REVISTA MILITAR de CIÊNCIA e TECNOLOGIA

Print version: ISSN 0102-3543

Volume XL - 1º Quarter 2023

ENGENHARIA MILITAR

Mobilidade, Contramobilidade e Proteção

REVISTA MILITAR de CIÊNCIA e TECNOLOGIA

REVISTA MILITAR DE CIÊNCIA E TECNOLOGIA - Volume XL
1º Trimestre de 2023

INSTITUTO MILITAR DE ENGENHARIA - IME
Praça General Tibúrcio, 80 - Praia Vermelha -
Rio de Janeiro-RJ - CEP 22.290-270 -
Tel.: (21) 2546-7115
Website: https://ebrevistas.eb.mil.br/CT - Email: rmct@ime.eb.br

BIBLIOTECA DO EXÉRCITO
EDITORA (BIBLIEx)
Palácio Duque de Caxias - Praça D. de Caxias, 25
3º andar - Ala Marcílio Dias
Centro - Rio de Janeiro-RJ
CEP 20.221-260
Tel.: (21) 2519-5707

ACESSE NOSSA REVISTA DIGITAL

Nossa capa:
Homenagem ao Quadro de Engenheiros Militares

EDITORIAL

REVISTA MILITAR de CIÊNCIA e TECNOLOGIA

The 19th century is known as the age of big machines, the 20th century as the age of information, and the 21st century as the age of innovation. But that lasted just over a decade, when it shifted into the age of quantum technologies. We are experiencing a total paradigm shift in science, technology, innovation (ST&I) and, more than ever, in war conflicts. The 21st century has seen no less than three generations of war, and new generations are likely to emerge. All of this has resulted from the incredible advances in ST&I applied to increasing both defense capabilities and combat power in the different domains of war and theaters of operations.

In 1915, British Admiral John Fisher declared that wars in the future would be won by inventions—nothing more visionary than that. The World War II elevated ST&I to the level of a subject of extreme relevance for guiding a country's destiny. To encourage this, prominent countries such as the United States have launched long-term strategic policies and actions. The report drawn up by Vannevar Bush, director of the Office of Scientific Research and Development, in 1945, at the behest of President Franklin D. Roosevelt, and the contributions of other visionaries helped define the new US model for ST&I development, replacing the one that allocated the majority of government investment to military research. In essence, in this new model basic scientific research, developed without concern for immediate application, takes on crucial importance due to the vast amount of knowledge produced which, used by third parties, would generate applications that would otherwise be impossible to envision. In it, public and private actors, both civilian and military, share responsibility for carrying out basic and applied scientific research and technological development. Today there are countless inventions and innovations that are of vital importance in the civilian world and are the direct result of war conflicts.

The Military Engineering Institute is located in this very context, a bicentennial corporate university whose primary objective is to meet the needs of the Brazilian Army, without neglecting dual applications that can contribute in some way to civil society.

In this context, the role played by the *Military Journal of Science and Technology (RMCT)*, a regular publication of the Brazilian Army, is essential. To achieve these objectives, this edition of RMCT marks the resumption of work, bringing it up to date with a new editorial board and the use of new technologies by hiring a specialized and excellent publisher. The journal is open to publishing adherent and good quality works by civilian and military authors. Accepting articles from both Brazil and abroad, RMCT publishes free of charge, with one of its objectives being to be an open access publication option.

In this issue, we present seven articles from different areas of engineering, with special emphasis on military engineering.

We wish you all a good read!

Dr. Eng. Fernando M. Araujo-Moreira
Editor-in-chief of RMCT

SUMMARY

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.12121.en

Identification of nonlinear systems by fitting LPV models with polynomial coefficients

Itamar Pifano Silva[1*], Roberto Ades[2], Paulo Cesar Pellanda[1], Marcelo de Araujo Oliveira[1]
[1] Military Institute of Engineering - Praça General Tibúrcio, 80, Praia Vermelha, Rio de Janeiro, RJ, Brazil, CEP 29.270-030
[2] Centro Federal de Educação Tecnológica – Av. Maracanã, 229, Maracanã, Rio de Janeiro, RJ, Brazil, CEP 20.271-110
*itamarps@ime.eb.br
roberto.ades@cefet-rj.br
pcpellanda@ieee.org
oliveira.marcelo@ime.eb.br

ABSTRACT: *This paper presents a method for the identification or tuning of LPV models with polynomial coefficients. The method is applicable to multivariable systems and to fit the behavior of nonlinear systems. It also presents an extension for multiple varying parameters. A Quarter-Car suspension model was used to illustrate the proposed method, and models with endogenous and exogenous parameters were adjusted.*

KEYWORDS: System identification; Nonlinear systems; LPV systems; Quasi-LPV; LPV model.

RESUMO: Este artigo apresenta um método para a identificação ou ajuste de modelos LPV com coeficientes polinomiais. O método é aplicável a sistemas multivariáveis e para a aproximação do comportamento de sistemas não lineares. Discute-se também a extensão para múltiplos parâmetros variantes. Um modelo de suspensão de um automóvel foi empregado para ilustrar o método proposto, tendo sido ajustados modelos com parâmetros endógenos e exógenos.

PALAVRAS-CHAVE: Identificação de sistema; Sistemas não lineares; Sistemas LPV; Quasi-LPV; Modelos LPV.

1. Introduction

In terms of controlling nonlinear systems, the classic gain scheduling approach, which uses linear design methods and their already consolidated tools, is among the approaches proposed in the literature. Although it is widely used, it does not guarantee stability and performance outside the operating points considered in linear designs, especially in cases where the parametric variation rate is high. The fact that there is a connection between the formulation of Linear Matrix Inequalities (LMIs) and the Lyapunov Theory has allowed the stability and performance criteria developed for linear systems to be extended to non-stationary linear systems, especially for the general class of Linear Parameter Varying (LPV) systems [1].

LPV control, with local or global stability and performance guaranteed in large envelopes of the operating domain of multivariable systems, has been presented as a real alternative to the classic gain scheduling approach. LPV controller gains are automatically programmed without the need for any ad hoc method or interpolation. Since the mid-1990s, LPV control techniques have evolved significantly through three distinct methods [2, 3], namely: polytopic approach, grid-based approach, and Linear Fractional Transformation (LFT). In polytopic approaches, models of some operating points are considered, in principle generated by the extreme values of the coordinates of the vector of varying parameters. The remaining operating points are obtained from the affine combination of these extreme models, which may not be true. The disadvantage of this type of approach is also related to conservatism, which will probably include a range of situations that, despite being considered, may not occur in practice. On the other hand, the optimization problem to be solved, initially of infinite dimension, becomes a problem with finite dimension and equal to the number of vertices of the polytope, since the other models are determined by the affine combination of the models at these vertices [4, 5, 6].

In grid-based approaches, the space of varying parameters is tested in a grid of values, considering

realistic trajectories. The advantage of this type of methodology is the reduction of conservatism in relation to the polytopic approach. Recursive and LMI algorithms of this type of methodology are found in the literature. On the other hand, it presents serious restrictions in relation to the number of varying parameters, that is, the computational effort grows exponentially and makes it impossible to treat LPV systems with more than two varying parameters. In addition, it also assumes that the system under analysis is relatively well-behaved, so that its dynamic can be approximated without a large increase in the grid density [7, 8, 6].

In the case of LFT approaches, the LPV model must be transformed into an LFT. Therefore, applications of this methodology are restricted to cases in which the LPV models present specific functions of the varying parameters, such as polynomial functions of the varying parameters. Once the LFT model has been obtained, μ synthesis and others can be used to calculate the controller [9, 10].

A wide variety of LPV control applications were initially developed in the aeronautical area, but applications are expanding to several other areas, and are validated by experiments or high-fidelity simulations, as presented in [11]. One of the main bottlenecks today in the application of LPV control techniques is the lack of methods for obtaining LPV models. This need has promoted the interest of the scientific community working in the area of system identification, in order to be able to produce models of nonlinear or nonstationary systems, with the ultimate objective of using existing LPV control methods.

System identification methods aim to obtain models from the measured signals of the inputs and outputs of a plant under study [12]. Basically, they can be classified into two fronts, depending on the structure of the model: parametric and nonparametric identification. The case of nonparametric identification involves an undetermined structure and, consequently, an a priori undetermined number of parameters. Nonparametric LPV identification has basically been divided into three main approaches: (a) the scattering function; (b) the least-squares support vector machine (LS-SVM), and (c) the one based on the Bayesian configuration, respectively [13, 14, 15] as cited in [16]. Regarding parametric prediction, the structure to be identified is previously established and a certain number of parameters must be adjusted.

The methods for identifying LPV systems [17] can be classified, as defined in [18], into two main areas, according to the mathematical representation used, LPV-IO (Input-Output) and LPV-SS (State Space).

LPV-SS methods adopt a discrete representation in state space or its equivalent LFR, which allows the representation of Multiple-Input Multiple-Output (MIMO) systems. More information about the approaches that use the LPV-SS structure can be found in [18].

This work uses the LPV-IO representation, which utilizes discrete time series models as its mathematical structure. In most cases, it uses the LTI error prediction configuration and is generally treated only for the Single-Input Single-Output (SISO) case. A discrete model in the form of time series can be represented as:

$$y(k) = -\sum_{i=1}^{n} a_i(\theta) q^i y(k) + \sum_{j=0}^{m} b_j(\theta) q^j u(k) + e(k) \qquad (1)$$

where q is the delay operator in the time domain, so that:

$$q^p y(k) = y(k-p) \qquad (2)$$

The variable $e(k)$ is the process noise, usually a white noise with zero mean, $n \geq m$ and the coefficients $\{a_i\}_{i=1}^{n}$ and $\{b_j\}_{j=0}^{m}$ are dependent on the parameter θ The prediction of the coefficients a_i and b_j in the model (1) can be performed as defined in [18]:

(i) Interpolation approach. The methods that use this approach are those originating from the classic concept of gain-scheduling, characterized by considering as the operating point specific values of the varying parameter that, once frozen, determines the error prediction structure of the LTI system, allowing the identification of local models. The global model is

obtained through the interpolation of local models, as developed in [19, 20].

(ii) Set association approach. In this case, the noise in the measured data is treated as deterministic uncertainty and, instead of a direct estimate of the coefficients, a set of their feasible values is calculated, as presented in [21, 22]. This feasible set represents the values of the coefficients that satisfy the model equation in (1) and a priori with an assumed error less than or equal to that of the noise in the measured data. A direct estimate of the coefficient is obtained by calculating the mean of the values in the feasible set. This approach generally uses non-convex programming methods.

(iii) Nonlinear programming approach. The coefficients $\{a_i\}_{i=1}^{n}$ and $\{b_j\}_{j=0}^{m}$ of the time series model in (1) are estimated using nonlinear programming methods to minimize the mean squared prediction error [13, 23]. The objective is to achieve a better prediction than that of linear regression methods. In some cases, this is done through a non-linear parameterization:

$$a_i(\theta) = \alpha_{i,0} + \alpha_{i,1} Z$$
$$b_j(\theta) = \beta_{j,0} + \beta_{j,1} Z$$

where $\alpha_{i,0}$, $\alpha_{i,1}$, $\beta_{j,0}$ and $\beta_{j,1} \in \mathbb{R}$ and the variable Z is the output of an artificial neural network that uses as inputs the output vector $[(y(k)\ y(k\text{-}1)\ \cdots)]^T$, the input vector $[(u(k)\ u(k\text{-}1)\ \cdots)]^T$ and the vector of the measures of the varying parameter $[(\theta(k)\ \theta(k\text{-}1)\ \cdots)]^T$ of the system to be identified. This approach, in most cases, uses a mixed procedure of linear and non-linear programming through linear regression methods combined with neural networks.

(iv) Linear regression approach. Structures of linear models of discrete time series are used, such as Autoregressive with Exogenous Inputs (ARX), widely used in the literature on the identification of LTI systems, which is part of the system identification package of the MATLAB® programming and numeric computing platform.

In the LPV case, coefficients $\{a_i\}_{i=1}^{n}$ and $\{b_j\}_{j=0}^{m}$ of (1) are polynomial functions of the varying parameter, such as:

$$a_i(\theta) = \alpha_{i,0} + \sum_{p=1}^{N_i} \alpha_{i,p} \theta^p$$
$$b_j(\theta) = \beta_{j,0} + \sum_{p=1}^{M_j} \beta_{j,p} \theta^p$$

In this way, it uses the concept of predicting the LTI error via Least-Squares (LS), recursive or not, as well as instrumental variables that lead to a better adjustment in the presence of noisy signals. As a result, a linear model in the parameters is obtained by linear regression, according to the precursor work of [24], an approach used in this work.

In [25] a method of LPV-IO identification is proposed using the linear regression approach, which seeks a parsimonious, nonparametric model structure that can capture the unknown dependence of the coefficients and as a function of the varying parameter in (1). This dependence can vary between polynomial, rational, or even discontinuous functions. To obtain an efficient solution, the article proposes the LS-SVM method, which leads to a construction of the model without the a priori information of order and delay of the system under study. It was originally developed as a class of supervised learning methods, as presented in [26, 27] as cited in [25], where it is used to obtain the model structure.

In [28], a method based on Instrumental Variable (IV) for bias correction was developed to identify SISO LPV-IO models, of the ARX type, from measurements of the output and the signal of the varying parameter, corrupted by noise. The noise process associated with the output is considered colored, of zero mean and with unknown distribution, while the measurements of the signal of the varying parameter are affected by a white Gaussian noise. The proposed method eliminates the bias resulting from the methods originating from LS when they neglect the measurement noise existing in the signal of the varying parameter. Thus, it provides a consistent estimate of LPV models with polynomial dependence of the varying parameter, whose instrument used only needs to be uncorrelated with the noise that corrupts the

output observations. In this way, an approximation for the noise-free varying parameter does not need to be calculated.

The work developed in [29] is an analysis of the one-step-ahead error prediction criterion used in LPV identification processes, with the aim of obtaining new kernel functions [27, 30] to be applied in LPV-IO identification processes of nonparametric Box-Jenkins models.

In [31] a bias correction scheme for closed-loop identification of LPV-IO models is presented, using the regression approach, caused by the correlation between the input signal that excites the process and the output noise. The proposed identification algorithm provides a consistent estimate of the open-loop model parameters when the output signal and the variable signal of the variable parameter are corrupted by measurement noise.

A nonparametric LPV-IO identification by regression is presented in [32] , using LS-SVM to estimate the skid angle of a passenger car, replacing the sensor used in commercial vehicles, due to its high cost. The problem of the article is inspired by [21], which uses the set association approach method, but the article uses the method proposed in [25].

The work of [33] presents a study of LPV-IO identification through the regression approach, in the search for a global model with a nonparametric structure without requiring much a priori information about the model order, using Hilbert space through Reproducing Kernel Hilbert Space (RKHS), which corresponds to a global quadratic optimization problem directly solvable with LMI constraints, for the selection of the structure of the parsimonious model to be identified.

An LPV-IO method by online regression is presented in [34], in which the analysis of the dynamics is performed in the domain of the varying parameter or as a function of it named by the authors as *causal regressor*, instead of the time domain. As a result, the coefficients to be identified are re-estimated using only present data and at least one previous estimate, that is, each coefficient prediction does not necessarily depend on its estimate at the previous instant, but on one of its past estimates at an instant when its associated causal regressors are similar to the present ones. The proposed criterion for measuring this similarity is the norm of the differences between the current and past causal regressors. The proposed method presented low computational cost for similar adjustment of the coefficients, compared to traditional offline methods.

In this work, some ideas presented in [24] were adopted, and the following additional developments were proposed: extension of the method to multivariate systems; obtainment of the solution from a batch of stored data instead of estimate by the recursive form; use of polynomials with independent degrees in each model coefficient; expansion with multiple varying parameters and the realization of the time lag in the model between the current output and the most recent input, which corresponds to the increase in the relative degree of the transfer function of the identified model. The proposed technique was also used to approximate nonlinear models in the Quasi-LPV format by LPV models with polynomial coefficients. To illustrate the proposed method, a car suspension system was explored in several situations, with parameter variations and nonlinearities, approximating them by LPV models.

Regarding the structure of this work, section 2 presents the definitions of LPV and Quasi-LPV systems, section 3 presents the proposed method with the extension to multivariable system and expansion to multiple varying parameters. Section 4 presents the problem addressed for a quarter-car suspension model with several nonlinear situations, depending on the type and quantity of variable parameters. Finally, section 5 presents the final considerations and conclusions.

2. LPV/Quasi-LPV Systems

An LPV or Quasi-LPV system is one whose matrices A, B, C and D in the state space representation are not constant and vary depending on parameters that are exogenous or endogenous to the system. These parameters, which alter the system dynamic, are

called variable parameters θ. An exogenous variable parameter is one that is external to the system; an endogenous parameter is characterized as being internal to the system, represented by one of the states of the system or a function of it. Both variable parameters must be measurable in order to portray the system. Depending on the variable parameters used in the system representation, the models can be classified as LPV or Quasi-LPV.

An LPV model is recognized by having all exogenous and measurable varying parameters. The following is the definition of LPV models.

Definition 1 – LPV Model [35], as cited in [36]: Given a compact subset $P \subset \mathbb{R}^d$, F_p represents the operator that maps $t \in \mathbb{R}^+$ onto a vector $\theta(t)$ of external parameters, whose components are continuous functions by parts $\forall t \in \mathbb{R}^+$ Consider also the continuous functions, $A{:}P \rightarrow \mathbb{R}^{n\times n}$, $B{:}P \rightarrow \mathbb{R}^{n\times w}$, $C{:}P \rightarrow \mathbb{R}^{\zeta\times n}$, and $D{:}P \rightarrow \mathbb{R}^{\zeta\times w}$. An LPV model of order n is defined as:

$$\begin{bmatrix}\dot{x}(t)\\ y(t)\end{bmatrix}=\begin{bmatrix}A(\theta(t)) & B(\theta(t))\\ C(\theta(t)) & D(\theta(t))\end{bmatrix}\begin{bmatrix}x(t)\\ u(t)\end{bmatrix}. \quad (3)$$

It can be added that the Linear Time Varying (LTV) model is a particular case of an LPV system, in which the dynamic matrices depend on the varying parameter $\theta(t) = t$.

A Quasi-LPV model is a nonlinear model that resembles the LPV model in (3). In this case, the vector $\theta(t)$ is composed of two types of varying parameters, exogenous and endogenous, both measurable.

Definition 2 – Quasi-LPV Model [36]: Be $\theta(t) \in P$ such that $\theta(t)=[\Omega(t)^T\ z(t)^T\ ]^T$, where $\Omega(t)$ corresponds to the vector of exogenous variables, similarly to the LPV system in (3), and $z(t)$ corresponds to the vector of endogenous variables, containing some states of the system or functions of these, which interfere in the system dynamic. The Quasi-LPV model can be defined by:

$$\begin{bmatrix}\dot{z}(t)\\ \dot{\eta}(t)\\ y(t)\end{bmatrix}=\begin{bmatrix}A_{11}(\theta(t)) & A_{12}(\theta(t)) & B_1(\theta(t))\\ A_{21}(\theta(t)) & A_{22}(\theta(t)) & B_2(\theta(t))\\ C_1(\theta(t)) & C_2(\theta(t)) & D(\theta(t))\end{bmatrix}\begin{bmatrix}z(t)\\ \eta(t)\\ u(t)\end{bmatrix},$$

where the state vector $x(t)=[z(t)^T\ \eta(t)^T\ ]^T$ and $\eta(t)$ represents the vector containing the states that do not interfere in the model matrices.

The following example depicts a mathematical manipulation to transform a nonlinear system into a Quasi-LPV model. The endogenous varying parameters considered are responsible for the nonlinearities of the system and must be measurable to characterize this representation.

According to [2], as an example, consider the nonlinear plant modeled by the following equations:

$$\dot{x}_1 = \text{sen}(x_1) + x_2$$
$$\dot{x}_2 = x_1x_2 + u$$

where u is the input of the system and considering x_1 and x_2 as the model states, $x = [x_1\ x_2]$ can be defined as the state vector. Thus, a Quasi-LPV representation of this nonlinear model could be

$$\dot{x} = \begin{bmatrix}\text{sen}(x_1)/x_1 & 1\\ x_2 & 0\end{bmatrix}x + \begin{bmatrix}0\\1\end{bmatrix}u.$$

This representation may not be adequate unless x_1 and x_2 are measurable and $x_1 \neq 0\ \forall t \in \mathbb{R}^+$ In this case, there are only the endogenous varying parameters and $\eta(t) = \emptyset$. Thus $x(t)=z(t)$ and $z(t)=[x_1\ x_2]^T$.

