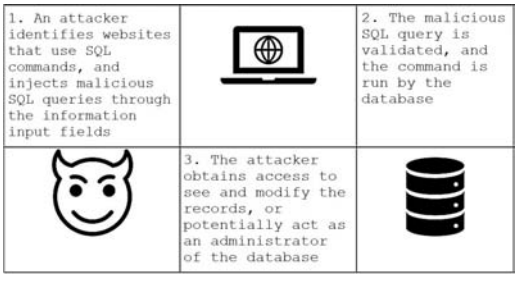


Figure 1. SQL injection process

Another way of representing the sequence of attacks is the one proposed by AVI Networks [17], which is presented in Figure 2.

Figure 2. SQL Injection attack sequence. Source [17]

According to Charania and Vyas [18], the SQL injection attack techniques may be classified as follows:

i) Tautologies, a type of attack that uses conditional queries and inserts SQL tokens in them, demonstrating to be always true.

ii) Illegal or logically incorrect queries, where the attackers use the error messages of the databases to find vulnerabilities in the applications.

iii) Queries with UNION, where the attackers inject infected queries over secure queries using the UNION operator and, therefore, recover information from the database.

iv) Queries with support or Piggy-backed: the attackers attach delimiters such as ";" to the original query and run them simultaneously, with the first being legitimate and the remaining false, but returning valuable information.

v) Stored procedures, a subset of precompiled queries, depending on which they are there will be different forms of attack.

vi) Blind injection, in which the developers hide error messages that may be useful for attackers to plan and execute an SQLIA attack. In this situation the attacker finds a static page, where true and false questions are made using SQL commands until the objective is attained.

vii) Timed attacks, which enable the attacker to observe the time required to execute a query. The attacker generates a big query using if-else sentences and, in this way, measures the amount of time spent by the page to load and determine if the injected sentence is true.

viii) Alternative coding, where ASCII and Unicode coding enable to evade the filter which scans "special characters" [19].

An evaluation of vulnerabilities by SQL injection may be undertaken with the use of technological tools for such purpose. Novaski [20] suggests the use of FOSS tools, of which 14 are proposed to be used: Arachni, Beef, Htcap, IronWASP, Metasploit, Skipfish, SQLMap, Vega, W3af, Wapiti, Wfuzz, XSSer, Xenotix and ZAP.

From this work it adds that only the tools IronWASP, Vega, ZAP and SQLMap detected the vulnerability of SQL injection, while the reflected XSS vulnerability was only detected by the tools ZAP and Xenotix.

It is indicated in their work that it was only possible to conduct a complete intrusion test in the SQL injection vulnerability, and it was necessary to apply three different tools for carrying out such test: i) wapitigetcookie, to obtain the session identifier; ii) Htcap to obtain points of input; and SQLMap to detect and explore the vulnerability.

Among the related works there are the ones shown in Table 1. This work, as opposed to those cited works, focus on testing security aspects in an application developed in Oracle APEX 5.

It is clear that, despite the time elapsed from the first time that the SQL injection attack appeared two decades ago [21], both the injection and the evasion and mitigation techniques are numerous. Information technologies are increasingly common in our environment and have notably affected our lifestyle because every time that the use and reliability of computers and computer systems increase, the threat on sensitive data also increases.

SQL injection vulnerabilities in web applications are surprisingly vast and are definitely a big threat for the security of the personal data stored in the web [21].

In practice, Cetin *et al.* [22] demonstrate that a GitHub automatic analysis shows that 15.7 % of the 120412 Java source files published contain code vulnerable to SQL identifier injection attacks (SQL-IDIA), also pointing out that, after a manual revision, they proved that 18939 Java files identified during the automatic analysis are vulnerable to this type of attacks.

Puneet [23] classifies SQL injection in two types: i) classic SQL injection and ii) advanced SQL injection.

Table 1. Works related with the use of SQLMap

Id.	Autors	Topic	Year	Objective
1	Nofal D.E. Amer A. A. [9]	SQL Injection Attacks Detection and Prevention Based on Neuro-Fuzzy Technique	2020	Conduct a work to detect and prevent SQL injection attacks, applying a fuzzy logic inference system.
2	O. Ojagbule H. Wimmer R. Haddad [24]	Vulnerability Analysis of Content Management Systems to SQL Injection Using	2018	Compare the vulnerabilities of SQL injection in the three most widely used content managers, considering the Nikto and SQLMap tools for such purpose.
3	F. Santin J. A. Oliveira V. Lago Machado [7]	Uso da ferramenta SQLMap para detecção de vulnerabilidades de SQL Injection	2017	Focus on describing the main risks to which web applications are subject, related to the SQL injection. They use the SQLMap tool for such purpose.
4	Badaruddin Bin Halib, Edy Budiman, Hario Jati Setyadi [10]	Técnicas de pirateo de servidores web con SQLMap en Kali Linux	2017	Propose a technique for hacking web servers using SQLMap in Kali Linux.
5	S. D. Axinte [25]	SQL injection Testing in Web Applications Using SQLMap	2014	The author conducts an analytical analysis of the SQL injection technique, and presents methods, tools and prevention actions.
6	Barinas, Alarcón, Callejas [1]	Vulnerabilidad de ambientes virtuales de aprendizaje utilizando SQLMap, RIPS, W3AF y Nessus*	2014	Analyzes the security aspects of virtual learning environments, security and vulnerability analysis tools.
7	A. Tajpour, S. Ibrahim, M. Masrom [26]	SQL Injection Detection and Prevention Techniques	2015	Propose attack and mitigation techniques against SQL injection attacks, comparing various types of them.