Assuming that one has only x_1 measurable, a more adequate representation could be:

$$\dot{x} = \begin{bmatrix}\text{sen}(x_1)/x_1 & 1\\ 0 & x_1\end{bmatrix}x + \begin{bmatrix}0\\1\end{bmatrix}u,$$

which allows rewriting the equation so that the state matrix only has dependence on the variable x_1, that is, $z(t) = x_1$ and $\eta(t) = x_2$.

3. Method proposed

The problem addressed here consists of adjusting an LPV model with polynomial coefficients, so that its output $\hat{y}$ approximates, according to some previously defined standard, the output of the nonlinear physical system. The proposal is that the model be discrete and in the form of a time series, whose coefficients can be polynomials dependent on the varying parameter $\theta = \theta(t)$. Initially, we are dealing only with the SISO case with a single varying parameter. To identify the

model, it is assumed that the system under study is previously monitored with sensors, so that the temporal data of its inputs, outputs and the varying parameter, even if continuous, are acquired according to a convenient sampling rate T.

In the same way as in [24], the class of discrete-time LPV models was adopted, parameterized as follows:

$$A(q,\theta)y(k) = B(q,\theta)u(k) \tag{4}$$

where q is the delay operator, as defined in (2), which leads to polynomials as a function of the varying parameter, according to:

$$A(q,\theta) = 1 + a_1(\theta)q + \cdots + a_n(\theta)q^n \tag{5}$$

$$B(q,\theta) = b_1(\theta)q^r + b_2(\theta)q^{r+1} + \cdots + b_m(\theta)q^{r+m-1} \tag{6}$$

In addition, it was considered that the varying parameter θ, although continuous, was transformed into a function of discrete time, that is, $\theta := \theta(kT) = \theta_k$, where T is the sampling period.

Thus, by (4), (5) and (6), the structure of the identified model, in the form of a time series, can be written as:

$$\begin{aligned} y_k = &-a_1 y_{k-1} - a_2 y_{k-2} - \cdots - a_n y_{k-n} + \\ &+ b_1 u_{k-r} + b_2 u_{k-r-1} + \cdots + b_m u_{k-r-m+1}, \end{aligned} \tag{7}$$

where $y_k = y(kT)$ and $u_k = u(kT)$ n, is the order of the model and represents the number of autoregressive terms of the output signal, m is the number of input terms, m is the delay between the current output and the most recent input considered, and $m + r - 1 \leq n$.

It was also considered that the coefficients of the model above have polynomial dependence in relation to the varying parameter $\theta = \theta_k$. Thus, $\forall i \in \{1,\cdots,n\}$ e $N_i \in \mathbb{N}$ and :

$$a_i = a_i(\theta) = \alpha_{i,0} + \alpha_{i,1}\theta + \alpha_{i,2}\theta^2 + \cdots + \alpha_{i,N_i}\theta^{N_i}. \tag{8}$$

Similarly, for the coefficients of the input variable u, $\forall j \in \{1, \cdots, m\}$ and $M_j \in \mathbb{N}$:

$$b_j = b_j(\theta) = \beta_{j,0} + \beta_{j,1}\theta + \beta_{j,2}\theta^2 + \cdots + \beta_{j,M_j}\theta^{M_j} \tag{9}$$

The varying parameter $\theta(t)$ is considered measurable, but may be out of phase with the current output, that is:

$$y_k = f(\theta_{k-\gamma}), \tag{10}$$

where $\gamma \in \{0,\cdots, n\}$. Usually, is adopted, that is, the determination of the current output depends on the value of θ at the previous instant.

Although the structure in (7) was adopted in the form of a time series, it is important to note that, defined in this way, it will have a one-to-one correspondence with the models in the form of state space, requiring only the use of a canonical realization. In this way, the calculation of eigenvalues of a model can be determined in the usual way, from the characteristic equation:

$$\det\,(A(\theta(t) - \lambda I) = 0.$$

Considering the examples in section 4, the vectors containing the orders of the polynomials in each coefficient of the autoregressive terms of the output and of the input terms are defined, as per (8) and (9):

$$\begin{aligned} N &= [N_1 \quad N_2 \quad \cdots \quad N_n] \in \mathbb{N}^n \\ M &= [M_1 \quad M_2 \quad \cdots \quad M_m] \in \mathbb{N}^m \end{aligned} \tag{11}$$

The number ξ of parameters to be identified can be calculated by means of:

$$\xi = n + m + \textstyle\sum_{i=1}^{n} N_i + \sum_{j=1}^{m} M_j. \tag{12}$$

Given the system (4), structured as per (5) to (12), ***Theorem 1*** shows how the polynomial coefficients to be identified are determined, from the resolution of a system of linear equations and the input, output and varying parameter data, all previously measured in the physical system.

Theorem 1 – Consider u_k, y_k and θ_k, with $k \in \{1,\cdots, p\}$, the measured data series referring, respectively, to the input, the output and the varying parameter of the system in (4). The polynomial coefficients in (5) and (6) of the LPV model can be determined by solving the following system of linear equations:

$$AX = \begin{bmatrix} A_\alpha & A_\beta \end{bmatrix} \begin{bmatrix} X_\alpha \\ X_\beta \end{bmatrix} = B, \tag{13}$$

where:

$$A_\alpha = \begin{bmatrix} Y_{n-1,N_1} & Y_{n-2,N_2} & \cdots & Y_{0,N_n} \\ Y_{n,N_1} & Y_{n-1,N_2} & \cdots & Y_{1,N_n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{p-1,N_1} & Y_{p-2,N_2} & \cdots & Y_{p-n,N_n} \end{bmatrix},$$

$$A_\beta = \begin{bmatrix} U_{n-r,M_1} & U_{n-r-1,M_2} & \cdots & U_{n-r-m+1,M_m} \\ U_{n-r+1,M_1} & U_{n-r,M_2} & \cdots & U_{n-r-m+2,M_m} \\ \vdots & \vdots & \ddots & \vdots \\ U_{p-r,M_1} & U_{p-r-1,M_2} & \cdots & U_{p-r-m+1,M_m} \end{bmatrix},$$

$$X_\alpha = -[\alpha_{1,0} \quad \alpha_{1,1} \quad \cdots \quad \alpha_{1,N_1} \quad \alpha_{2,0} \quad \cdots \quad \alpha_{n,N_n}]^T,$$
$$X_\beta = [\beta_{1,0} \quad \beta_{1,1} \quad \cdots \quad \beta_{1,M_1} \quad \beta_{2,0} \quad \cdots \quad \beta_{m,M_m}]^T,$$
$$B = [y_n \quad y_{n+1} \quad y_{n+2} \quad \cdots \quad y_p]^T \in \mathbb{R}^{p-n+1},$$
$$Y_{k,p} = y_k\begin{bmatrix}1 & \theta_{k-\gamma} & \theta_{k-\gamma}^2 & \cdots & \theta_{k-\gamma}^p\end{bmatrix} \in \mathbb{R}^{p+1},$$
$$U_{k,p} = u_k\begin{bmatrix}1 & \theta_{k-\gamma} & \theta_{k-\gamma}^2 & \cdots & \theta_{k-\gamma}^p\end{bmatrix} \in \mathbb{R}^{p+1}.$$

Demonstration: The system of linear equations in (13) comes directly from (7). For a given instant of time $t = kT$, equation (7) can be rewritten with the help of (8) and (9) as follows:

$$y_k = -(\alpha_{1,0} + \cdots + \alpha_{1,N_1}\theta^{N_1})y_{k-1} - \cdots$$
$$-(\alpha_{n,0} + \cdots + \alpha_{n,N_n}\theta^{N_n})y_{k-n} +$$
$$+(\beta_{1,0} + \cdots + \beta_{1,M_1}\theta^{M_1})u_{k-r} + \cdots$$
$$+(\beta_{m,0} + \cdots + \beta_{m,M_m}\theta^{M_m})u_{k-r-m+1},$$

so that:

$$y_k = [Y_{k-1,N_1} \cdots Y_{k-n,N_n} U_{k-r,M_1} \cdots$$
$$\cdots U_{k-r-m+1,M_m}]\begin{bmatrix} X_\alpha \\ X_\beta \end{bmatrix}.$$

Considering that $k \in \{n, n+1, \cdots, p\}$, we arrive at the system of linear equations in (13).

The linear system resulting from ***Theorem 1*** is overdetermined. Several methods can be used to solve it, such as pseudo-inverse, partial scaling and pivoting, Gauss-Jordan and others. An alternative is to use nonlinear programming methods, seeking the minimization of $\|AX - B\|$, which is a convex problem [37]. However, in this case, by means of mathematical transformations, it is also possible to use linear programming methods, such as LMI resolution packages [38].

In the case of a multivariable system with ω inputs and ζ outputs, it is assumed that each of the outputs can be identified independently, that is, that the original problem can be decomposed into a set of ζ Multiple-Input Single-Output (MISO) problems with ω independent inputs. Thus, for each output i, with $i \in \{1,2,\cdots,\zeta\}$

$$\mathrm{A}_i(q,\theta)y_{i,k} = \mathrm{B}_1(q,\theta)u_{1,k} + \cdots + \mathrm{B}_\omega(q,\theta)u_{\omega,k}, \quad (14)$$

Where $y_{i,k}$ represents the output y_i at instant $t = kT$ Using ***Theorem 1*** and developing (14), analogously to the SISO case, it is possible to determine a system of linear equations to calculate the coefficients dependent on the varying parameter θ in A_i , B_1 ,$\cdots$, B_ω.

With multiple inputs, the vector **M** in (11) is transformed into a matrix, with each row presenting the degrees of the polynomial expansions of the terms of each input. The number of coefficients ξ to be identified in (12), considering a single varying parameter θ, will be:

$$\xi = n + m + \sum_{i=1}^{n} N_i + \sum_{i=1}^{\omega}\sum_{j=1}^{m} M_{i,j}. \quad (15)$$

3.1 Adjustment Indices

The assessment of the adjustment error between the behavior of the model and the real dynamic system is carried out in two parts. In the first part, called the coefficient adjustment error, the error in the calculation of the coefficients during the identification process is tested, based on the data measured at the input and output of the physical system. Using (13) and that $v = p - n + 1$, the adjustment error vector is defined as:

$$\begin{bmatrix} e_{a,1} \\ e_{a,2} \\ \vdots \\ e_{a,v} \end{bmatrix} = \begin{bmatrix} b_1 - a_{11}x_1 - \cdots - a_{1v}x_v \\ b_2 - a_{21}x_1 - \cdots - a_{2v}x_v \\ \vdots \\ b_v - a_{v1}x_1 - \cdots - a_{vv}x_v \end{bmatrix},$$

and the adjustment indexes of the coefficients as:

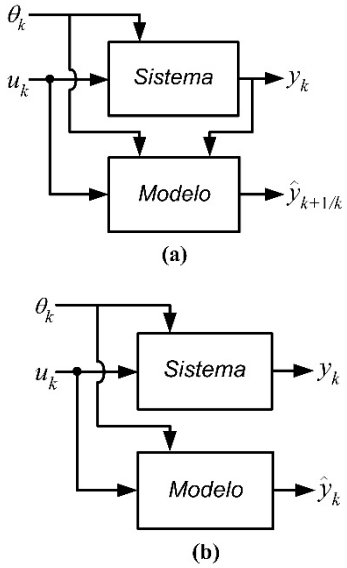
$$J_2^a = \|E_a\|_2 = \sqrt{|e_{a,1}|^2 + |e_{a,2}|^2 + \ldots + |e_{a,v}|^2}, \quad (16)$$

$$J_{\infty}^{a} = \left\| E_a \right\|_{\infty} = \max\{|e_{a,1}|, |e_{a,2}|, \ldots, |e_{a,v}|\}, \quad (17)$$

It can be seen that B is the vector containing the measured outputs of the system. In this case, the error for generating the current output is evaluated, considering that the previously measured outputs are available. This corresponds to the one-step-ahead prediction of the output, that is, $\hat{y}_{k+\frac{1}{k}}$, as illustrated in Figure 1(a). It is also equivalent to saying that the model uses the previous output measurements of the system in real time for the prediction. It is important to note that the use of the model in this format should not be confused with recursive prediction, since its coefficients are already determined and will not be adjusted during operation.

Once the model has been identified, it is possible to evaluate the prediction error independently of the system, based on a simulation, which is much more rigorous. A new input signal is adopted for validation, with the same initial conditions for the system and model. In this case, the model is considered autonomous, that is, its output $\hat{y}_k$ is generated exclusively from the input provided and the trajectory of the varying parameter θ without interference from the system output y, as shown in Figure 1(b).

Figure 1 - Simulation of the predicted outputs: (a) one step ahead; (b) independent.



The error for a horizon of h simulation periods is evaluated as follows:

$$E_s = \begin{bmatrix} e_{s,1} \\ e_{s,2} \\ \vdots \\ e_{s,h} \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \quad \vdots \\ y_h - \hat{y}_h \end{bmatrix} = y - \hat{y},$$

and the simulation or validation indices by:

$$J_2^s = \left\| E_s \right\|_2 = \sqrt{|e_{s,1}|^2 + |e_{s,2}|^2 + \ldots + |e_{s,h}|^2}, \quad (18)$$

$$J_{\infty}^{s} = \left\| E_s \right\|_{\infty} = \max\{|e_{s,1}|, |e_{s,2}|, \ldots, |e_{s,h}|\}, \quad (19)$$

It should be noted that, using the same initial conditions, input signals, sampling rate and sample size in the one-step-ahead simulation and in the free simulation, we obtain $J_2^a \leq J_2^s$.

3.2 Expansion by Multiple Varying Parameters

In the situation where there are multiple varying parameters $\theta(t) = [\theta_1(t)\ \theta_2(t) \cdots \theta_d(t)]^T$, the expansion of the model coefficients can be performed in analogy with (8) and (9), also considering the cross terms of the varying parameters. Thus, for the case of two varying parameters, the coefficients of the model output terms and, equivalently, for the input ones, would have the following format:

$$\begin{aligned} a_i = a_i(\theta_1, \theta_2) = \alpha_{i,0} + \alpha_{i,1}\theta_1 + \alpha_{i,2}\theta_1^2 + \cdots + \\ + \alpha_{i,N_i}\theta_1^{N_i} + \alpha_{i,N_i+1}\theta_2 + \cdots + \alpha_{i,2N_i}\theta_2^{N_i} + \\ + \alpha_{i,2N_i+1}\theta_1\theta_2 + \alpha_{i,2N_i+2}\theta_1\theta_2^2 + \alpha_{i,2N_i+3}\theta_1^2\theta_2 + \\ + \cdots + \alpha_{i,\frac{N_i(N_i+3)}{2}}\theta_1^{N_i-1}\theta_2 \end{aligned}$$

Also in this case, the number of parameters ξ to be identified is significantly impacted by the number of components d of the vector $\theta(t)$, which makes the proposed methodology unfeasible for $d \gg 1$. For $d=2$:

$$\begin{aligned} \xi = n + m + \sum_{i=1}^{n} N_i + \sum_{i=1}^{\omega}\sum_{j=1}^{m} M_{i,j} + \\ + \sum_{i=1}^{n} \gamma_i \frac{N_i(N_i-1)}{2} + \sum_{i=1}^{\omega}\sum_{j=1}^{m} \rho_{i,j} \frac{M_{i,j}(M_{i,j}-1)}{2} \end{aligned}$$

where for $i \in \{1, \cdots, n\}$: $\gamma_i = \begin{cases} 0, N_i \leq 1 \\ 1, N_i > 1 \end{cases}$;

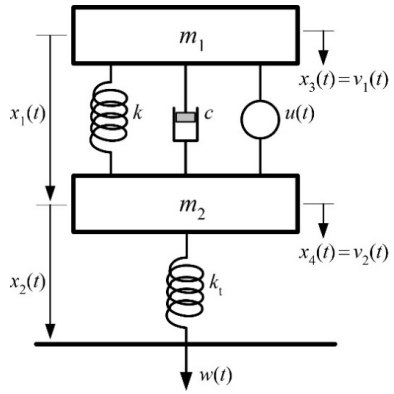
and for $i \in \{1, \cdots, w\}, j \in \{1, \cdots, m\}$, $\rho_{i,j} = \begin{cases} 0, M_{i,j} \leq 1 \\ 1, M_{i,j} > 1 \end{cases}$.

4. Identification of LPV models in a quarter-car suspension system.

This section uses the fourth-order car active suspension model, as per [39] as cited in [40]. This model, in each subsection below, underwent new considerations in relation to the varying parameter, which led to different degrees of nonlinearity and enabled a more detailed and extended analysis of the proposed method.

Figure 2 illustrates the physical model of the quarter-car active suspension. The constant m_1 represents the damped mass of a quarter of the car and m_2 the undamped mass of a wheel and tire assembly. The upper spring, with elastic constant k and the shock absorber, with damping constant c, represent the very suspension of the car. The lower spring, with elasticity constant k_t, refers to the damping generated by the deformation of the tire during the movement of the car on the road. The disturbance $w(t)$ represents the excitation input of the model and consists of a vertical velocity signal due to irregularities found on the road. The input $u(t)$ represents the actuation force produced by the active suspension mechanism. Its purpose is to isolate vibrations in the mass m_1, in addition to providing greater car adhesion to the road.

Figure 2 - Physical quarter-car active suspension model



For the mathematical modeling of the assembly, the state variables of the system can be defined as in [39]:
x_1: distance between the masses m_1 and m_2, from the equilibrium position;
x_2: distance between the wheel axle and its base, also from the equilibrium position;
x_3:vertical speed $v_1(t)$ of the body in relation to the inertial reference;
x_4: vertical speed $v_2(t)$ of the wheel axle in relation to the inertial reference.
Based on the definition of the states presented, it is possible to conclude that:

$$\dot{x}_1 = x_4 - x_3, \tag{20}$$

And also,

$$w(t) = \dot{x}_2 + x_4 \quad \text{or} \quad \dot{x}_2 = w(t) - x_4, \tag{21}$$

Applying Newton's second law to the mass m_1 and considering that there is a linear dependence of the force on the speed in the shock absorber, through the constant c:

$$m_1\dot{x}_3 = kx_1 + c\dot{x}_1 + u. \tag{22}$$

From (20) in (22):

$$\dot{x}_3 = \frac{k}{m_1}x_1 - \frac{c}{m_1}x_3 + \frac{c}{m_1}x_4 + \frac{1}{m_1}u(t). \tag{23}$$

Applying Newton's second law to the mass m_2:

$$m_2\dot{x}_4 = -kx_1 + k_t x_2 + cx_3 - cx_4 - u(t),$$

$$\dot{x}_4 = -\frac{k}{m_2}x_1 + \frac{k_t}{m_2}x_2 + \frac{c}{m_2}x_3 - \frac{c}{m_2}x_4 - \frac{1}{m_2}u(t), \tag{24}$$

In addition, the acceleration of the damped mass m_1 and the state x_2 were considered as output variables [40]. Thus, it is possible to write the mathematical model in state space form, according to:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \\ \frac{k}{m_1} & 0 & -\frac{c}{m_1} & \frac{c}{m_1} \\ -\frac{k}{m_2} & \frac{k_t}{m_2} & \frac{c}{m_2} & -\frac{c}{m_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m_1} \\ -\frac{1}{m_2} \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} w(t)$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{k}{m_1} & 0 & -\frac{c}{m_1} & \frac{c}{m_1} \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} \frac{1}{m_1} \\ 0 \end{bmatrix} u(t). \quad (25)$$

where $y(t) = [y_1(t)\, y_2(t)]^T$, $y_1(t)$ is the acceleration of the mass m_1 and $y_2(t)$ the displacement of the mass m_2. The model outputs, according to [40], are related to the acceleration of the damped mass x_3 and the deformation of the tire, x_2.

The nominal values of the parameters adopted [39] were: $m_1 = 288{,}9$ kg; $m_2 = 28{,}58$ kg; $c = 850\frac{\text{Ns}}{\text{m}}$; $k = 10.000\frac{\text{N}}{\text{m}}$; and $k_t = 155.900\frac{\text{N}}{\text{m}}$. In all cases discussed below, the sampling period T=0,0025s and simulation duration of 2 s were used, which totals 800 periods. In this article, the models were determined with in γ=0 em (10).

In the case of active suspension, considering all measurable states, the control law $u(t) = Kx(t)$ was adopted, where $x(t)$ is the state vector, knowing that the value of K employed was obtained in order to maintain a compromise between passenger comfort and tire adhesion to the road. The value of K presented in [40] and used in this work was:

$$K = 10^3 \times [-9{,}9997 \quad -0{,}0002 \quad +0{,}8325 \quad -0{,}8461]$$

For the simulation of the data to be used in the identification, it was considered that the base of the tire, in contact with the road, is subjected to a disturbance at speed w(t) of the form:

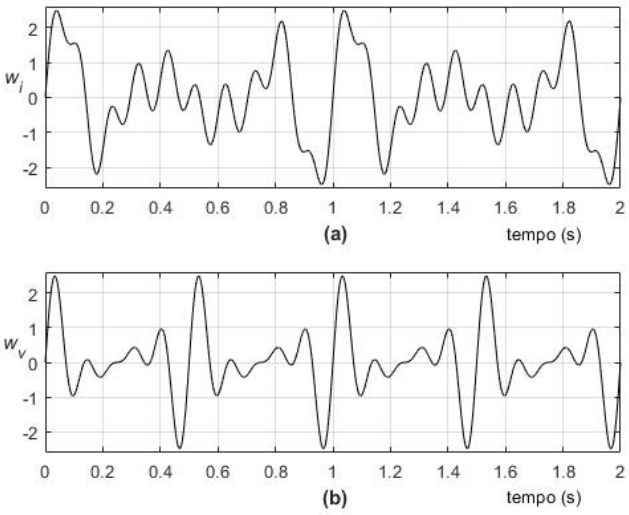
$$w(t) = w_i(t) = 0.9\text{sen}(6\text{p}t) + 0.5\text{sen}(10\text{p}t) + \\ + 0.75\text{sen}(8\text{p}t) + 0.6\text{sen}(20\text{p}t), \quad (26)$$

and for the validation of the models, the same signal as in [40] was used, represented by the equation:

$$w(t) = w_v(t) = 0.6\text{sen}(8\text{p}t) + 0.75\text{sen}(12\text{p}t) + \\ + 0.9\text{sen}(16\text{p}t) + 0.5\text{sen}(20\text{p}t), \quad (27)$$

Figure 3 shows the graphs of the input signals $w(t)$ used for identification and validation.

Figure 3 - Input signals for: (a) identification w_i; (b) validation w_v.



4.1 LTI Models

In this item and the following, it was considered that the mass $m_1(\theta) = 288{,}9 + 100\theta$ and the signal of the varying parameter $\theta(t) = 0{,}5t$ for $0 \leq t \leq 2$s, in the identification phase. In the validation of the model, another signal was used for the varying parameter, that is, $\theta(t) = \text{sen}(0{,}5\pi t)$ Thus, the mass presented values in the range of $288{,}9 \leq m_1 \leq 388{,}9$, which could correspond to the addition of passengers and luggage in the undamped mass. This consideration is quite conservative in terms of the variation rate of the varying parameter, taking into account a 2-second simulation period. In other words, a variation rate of the parameter slightly higher than what can happen in practice, but respecting that every real physical system is a low-pass system.

Figures 4 and 5 show, respectively, the system outputs in the cases of passive and active suspension as a function of the excitation with the validation signal and $w_v(t)$ e $m_1(\theta)$.

Figure 4 - Output signals for validating the passive suspension model: (a) y_1; (b) y_2.

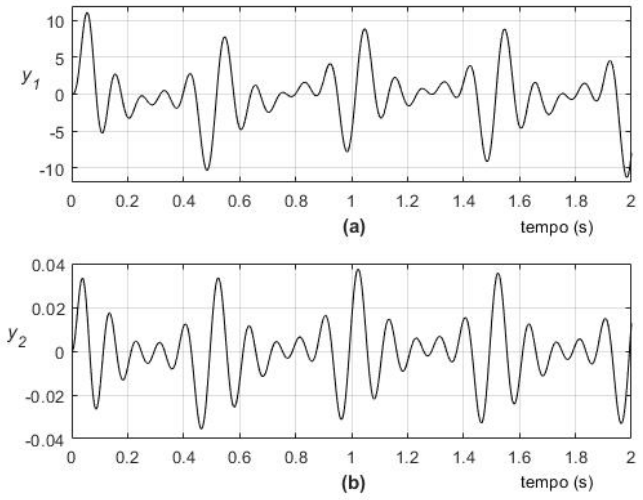


Figure 5 - Output signals for validating the active suspension model: (a) ; (b) .

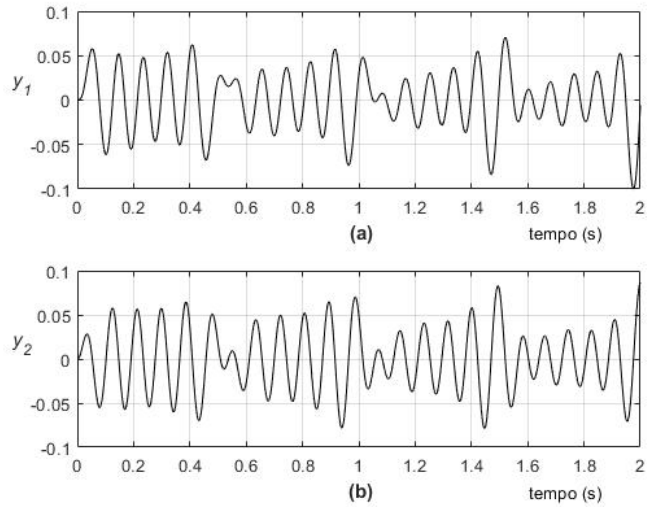


Table 1 shows the characteristics of four LTI models that were adjusted for outputs y_1 and y_2 in the cases of passive and active suspension, remembering that $m_1 = m_1(\theta)$.

It is worth mentioning [40] that the tire deformation x_2, when compared to the outputs y_2 in Figures 4(b) and 5(b), in the active suspension, the excursion was approximately double, while the acceleration y_1 in Figure 5(a) of the damped mass m_1 was of the order of 1% of that occurred for the passive suspension in Figure 4(a).

Table 1 - Characteristics of the identified LTI models.

	Models			
	M1	**M2**	**M3**	**M4**
susp.	passive	passive	active	active
output	y_1	y_2	y_1	y_2
n/m/r	4/4/1	4/4/1	2/2/1	2/2/1
N	[0 0 0 0]	[0 0 0 0]	[0 0]	[0 0]
M	[0 0 0 0]	[0 0 0 0]	[0 0]	[0 0]
ξ	8	8	4	4
J_2^a	3.4697e–3	5.3165e–7	1.2436e–3	2.8967e–8
J_∞^a	6.0661e–4	8.3421e–8	1.6489e–4	2.5778e–9
J_2^s	38.274	1.0384e–2	1.5798e–1	1.0063e–5
J_∞^s	4.2958	6.8445e–4	2.4096e–2	8.5410e–7
max| λ |	0.9952	0.9970	0.9994	0.9998

Table 1 shows that the elements of the vectors **N** and **M** in (11) were zero, which corresponds to the adjustment of the LTI models. The values of J_2^a and J_∞^a in this table and in the following ones correspond to the cost of one-step-ahead prediction, according to Figure 1(a), using the validation input $w_v(t)$. It is also worth mentioning that the LTI M2 and M4 models of the outputs y_2, for the passive and active suspension, had satisfactory adjustments, and these outputs are no longer addressed in the next cases. The graphs of M2 and M4 are exactly those shown in Figures 4(b) and 5(b).

Figure 6(a) shows the outputs of the system, in blue, and of the identified LTI M1 model, in red, for passive suspension. Figure 6(b) illustrates the absolute value of the error between these output signals. Figure 7 is the equivalent of Figure 6 for the active suspension case.

Figure 6 - (a) Outputs y_1 of the system (blue) and the LTI M1 model (red) with passive suspension; (b) absolute value of the error between these outputs.

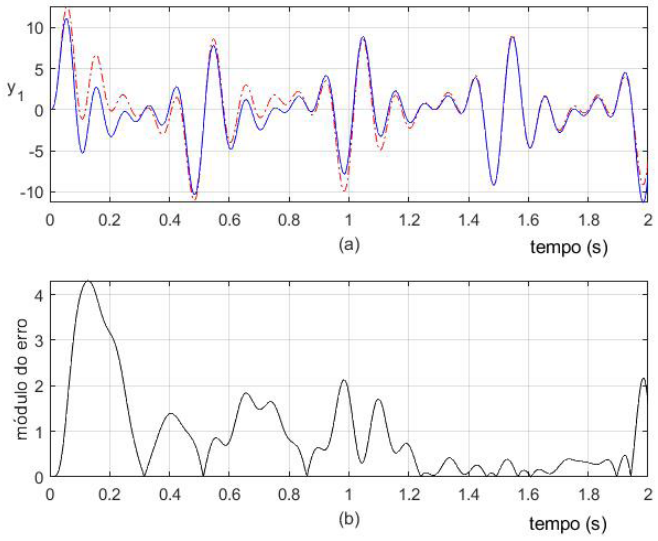
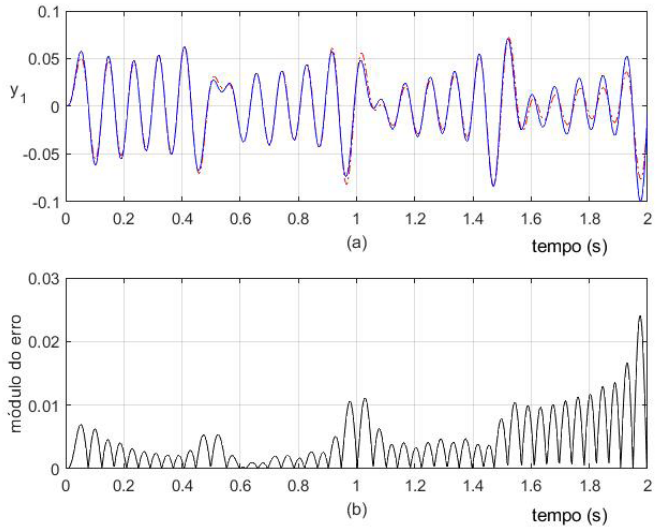


Figure 7- (a) Outputs y_1 of the system (blue) and the LTI M3 model (red) with active suspension; (b) absolute value of the error between these outputs.



It should be noted that the maximum values of the adjustment errors presented in Figures 6(b) and 7(b) correspond to the respective values of J_∞^s in Table 1 for M1 and M3.

4.2 LPV models with one exogenous parameter

In this case, the LPV M5 and M6 models in Table 2 were identified as the best models that reproduced, respectively, the behavior of the outputs y_1 of the car passive and active suspension, considering that the mass m_1 varies over time.

Table 2 - Characteristics of the models identified in 4.2 and 4.3.

	Models			
	M5	**M6**	**M7**	**M8**
susp.	passive	active	passive	passive
n/m/r	4/4/1	2/2/1	4/4/1	4/4/1
N	[1 0 0 0]	[1 0]	[0 0 0 0]	[1 0 0 0]
M	[2 0 2 2]	[0 1]	[0 0 0 0]	[1 2 1 1]
ξ	15	6	8	21
J_2^a	5,6473e–4	3,1939e–4	3,2598e–3	3,5658e–4
J_∞^a	1,0453e–4	6,0073e–5	4,0924e–4	5,6716e–5
J_2^s	3,1698	9,1048e–2	39,014	2,6949
J_∞^s	3,4643e–1	9,2476e–3	3,8391	2,4190e–1
max| λ |	0.9966	0.9995	1.0012	0.9980

Comparing the values of J_2^s of M5 and M6 in Table 2 with their corresponding M1 and M3 in Table 1, it can be seen that their adjustments are significantly better, but logically at the expense of the increase in the number of parameters.

Figure 8 presents the absolute values of the error between the outputs y_1 of the system and the validated models, both for passive suspension and active suspension. It is worth noting that the maximum absolute values of these errors in the graphs of Figure 8 correspond to the values of J_∞^s in Table 2. Comparing the graphs of Figures 6(b) and 8(a), it can be seen that the adjustment error of the LPV model is in the range of 8% of the values presented by the LTI model. Similarly, for the case

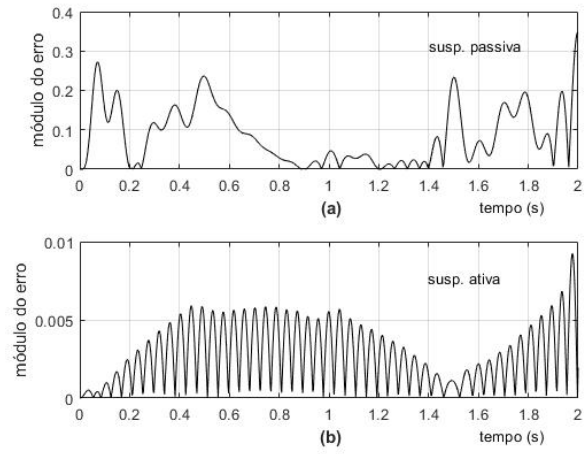
of active suspension with the output y_1, comparing the graphs of Figures 7(b) and 8(b), it can be seen that the adjustment error of the LPV model presents values lower than 40% of those of the LTI model.

An important characteristic observed in LPV models is that their coefficients depend on θ and, consequently, vary over time. For this reason, it is observed that their eigenvalues also change during the simulation. Thus, since the models are discrete, it is desired that the eigenvalues $\lambda_i = \lambda_i(t)$ meet the following condition:

$$\max_{0 \le t \le 2s} |\lambda_i(t)| < 1.$$

in order to maintain the stability of the model. However, it was found that small exceedance of this limit do not always cause mismatches between the model output and that of the plant. Another aspect that deserves to be mentioned refers to the significant increase in the number of parameters to be adjusted in LPV models. Finally, it was observed that the adjustment of the model is highly dependent on the trajectory of the parameter θ and its speed.

Figure 8 - Absolute error value of the output y_1 for the LPV models: (a) M5 and (b) M6.



4.3 LPV models with two exogenous parameters

In this case, it was considered that the system has two independent exogenous varying parameters, which parameterize the mass m_1 and the spring constant k, as follows:

$$m_1(\theta_1) = 288{,}9 + 100\theta_1,$$
$$k(\theta_2) = 9.000 + 2.000\theta_2,$$

where for $0 \le t \le 2$ s:

$$\theta_1 = \theta_1(t) = 0{,}5t,$$
$$\theta_2 = \theta_2(t) = \text{sen}(0{,}5\pi t).$$

Thus, the parameters m_1 and k varied during the simulation in the following intervals:

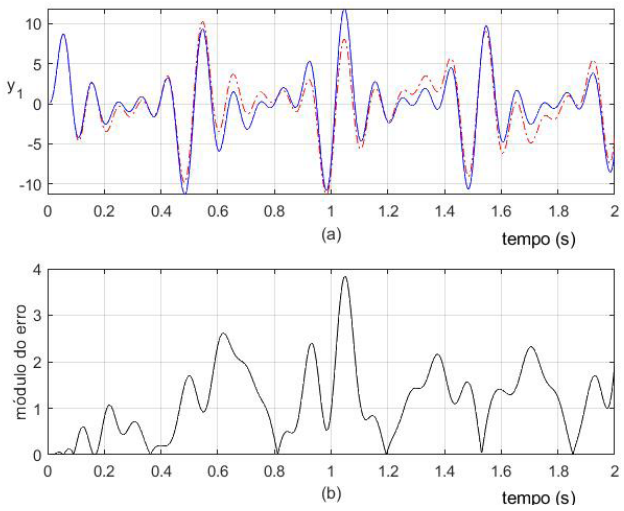
$$288{,}9 \le m_1 \le 388{,}9;$$
$$9.000 \le k \le 11.000.$$

The parameters θ_1 and θ_2 above were used during the identification. In the validation, θ_1 = sen $(0{,}5\pi t)$ and $\theta_2 = 0{,}5t$.

Table 2 presents the characteristics of two models, M7 and M8, adjusted for the output $y_1(t)$, in the case of passive suspension with the simultaneous variation of m_1 and k. M7 is of the LTI type, while M8 is the LPV for the same data. Figure 9 presents the output estimated by M7, in red, and the corresponding system output. Table 2 shows that the adjustment cost J_2^s of M8 was less than 7% of that presented by M7. It is worth mentioning that the output of M8 coincides with the system output, in blue, in Figure 9.

The identification for active suspension was not performed, since the controller k in [39] leads the closed-loop system to instability with the simultaneous variation of m_1 and k.

Figure 9 - (a) Outputs y_1 of the system (blue) and of the LTI M7 model (red) with passive suspension; (b) absolute value of the error between these outputs.



4.4 LPV models with one endogenous parameter

In this case, the mass m_1 = 288,9 kg was considered, fixed at its nominal value, but the spring model was replaced by a more realistic one [41], which considers the elastic constant k to vary from a certain deformation, as shown in Figure 10. Since the spring deformation is the state x_1, then k becomes dependent on it, that is:

$$k = k(x_1) = \begin{cases} 10^4 \frac{\text{N}}{\text{m}}, & |x_1| \leq 0{,}08 \text{ m} \\ \frac{800+10^5(|x_1|-0{,}08)}{|x_1|} \frac{\text{N}}{\text{m}}, & |x_1| > 0{,}08 \text{ m} \end{cases}$$

and (25) becomes a nonlinear Quasi-LPV model, with $\theta = x_1$.

Figure 10 - Suspension spring force versus deformation.

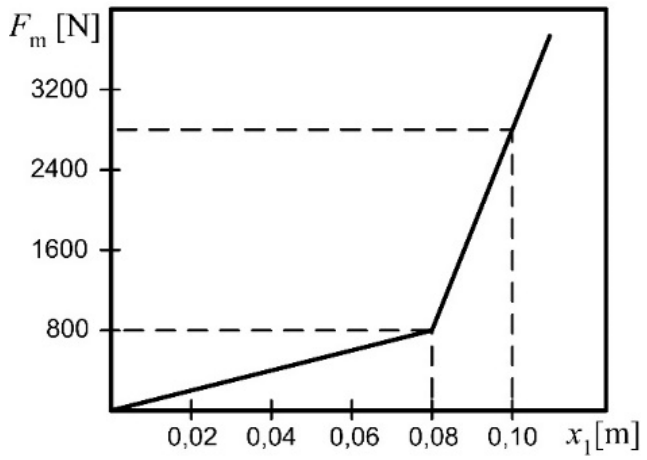


Figure 11 illustrates the behavior of x_1 and k over time for the input signal w_i. When the deformation x_1 exceeds 0.08 m, characterized by the dashed lines in Figure 11(a), the spring constant k becomes variable, as shown in Figure 11(b).

For this case, M9 and M10 from Table 3 were determined, which are distinguished by their parametric structure and the number of parameters. The adjustments of these LPV models were not as good as those identified in the previous cases.

Figure 11 - Temporal evolution for the identification input: (a) state x_1; (b) (b) spring constant.

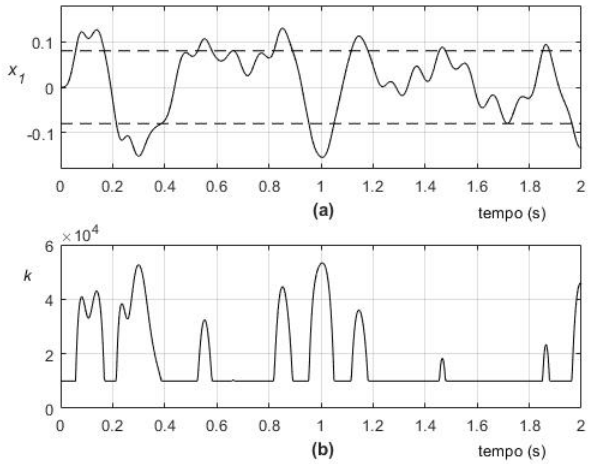


Table 3 - Characteristics of the LPV models identified for passive suspension in 4.4 and 4.5.

	Models			
	M9	**M10**	**M11**	**M12**
n/m/r	4/2/3	2/2/1	4/4/1	4/5/0
N	[5 0 3 0]	[2 2]	[0 3 1 2]	[1 0 0 0]
M	[3 4]	[5 2]	[4 2 1 0]	[0 0 3 1 1]
x	21	15	45	24
J_2^a	4.1019	4.7566	8.0199	9.8109
J_∞^a	1.7896	1.8838	2.8412	3.3706
J_2^s	57.719	61.040	52.349	63.038
J_∞^s	6.4872	6.2101	7.5600	8.5038
max| λ |	1.0104	1.0067	1.1402	0.9294

Figure 12 shows the system response, in blue, and that of the validated M9 model, in red. Nevertheless, it was observed that for the one-step-ahead prediction, this model presents a significant improvement in its adjustments, since its estimated responses practically coincide with that of the system, in blue, in this figure. Table 3 confirms this information based on the

values of J_2^a presented by M9 and M10. Figure 12(b) illustrates the temporal evolution of the maximum eigenvalue module. It is interesting to note that in some time intervals, this exceeds the unitary value.