1.1.1. Classic SQL injection

The basic injection techniques, suggested by [23] are summarized as follows:

a) Piggy Backed Queries

The intention of the attack is primarily the denial of service. The database receives multiple queries in which, during the execution, the normal query operates as in a normal case, while the second query adheres to the first to attain the attack. An example of this attack may be the following:

```
select cliente from cuentas where
login_id = "admin" AND pass = ‘123’;
DELETE FROM accounts WHERE
ClienteNombre = ’Francisco’;
```

After the execution of the first query, the interpreter detects the semicolon ";" and executes the second query together with the first, eliminating all the data of the client "Francisco". This type of malicious data may be protected by first determining the correct SQL query by means of the appropriate validation or using appropriate detection techniques, as it is the static analysis, which does not need the supervision of the run time.

b) Stored procedure

The intention of attack is summarized as escape authentication and denial of service. Mistakenly, IT professionals think that the stored SQL procedures are a remedy for the SQL injection [17], since they are placed in front of the databases and the security characteristics are not directly applicable. The stored procedures do not use standard structured query language, they use their own script languages that do not have the same vulnerability as SQL, but keep other diverse vulnerabilities related with the scripting language. For example, it may be indicated the following:

```
CREATE PROCEDURE Info_usuario @usuario
varchar2 @password varchar2 @idcliente
int AS BEGIN EXEC(‘Select info_cliente
from tabla_cliente where username=’
"+@usuario " ’ and pass = ’
"+@password " ’ GO
```

Any malicious user may enter malicious data in the username and password fields. A simple entered command may destroy the whole database or cause a denial of service. In this way, [23] suggests that critical information is not collected in the stored procedures.

c) Union query

It is a type of attack that uses the union operator (U) while inserting the SQL query. The two SQL queries, normal and harmful, are joined together using this operator. The example shows how it is proceeded, visualizing that the second query is malicious and the following text (-) is not taken into account, since it is converted to a comment by the SQL Analyzer.

```
select * from cuentas where id=’212’
UNION select * from factura where
usuario=’admin’-’ and password=’pass’
```

1.1.2. d) Alternative coding

With respect to this type of attack, the attacker changes the SQL injection pattern so that it is not detected by common detection and prevention techniques. In this method, the attacker uses hexadecimal, Unicode, octal and ASCII code representation in the SQL instruction, to avoid being detected due to the use of coded chains.

1.1.3. Advanced SQL injection

The advanced SQL injection techniques suggested by [23], are summarized as follows:

a) Deep Blind SQL Injection Attack

In a great number of web applications the visualization of mysql errors or another SQL is disabled. In this attack, the information is inferred by means of true/false questions. If the injection point is absolutely blind, then the only way to attack is through the use of the WAIT FOR DELAY or BENCHMARK [23] command.

b) Fast flux SQL Injection Attack

The objective of the attack is the extraction of data or the identity theft through phishing. A host that carries out phishing may be easily detected by tracking its IP address or through the identification of its domain name. However, according to [23] and [27], the protection systems of many web hostings may suspend the service due to the massive traffic generated, thus cancelling out the criminal purposes. In this manner, to avoid this problem, attackers apply technique of the Fast Flux, which is a DNS technique to hide the phishing and malware distribution sites behind a network of constant change.

1.1.4. c) Compounded SQL injection attacks

It is a mix of two or more attack techniques, generating an effect greater than the indicated with the previously described techniques. Compounded SQL injection, as it is known in the dark world, derives from the mix of the SQL attack and other web applications attacks as, for example, the SQL injection attack + the distributed denial of service (DDoS) attacks. Based on what is exposed by [23] citing [28], the code to perform this type of attack would be:

```
http://exploitable-web.com/link.php?id=1’
union select 1,2,tab1,4 from
(select decode(encode(convert
(compress(post)
using latin1),
des_encrypt
(concat(post,post,post,post),8)),
des_encrypt(sha1(concat(post,post,
post,post)),9))
as tab1 from table_1)a-
```

Another way of combining an attack is mixing an SQL injection with insufficient authentication. This attack is exploitable when the parameters of security have not been initialized where the application fails when identifying the location of the user, the service or the application. This enables the attacker to access to privileged information without verifying the identity of the user.

In this manner, this type of attack is relatively simpler than with any other type of attack [23], where the first step is locating a website that has insufficient authentication.

2. Materials and methods

The purpose of the security analysis was to evaluate the security of an application developed in Oracle APEX, platform which es being developed in the UDA ERP software of the Universidad del Azuay. With this premise, it was configured a laboratory considering the materials and methods which are described in the following.

For the tests, it was considered the KALI suite and it was utilized the SQLMap tool, which is based in Free and Open Source, developed under a GNU GPLv2 license by Miroslav Stampar and Bernardo Damele, considering that, according to Charania and Vyas [18], SQLMap supports, among others, the Oracle database, which is the one used by the UDA ERP software.

It also states that SQL support six injection techniques: Boolean-based blind, time-based blind, error-based, UNION query-based, out-of-band and stacked queries.

The previously mentioned techniques take part in the testing parameters which are included in SQLMap, which are automatically applied. Using BurpSuite, the capturing of session cookies is adjusted, applying the configuration of a local proxy (127.0.0.1:8080) with the purpose of capturing the POST requests, which will be further used with SQLMap. Before executing the tests, the dependencies and packages of the suite were updated.

Acknowledging that SQLMap is a tool that enables exploring database servers, for its use it is important to point to the URL address that contains the SQL script. The structure "sqlmap -u URL –[parameters]" is the common sentence, where the -dbs parameter will enable obtaining the database. After detecting the vulnerability, it should be used the -D parameter and the name of the database which will be analyzed. If the result obtained is effective, the -tables parameter will enable recovering all the tables from the specified database.

With the purpose of identifying the vulnerabilities, three tests were conducted, in each of which the analysis was increased and aspects such as level of aggressiveness in the tests, use of cookies of established sessions and evasion to identification systems were varied.

The first test consisted in listing the databases; in the second, it was increased the degree of aggressiveness and the number of tests to obtain information from the databases; and in the third attack it was pretended to use a random agent evading the proxies with the unique purpose of capturing a session cookie, element which is used as the base for automatic tests, in which a valid session of an active user is simulated.

3. Results

3.1. First test

The evaluation of vulnerabilities of the database by executing the command root@kali: /sqlmap-dev# python3 sqlmap.py -u –dbs, gave the result expressed in Table 2, considering that the attack generated its own cookie for evaluation: ('USUARIO=ORA_WWV-cUs...UNNyk6flfB'). The tests starts at 13:03:52 on 2020-03-04 and ends at 13:04:22 /2020-03-04/

Table 2. Results of the first test

Test	Description
Heuristic analysis	The heuristic analysis detected that the target is protected by some type of WAF/IPS.
URL Content	The 'p' parameter of the GET method does not seem to be dynamic. The basic heuristic tests conducted indicate that the 'p' parameter is not injectable.
SQL injection	It is not vulnerable.
Test 'AND boolean-based blind-WHERE or HAVING clause'	Reflective values found and filtered out. It means that there are "reflective" values within the response that contains (parts of) the useful load. This is significantly bad in some cases, especially in Boolean injections.

3.2. Second test

In the second test it was executed the command with options: python3 sqlmap-dev/sqlmap.py -u "http://172.16.1.87:8080/ords/f?p=502" –level=5 –risk=3 -dbs -a –tamper=between. The test starts at 12:30 on 2020-03-05 and ends at 15:18 on 2020-03-05.

With the parameters chosen the intensity of the attack is increased, as well as the level and the number of tests. The results are expressed in Table 3.

Table 3. Results of the second test

Test	Description
Heuristic analysis	The heuristic analysis detected that the target is protected by some type of WAF/IPS.
URL Content	The 'p' parameter of the GET method does not seem to be dynamic. The basic heuristic tests conducted indicate that the 'p' parameter is not injectable.
SQL injection	It is not vulnerable.
Test 'AND boolean-based blind – WHERE or HAVING clause'	Reflective values found and filtered out. It means that there are "reflective" values within the response that contains (parts of) the useful load. This is significantly bad in some cases, especially in Boolean injections.
Test UNION con consulta NULL y Método heurístico con parámetro 'User-Agent'	The 'p' parameter of the GET method does not seem to be dynamic. The basic heuristic tests conducted indicate that the 'p' parameter is not injectable. The User-Agent parameter is not injectable. The 'Referer' parameter does not seem to be dynamic. In the heuristic analysis the 'Referer' parameter does not seem to be injectable. The HOST parameter does not seem to be dynamic. In the heuristic analysis the 'Host' parameter does not seem to be injectable.

3.3. Third test

It is captured the cookie ORA_WWV-W7Hhdq_v8DH8Oli2Fp4IsyM and it is proceeded to use it while the application is active.

It is executed the command python3 sqlmap-dev/sqlmap.py -u "http://172.16.1.87:8080/ords/f?p=502:1:1347964822807:::::" –tables –cookie=ORA_WWV-W7Hhdq_v8DH8Oli2Fp4IsyMR –random-agent –ignore-proxy –level 5, including a random agent and ignoring the proxies, because the latter is only used with the purpose of capturing cookies, as well as increasing the level of analysis to the maximum. The module of identification of tables is added. The analysis starts on 2020-03-16 at 12:21:44. Table 4 reflects the results obtained.

4. Discussion

The attacks common to computer systems occur by viruses, worms and human adversaries [3]. The detection of an attack by SQL injection may occur when it is usual to check log verifications, access registers, intrusion detection, among others [7], [15], to which it is added the application of the principle of defense in depth applying tools such as IDS or WAFs [3]. The continuous evaluations to the applications that are developed and their certifications, before passing to production environments, become another fundamental practice, aspect which is usually omitted in the organizations due to the attempt to publish as soon as possible.

Table 4. Results of the third test

Test	Description
Connection with the URL	The connection requests to redirect to a new URL generated randomly. This is not accepted since a connection already established is being used.
Heuristic analysis	The WAF/IPS is evaded.
URL Content	The 'p' parameter of the GET method does not seem to be dynamic. The basic heuristic tests conducted indicate that the 'p' parameter is not injectable.
SQL injection	It is not vulnerable.
Test 'MySQL Boolean-based blind – Parameter replace (MAKE_SET)'	Reflective values found and filtered out. It means that there are "reflective" values within the response that contains (parts of) the useful load. This is significantly bad in some cases, especially in Boolean injections.

The SQL injection is not a technique which is applied by pressing a button. Knowledge about SQL language is required, Clarke [8] adds that the use of tools to carry out this type of attacks is important, because they enable to automate the attack. The combination of tools enables obtaining more precise results. In agreement with Satin *et al.* [7] and Ojagbule *et al.* [5], the application of the SQLMap tool for the analysis was chosen due to its popularity, availability and the access to its diverse distributions, verifying that tools based on FOSS enable getting results as interesting as using paying instruments.

It has been evidenced that in the exploits database (https://www.exploit-db.com), the last mechanism to violate a system made in APEX was released on April 16th, 2009.

As opposed to the work conducted by Clarke (2009) in which an application deliberately vulnerable (damn vulnerable website) is used, developed in PHP with the MySQL data engine whose objective is providing a test platform to professionals that require testing their levels of skills and knowledge.

The result of the 'AND Boolean-based blind – WHERE or HAVING clause' test can be exemplified in a static page, except with a small part where it is reflected the value of the parameter tested (the same where the SQLMap carries out the injection). In case that such "reflective" content is not detected and neutralized, there is a considerable potential that the response appears as a change due to the useful loads of SQL injection utilized (for example, AND 2>3). Therefore, the risk of detecting false positives (or false negatives in some cases) arises.

In the first execution, the tool ends with the argument that all the parameters evaluated do not seem injectable, which suggests increasing the level and the risk if it is desired to conduct more tests.