It is worth noting that since the state x_1 is the varying parameter itself, θ will be changed with the modification of the identification input to the validation input.

4.5 LPV models with two endogenous parameters

In this case, it was assumed that the suspension is subjected to high speeds. Therefore, a more realistic behavior for the force generated by the shock absorber is one that presents a cubic relationship with the compression or expansion speed. The mathematical model no longer has a linear behavior, but can be transformed into the Quasi-LPV form, as shown below. Equation (22) can be rewritten as:

$$kx_1 + c(\dot{x}_1)^3 + u = m_1\dot{x}_3. \tag{27}$$

From (20):

$$(\dot{x}_1)^3 = (x_4 - x_3)^3 = x_4^3 - 3x_4^2x_3 + 3x_4x_3^2 - x_3^3. \tag{28}$$

From (28) in (27):

$$\dot{x}_3 = \frac{k}{m_1}x_1 - \frac{c}{m_1}(3x_4^2 + x_3^2)x_3 + \frac{c}{m_1}(3x_3^2 + x_4^2)x_4 + \frac{u(t)}{m_1}. \tag{29}$$

Developing (24) in an analogous way, we arrive at:

$$\dot{x}_4 = \frac{k}{m_2}x_1 + \frac{k_t}{m_2}x_2 + \frac{c}{m_2}(3x_4^2 + x_3^2)x_3 + \\ -\frac{c}{m_2}(3x_3^2 + x_4^2)x_4 - \frac{u(t)}{m_2} \tag{30}$$

Using (20), (21), (29) and (30), we arrive at the high-speed active suspension model:

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\\ \dot{x}_3\\ \dot{x}_4\end{bmatrix} = \begin{bmatrix}0 & 0 & -1 & 1\\ 0 & 0 & 0 & -1\\ \frac{k}{m_1} & 0 & -\frac{c}{m_1}(3x_4^2 + x_3^2) & \frac{c}{m_1}(3x_3^2 + x_4^2)\\ -\frac{k}{m_2} & \frac{k_t}{m_2} & \frac{c}{m_2}(3x_4^2 + x_3^2) & -\frac{c}{m_2}(3x_3^2 + x_4^2)\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix} + \\ + \begin{bmatrix}0\\ 0\\ \frac{1}{m_1}\\ -\frac{1}{m_2}\end{bmatrix}u(t) + \begin{bmatrix}0\\ 1\\ 0\\ 0\end{bmatrix}w(t) \tag{31}$$

$$\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}\frac{k}{m_1} & 0 & -\frac{c}{m_1}(3x_4^2 + x_3^2) & \frac{c}{m_1}(3x_3^2 + x_4^2)\\ 0 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix} + \begin{bmatrix}\frac{1}{m_1}\\ 0\end{bmatrix}u(t)$$

From the model in (31), it can be seen that the states x_3 and x_4 compose the matrices A and C of the dynamic, transforming it into a nonlinear Quasi-LPV model.

Figure 12 - (a) Outputs of the system (blue) and the M9 model (red) with passive suspension; (b) Maximum eigenvalue modulus during validation.

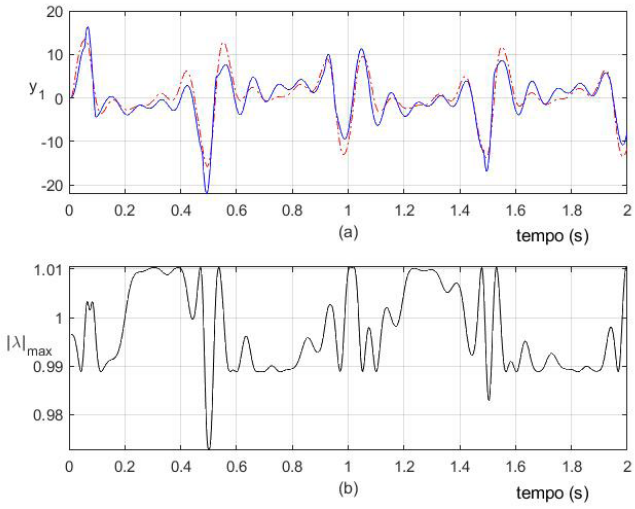


Figure 13 shows the temporal evolution of the states x_3 and x_4 in view of the application of the validation input. In this case, these states will be considered the varying parameters of the LPV model.

Table 3 presents the characteristics of M11 and M12, adjusted for this problem. Although M12 has a worse performance than M11 in terms of the cost J_2^s, it was selected due to the parsimony criterion, as it presents a significantly smaller number of parameters.

Even though the adjustment was not perfect, it is still much better than that of the corresponding LTI model, which has $J_2^s = 90,21$. This model was not presented in the tables, but Figure 14 shows the system output curve (blue), of M11 (black) and the output of this LTI model (red).

Figure 13 - Evolution of the state with the validation input: (a) x_3; (b) x_4.

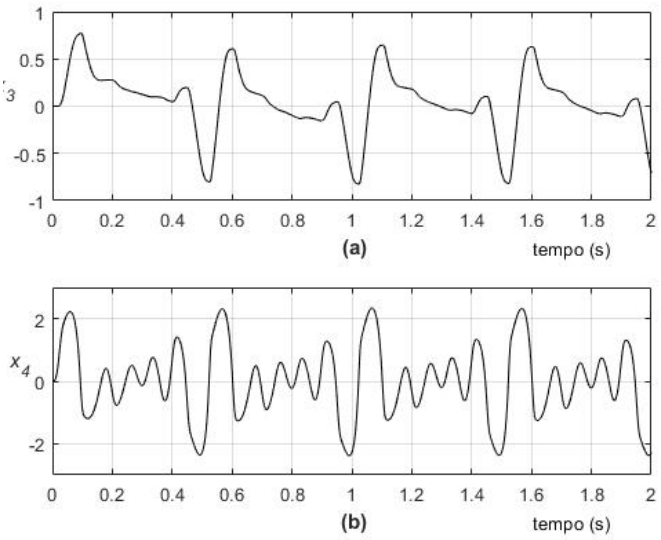
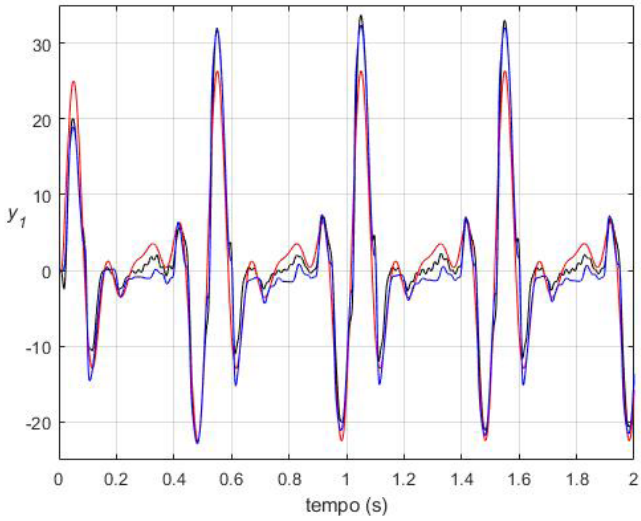


Figure 14 - Outputs of the passive suspension for the system (blue), M11 (black) and LTI (red).



5. Conclusions

In this article, a multivariate method for identifying LPV models with polynomial coefficients was presented. Among the applications of the method, in addition to the identification itself, there is the possibility of approximating nonlinear models by LPV models, with the purpose of applying LPV control techniques. The results were explored through an example related to a car suspension. Several cases were addressed, with multiple endogenous and exogenous varying parameters, seeking to approximate even nonlinear Quasi-LPV systems. Some contributions in relation to the ideas of [24] were implemented, such as the extension of the method to multivariate systems, obtaining the solution from a data batch, the use of polynomials with independent degrees per parameter, the possibility of performing the time lag between the current output and the most recent input in the model, as well as the extension of the method to multiple varying parameters. Although the adjustments of the LPV models were significantly better than those of the LTI models, on the other hand, there was also a large increase in the number of parameters to be determined. It was also observed that the graph of the maximum module of the eigenvalues of the model over time is an important tool for analyzing the quality of the identified model and its chosen parametric structure. It is worth remembering that the one-step-ahead prediction showed excellent results for all the models presented, although there is a dependence on the output measurements of the system in real time. Finally, it was verified, mainly in the approximation of nonlinear systems, the strong dependence of the adjustment of the model on the trajectory of the variable parameter and its variation rate.

Acknowledgments

This article was prepared with the support from the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

References

[1] SIMÕES, A.M. Controle linear a parâmetros variáveis de sistemas não lineares. Dissertação (Mestrado) — Instituto Militar de Engenharia, Rio de Janeiro, 2004.
[2] RUGH, W. J.; SHAMMA, J. S.; Survey paper research on gain scheduling. Automatica, v.36, n.10, p.1401-1425, 2000.
[3] MOHAMMADPOUR, J.; SCHERER, C. W. (Ed.). Control of Linear Parameter Varying Systems with Applications. New York: Springer, 2012.
[4] BECKER, G.; PACKARD, A. K. Robust performance of linear parametrically va-rying systems using parametrically dependent linear feedback. Systems & Control Letters, v.23, n.3, p.205-215, 1994.
[5] APKARIAN, P.; GAHINET, P.; BECKER, G.; Self-scheduled H∞ control of linear parameter-varying systems: A design example. Automatica, 31, n.9, p.1251-1261, 1995.
[6] PELLANDA, P. C. ; Commande de Systèmes Instationnaires: Séquencement de Compensateurs et Commande LPV. Thèse (Docteur) — L'Ecole Nationale Supérieure de L'Aéronautique et de L'Espace, Toulouse, 2001.
[7] WU, F.; Control of linear parameter varying systems. Thesis (PhD, University of Berkeley, Berkeley, 1995.
[8] APKARIAN, P.; ADAMS, R. J.; Advanced gain-scheduling techniques for uncer-tain systems. IEEE Transactions on Control Systems Technology, v.6, n.1, 1998.
[9] APKARIAN, P.; GAHINET, P.; A convex characterization of gain-scheduled H∞ controllers. IEEE Transactions on Automatic Control, v.40, n.5, p.853-864, 1995.
[10] ZHOU, K.; DOYLE, J. C.; GLOVER, K.; Robust and optimal control. London: Pearson, 1996.
[11] HOFFMANN, C.; WERNER, H.; A survey of linear parameter-varying control applications validated by experiments or high-fidelity simulations. IEEE Transacti-ons on Control Systems Technology, v.23, n.2, p.416-433, 2015.
[12] LJUNG, L.; System Identification: Theory for the User Second Edition. London: Prentice Hall PTR, 1999.
[13] HSU, K.; VINCENT, T. L.; POOLLA, K.; Nonparametric methods for the identification of linear parameter varying systems. In: Proceedings of the International Symposium on Computer-Aided Control System Design, San Antonio, p.846-851, 2008.
[14] PIGA, D.; TÓTH, R.; LPV model order selection in an LS-SVM setting. In: 52nd IEEE Conference on Decision and Control, 2013.
[15] GOLABI, A.; MESKIN, N.; TÓTH, R.; MOHAMMADPOUR, J.; A bayesian approach for estimation of linear regression LPV models. In: IEEE Conference on Decision and Control, 53, p.2555-2560, 2014.
[16] DARWISH, M. A. H.; COX, P. B.; PROIMADIS, I.; PILLONETTO, G.; TÓTH, R.; Prediction-error identification of LPV systems: A nonparametric gaussian regression approach. Automatica, v.97, p.92-103, 2018.
[17] SANTOS, P. L.; PERDICOÚLIS, T. P. A.; NOVARA, C.; RAMOS, J. A.; RIVERA, D. E.; Linear parameter-varying system identification: new developments and trends. Hackensack: World Scientific, 2012.
[18] TÓTH, R.; Modeling and identification of linear parameter-varying systems. Berlin: Springer, 2010.
[19] FUJIMORI, A.; LJUNG, L.; Model identification of linear parameter varying air-craft systems. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, v.220, n.4, p.337-346, 2006.
[20] GHOSH, D.; BOMBOIS, X.; HUILLERY, J.; SCORLETTI, G.; MERCÈRE, G.; Optimal identification experiment design for LPV systems using the local approach, Automatica, v.87, p.258-266, 2018.
[21] CERONE, V.; PIGA, D.; REGRUTO, D.; Set-membership LPV model identification of vehicle lateral dynamics. Automatica, v.47, n.8, p.1794-1799, 2011.
[22] CERONE, V.; PIGA, D.; REGRUTO, D.; A convex relaxation approach to set-membership identification of LPV systems. Automatica, v.49, n.9, p.2853-2859, 2013.
[23] LAURAIN, V.; GILSON, M.; TÓTH, R.; GARNIER, H.; Refined instrumental va-riable methods for identification of LPV Box–Jenkins models. Automatica, v.46, n.6, p.959-967, 2010.
[24] BAMIEH, B.; GIARRÉ, L.; Identification of linear parameter varying models. International Journal of Robust and Nonlinear Control, v.12, n.9, p.841-853, 2002.
[25] TÓTH, R.; LAURAIN, V.; ZHENG, W. X.; POOLLA, K.; Model structure learn-ing: A support vector machine approach for LPV linear-regression models. In: IEEE Conference on Decision and Control and European Control Conference. 50, p.3192-3197, 2011.
[26] VAPNIK, V. N.; Statistical Learning Theory. Hoboken: John Wiley & Sons, 1998.
[27] SCHÖLKOPF, B.; SMOLA, A. J.; Learning with kernels: support vector machi-nes, regularization, optimization, and beyond. Cambridge, MIT Press, 2002.

[28] PIGA, D.; COX, P.; TÓTH, R.; LAURAIN, V.; LPV system identification under noise corrupted scheduling and output signal observations. Automatica, v.53, p.329-338, 2015.
[29] DARWISH, M. A. H.; COX, P. B.; PROIMADIS, I.; PILLONETTO, G.; TÓTH, R.; Prediction-error identification of LPV systems: A nonparametric Gaussian regression approach. Automatica, v.97, p.92-103, 2018.
[30] PILLONETTO, G.; DINUZZO, F.; CHEN, T.; de Nicolao, G.; LJUNG, L.; Kernel methods in system identification, machine learning and function estimation: A sur-vey. Automatica, v.50, n.3, p.657-682, 2014.
[31] MEJARI, M.; PIGA, D.; BEMPORAD, A.; A bias-correction method for closed-loop identification of linear parameter-varying systems. Automatica, v.87, p.128-141, 2018.
[32] BRESCHI, V.; FORMENTIN, S.; RALLO, G.; CORNO, M.; SAVARESI, S. M.; Vehicle sideslip estimation via kernel based LPV identification: Theory and experi-ments. Automatica, v.122, 2020.
[33] LAURAIN, V.; TÓTH, R.; PIGA, D.; DARWISH, M. A. H.; Sparse RKHS estimation via globally convex optimization and its application in LPV-IO identification. Automatica, v.115, 2020.
[34] ALEGRIA, E. J.; GIESBRECHT, M.; BOTTURA, C. P.; Causal regression for online estimation of highly nonlinear parametrically varying models. Automatica, v.125, 2021.
[35] BECKER, G. S.; Quadratic stability and performance of linear parameter de-pendent systems. Ph.D. thesis, University of California-Berkeley, Berkeley, 1993.
[36] MARCOS, A.; BALAS, G.J.; Development of linear-parameter-varying models for aircraft. Journal of Guidance, Control,and Dynamics. v.27, n.2, p.218-228, 2004.
[37] AGUIRRE, L. A.; Introdução à Identificação de Sistemas. Técnicas Lineares e Não Lineares: Teoria e Aplicação. Belo Horizonte: Editora UFMG, 2015.
[38] LOFBERG, J.; Yalmip: a toolbox for modeling and optimization in MATLAB. In: IEEE International Conference on Robotics and Automation, p.284-289, 2004.
[39] EL-MADANY, M. M.; AL-MAJED, M. I.; Quadratic synthesis of active controls for a quarter-car model. Journal of Vibration and Control, v.7, n.8, p.1237-1252, 2001.
[40] PALHARES, R. M.; GONÇALVES, E. N.; Desigualdades matriciais lineares em controles. In: AGUIRRE, L. A. Enciclopédia de automática, controle e automação, São Paulo: Editora Blucher, 2007.
[41] KARNOPP, D. C.; MARGOLIS, D. L.; ROSEMBERG, R. C.; System Dynamics. Modeling, Simulation, and Control of Mechatronic Systems. John Wiley & Sons, 2012.

Appendix A - Model coefficients

Table A1 - Coefficients of the identified models.

	M1	M2	M3	M4
1	–3.8899	–3.8875	–1.9650	–1.9657
2	5.7064	5.6979	9.9883e–1	9.9966e–1
3	–3.7426	–3.7329	2.0657e–4	2.4858e–3
4	9.2613e–1	9.2258e–1	1.8604e–4	–2.4850e–3
5	–1.4126e–2	2.4678e–3		
6	1.2122e–1	–7.2040e–3		
7	–1.9795e–1	7.0104e–3		
8	9.0897e–2	–2.2742e–3		

	M5	M6	M7	M8
1	–3.8854	–1.9650	–3.9020	–3.8902
2	–1.6614e-6	–4.2788e–5	5.7371	1.5265e–6
3	5.6915	9.9900e–1	–3.7683	8.0307e–7
4	–3.7266	1.9014e–4	9.3315e–1	5.7055
5	9.2050e–1	2.6350e–4	1.3017e–1	–3.7401
6	4.7693e–2	–1.1792e–4	–3.1863e–1	9.2488e–1
7	–1.0689e–2		2.4856e–1	3.7301e–2
8	2.3814e–3		–6.0094e–2	1.1522e–2
9	–4.4441e–2			1.1000e–2
10	–5.1441e–2			–3.8458e–2
11	3.1428e–2			–1.0157e–2
12	–7.0396e–3			–4.3035e–5
13	4.8188e–2			–3.5744e–2
14	–2.0730e–2			–4.7985e–6
15	4.6544e–3			–5.4108e–7
16				–3.3115e–2
17				–1.3655e–2
18				3.8937e–2
19				3.4275e–2
20				1.2319e–2
21				–1.4199e–2

	M9	M10	M11	M12
1	–2.1451	–1.8308	–2.6426	–2.8394
2	1.9944e-1	8.0398e–2	3.0743	–1.9056e–2
3	–3.2760	–4.5524	–1.9459e–1	2.1094e–3
4	–24.188	8.3022–1	–1.2791e–2	3.4121
5	28.608	–5.1005e–2	–2.7967e–2	–2.0704
6	536.92	4.9448	–1.2417e–1	5.2334e–1
7	1.4962	8.8089e–1	–4.5444e–2	–255.01
8	–3.3536e–1	–9.3699e–1	3.4387e–3	1024.2
9	–1.2673e–1	–5.1392e–1	9.8325e–2	–1548.7
10	2.9411	70.379	–1.2912e–2	-5.6339
11	10.771	–850.63	3.6693e–2	1.1043e–1
12	–1.5032e–2	–4000.2	–1.6806	–1.0481e–1
13	8.7969e–1	–8.7297e–1	4.6561e–1	1.8862e–1
14	–5.7447e–1	6.8925e–1	2.1312e–1	2.5113e–2
15	15.912	16.264	3.3240e–1	7.4621e–3
16	144.42		–2.3406e–1	–1.0500e–1
17	–8.5403e–1		–2.3056e–2	–4.1403e–2
18	7.6552e–1		–1.0547e–1	9.2937e–2
19	–5.0067		2.7712e–2	1045.4
20	–176.37		–6.0313e–2	11.942
21	–565.46		–8.8240	–5.4097e–1
22			–3.6792	-265.75
23			3.8992	-6.4258
24			–1.4866e–1	3.4988e–1
25			–3.3907e–1	
26			–9.5981e–2	
27			7.8418e–1	
28			2.0652e–2	
29			–3.9837e–2	
30			–2.6922	
31			–1.1238e–1	
32			2.9413e–1	
33			2.7791e–1	
34			–5.6398e–1	
35			5.4148e–1	
36			26.288	

M9	M10	M11	M12
37		6.5098	
38		–3.2759	
39		1.4140e–1	
40		–6.5476e–1	
41		1.9229	
42		–27.293	
43		–2.5943	
44		–1.2012e–1	
45		10.132	

The coefficients of the models presented in Table A1 are arranged in the order shown in (7), that is, a_1, a_2, a_3, ..., b_1, b_2, b_3, ... As each coefficient has several parameters, they are in the sequence shown in item 3.2.

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.9484.en

Production and characterization of HA, HA/SiO_2, and KNN biomaterial granules

Pedro de Farias Vanzan[1], Suzana Noronha Ferreira Ribeiro[1], Amal Elzubair[1], Marcelo Henrique Prado da Silva*[1]
[1]Military Institute of Engineering (IME)
Praça General Tibúrcio, 80 – Urca, Rio de Janeiro (RJ), Brazil
*marceloprado@ime.eb.br

ABSTRACT: *Repair and replacement of damaged or lost tissue is a problem of utmost importance for contemporary medicine. Hydroxyapatite, $Ca_{10}(PO_4)_6(OH)_2$, (HA), is the bioceramic most similar to the inorganic phase of bone tissue, besides being bioactive. Hydroxyapatites with partial ionic substitutions and other bioactive bioceramics are adopted as synthetic bone grafts. These grafts can be produced in different forms, such as microspheres, porous scans, or granules. In this work, HA partially replaced with biogenic silica and sodium-potassium niobates (KNN) were synthesized and processed in the form of granules. The materials were calcined, pressed, sintered, ground and then sieved. The samples were characterized by X-ray diffraction (XRD), scanning electron microscopy (SEM) and energy dispersion spectroscopy (EDS). The results showed the presence of micrometric and manometric pores in the HA samples, with higher porosity in the samples with higher percentage of silica.*

KEYWORDS: *Hydroxyapatite. Piezoelectricity. Biomaterials. Granules.*

RESUMO: *O reparo e a substituição de tecido lesado ou perdido é um problema de extrema importância para a medicina contemporânea. A hidroxiapatita, $Ca_{10}(PO_4)_6(OH)_2$, (HA), é a biocerâmica mais similar à fase inorgânica do tecido ósseo, além de ser bioativa. Hidroxiapatitas com substituições iônicas parciais e outras biocerâmicas bioativas são adotadas como enxertos ósseos sintéticos. Estes podem ser produzidos em diferentes formas, tais como microesferas, arcabouços porosos ou grânulos. Neste trabalho, HA parcialmente substituída com sílica biogênica e niobato de sódio e potássio (KNN) foram sintetizados e processados sob a forma de grânulos. Os materiais foram calcinados, prensados e sinterizados, posteriormente sendo moídos e peneirados. As amostras foram caracterizadas por difração de raios X (DRX), microscopia eletrônica de varredura (MEV) e espectroscopia por dispersão de energia (EDS). Os resultados mostraram a presença de poros micrométricos e nanométricos nas amostras de HA, com maior porosidade naquelas com maior porcentagem da sílica.*

PALAVRAS-CHAVE: *Hidroxiapatita. Piezeletricidade. Biomateriais. Grânulos.*

1. Introduction

Biomaterials are biocompatible materials, either natural or synthetic, and can belong to any class, including metals, ceramics, polymers or composites. They must be capable of partially or completely replacing organs or functions of the human body. Metallic biomaterials are used in structural applications in which tensile, compressive, flexural and torsional strength are required. For example, orthopedic and dental implants are designed so that metals or metal alloys fulfill their function based on the application and the host site [1]. Non-structural metallic biomaterials include stents with a shape memory effect, which are flexible metallic structures inserted into the human body after being plastically deformed. When body temperature is reached, these stents return to their original shape [2].

Biocompatible ceramics, or bioceramics, are used in various applications. Inert types, or bioinerts, are employed when chemical inertness is a fundamental requirement. These materials are generally very hard and resistant to abrasion. Alumina (Al_2O_3) and zirconia (ZrO_2) are examples of bioinert bioceramics used in femoral head prostheses, where chemical inertness, abrasion resistance, and hardness are crucial [3]. On the other hand, bioactive bioceramics can chemically bond to bone. These are used as bioactive coatings on metallic implants, providing a chemical bonding interface between bone tissue and the metallic implant, which bioceramics alone would not fulfill [4].

Bioglasses, or biocompatible glasses, are often categorized with bioceramics since they are formed from both metallic and non-metallic elements. When subjected to thermal treatments, bioglasses can become partially ceramic, glass-ceramic, or even fully crystallized, forming bioceramics. Bioactive bioglasses can also be used as coatings for metal implants.

Among bioactive bioceramics, research on calcium phosphate-based materials has shown

promising results for bone defect repair. Bioactive bioceramics are used in trauma treatments for bone regeneration, healing, and reconstitution. Their biocompatibility is a direct result of their chemical composition. Hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$) is the most commonly used bioceramic as a synthetic bone graft because it is bioactive, meaning it can chemically bind to bone. Additionally, ionic substitutions can alter its rate of bioreabsorption in biological media. The synthesis of hydroxyapatites with ionic substitutions is a well-studied area, as human bone is a composite material based on collagen and hydroxyapatite nanoparticles, with substitutions of ions such as Zn^{2+}, K^+, Mg^{2+}, and Na^+ [5].

Bioglasses added to hydroxyapatite (HA) during the sintering process result in glass-reinforced hydroxyapatites. If these bioglasses are silicatic, the silicon or silicate group can integrate into the hydroxyapatite structure [6,7]. Silicon dioxide or silica (SiO_2), in its amorphous form, has been used in medical and dental applications because the presence of silica contributes to bioactivity [8]. This is directly related to the material's ability to form bridges in the presence of silica in biological media [8].

Other considerations related to the sintering of glassreinforced hydroxyapatites involve the presence of liquid phase, the possibility of ionic substitutions during the process, and the decomposition of hydroxyapatite into more soluble phases *in vitro* and *in vivo*.

Sintering in the presence of a liquid phase is a key factor in producing hydroxyapatite composites reinforced with bioglass. The reinforcement results from the increased densification achieved by adding bioglass powder. When bioglass melts, it wets the hydroxyapatite particles, and capillary forces promote densification. Additionally, the presence of a liquid phase enhances diffusional mechanisms, as diffusion is much more effective in the liquid state than in the solid state. Since densification occurs through diffusional mechanisms that bring ceramic powder particles closer together, GR-HA (glass-reinforced hydroxyapatite) samples tend to be denser than HA (pure hydroxyapatite) samples pressed and sintered under the same conditions.

Sintering additives can interact with the primary ceramic powder in several ways: they may form a new phase, promote partial ionic substitutions in the main phase, or remain as a non-soluble second phase within the primary phase [9]. The hydroxyapatite structure allows for substitutions of both anionic and cationic groups. Thus, using bioglasses as sintering additives (or reinforcements) to hydroxyapatite powder is a strategy for synthesizing partially substituted hydroxyapatites [10]. Depending on the bioglass composition, cationic and/or anionic groups can cause partial or total substitutions in the HA structure [10].

The properties of hydroxyapatite, such as bioactivity, thermal stability, and bioresorption rate *in vivo* and *in vitro*, are directly related to its crystallinity and the presence of cationic and anionic substitutions in the crystal lattice [10]. Cationic and anionic substitutions affect hydroxyapatite crystallinity by creating defects in the lattice. Substitutions with F^- and $(CO_3)^{2-}$ ions have opposing effects on crystallinity: replacing $(OH)^-$ groups with F^- anions increases crystallinity, while substitutions of $(OH)^-$ groups with $(CO_3)^{2-}$ anions decrease it. This effect is also observed when $(CO_3)^{2-}$ groups replace $(PO_4)^{3-}$ groups [10]. This explains why fluorapatites are more crystalline than carbonate apatites.

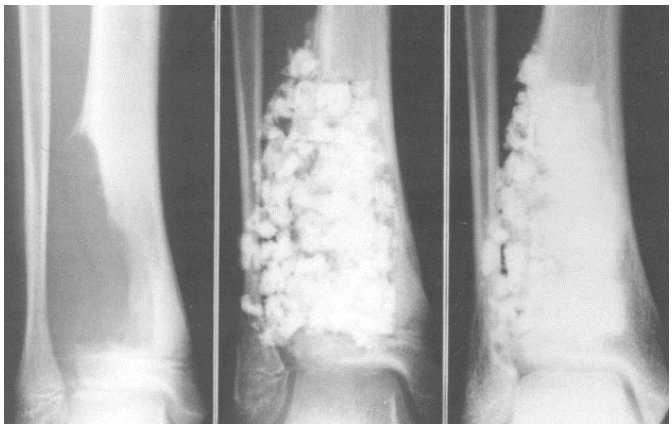
Among the biologically relevant elements, silicon is notable for promoting biomineralization and playing a crucial role in bone metabolism [8, 11]. Silicon is therefore studied for partial substitution in the structure of hydroxyapatite and bioglasses. In hydroxyapatite, silicate groups $(SiO_4)^{4-}$ partially replace phosphate groups $(PO_4)^{3-}$. Many authors set a limit of 5% w/w of Si (approximately 1.7 mol Si/1 mol HA). Silicon additions introduce defects in the HA structure, leading to decomposition into phases such as tricalcium phosphate ($Ca_3(PO_4)_2$) after sintering [11].

The crystallinity of hydroxyapatite affects its properties, including bioactivity. More crystalline materials are less bioactive and have lower bioresorption rates *in vivo* and *in vitro*. For this reason, fluorapatites are less bioactive and have lower

bioresorption rates [12]. Thermal stability is also dependent on crystallinity: the more crystalline the HA, the more thermally stable it is. Therefore, sintering hydroxyapatites with varying degrees of crystallinity at the same temperature can yield different phases. Highly crystalline hydroxyapatites are more thermally stable, resulting in higher thermal decomposition temperatures. Conversely, substitutions that reduce crystallinity lower the thermal decomposition temperature. For example, sintering fluorapatite and carbonate apatite at the same temperature can produce highly crystalline fluorapatite as a single phase in the fluorapatite sample, and a mixture of hydroxyapatite and tricalcium phosphate in the carbonate apatite sample [13]. These effects illustrate that GR-HA composites can be more or less bioactive than pure HA and can exhibit different bioresorption rates *in vivo* and *in vitro*. They can consist of various phases such as HA, partially substituted HA, orthorhombic tricalcium phosphate (α-TCP), or rhombohedral tricalcium phosphate (β-TCP), among others [14]. Comparing hydroxyapatite with tricalcium phosphates, the order of bioresorption *in vivo* and *in vitro* is α-TCP > β-TCP > HA.

Ceramics based on biphasic calcium phosphates, containing HA + β-TCP, or three-phase ceramics, containing HA + α-TCP + β-TCP, are strategies for developing alloplastic (synthetic) bone grafts. Bone grafts can be autogenous (from the patient), allogenous (from donors of the same species), xenogenic (from animals of another species), or alloplastic/synthetic [15]. Alloplastic grafts have the advantage over allografts and xenografts of not being vectors for pathogen transmission. Autogenous grafts are considered the gold standard for bone repair because they carry growth factors from the donor. These grafts are osteoinductive, meaning they can induce differentiation of undifferentiated cells into osteoblasts. However, extensive lesions would require creating a large defect in the donor, which can result in morbidity [16]. The bioceramics most commonly used as alloplastic grafts include HA, biphasic calcium phosphate (HA + TCP), and bioglasses.

Fig. 1 - Bone regeneration with the help of HA-containing graft. Source [17]**.**



(a) (b) (c)

Figure 1 illustrates the use of an alloplastic graft in repairing a bone defect. It presents, in (a), a bone defect; in (b), the defect shortly after being filled with HA granules; and, in (c), the radiographic image of the defect immediately after surgery and three years later [17]. The latest generation of alloplastic grafts includes granules and porous scaffolds made from bioactive, bioresorbable, and preferably osteoinductive bioceramics. Recent studies highlight piezoelectric bioceramics as a new alternative to calcium phosphates and bioglasses [18]. *In vivo* studies suggest that applying direct current electrical stimulation to an implant site enhances early-stage implant osseointegration and interfacial resistance, and promotes bone tissue formation. Piezoelectric materials can mimic the bone's ability to generate electrical potentials under mechanical load without requiring an external energy source [18]. Bioactive and piezoelectric bioceramics, such as sodium niobates (KNN), (which are potassium) combine bioactivity and piezoelectricity, making KNN a promising candidate for alloplastic grafting [18].

In this study, pure hydroxyapatite granules (HA0), hydroxyapatite with biogenic silica (HA5 and HA10), and KNN were produced and characterized.

2. MATERIALS AND METHODS

The materials used in this study included: HA powders, silica (SiO_2), KNN powders calcined at different temperatures, and polyethylene wax beads (Licowax PE 830 from Clariant). The silica was sourced from the freshwater sponge *Metania reticulata*, collected in the Amazon region. The HA powders were mixed with 5% to 10% by weight of silica.

The hydroxyapatite (HA) was synthesized via precipitation in an aqueous medium from a precursor solution containing (Ca^{2+}) cations and (PO_4^{3-}) anions, with pH control at room temperature. The synthesis method is described by [19]. It involves aqueous reactions between phosphate and calcium precursors with controlled pH and temperature. A suspension of 0.5 M calcium hydroxide ($Ca(OH)_2$) was prepared under magnetic stirring. Simultaneously, a 1 M solution of lactic acid ($C_3H_6O_3$) was prepared and added to the calcium hydroxide suspension. Then, a 0.3 M solution of phosphoric acid (H_3PO_4) was prepared and slowly added to the mixture at a rate of 8 mL/min. The resulting solution, known as the precursor solution, was stirred for 24 hours. Sodium hydroxide (NaOH) was then added to adjust the pH to 12, allowing the precipitation of hydroxyapatite. The precipitate was aged for 24 hours without stirring, filtered using filter paper and a vacuum system, and resuspended in deionized water to remove excess NaOH, achieving a pH of 7. The precipitate was then dried in an oven at 60°C for 24 hours.

To incorporate silica into hydroxyapatite, a composite was produced using hydroxyapatite obtained by the precipitation method and silica from calcined freshwater sponge. Five different percentages of silica were used to create a ceramic slip composed of hydroxyapatite, biogenic silica, water, and polyethylene glycol (PEG). The resulting paste was applied to polymeric sponges, which served as a porous framework for the ceramic paste. The sponges were immersed in the paste for 12 hours, then heat-treated with a heating rate of 0.5°C/min up to 550°C, where they were held for 2 hours to remove PEG and other organic additives.

KNN was synthesized using the sol-gel method, chosen for its simplicity and the homogeneity of the final material, as described by Jigong et al. [8]. The raw materials included potassium carbonate (K_2CO_3), sodium carbonate (Na_2CO_3), niobium(V) oxide (NbO_5), and citric acid ($C_6H_8O_7 \cdot H_2O$). The carbonates were dissolved in deionized water and stirred for 20 minutes. Citric acid was then added to the carbonate solution. An ammonia solution was added to adjust the pH and form the sol. Niobium(V) oxide was introduced into the sol, and the mixture was ground with alumina spheres for 8 hours. The resulting paste (Figure 3) was dried at 120°C to form xerogel, which was then calcined at 650°C, 700°C, and 750°C for 2 hours to obtain the powders.

The powders were mixed with polyethylene (PE) wax beads in a volumetric ratio of 65% (v/v) powder to 35% (v/v) wax beads. The mixture was pressed at 50 MPa for 1 minute. The green bodies of the samples were sintered with a heating ramp up to 550°C to eliminate the PE beads, with a heating rate of 0.5°C/min and a 2-hour hold at 550°C. Sintering continued at a rate of 3°C/min up to 1100°C, with a 1-hour hold at 1100°C. The sintered pellets were ground and sieved to granule sizes between 0.18 mm and 0.71 mm. Finally, the structure of the granules was characterized by X-ray diffraction (XRD) using an X'PERT PRO MRD diffractometer (PANalytical) with a cobalt-CoKα source (λ=1.789 Å), a current of 30 mA, and a voltage of 40 kV, scanning from 10° to 80° 2θ with a step size of 0.02 seconds. Qualitative phase analysis was performed using HighScore Plus software. The morphology and chemical composition of the granules were assessed using a Field Emission Gun Scanning Electron Microscope (FEG-SEM), model QUANTA 250 FEG from FEI, coupled with an energy dispersive spectrometer (EDS), model Bruker 6-60.

3. RESULTS AND DISCUSSION

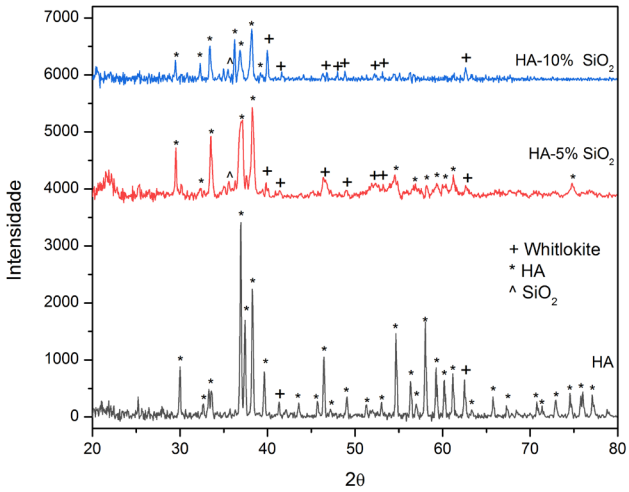
3.1 XRD analysis

Figure 2.4 presents the results of the X-ray diffraction (XRD) analyses for hydroxyapatite (HA) samples, both pure and partially substituted with SiO_2, after sintering at 1100°C. The analysis revealed that the samples without biogenic silica addition exhibited a single phase of hydroxyapatite (HA). In contrast, the samples

containing 5% w/w and 10% w/w SiO_2 displayed additional phases of whitlockite ($Ca_3(PO_4)_2$) and silica (SiO_2). These findings demonstrate the effect of biogenic silica on the thermal stability of HA. Specifically, the addition of silica led to phase transformations upon sintering at 1100°C, whereas the pure HA samples retained a single phase of hydroxyapatite at the same temperature. This observation supports previous studies indicating that silica influences the phase stability of HA.

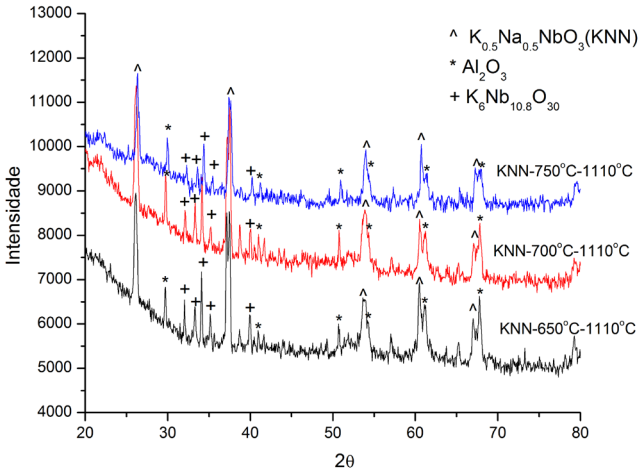
It was also observed that the HA phase showed less intense and wider peaks in the samples that received biogenic silica. This result corroborates the effect of ionic substitutions with $(SiO_4)^{4-}$ groups in the crystallinity of HA [11]. These authors report the creation of defects in the structure of HA, due to the substitutions of $(PO_4)^{3-}$ groups by $(SiO_4)^{4-}$ groups. In fact, [21] reported the creation of $(OH)^-$ group vacancies due to the need for load rebalancing, since there is a difference between the loads of the $(PO_4)^{3-}$ groups and $(SiO_4)^{4-}$ groups. The variation in ionic radii between these groups also induces stress in the HA lattice, contributing to the observed reduction in crystallinity in the samples with ionic substitutions.

Fig. 2 - X-ray diffractogram of HA samples with and without addition of biogenic silica.



The XRD analysis of KNN samples, shown in Figures 3 to 5, identified three phases: $K_{0.5}Na_{0.5}NbO_3$, $K_6Nb_{10.8}O_{30}$ (KNN) and Al_2O_3, with KNN being the predominant phase. Some peaks corresponding to alumina (Al_2O_3) were also observed, likely due to the grinding process during synthesis.

Fig. 3 - X-ray diffractograms of KNN samples.



3.2 SEM analysis

Figures 4 to 9 present the scanning electron microscopy with field emission gun (SEM-FEG) analysis of the HA samples. These images confirm that the granule sizes fall within the predicted range. Notably, an increase in porosity is observed with higher percentages of silica, as illustrated in Figures 7 to 9.

Fig. 4 - SEM micrograph of the HA0 sample at 100× magnification.