If it is suspected of any type of protection mechanism involved (for example, WAF), it could be used the option '–tamper' (e.g. '–tamper=space2comment') and/or change it by '–random-agent'. The adjustment in test 3 with the configuration of the LEVEL=5 parameter was necessary for SQLMap to carry out the vulnerability test of cookies.

According to [21], the detection and prevention becomes a difficult task if the concept of this type of attacks is not appropriately understood. Carrying out binary evaluations, as proposed by [29], might be considered as a mitigation alternative, since it is an extremely automated method that detects and blocks SQL injection attacks in web applications.

When facing blind SQL injection attacks, the most popular technique is AMNESIA (Analysis and Monitoring for Neutralizing SQL-injection Attacks) [30], which is a tool only applicable to protect applications based on JAVA and which use monitoring in runtime [21]. This tool uses machine learning algorithms to provide prevention and detection mechanisms of blind SQL injection threats. Another prevention mechanism is the pattern identification algorithm, proposed by Aho-Corasick [31], which has two phases: i) a static phase and ii) a dynamic phase.

According to [32], in the static phase the SQL queries generated by the user are compared with a list of patterns that contain a sample of the best-known attack patterns. If the SQL sentence agrees exactly with one of the patterns given in the list of static patterns, it means that an SQL attack is being attempted.

Another alternative proposed by [31] citing [33] is the SQLRand, where the basic idea is to generate SQL sentences using random commands in which the query template within the application may be randomized. In this way, the SQL commands that are injected by malicious users are not coded, because the proxy does not recognize the commands injected, thus causing that the attack is not carried out.

In addition, [31] mentioning [34] indicates that there exists an additional method known as Query Tokenization Method, in which a token is generated of both the original query and the query with injection. Then, the tokens resulting from this process are stored in an array. The lengths of array obtained from the original query and from the query with injection are compared, and if there is a coincidence there is no attempt of SQL injection, otherwise, it is an attack.

The proposal [35] is added to the list of mitigation options, which consists in a grammar tree validation approach, represented in a sentence. To grammatically analyze a sentence requires knowledge of the grammar of the language in which the sentence is written. In such a way, when the attacker injects a malicious SQL query, then the grammar tree of the original query and of the query with injection do not coincide. In this technique, the sentence in particular and the original sentence are compared in runtime.

It is also important to indicate that secure coding is crucial for the design of software and computing systems, aspect which is omitted by developers [12] largely due to the lack of knowledge of secure coding standards, negligence and performance loss, to which it is added usability situations [36].

5. Conclusions

One of the most notable features of Oracle APEX is that of creating sessions with cookies and URL links to the software with random data. The tests conducted with the execution of the different options of the commands indicated in this technique, did not enable to undertake the software with the SQL injection technique to a solution developed in this platform. Although a cookie may be captured with Burp Suite, to decipher it takes a considerable time. However, during that time, Oracle APEX already dynamically generated a new cookie, making an attack through this technique basically impossible.

The contribution of this paper has been to evaluate different SQL injection techniques to emphasize on safe code overwriting, optimize the labels generated by default, thus improving the level of security of an application developed in Oracle APEX, without forgetting that tests have been developed in a time span, without exempting latent vulnerabilities that may appear on day 0.

References

[1] A. Barinas López, A. C. Alarcón Aldana, and M. Callejas Cuervo, "Vulnerabilidad de ambientes virtuales de aprendizaje utilizando SQLMAP, RIPS, W3AF y Nessus," *Ventana Informática*, no. 30, pp. 247–260, 2014. [Online]. Available: https://doi.org/10.30554/ventanainform.30.276.2014

[2] S. Mohammadi and A. Namadchian, "Anomaly-based Web Attack Detection: The Application of Deep Neural Network Seq2Seq With Attention Mechanism," *The ISC International Journal of Information Security*, vol. 12, no. 1, pp. 44–54, 2020. [Online]. Available: http://doi.org/10.22042/ISECURE.2020.199009.479

[3] K. L. Ingham, A. Somayaji, J. Burge, and S. Forrest, "Learning DFA representations of HTTP for protecting web applications," *Computer Networks*, vol. 51, no. 5, pp. 1239–1255, 2007, from Intrusion Detection to Self-Protection. [Online]. Available: https://doi.org/10.1016/j.comnet.2006.09.016

[4] B. Dwan, "The Computer Virus – From There to Here.: An Historical Perspective." *Computer Fraud & Security*, vol. 2000, no. 12, pp. 13–16, 2000. [Online]. Available: https://doi.org/10.1016/S1361-3723(00)12026-3

[5] O. Ojagbule, H. Wimmer, and R. J. Haddad, "Vulnerability Analysis of Content Management Systems to SQL Injection Using SQLMAP," in *SoutheastCon 2018*, 2018, pp. 1–7. [Online]. Available: https://doi.org/10.1109/SECON.2018.8479130

[6] C. Kruegel, G. Vigna, and W. Robertson, "A multi-model approach to the detection of web-based attacks," *Computer Networks*, vol. 48, no. 5, pp. 717–738, 2005, web Security. [Online]. Available: https://doi.org/10.1016/j.comnet.2005.01.009

[7] F. Santin, J. A. Oliveira de Figueiredo, and V. Lago Machado, "Uso da ferramenta sqlMap para detecção de vulnerabilidades de SQL Injection," in *Anais do EATI - Encontro Anual de Tecnologia da Informação*, 2017. [Online]. Available: https://bit.ly/340cKP6

[8] J. Clarke, *SQL Injection Attacks and Defense (Second Edition)*, second edition ed., J. Clarke, Ed. Boston: Syngress, 2012. [Online]. Available: https://doi.org/10.1016/B978-1-59-749963-7.00012-8

[9] D. E. Nofal and A. Amer, *SQL Injection Attacks Detection and Prevention Based on Neuro-Fuzzy Technique*. Springer, Cham, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-31129-2_66