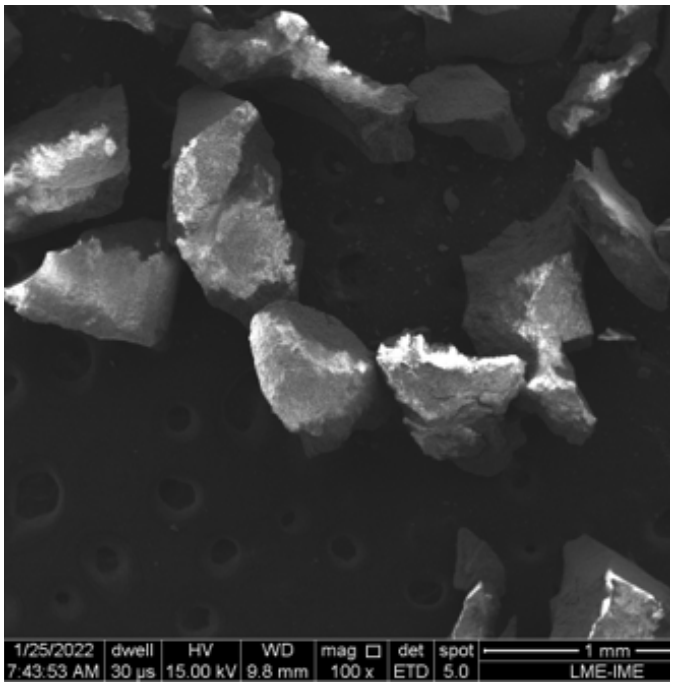


Fig. 5 - SEM micrograph of the HA5 sample at 100× magnification.

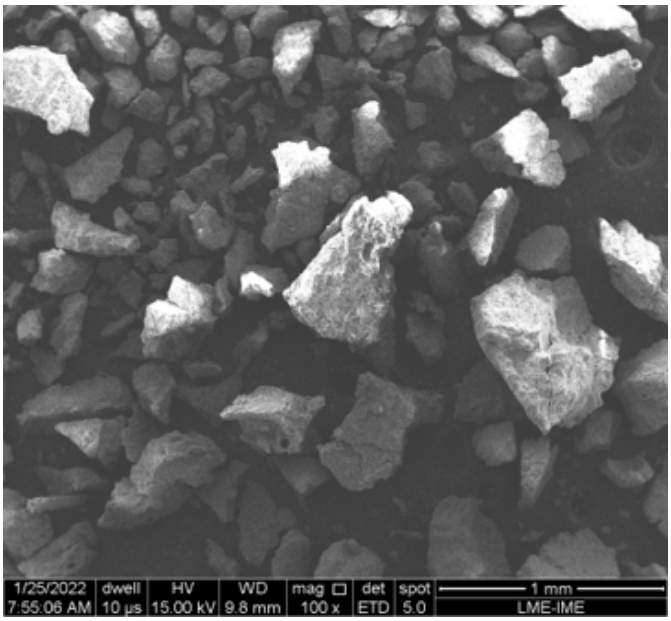


Fig. 6 - SEM micrograph of the HA10 sample at 100× magnification.

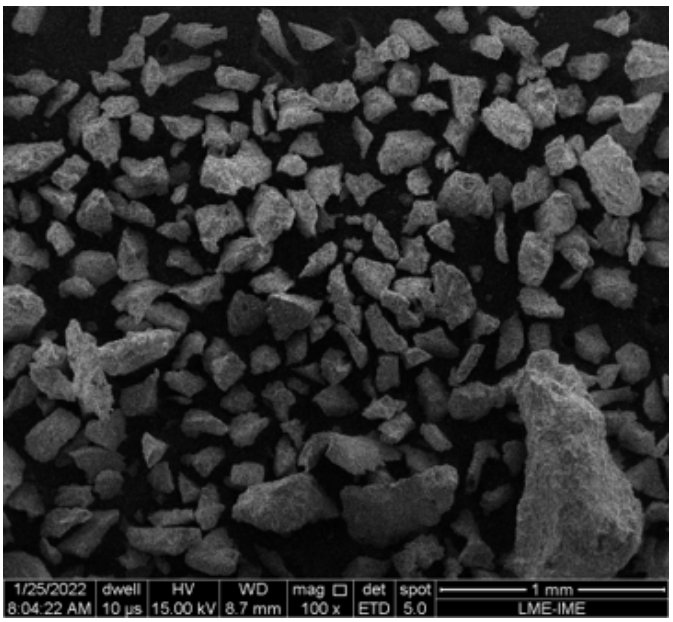


Fig. 7 - SEM micrograph of the HA0 sample at 5000× magnification

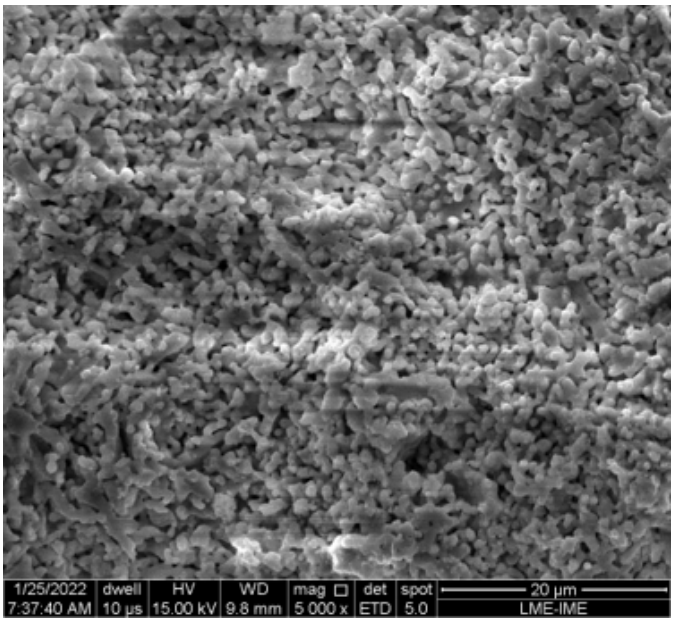


Fig. 8 - SEM micrograph of the HA5 sample at 5000× magnification.

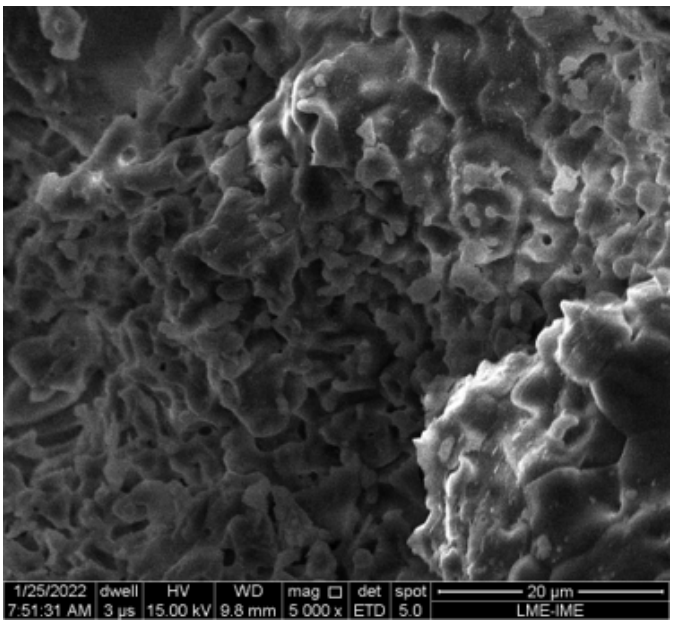


Fig. 9 - SEM micrograph of the HA10 sample at 5000× magnification.

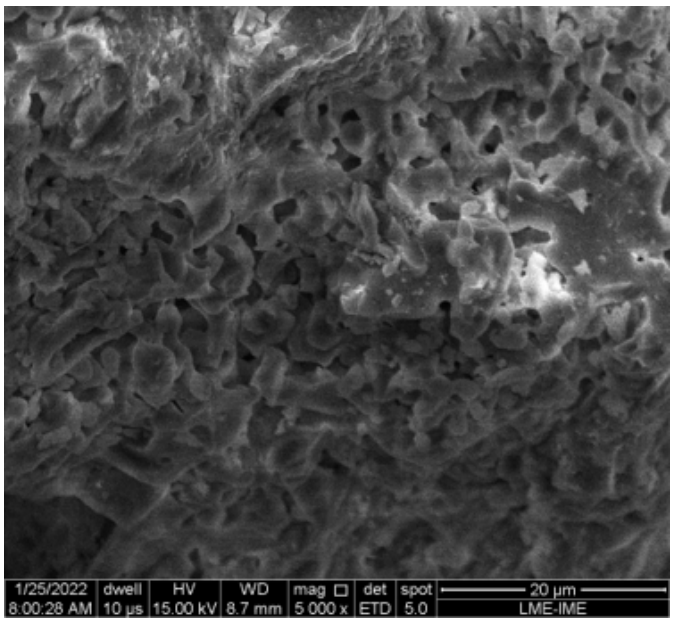
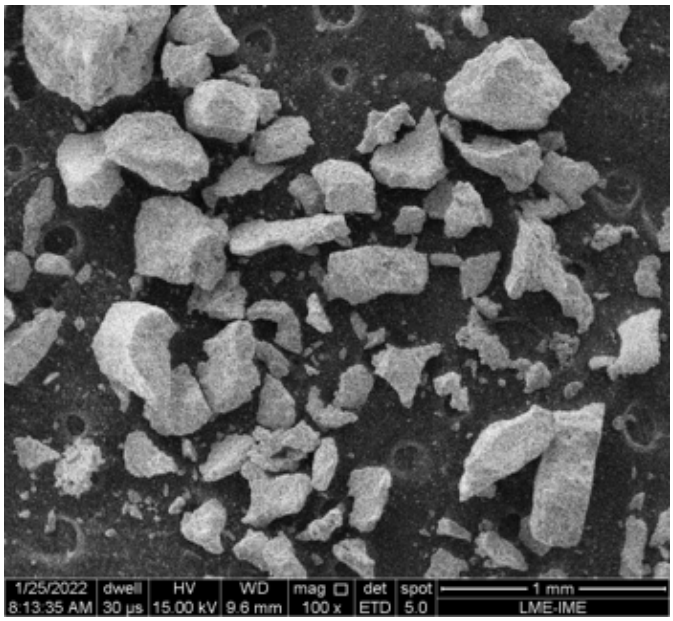


Fig. 10 - SEM micrograph of the KNN650 sample at 100× magnification.



The results of SEM analysis of KNN samples are presented in Figures 10 to 15. Figures 10 to 12 also confirmed the particle size range, in accordance with the sieving process. Figures 12 to 15 revealed changes in the morphologies of the granules, with an increase in the calcination temperature.

Fig. 11 - SEM micrograph of the KNN700 sample at 100× magnification.

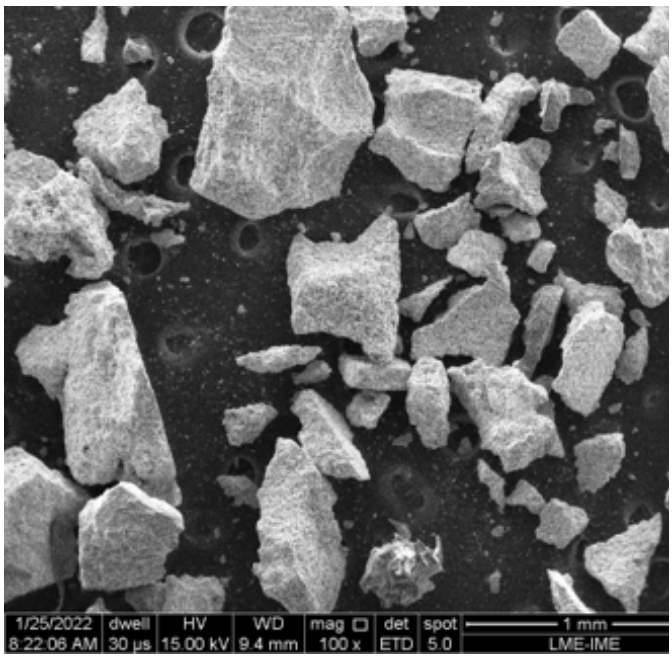


Fig. 12 - SEM micrograph of the KNN750 sample at 100× magnification.

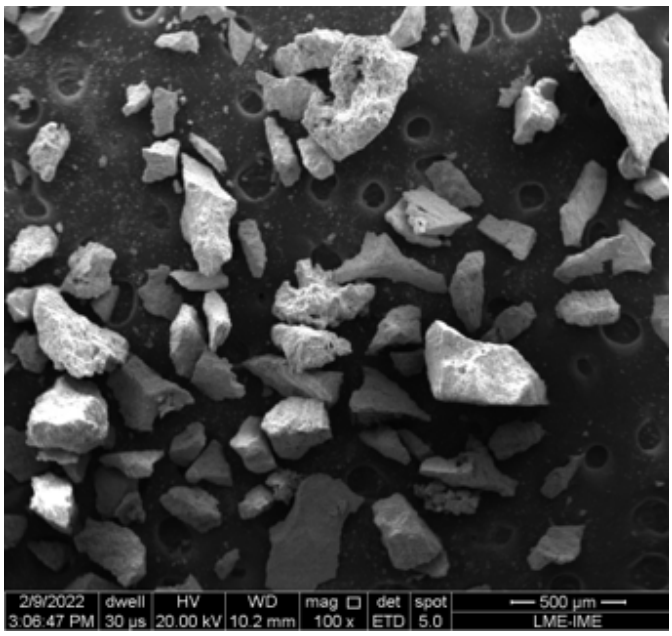


Fig. 13 - SEM micrograph of the KNN650 sample at 5000× magnification.

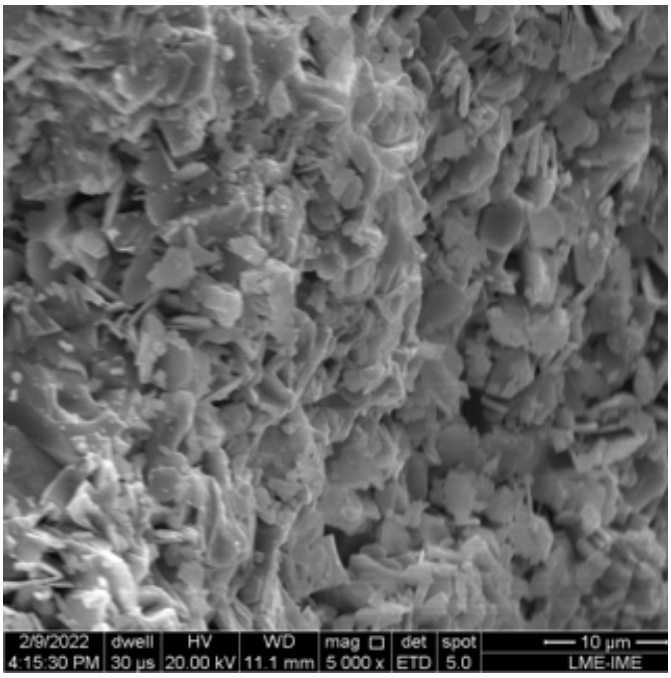


Fig. 14 - SEM micrograph of the KNN700 sample at 5000× magnification.

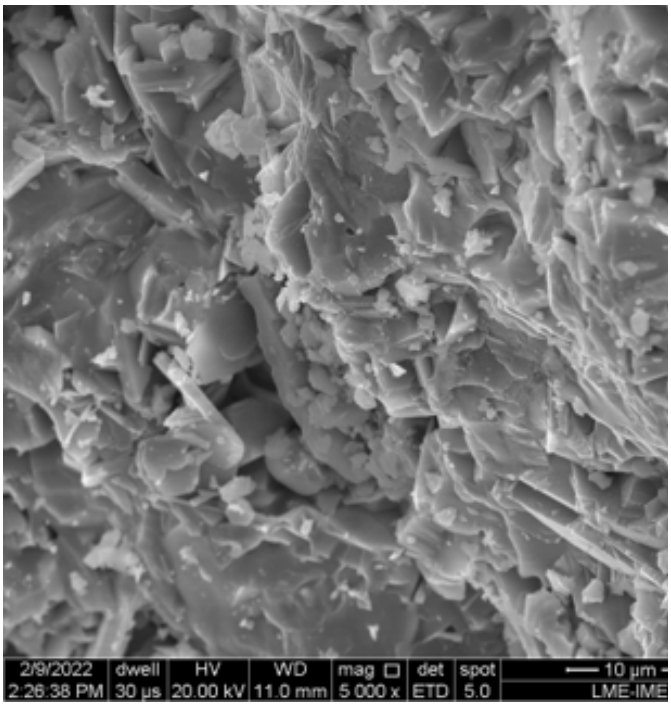
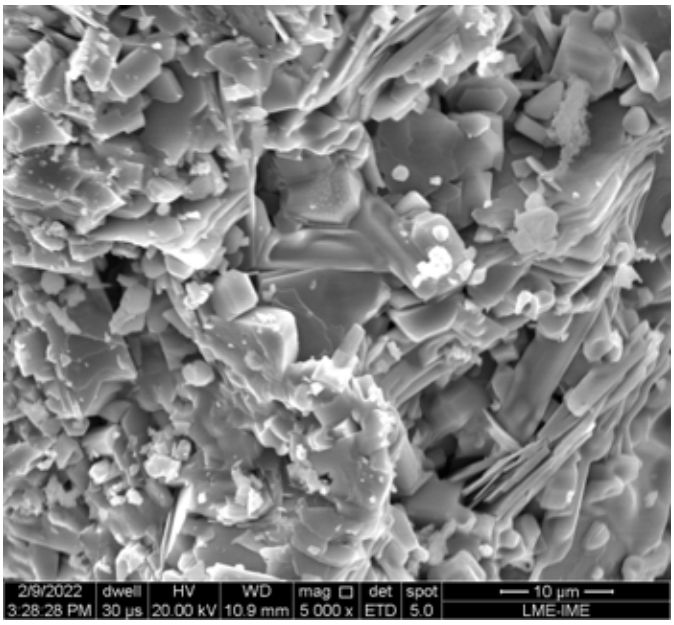


Fig. 15 - SEM micrograph of the KNN750 sample at 5000× magnification.



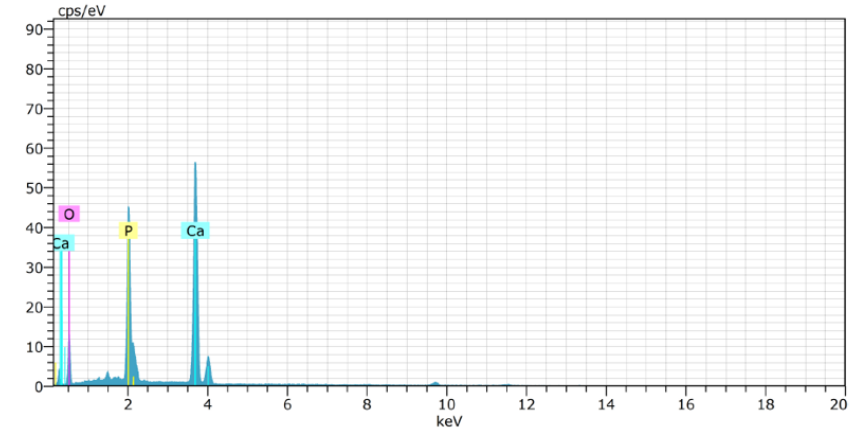
3.2 EDS analysis

The energy-dispersive X-ray spectroscopy (EDS) analysis of the HA samples confirmed the presence of calcium (Ca), phosphorus (P), and oxygen (O), as shown in the spectrum in Figure 16. The table included in this figure provides atomic and weight fractions. It is important to note that EDS is a semiquantitative technique, which means the results should be considered with this limitation in mind. Nonetheless, the EDS results are consistent with those obtained from the XRD analysis of the HA samples.

For HA samples with added biogenic silica, the EDS analysis revealed the presence of calcium (Ca), phosphorus (P), oxygen (O), and silicon (Si), as illustrated in Figures 17 and 18. These findings support the XRD results, which indicated the presence of the SiO_2 phase in these samples.

In the KNN samples, EDS analysis detected the presence of potassium (K), sodium (Na), niobium (Nb), oxygen (O), and aluminum (Al), as shown in Figure 19. This corroborates the XRD results, which identified the Al_2O_3 phase in the samples. The presence of aluminum (Al) and oxygen (O) in the EDS spectra further confirms this phase.