[10] B. Bin Halib, E. Budiman, and H. Jati Setyadi, "Teknik Hacking Web Server Dengan SQLMAP Di Kali Linux," *Jurnal Rekayasa Teknologi Informasi*, vol. 1, no. 1, pp. 67–72, 2017. [Online]. Available: http://dx.doi.org/10.30872/jurti.v1i1.642

[11] OWASP. (2017) lobally recognized by developers as the first step towards more secure coding. [Online]. Available: https://bit.ly/2JTb9DF

[12] S. Kals, E. Kirda, C. Kruegel, and N. Jovanovic, "SecuBat: A Web Vulnerability Scanner," in *Proceedings of the 15th International Conference on World Wide Web*, ser. WWW '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 247–256. [Online]. Available: https://doi.org/10.1145/1135777.1135817

[13] J. Fonseca, M. Vieira, and H. Madeira, "Testing and Comparing Web Vulnerability Scanning Tools for SQL Injection and XSS Attacks," in *13th Pacific Rim International Symposium on Dependable Computing (PRDC 2007)*, 2007, pp. 365–372. [Online]. Available: https://doi.org/10.1109/PRDC.2007.55

[14] E. B. Setiawan and A. Setiyadi, "Web vulnerability analysis and implementation," *IOP Conference Series: Materials Science and Engineering*, vol. 407, p. 012081, sep 2018. [Online]. Available: https://doi.org/10.1088%2F1757-899x%2F407%2F1%2F012081

[15] J. Atoum and A. Qaralleh, "A hybrid technique for SQL injection attacks detection and prevention," *International Journal of Database Management Systems (IJDMS*, vol. 6, no. 1, pp. 21–28, 2014. [Online]. Available: http://doi.org/10.5121/ijdms.2014.6102

[16] D. Herrmann and H. Pridöhl, *Basic Concepts and Models of Cybersecurity*, 2020, vol. 21. [Online]. Available: https://doi.org/10.1007/978-3-030-29053-5_2

[17] AVI Network. (2020) SQL Injection Attack. [Online]. Available: https://bit.ly/3mb96YF

[18] P. Ramasamy and S. Abburu, "SQL Injection Attack: Detection and Prevention," *International Journal of Engineering Science and Technology*, vol. 4, no. 4, pp. 1396–1401, 2016. [Online]. Available: https://bit.ly/3n7aSeV

[19] XS Code. (2020) XS:Code. [Online]. Available: https://bit.ly/37MYc6s

[20] D. Novski Neto, "Web (eternamente) revisitada : análise de vulnerabilidades web e de ferramentas de código aberto para exploração," 2019. [Online]. Available: https://bit.ly/37VrNui

[21] V. K. Gudipati, T. Venna, S. Subburaj, and O. Abuzaghleh, "Advanced automated SQL injection attacks and defensive mechanisms," in *2016 Annual Connecticut Conference on Industrial Electronics, Technology Automation (CT-IETA)*, 2016, pp. 1–6. [Online]. Available: https://doi.org/10.1109/CT-IETA.2016.7868248

[22] C. Cetin, D. Goldgof, and J. Ligatti, "SQL-Identifier Injection Attacks," in *2019 IEEE Conference on Communications and Network Security (CNS)*, 2019, pp. 151–159. [Online]. Available: https://doi.org/10.1109/CNS.2019.8802743

[23] J. P. Singh, "Analysis of SQL Injection Detection Techniques," 2016. [Online]. Available: https://bit.ly/375XeDh

[24] O. Ojagbule, H. Wimmer, and R. J. Haddad, "Vulnerability Analysis of Content Management Systems to SQL Injection Using SQLMAP," in *SoutheastCon 2018*, 2018, pp. 1–7. [Online]. Available: https://doi.org/10.1109/SECON.2018.8479130

[25] A. Ciampa, C. A. Visaggio, and M. Di Penta, "A Heuristic-Based Approach for Detecting SQL-Injection Vulnerabilities in Web Applications," in *Proceedings of the 2010 ICSE Workshop on Software Engineering for Secure Systems*, ser. SESS '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 43–49. [Online]. Available: https://doi.org/10.1145/1809100.1809107

[26] R. Alsahafi, "SQL Injection Detection and Prevention Techniques," *International Journal of Scientific & Technology Research*, vol. 8, no. 1, pp. 182–185, 2019. [Online]. Available: https://bit.ly/2W24Ksp

[27] L. Wichman, "Mass SQL injection for malware distribution," SANS Institute, Tech. Rep., 2011. [Online]. Available: https://bit.ly/2Ke3ks0

[28] JAVANICUS. (2016) Posts Related to Web-Pentest-SQL-Injection. [Online]. Available: https://bit.ly/2IEFUMc

[29] V. Sunkari and C. V. Guru rao, "Protect Web Applications against SQL Injection Attacks Using Binary Evaluation Approach," *International Journal of Innovations in Engineering and Technology (IJIET)*, pp. 484–490, 2016. [Online]. Available: https://bit.ly/377eVSR

[30] W. G. J. Halfond and A. Orso, "AMNESIA: Analysis and Monitoring for NEutralizing SQL-Injection Attacks," in *Proceedings of the 20th IEEE/ACM International Conference on Automated Software Engineering*, ser. ASE '05. New York, NY, USA: Association for Computing Machinery, 2005, pp. 174–183. [Online]. Available: https://doi.org/10.1145/1101908.1101935

[31] M. A. Prabakar, M. KarthiKeyan, and K. Marimuthu, "An efficient technique for preventing SQL injection attack using pattern matching algorithm," in *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*, 2013, pp. 503–506. [Online]. Available: https://doi.org/10.1109/ICE-CCN.2013.6528551

[32] G. Yiğit and M. Arnavutoğlu, "SQL Injection Attacks Detection & Prevention Techniques," *International Journal of Computer Theory and Engineering*, vol. 9, no. 5, pp. 351–356, 2017. [Online]. Available: https://bit.ly/3qKrEm5

[33] S. W. Boyd and A. D. Keromytis, "Boyd s.w., keromytis a.d." in *International Conference on Applied Cryptography and Network Security*, 2004, pp. 292–302. [Online]. Available: https://doi.org/10.1007/978-3-540-24852-1_21

[34] L. Ntagwabira and S. L. Kang, "Use of Query tokenization to detect and prevent SQL injection attacks," in *2010 3rd International Conference on Computer Science and Information Technology*, vol. 2, 2010, pp. 438–440. [Online]. Available: https://doi.org/10.1109/ICCSIT.2010.5565202

[35] G. Buehrer, B. W. Weide, and P. A. G. Sivilotti, "Using Parse Tree Validation to Prevent SQL Injection Attacks," in *Proceedings of the 5th International Workshop on Software Engineering and Middleware*, ser. SEM '05. New York, NY, USA: Association for Computing Machinery, 2005, pp. 106–113. [Online]. Available: https://doi.org/10.1145/1108473.1108496

[36] F. D. Nembhard, M. M. Carvalho, and T. C. Eskridge, "Towards the application of recommender systems to secure coding," *EURASIP Journal on Information Security*, vol. 2019, no. 1, p. 9, Jun. 2019. [Online]. Available: https://doi.org/10.1186/s13635-019-0092-4

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
pISSN: 1390-650X / eISSN: 1390-860X

Guidelines for publication in Ingenius Journal

1. General Information

Ingenius is a scientific publication of the *Universidad Politécnica Salesiana* of Ecuador, published since January 2007, with a fixed biannual periodicity, specialized in Mechanical Engineering, Electrical Engineering, Electronics, Computer Science and its integration in what is now known as Mechatronics; these lines of action strengthen areas such as automation, control, robotics, among others..

It is a scientific journal, which uses the peer-review system, under double-blind review methodology, according to the publication standards of the Institute of Electrical and Electronics Engineers (IEEE). Compliance with this system allows authors to guarantee an objective, impartial and transparent review process, which facilitates the publication of their inclusion in reference databases, repositories and international indexing.

Ingenius is indexed in the directory and selective catalog of the Regional Online Information System for Scientific Journals of Latin America, the Caribbean, Spain and Portugal (Latindex), in the Directory of Journals of Open Access DOAJ, In the Information Matrix for the Analysis of Journals, MIAR, In the Ibero-American Network of Innovation and Scientific Knowledge, REDIB and in repositories, libraries and specialized catalogs of Latin America.

The journal is published in a double version: printed (ISSN: 1390-650X) and digital (e-ISSN: 1390-860X), in Spanish, each work being identified with a DOI (Digital Object Identifier System. The articles sent to Ingenius magazine must comply with the following criteria:

2. Scope and policy

2.1. Theme

Original contributions in Mechanical Engineering, Electrical and Electronic Engineering, Computer Science and its integration in what is now known as Mechatronics, as well as related areas: Automation, Control, Domotics, Robotics in their different fields of action and all those related disciplines with the same central theme.

All the work carried out by national or foreign researchers may be published once they meet the required scientific quality criteria.

2.2. Contributions

Ingenius Journal preferably publishes articles related to empirical research, and also reports of technological development, proposals for models and innovations, products for the elaboration of graduate and postgraduate thesis that contribute to the field of science and technology, as well as select revisions of literature. (state-of-the-art).

- **Research:** 5,000 to 6,500 words of text, including title, abstracts, descriptors, charts and references.
- **Reports:** 5,000 to 6,500 words of text, including title, abstracts, charts and references.
- **Reviews:** 6,000 to 7,000 words of text, including charts and references. Current, selective and justified references, would be specially valued from among 40 works

The Ingenius Journal publishes original and unpublished works written in Spanish and English, they may not have been published

through any printed or electronic media, nor be in the process of arbitration or publication.

Every article will be subjected to a rigorous arbitration process; the evaluation of the article will be made according to criteria of originality, relevance, relevance, contributions, scientific rigor and compliance with established editorial guidelines.

Being an arbitrated publication, the Editorial Board approves its publication based on the concept of specialized pairs. The reception of a document does not imply commitment of publication.

It is essential to present a letter of presentation and grant of rights which can be downloaded from: <
urlhttps://goo.gl/ZNkMRD>.

Contributions must be exclusively sent and through the OJS (Open Journal System) <`https://goo.gl/JF7dWT`>. In which all authors must previously register as a user. For any consultation of the procedure you should contact:

<revistaingenius@ups.edu.ec>,
<jcalle@ups.edu.ec> ó
<mquinde@ups.edu.ec>.

3. Presentation and structure of the manuscripts

For those works that are empirical investigations, the manuscripts will follow the IMRDC structure (Introduction, Materials and Methods, Results and Discussion and Conclusions), being optional the Notes and Supports. Those papers that, on the contrary, deal with reports, studies, proposals and reviews may be more flexible in their epigraphs, particularly in material and methods, analysis, results, discussion and conclusions. In all typologies of works, references are mandatory.

Articles may be written on Microsoft Word (.doc or .docx) or LaTeX(.tex). The template to be used can be downloaded from the journal's website, a, <`https://goo.gl/gtCg6m`>, while for LaTeX in <`https://goo.gl/hrHzzQ`>, it is necessary that the file be anonymised in Properties of File, so that the author(s) ID is not displayed.

Figures, Graphs and/or Illustrations, as well as Charts shall be numbered sequentially including an explanatory description for each. The equations included in the article must also be numbered; the figures, charts and equations must be cited in the text.

Use space after point, commas and question marks.

Use "enter" at the end of each paragraph and title heading. Do not use .[en]ter.[a]nywhere else, let the word processor program automatically break the lines.

Do not center headings or subheadings as they should be aligned to the left.

Charts must be created in the same program used for the document body, but must be stored in a separate file. Use tabs, not spaces, to create columns. Remember that the final size of printed pages will be 21 x 28 cm, so the tables must be designed to fit the final print space.