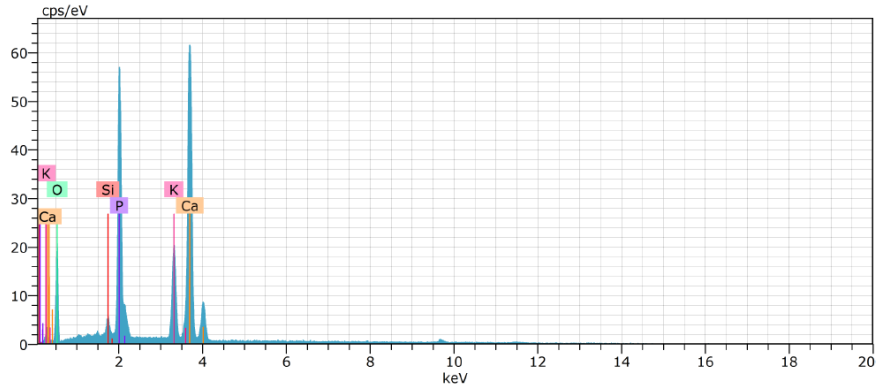
Fig. 16 - EDS spectrum of the HA0 sample.



Area 84 Date:2/16/2022 12:05:44 AM HV:20.0kV Puls th.:23.00kcps

El	AN	Series	unn. C [wt.%]	norm. C [wt.%]	Atom. C [at.%]	Error (1 Sigma) [wt.%]
Ca	20	K-series	36.56	53.83	36.70	1.10
O	8	K-series	18.56	27.32	46.66	2.67
P	15	K-series	12.81	18.85	16.63	0.53
		Total:	67.92	100.00	100.00	

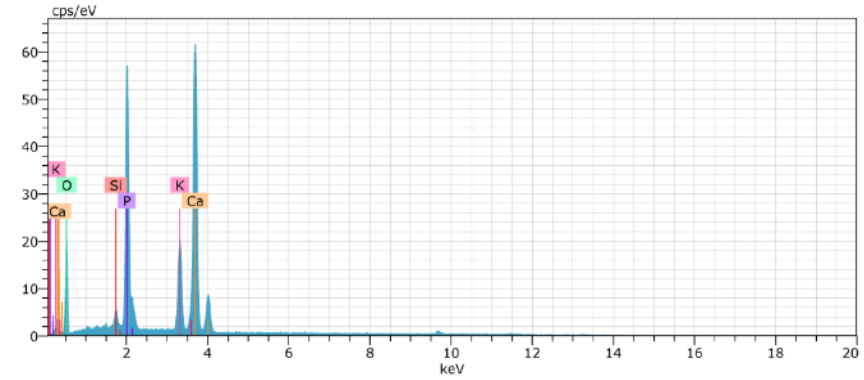
Fig. 17 - EDS spectrum of the HA5 sample.



Area 5Date:1/25/2022 8:51:38 PM HV:20.0kV Puls th.:29.55kcps

El	AN	Series	unn. C [wt.%]	norm. C [wt.%]	Atom. C [at.%]	Error (1 Sigma) [wt.%]
Ca	20	K-series	28.44	43.58	29.12	0.86
O	8	K-series	19.16	29.35	49.14	2.68
P	15	K-series	11.07	16.95	14.66	0.46
K	19	K-series	6.22	9.54	6.53	0.22
Si	14	K-series	0.38	0.58	0.55	0.04
		Total:	65.27	100.00	100.00	

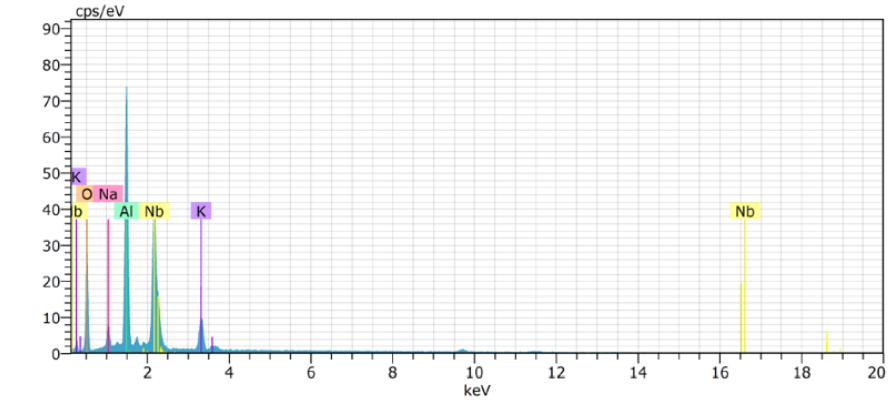
Fig. 18 - EDS spectrum of the HA10 sample.



Area 5Date:1/25/2022 8:51:38 PM HV:20.0kV Puls th.:29.55kcps

El	AN	Series	unn. C [wt.%]	norm. C [wt.%]	Atom. C [at.%]	Error (1 Sigma) [wt.%]
Ca	20	K-series	28.44	43.58	29.12	0.86
O	8	K-series	19.16	29.35	49.14	2.68
P	15	K-series	11.07	16.95	14.66	0.46
K	19	K-series	6.22	9.54	6.53	0.22
Si	14	K-series	0.38	0.58	0.55	0.04
		Total:	65.27	100.00	100.00	

Fig. 19 - EDS spectrum of the KNN650 sample.



Area 83 Date:2/15/2022 11:59:29 PM HV:20.0kV Puls th.:25.84kcps

El	AN	Series	unn. C [wt.%]	norm. C [wt.%]	Atom. C [at.%]	Error (1 Sigma) [wt.%]
Nb	41	L-series	21.74	37.04	11.58	0.83
O	8	K-series	17.82	30.37	55.13	2.37
Al	13	K-series	13.99	23.84	25.66	0.70
K	19	K-series	3.85	6.57	4.88	0.15
Na	11	K-series	1.28	2.18	2.75	0.11
		Total:	58.68	100.00	100.00	

4. Conclusions

When comparing the HA granules, an increase in porosity was observed with higher silica content. This increased porosity enhances the specific surface area, which is beneficial for graft applications, as it boosts the bioactivity of the material and facilitates the permeation of body fluids. For the KNN granules, analysis revealed that no phase transformations occurred at the various calcination temperatures. After sintering at 1100ºC, the samples consistently exhibited three phases: $K_{0.5}Na_{0.5}NbO_3$, $K_6Nb_{10.8}O_{30}$ (KNN), and Al_2O_3. The presence of alumina is attributed to the milling process. The observed low density and alumina contamination could potentially impact the piezoelectric properties of the KNN granules, suggesting that the processing route may need further optimization.

Acknowledgements

This work was supported by the National Council for Scientific and Technological Development (CNPq), the Brazilian Metallurgy and Mining Company (CBMM), and the Coordination for the Improvement of Higher Education Personnel (CAPES).

References

[1] GORIAINOV, V.; COOK, R.; LATHAM, J. M.; DUNLOP, D. G.; OREFFO, R. O. C. Bone and metal: An orthopaedic perspective on osseointegration of metals. Acta Biomaterialia, v. 10, n. 10, p. 4043-4057, 2014.
[2] MAY, A.; SELMAIER, M.; HOCHBERGER, J.; GOSSNER, L.; MÜHLDORFER, S. HAHN, E. G., ELL, C. Memory metal stents for palliation of malignant obstruction of the oesophagus and cardia. Gut, v. 37, n. 3, p. 309-313, 1995.
[3] ALVES, H. L. R., STAINER, D., BERGMANN, C. P. Método alternativo para fabricação de cabeça femoral cerâmica para próteses de quadril. Revista Brasileira de Engenharia Biomédica, v. 20, n. 2-3, p. 81-88, 2004.
[4] RIGO, E. C. D. S.; OLIVEIRA, L. C.; SANTOS, L. A.; BOSCHI, A. O.; CARRODEGUAS, R. G. Implantes metálicos recobertos com hidroxiapatita. Revista Brasileira de Engenharia Biomédica, v. 15, n. 1-2, p. 21-29, 1999.
[5] HENCH, L. L. Biomaterials: a forecast for the future. Biomaterials, v. 19, n. 16, p. 1419-1423, 1998.
[6] FARZADI, A.; SOLATI-HASHJIN, M.; BAKHSHI, F. Synthesis and characterization of hydroxyapatite/β-tricalcium phosphate nanocomposites using microwave irradiation. Ceramics Internacional, v. 37, n. 1, p. 65-71, 2011.
[7] PUPIO, F. SANTOS, R.B.M., CAMARGO, N.H.A., DELIMA, S.A., GEMELLI, E. Síntese e caracterização de pós nanocompósitos fosfato de cálcio/SiO2n para aplicações biomédicas. In: Congresso Latino Americano de Órgãos Artificiais e Biomateriais, 5., 2008, Ouro Preto: COLAOB', 2008. v. 1, p. 1-6.
[8] BELLINI, O. J. Síntese e caracterização de uma matriz óssea de fosfato de cálcio e nanocompósitos de fosfato de cálcio/SiO2n para substituição e regeneração óssea. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis, 2007.
[9] GERMAN, R. M.; RABIN, B. H. Enhanced Sintering Through Second Phase Additions. Powder Metallurgy, v. 28, n. 1, p. 7-12, 1985.
[10] ROCHA, D. N.; SILVA, M. H. P.; GOBBO, L. A. Produção de hidroxiapatita parcialmente substituída com nióbio. Revista Militar de Ciência e Tecnologia, p. 11-17, 2012.
[11] SZURKOWSKA, K.; KOLMAS, J. Hydroxyapatites enriched in silicon–Bioceramic materials for biomedical and pharmaceutical applications Progress in Natural Science: Materials International, v. 27, n. 4, p. 401-409, 2017.
[12] CHENG, K.; WENG, W.; QU, H.; DU, P.; SHEN, G.; HAN, G. et al. Sol-gel preparation and in vitro test of fluorapatite/hydroxyapatite films. Journal of Biomedical Materials Research, v. 69B, n. 1, p. 33-37, 2004.
[13] DJOUALLAH, S.; BELHOUCHET, H., KENZOUR, A.; KHERIFI, D. Sintering behavior of fluorapatite--based composites produced from natural phosphate and alumina. Ceramics International, v. 47, n. 3, p. 3553-3564, 2021.
[14] DUCHEYNE, P.; RADIN, S.; KING. L. The effect of calcium phosphate ceramic composition and structure on in vitro behavior. I. Dissolution. Journal of Biomedical Materials Research, v. 27, n. 1, p. 25-34, 1993.

[15] GIRÒN BASTIDAS, J.; BALDIN, E. K. K.; GODOI, T. P. M.; OLIVEIRA, L. S. D.; MACHADO, G. M.; MALFATTI, C. D. F.; PRANKE, P. H. L. Biomaterials for bone regeneration: an orthopedic and dentistry overview. Brazilian Journal of Medical and Biological Research, v. 54, n. 9, 2021
[16] BAUER, T. W.; MUSCHLER, G. F. Bone graft materials. An overview of the basic science. Clinical Orthopaedics and Related Research, v. 371, p 10-27.
[17] OONISHI, H. Orthopaedic applications of hydroxyapatite. Biomaterials, v. 12, n. 2, p. 171-178, 1991.
[18] POON, K. K.; WURM, M. C.; EVANS, D. M.; EINARSRUD, M. A.; LUTZ, R.; GLAUM, J. Biocompatibility of (Ba,Ca)(Zr,Ti)O_3 piezoelectric ceramics for bone replacement materials. Journal of Biomedical Materials Research, v. 108B, p. 1295-1303.
[19] PRADO DA SILVA, M. H., NAVARRO DA ROCHA, D. BR n PI 020110137091, 2011.
[20] JIGONG, H.; ZHIJUN, X. U.; RUIQING, C.; YANJIE, Z.; QIAN, C.; PENG, F. et al. Characterization of ($K_{0.5}Na_{0.5}$) NbO_3 powders and ceramics prepared by a novel hybrid method of sol–gel and ultrasonic atomization. Materials and Design, v. 31, n. 6, p. 3146–3150, 2010.
[21] GIBSON, I. R.; BEST, S. M.; BONFIELD, W. Chemical characterization of silicon-substituted hydroxyapatite. Journal of Biomedical Materials Research, v. 44, p. 422-428, 1999.

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.9431.en

Command and control as organizational innovation in the provision of public health services

Sylvio J Souza J[1], Fabio O Paula[1], Paulo C Pellanda[2]
[1] IAG – Business School PUC-Rio R. Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ, Brazil
[2] Section of Electrical Engineering – Military Institute of Engineering, Praça General Tibúrcio, 80, 22290-270, Praia Vermelha, Rio de Janeiro, RJ, Brazil
*sylvio.souza@phd.iag.puc-rio.br

ABSTRACT: *The Command and Control field has gone through a series of evolution in recent decades as a result of the adoption of the concepts of Network-Centered Operations Doctrine. This evolution entitles Command and Control to manage projects other than the usual one, which would be the military field. At the same time, the world faces the problem of efficiency and effectiveness in the delivery of public services, particularly in the area of health, characterized by divergences between what citizens want as users of services and what governments provide. Research has been developed seeking to elucidate the issue of the poor performance of public administration in meeting social demands. Organizational innovation presents itself as a promising area for the development of ways to maximize the provision of public services. In this context, the article presents a proposal for Organizational Innovation that applies Command and Control in the Management of Public Service Provision Operations.*

KEYWORDS: Provision of public service. Organizational innovation. Command and control. Doctrine centered on networks.

RESUMO: *A área de Comando e Controle vem sofrendo uma série de evoluções nas últimas décadas, em decorrência da adoção de conceitos de Doutrina de Operações Centradas em Redes. Essa evolução credencia a aplicação de Comando e Controle no gerenciamento de empreendimentos distintos do usual, que seria o campo militar. Paralelamente, o mundo se depara com o problema da eficiência e da eficácia na prestação dos serviços públicos, notadamente na área da saúde, caracterizado pelas divergências entre o que os cidadãos esperam, como usuários dos serviços, e o que os governos entregam. Pesquisas têm sido desenvolvidas buscando elucidar a questão do fraco desempenho da administração pública em atender às demandas sociais. A inovação organizacional se apresenta como uma área promissora para apresentar caminhos que maximizem a prestação dos serviços públicos. Neste contexto, o artigo apresenta uma proposta de Inovação Organizacional que aplica Comando e Controle no gerenciamento das Operações de Prestação de Serviços Públicos.*

PALAVRAS-CHAVE: Prestação de serviço público. Inovação organizacional. Comando e controle. Doutrina centrada em redes.

1. Introduction

This article aims to present a proposal for an innovative solution, in the form of an Organizational Innovation (OI), for the immediate treatment of failures occurring during the Provision of Public Health Services (PSP) and that contribute to unsatisfactory results, based on the concepts of Command and Control (C2) and Network Centered Operations Doctrine (DOCR), a typically military application area. This article is based on a Doctoral Thesis [1] developed at the Military Institute of Engineering (IME-RJ).

C2 has a reputation for being a mysterious subject, even for those who may be or have been a professional in the field. The words command and control, individually or together, have different meanings when used by different communities [2].

C2 techniques have undergone a series of evolution in recent decades, due to the adoption of DOCR concepts [3]. In parallel, the current perception denotes that the PSP in Brazil does not present the necessary quality to meet the needs of the population, both in scope and quality, as neither the scope nor the standard is satisfactory.

This perception is supported by research produced in the academic environment, covering various areas of PSP, such as transportation [4], health [5] [6] [7], and education [8][9], and also by evidence constantly reported in the media, ranging from lack of hot food in state schools in Rio de Janeiro [10] to a patient seen on the floor of the largest hospital in Fortaleza [11]. Numerous other press reports have evidenced similar problems [12] [13][14][15][16].

Based on this evidence, among many others reported daily, it is undeniable that the quality of public

service in Brazil is insufficient, requiring several actions to improve it. Actions to achieve the necessary improvements may consist of changes in the form of management and execution of Public Service Provision Operations (OPSP), which can be done from the implementation of OI.

In recent years, the implementation of OI in the public sector has received increasing attention, both for providing better levels of quality of public services and for influencing the capacity of the private sector to innovate [17]. While, on the one hand, news reports or scientific works emphasize the issue of problems with PSP, on the other hand, the specialized literature contains numerous publications on OI that can mitigate this problem [18][19][20].

Following this trend, the authors propose an OI in the area of health PSP that, based on the application of concepts and processes of C2 and DOCR, provides a reduction in the occurrence of failures or the reduction of their consequences. A pilot project developed between the Military Institute of Engineering (IME) and the Secretariat of Health of the State of Rio de Janeiro (SES-RJ) was implemented as part of the research process.

The purpose of the pilot project was to analyze and validate the model developed in the first cycle of action research of the research process, and its results were used in the improvement of the model, which occurred in the second cycle of action research. A summary of its implementation and the results obtained is presented at the end of the article

This article is structured as follows: the theoretical basis is presented in Section 2; Section 3 is dedicated to the description of the method used; section 4 discusses the results obtained, presenting the proposed models; and section 5 contains the summary and main results of the pilot project implementation.

2. Theoretical Foundation

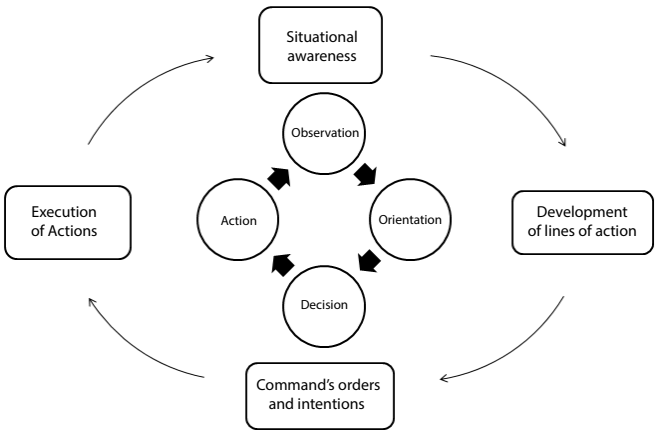
2.1 C2 and DOCR – C2

C2 is not an end in itself, but a direction for value creation, such as fulfilling a mission. Specifically, C2 centralizes efforts of several entities, individual or organizational, and resources, including information, aiming at the accomplishment of some task or objective [2].

For C2 processes to be effective, their cycle, which is the sequence in which C2 actions are performed, must unfold repeatedly and as quickly as possible in each of them, throughout the time in which the operations are necessary [21][22]. These actions are known by the acronym OODA – Observation, Orientation, Decision, Action. The concept of this cycle was introduced by John Boyd, an American military strategist active in the twentieth century [23].

The *observation* represents the continuous monitoring of the operating environment, so to detect any change in the course of events that may require an intervention. *Orientation* means the projection of the new reality to be achieved in the operating environment, the new desired situation. The *decision* corresponds to the conduct to be developed so that the new desired reality is effectively achieved. Finally, the *actions* resulting from the decision taken are implemented. A new cycle then resumes, with the resumption of observation – FIG. 1.

Fig. 1 - OODA Cycle.



Source: Adapted from [22].

The C2 functions comprise the set of activities for carrying out the PDCC – Planning, Direction, Control and Coordination – of C2 actions, which are part of the aforementioned OODA cycle [21][22]. According to [2], C2 functions are applicable not only in

military enterprises, but also in civil enterprises. Viewed as a whole, the expression C2 covers the following functions:

- Establish intention (the goal or objective);
- Determine roles, responsibilities and relationships;
- Establish rules and restrictions;
- Monitor and evaluate the situation and progress.

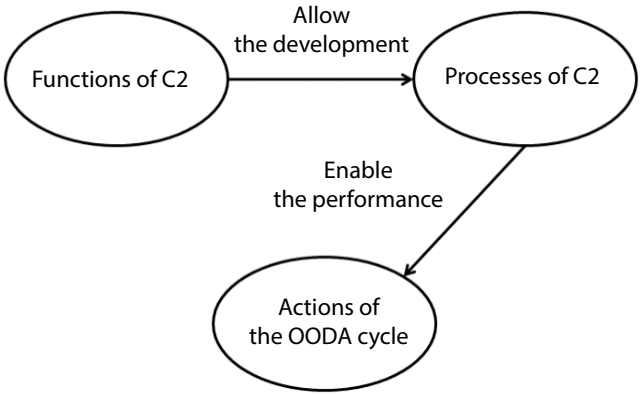
Although 'command' and 'leadership' can refer to the same person, in practice, the possibility must be allowed that these functions can be performed by different individuals or sectors. However, in view of the fact that 'command' usually refers to the figure of the leader, the following functions allow us to consider this aspect:

- Inspire, motivate, and generate confidence;
- Train and educate.

Finally, the question of the resources used in carrying out the mission cannot be overlooked. Without the necessary resources, the mission cannot be completed. Therefore, the last function considered by [2] is 'provisioning'.

The *functions* of C2 enable the development of C2 *processes*. They are what enable to carry out C2, as defined above, with the actions of the OODA cycle [21] [22]. FIG.2 illustrates this relationship.

Fig. 2 - The relationship between C2 functions, processes and actions.