3.1. Structure of the manuscripts

3.1.1. Presentation and cover letter

1. **Título (español) / Title (inglés):** Concise but informative, in Spanish on the front line and in English on the second, when the article is written in Spanish and vice versa if it is written in English.
2. **Authors and affiliations:** Full name and surname of each author, organized by order of priority and their institutional affiliation with reference to the end of the first sheet, where it must include: Dependency to which belongs within the

institution, Institution to which he/she belongs, country, ORCID. A maximum of 5 authors will be accepted, although there may be exceptions justified by the complexity and extent of the topic.

3. **Abstract (Spanish) / Abstract (English):** It will have a maximum extension of 230 words, first in Spanish and then in English. : 1) Justification of the topic; 2) Objectives; 3) Methodology and sample; 4) Main results; 5) Main conclusions.
4. **Keywords (Spanish) / Keywords (English):** 6 descriptors must be presented for each language version directly related to the subject of the work. The use of the key words set out in UNESCO's Thesaurus will be positively valued.
5. **Presentation (Cover Letter):** A statement that the manuscript is an original contribution, not submission or evaluation process in another journal, with the confirmation of the signatory authors, acceptance (if applicable) of formal changes in the manuscript according to the guidelines and partial assignment of rights to the publisher, according to the format established in: <https://goo.gl/ZNkMRD>

3.1.2. Manuscript

1. **Título (español) / Title (inglés):** Concise but informative, in Spanish on the front line and in English on the second, when the article is written in Spanish and vice versa if it is written in English.
2. **Authors and affiliations:** Full name and surname of each author, organized by order of priority and their institutional affiliation with reference to the end of the first sheet, where it must include: Dependency to which belongs within the institution, Institution to which he/she belongs, country, ORCID. A maximum of 5 authors will be accepted, although there may be exceptions justified by the complexity and extent of the topic.
3. **Abstract (Spanish) / Abstract (English):** It will have a maximum extension of 230 words, first in Spanish and then in English. : 1) Justification of the topic; 2) Objectives; 3) Methodology and sample; 4) Main results; 5) Main conclusions.
4. **Keywords (Spanish) / Keywords (English):** 6 descriptors must be presented for each language version directly related to the subject of the work. The use of the key words set out in UNESCO's Thesaurus will be positively valued.
5. **Introduction:** It should include the problem statement, context of the problem, justification, rationale and purpose of the study, using bibliographical citations, as well as the most significant and current literature on the topic at national and international level.
6. **Material and methods:** It must be written so that the reader can easily understand the development of the research. If applicable, it will describe the methodology, the sample and the form of sampling, as well as the type of statistical analysis used. If it is an original methodology, it is necessary to explain the reasons that led to its use and to describe its possible limitations.
7. **Analysis and results:** It will try to highlight the most important observations, describing, without making value judgments, the material and methods used. They will appear in a logical sequence in the text and the essential charts and figures avoiding the duplication of data.
8. **Discussion and Conclusions:** It will summarize the most important findings, relating the observations themselves to relevant studies, indicating contributions

and limitations, without adding data already mentioned in other sections. It should also include deductions and lines for future research.

9. **Supports and acknowledgments (optional):** The Council Science Editors recommends the author (s) to specify the source of funding for the research. Priority will be given to projects supported by national and international competitive projects.
10. **The notes (optional):** will go, only if necessary, at the end of the article (before the references). They must be manually annotated, since the system of footnotes or the end of Word is not recognized by the layout systems. The numbers of notes are placed in superscript, both in the text and in the final note. The numbers of notes are placed in superscript, both in the text and in the final note. No notes are allowed that collect simple bibliographic citations (without comments), as these should go in the references.
11. **References:** Bibliographical citations should be reviewed in the form of references to the text. Under no circumstances should references mentioned in the text not be included. Their number should be sufficient to contextualize the theoretical framework with current and important criteria. They will be presented sequentially in order of appearance, as appropriate following the format of the IEEE.

3.2. Guidelines for Bibliographical references

Journal articles:

[1] J. Riess, J. J. Abbas, "Adaptive control of cyclic movements as muscles fatigue using functional neuromuscular stimulation". IEEE Trans. Neural Syst. Rehabil. Eng vol. 9, pp.326–330, 2001. [Onine]. Available: https://doi.org/10.1109/7333.948462

Books:

[1] G. O. Young, "Synthetic structure of industrial plastics" in Plastics, 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw–Hill, 1964, pp. 15–64.

Technical reports:

[1] M. A. Brusberg and E. N. Clark, "Installation, operation, and data evaluation of an oblique–incidence ionosphere sounder system," in "Radio Propagation Characteristics of the Washington–Honolulu Path," Stanford Res. Inst., Stanford, CA, Contract NOBSR–87615, Final Rep., Feb. 1995, vol. 1

Articles presented in confeences (unpubished):

[1] Vázquez, Rolando, Presentación curso "Realidad Virtual". National Instruments. Colombia, 2009.

Articles of memories of Conferences (Published):

[1] L. I. Ruiz, A. García, J. García, G. Taboada. "Criterios para la optimización de sistemas eléctricos en refinerías de la industria petrolera: influencia y análisis en el equipo eléctrico," IEEE CONCAPAN XXVIII, Guatemala 2008.

Thesis:

[1] L.M. Moreno, "Computación paralela y entornos heterogéneos," Tesis doctoral, Dep. Estadística, Investigación Operativa y Computación, Universidad de La Laguna, La Laguna, 2005.

Guidelines:

[1] IEEE Guide for Application of Power Apparatus Bushings, IEEE Standard C57.19.100–1995, Aug. 1995.

Patents:

[1] J. P. Wilkinson, "Nonlinear resonant circuit devices," U.S. Patent 3 624 125, July 16, 1990.

Manuals:

[1] Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

Internet resources:

[1] E. H. Miller, "A note on reflector arrays" [Online]. Available. https://goo.gl/4cJkCF

3.3. Epigraphs, Figures and Charts

The epigraphs of the body of the article will be numbered in Arabic. They should go without a full box of capital letters, neither underlined nor bold. The numbering must be a maximum of three levels: 1. / 1.1. / 1.1.1. At the end of each numbered epigraph will be given an enter to continue with the corresponding paragraph.

The charts must be included in the text according to order of appearance, numbered in Arabic and subtitled with the description of the content, the subtitle should go at the top of the table justified to the left.

Figures can be linear drawings, maps or black and white halftone or color photographs in 300 dpi resolution. Do not combine photographs and line drawings in the same figure.

Design the figures so that they fit eventually to the final size of the journal 21 x 28 cm. Make sure inscriptions or details, as well as lines, are of appropriate size and thickness so that they are not illegible when they are reduced to their final size (numbers, letters and symbols must be reduced to at least 2.5 mm in height After the illustrations have been reduced to fit the printed page). Ideally, the linear illustrations should be prepared at about a quarter of their final publication size.

Different elements in the same figure should be spelled a, b, c, etc.

Photographs should be recorded with high contrast and high resolution. Remember that photographs frequently lose contrast in the printing process. Line drawings and maps should be prepared in black.

The text of the figures and maps must be written in easily legible letters.

If the figures have been previously used, it is the responsibility of the author to obtain the corresponding permission to avoid subsequent problems related to copyright.

Each figure must be submitted in a separate file, either as bitmap (.jpg, .bmp, .gif, or .png) or as vector graphics (.ps, .eps, .pdf).

4. Submission process

The manuscript must be sent through the OJS system of the journal, <`https://goo.gl/JF7dWT`>,the manuscript should be uploaded as an original file in .pdf without author data and anonymized according to the above; In complementary files the complete manuscript must be loaded in .doc or .docx (Word file), that is to say with the data of the author (s) and its institutional ascription; Also the numbered figures should be uploaded in independent files according to the corresponding in the manuscript (as bitmap .jpg, .bmp, .gif, or .png or as vector graphics .ps, .eps, .pdf). It is also obligatory to upload the cover letter and grant of rights as an additional file.

All authors must enter the required information on the OJS platform and only one of the authors will be responsible for correspondence.

Once the contribution has been sent the system will automatically send the author for correspondence a confirmation email of receipt of the contribution.

5. Editorial process

Once the manuscript has been received in OJS, a first check by the editorial team of the following points:

- The topic is in accordance with the criteria of the journal.
- Must have the IMRDC structure.
- Must be in the INGENIUS format.
- Must use the IEEE citation format.
- All references should be cited in the text of the manuscript as well as charts, figures and equations.
- The manuscript is original; for this, software is used to determine plagiarism.

The assessment described above can take up to 4 weeks.

If any of the above is not complete or there is inconsistency, an email will be sent to the author to make the requested corrections.

The author will make the corrections and resend the contribution through an email in response to the notification and will also upload the corrected manuscript into OJS supplementary files.

The editorial team will verify that the requested corrections have been incorporated, if it complies, the manuscript will start the second part of the process that may be followed by the author through OJS, otherwise the author will be notified and the manuscript will be archived.

The second phase of the process consists of the evaluation under the methodology of double-blind review, which includes national and foreign experts considering the following steps:

- The editor assigns two or more reviewers for the article.
- After reviewing the article, the reviewers will submit the evaluation report with one of the following results.
  - Publishable
  - Publishable with suggested changes
  - Publishable with mandatory changes
  - Non publishable
- The editor once received the evaluation by the reviewers will analyze the results and determine if the article is accepted or denied.
- If the article is accepted, the author will be notified to make corrections if required and the corresponding editorial process will be continued.
- If the article is denied, the author will be notified and the manuscript will be archived.
- In the two previous cases the result of the evaluation of the reviewers and their respective recommendations will be sent.

The second phase of the process lasts at least 4 weeks, after which they will be notified to the author giving instructions to continue with the process.

6. Publication

The INGENIUS Journal publishes two issues per year, on January 1st and July 1st, so it is important to consider the dates for sending the articles and their corresponding publication. Articles received until October will be considered for the January publication and those received until April for the July publication.

UNIVERSIDAD POLITÉCNICA SALESIANA DEL ECUADOR
Javier Herrán Gómez, sdb
Rector

Turuhuayco 3-69 y Calle Vieja
Postal code 2074
Cuenca, Ecuador
Teléfono: (+593 7) 205 00 00
Fax: (+593 7) 408 89 58
Email: srector@ups.edu.ec

Exchange

Exchange with other periodicals is accepted.

Address:
Secretaría Técnica de Comunicación
Universidad Politécnica Salesiana
Turuhuayco 3-69 y Calle Vieja
Postal code 2074
Cuenca, Ecuador
Phone: (+593 7) 205 00 00 Ext. 1182
Fax: (+593 7) 408 89 58
Email: rpublicas@ups.edu.ec
www.ups.edu.ec
Cuenca – Ecuador

INGENIUS, Journal Science of Technology,
Semester publication, N.° 25, january/june 2021
John Calle Sigüencia, Editor in chief
revistaingenius@ups.edu.ec

Printed
Centro Gráfico Salesiano: Antonio Vega Muñoz 10-68 y General Torres.
Phone: (+593 7) 283 17 45
Cuenca – Ecuador
Email: centrograficosalesiano@lns.com.ec

OTHER PERIODIC PUBLICATIONS OF THE UNIVERSITY

UNIVERSITAS, Journal of Social and Human Sciences.

LA GRANJA, Journal of Life Sciences.

ALTERIDAD, Journal of Education.

RETOS Journal of Administration Sciences and Economics.

UTOPÍA, University Youth Ministry Magazine.

SOPHIA, Collection of Philosophy of Education.

ABYA YALA | UNIVERSIDAD POLITÉCNICA SALESIANA

ECUADOR SALESIANOS DON BOSCO