For C2 processes to achieve the primary objectives, it is essential that the perception of reality and reality itself are in perfect harmony. This attunement is called *situational awareness* [21][22]. The structural framework that provides the support for all these activities and objectives to be carried out and achieved is called the C2 structure, which covers one or more – depending on the scope – C2 centers. Such centers are the operations centers designed to provide the links between the command structure with the upper and subordinate echelons [21][22].

C2 system is the set of facilities, equipment, information systems, communications, doctrine, procedures, and personnel essential for the commander to plan, direct and control the actions of their organization to achieve a certain purpose [21][22]. The initial conditions for a given mission to occur may vary over time. These are factors that need to be observed for the mission to be carried out within the desired parameters, producing the expected objectives.

Once the mission starts, objectives, rules, responsibilities, resource allocation and everything else can be changed. It is up to the 'control' component to carry out this verification: if something has changed without proper authorization, preventing the mission from being carried out, or if something needs to change. The only exception concerns the definition of objectives (intention), which is up to the 'command' component. The 'command' aims to return the mission to the path of success, based on the checks of the 'control' component [2].

2.2 C2 and DOCR – DOCR

Military organizations are, by their very nature, resistant to change, and deal with information, traditionally, in three ways: *commands*, which serve to define the specific task at hand – directives and guidelines; *intelligence*, which provides information about the environment in which the task is to be performed; and *doctrine*, which provides the rules of the game or standard operating procedures [24]

Changes, particularly changes that can affect relationships between organizations and between commanders and their subordinates, present significant risks and therefore generate considerable concern. The explosion of information-related technologies has unleashed a virtual tide of

change that profoundly affects organizations and individuals across multiple dimensions. The military is no exception [24].

Consequently, traditional military C2 has been challenged by a number of modern problems, such as environmental complexity, dynamism, new technologies and competition capable of exploiting the weaknesses of an organizational paradigm that has been dominant since the industrial revolution. The conceptual response to these challenges is a new type of C2 organization called *Network Enabled Capability* (NEC) [3], which incorporates DOCR.

Traditional military architectures have been organized in a very hierarchical way, but modern innovations in communications and computer technology provide an extensive range of other possible structures. In parallel, the emerging emphasis on operations not necessarily focused on war has required more flexible organizational structures. This new scenario requires the use of formal techniques for the evaluation of a set of possible organizational structures, and Social Network Analysis techniques are an obvious choice for such analyses [25].

In the industrial era, a period that began with the Industrial Revolution and lasted until the 20th century, interaction models were designed to ensure control from the center. Thus, the flow of information always followed the "chain of command" or the management structure of the enterprise. This pattern of interaction can be recognized in military models, in which official correspondence is addressed to unit commanders. This practice reinforces the tradition that the flow of information must follow along the command line and that all information within a unit is of interest and belongs to the unit commander. Thus, in the industrial era, communication mechanisms imitated the organizational structure of the organization [2].

The information age, also known as the computer age or digital age, is a historical period beginning in the twentieth century and characterized by the rapid change from traditional industry, which the Industrial Revolution provided, to an economy strongly based on Information Technology (IT).

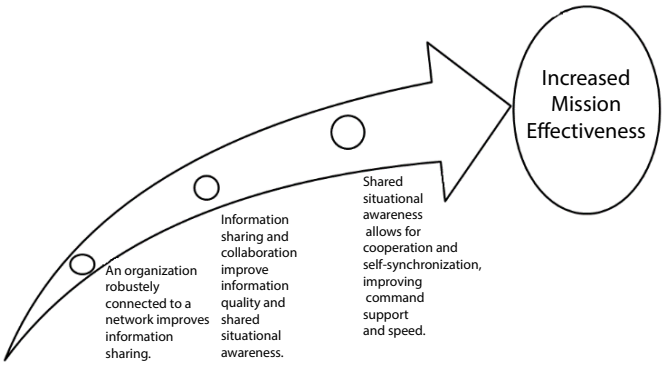
Traditional interaction patterns developed in the age of industry have undergone changes to adapt to new paradigms, transforming from centralized models to networked models. In the case of *Network Centric Warfare* (NCW) and the *Edge Organization* (a relatively new organizational form that distributes knowledge and power to its limits and allows members and organizational units to self-organize and self-synchronize their activities) [26], social networks are the ones that will be activated by any available mechanisms: mail, telephone, videoconference, local area networks, wide area networks, world wide web, etc. Social networks also depend on cooperation: the willingness to work together and collaborate when appropriate [2].

The DOCR is adopted by the U.S. Department of Defense for the implementation of its C2 systems. This doctrine has transformed and continues to transform C2 processes. It proposes the massive sharing of information through communication networks, leading to a natural self-synchronization of actions between the actuating forces and, thus, optimizing the OODA cycle and the subsequent increase in the effectiveness of the mission to be carried out [27].

According to [28], the value chain of Network Centered Operations (OCR) is reflected in a set of hypotheses that advocate, briefly, that a robustly networked force contributes to increasing mission effectiveness, according to the causal connections presented in Fig. 3. The massive sharing of information via communication networks leads to a natural self-synchronization of actions, optimizing the OODA cycle and increasing the effectiveness of the mission to be carried out [29].

This work considers that the systematic and structured observation of the complex PSP environment, based on the adapted application of the elements of C2 and DOCR, can meet the need that, according to [30], governments have to share power and be more flexible in their control mechanisms, reducing bureaucratic procedures that can impair the efficient provision of service to the public.

FIG.3 - Network-Centered Operations



Source: Made from [27] and [28].

2.3 PSP

Twenty-first-century public administration is undergoing dramatic changes, especially in advanced economies, but also in many parts of the developing world. Globalization and pluralization of service provision are the driving forces behind these changes. The political problems faced by governments are increasingly complex, perverse, and global rather than simple, linear, and nationally focused. However, the prevailing paradigms through which public sector reform is designed and implemented are relatively static and do not fully encompass the meaning or implications of these broader changes. While public sector reforms in the developing world are influenced by policy experiments and organizational practices originating in *Organisation for Economic Co-operation and Development* (OECD) countries, they tend to operate within the traditional paradigm of public administration. As a result, there is often a discrepancy between the momentum of public sector reform efforts in developing country contexts and broader changes in the nature of governance and contemporary approaches to public management based on OECD experience [31].

Contemporaneously, Public Governance (GP) has been proclaimed as a new paradigm, distinct from New Public Management (NGP) and orthodox bureaucratic public administration. It applies in new times and contexts, marked by pluralism, complexity, ambiguity, and fragmentation, unlike the contexts of the old public administration (the birth and apogee of the developmentalist welfare state) and the NGP (the neoliberal wave of the 1980s and 1990s). GP is capacity for governance; it is collaborative network governance; it is results-oriented governance [32][33][34].

An important aspect to be considered in the PSP scenario concerns the issue of efficiency and effectiveness. A reference to a quote by Peter Druker is made in [35]: the judgment of a manager's performance must be through the twin criteria of effectiveness – the ability to do the 'right' things – and efficiency – the ability to do things 'right'.

In [36] a framework is presented that discusses two general questions: whether the public administrator is 'doing the right things', that is, delivering services consistent with the needs of the citizen; and whether it is doing 'the things right', that is, providing quality services at the lowest cost. To answer these questions, in [36] two empirical tests are performed, one of effectiveness and the other of efficiency. The effectiveness test asks to what extent public programs meet established goals, while the efficiency test asks whether services are provided at the lowest possible cost.

In parallel to the process of evolution of the PSP, changes also occurred on the citizen's side. Society's pressure for quality public services is growing, and the pressure is more perceived in relation to those services related to the direct provision to the citizen. This situation became possible mainly from 1995, when the Brazilian national scenario was faced with the entry into the agenda of the reform of the State apparatus, and, within this context, of the administrative reform. With the approval of Constitutional Amendment No. 19, of June 4, 1998 [37], the adoption of the Principle of Efficiency in the Brazilian public administration was established.

PSP is an essential function of governments [30], characterized by the delivery of the product, made directly by the government or by a representative, to citizens. It can reach many stakeholders,

as it involves not only government agencies, but also inter- and intra-governmental agencies. As a consequence, there is a risk of lack of coordination and conflict of interest. Since there are many stakeholders, one needs to make an adjustment in the expectations of interactions [30].

Because it is very heterogeneous, the public sector has particular challenges in ensuring its effectiveness and efficiency. Three main dimensions can be used to classify public organizations: area (health, education, etc.), level or sphere of government (federal, state, municipal) and type of institution (policy formulation/central administration, regional agency, service provider unit) [17].

For [38], governments, especially in developing countries, do not yet have mature organizations to implement efficient public policies related to the direct provision of services to citizens. The aforementioned author reports that, in addition to politics, implementation issues can be an important focus of the research. Among the questions pointed out by [38], one is of special interest to the research aims of this study: what can be done to avoid failures or correct them more promptly?

Concern for service provision operations is also present in the work of [39]. For the author, operations management is essential to maximize the production of goods and services in the public service, while quality management ensures the satisfaction of beneficiaries beyond their expectations.

The proposal dealt with in the article seeks precisely to act on the effectiveness and efficiency of OPSP, acting on one of its variables, as seen below, through the improvement of service delivery management, in the form of an OI, positioning itself according to several publications that seek to mitigate the PSP quality problem through innovations (e.g. [19][20][40]).

Briefly, the proposed models for the application of C2 and DOCR in the PSP environment seek to: mitigate the discrepancy between the impulses of efforts for reforms and broader changes [31]; integrate with the precepts of the GP [32][33] [34]; maximize the efficiency and effectiveness of the PSP **[35]** [36] [37]; enhance interactions and reduce conflicts between those interested in the PSP [30]; manage the PSP of regional agencies and providing units of the three levels of government [17]; avoid failures and correct them more promptly [38]; and manage OPSP to maximize the production of goods and services in the SP [39].

2.4 Open OI in the Public Sector

Innovation is more than a new idea or an invention. An innovation requires implementation; that it be put to active use or made available for use by other companies, individuals, or organizations. The economic and social impacts of inventions and ideas depend on the diffusion and acceptance of related innovations [41].

According to [42], a new good, service, or process would be an innovation only if it is connected to the market. For the author, this connection means that there is no sense in innovating if there is no reflection in the market, or, in other words, if the company does not obtain an increase in sales of its products or services. Specifically in the case of public services, the increase in sales does not make sense; however, the increase in quality of service provision is still a connection with the "market".

Over time, the understanding of the process of generating innovation has deepened, showing that, on the one hand, the meaning of innovation triggers has changed, and on the other, the different phases of the innovation process have also been substantially redefined. Another novelty is the change in the understanding of the innovation process as a linear sequence of different phases for an integrated view of the process, with the replacement of simple, linear and sequential types by increasingly complex patterns. This means that the individual phases overlap and that there is feedback between them. In terms of knowledge and technology transfer, interactive models enrich both basic research and applied research and development [43].

External sources of knowledge and R&D play a central role in the development of innovation, being

complementary in companies in more technologically intense industrial areas, which usually have greater absorption capacity. In companies in low-tech industrial sectors, both strategies are effective for improving innovation performance, but the combination does not have significant effects [44].

OI, as mentioned in [45], refers to the creation or adoption of a new idea or behavior for the organization, accompanied by new processes and a new information structure, which can be caused endogenously or exogenously. This definition is consistent with the definition of the Oslo Manual [41], which defines as a minimum requirement for an innovation to be identified the existence of one or more characteristics that differ considerably from those that characterized the product or business processes previously offered or used by the company.

OI is seen as the implementation of a new organizational method in company business practices, workplace organization, or external relations. Therefore, it can be concluded that an OI has a comprehensive scope, contemplating the development of new processes and business models and the achievement of new results from the perspective of creating innovation. In addition, it includes the development of skills, as well as processes of creation and transfer of knowledge [46].

In particular, an OI is the introduction of new organizational methods for managing the business both internally and in relation to external agents [41][47], and represents one of the most important and sustainable sources of competitive advantage for companies.

An Open Innovation (IA) is a concept based on the conviction that companies can and should look for ideals and ways to create innovations not only within the limits of their structures, but also in their environment, formed by their external partners – customers, other companies, etc. [48].

The increasing complexity of products and services, coupled with rapidly changing market demands or increasing pressure from various social groups are trends that force companies to adopt new practices to remain competitive. External sources of information and integration in the context of IA is a practice that can leverage success [49].

The path to the development of new technologies and new knowledge makes companies increasingly dependent on external knowledge and technology, which may be in the public domain or belong to other companies, individuals or research institutions [50].

This study seeks, in short, to solve the issue of low efficiency and effectiveness of the PSP in Brazil, through the proposal of an open OI, which is based on the introduction of concepts and processes of C2 and DOCR in the OPSP environment.

Thus, the proposal is characterized, from the point of view of OI: as an attempt to increase the quality of service provision, representing a connection with the "market" [42]; by the application of external sources of knowledge and R&D [44]; by symbolizing the exogenous adoption of a new behavior accompanied by new processes and a new information structure [45]; by the introduction of new organizational methods for business management [41][47]; as a search to meet the growing pressure exerted by PSP's client social groups [49]; and by portraying, in the form of an open innovation [48], the application of research institution knowledge in the development of skills and practices within the public company [50].

3. Method

Two approaches were used in the development of this research. The first (general) approach concerns the scientific method related to the stages of development and validation of the research itself, which consists of identifying and studying the theoretical framework; analyzing the data; proposing a first version of the solution; in the implementation, evaluation and evolution of a pilot project; in a new round of theoretical study, due to the incorporation of a new theoretical framework; and in the formulation of a final proposal. As for the second (specific) approach, it concerns the technique and modeling tools used on the PSP environment, for

understanding the problem and developing and presenting the solution.

Regarding the first methodological approach (general), the research was developed according to the Action Research method [51][52], involving the State Department of Health of RJ (SES-RJ), in partnership with the Military Institute of Engineering (IME). SES-RJ was responsible for the role of the institution where the problem is detected and IME for the role of the holder of expertise in the area of knowledge applied in the proposal – C2 and DCR. During the development of the proposed model, a pilot project was implemented in SES-RJ. This combination of roles of the two institutions involved guided the choice of the action research method.

The development of the research had two cycles of action research. The first cycle consisted of three stages: exploratory, experimental and validation. In the exploratory stage, research related to the theoretical framework was carried out, the diagnosis of the problem and the model for the solution were developed. The experimental stage consisted of the implementation of the proposed model (pilot project) and its assisted operation. The validation stage consolidated the data from the pilot project and supported the second cycle of action research.

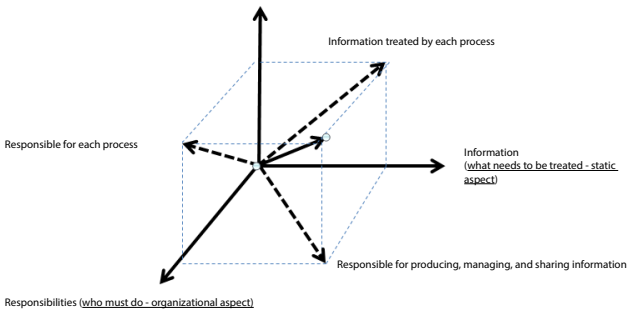
In the second cycle, only the exploratory stage was carried out and consisted of the deepening of the bibliographic review and the evolution of the proposed model. It is intended to complete the second cycle in the future, by performing a second experimental step followed by the validation step.

Regarding the second (specific) methodological approach, the modeling of the problem and the proposed solution followed a specific methodology based on the use of systems/software engineering techniques and tools [53][54]. This methodology is based on the use of diagrams representative of different views of the object studied and is also organized in cycles (conception, elaboration, construction and transition) and was used to develop the model and define the implementation of C2 and DOCR processes and concepts in the PSP environment. Throughout the steps of the two action research cycles, several modeling cycles occurred.

4. Results

The proposed OI suggests the adoption of processes, organizational structures (institutional actors with well-established responsibilities), and information structures for the incorporation of C2 and DOCR in the treatment of failures that occur during the PSP. The three dimensions covered are integrated, and the one resulting from the integration defines 'who' (actor) does 'what' (process), 'what is treated' (information) and who is responsible for the information, always related to the treatment of failures (FIG. 4 – Proposed three-dimensional model) and the use of C2 and DOCR processes and concepts.

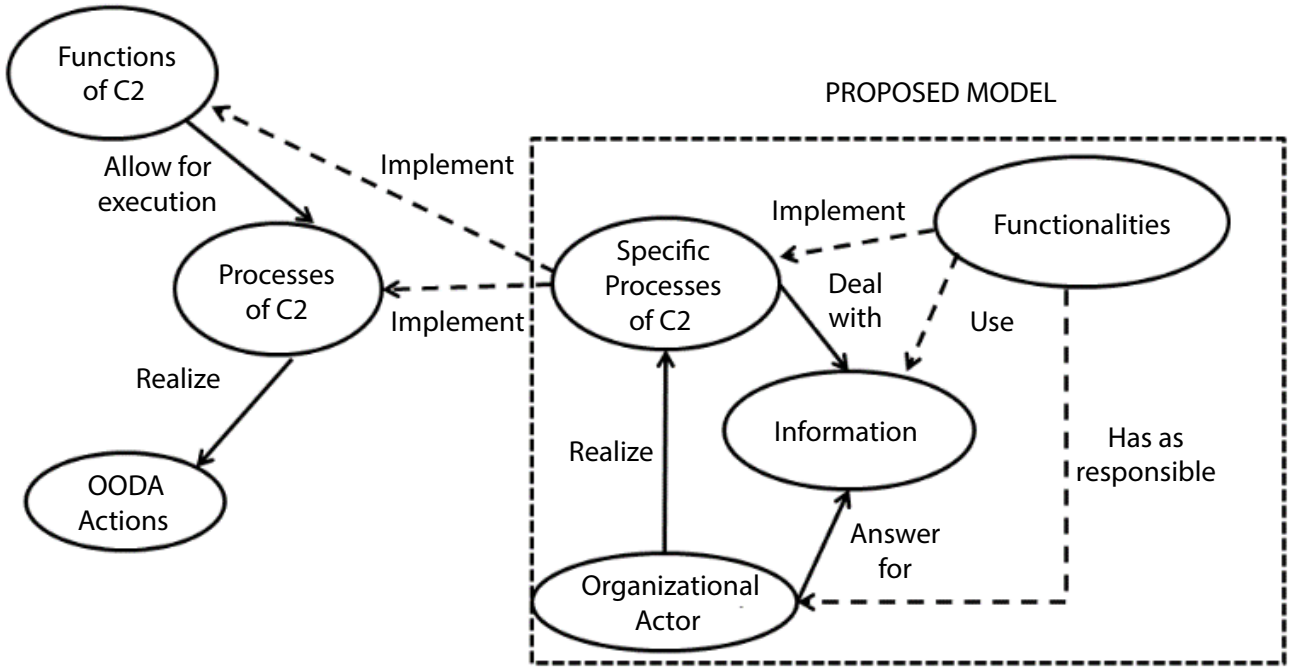
Figure 4 - Proposed three-dimensional model.



To perform the OODA cycle, C2 processes must be performed cyclically (FIG. 5). The Functions of C2 [2] [21][22] promote the framework so that the processes of C2 can be performed, as shown in FIG. 5.

Thus, the proposed model should provide specific C2 processes that implement C2 Functions and Processes, and that support the performance of the OODA cycle Actions. The execution of the specific processes of the proposed model is done by the functionalities of the system and is only possible with the proper treatment of information and the correct identification of the organizational actor responsible for the execution of the specific processes and the management of the information, as presented in FIG. 5.

Fig. 5 - Proposed model and the implementation of C2 and DOCR.



The proposed model addresses the research problem by acting on the response time elapsed between the occurrence of a failure and its total solution.

- When a failure occurs, corrective actions must be taken to resolve the problem. One variable can be identified: the response time. This variable can be divided into three parts:
- Perception time (*tp*) – time elapsed between the occurrence of the failure and its perception.
- Triggering time (*td*) – time elapsed between the perception of the failure and the triggering of the corrective action;
- Execution time (*te*) - time elapsed between the beginning of the corrective action and the complete solution.

Therefore, the response time (*tr*) is equal to *tp* + *td* + *te*.

Since efficiency is the normative measure of resource use in a given process [35], and since time is a resource used by processes, it is acceptable to assume that the *tr* used to correct a failure is a factor that inversely influences the efficiency value: the higher the *tr*, the lower the efficiency.

Since effectiveness is the normative measure of the achievement of the results of a given process [35], and since *tr* is the time interval between the moment of expectation of the result (when the failure is perceived) and the moment when the effective result is achieved, it is plausible to consider that *tr* also influences the value of effectiveness in an inversely proportional manner: the longer the time elapsed to achieve the results, the lower the value of this result.

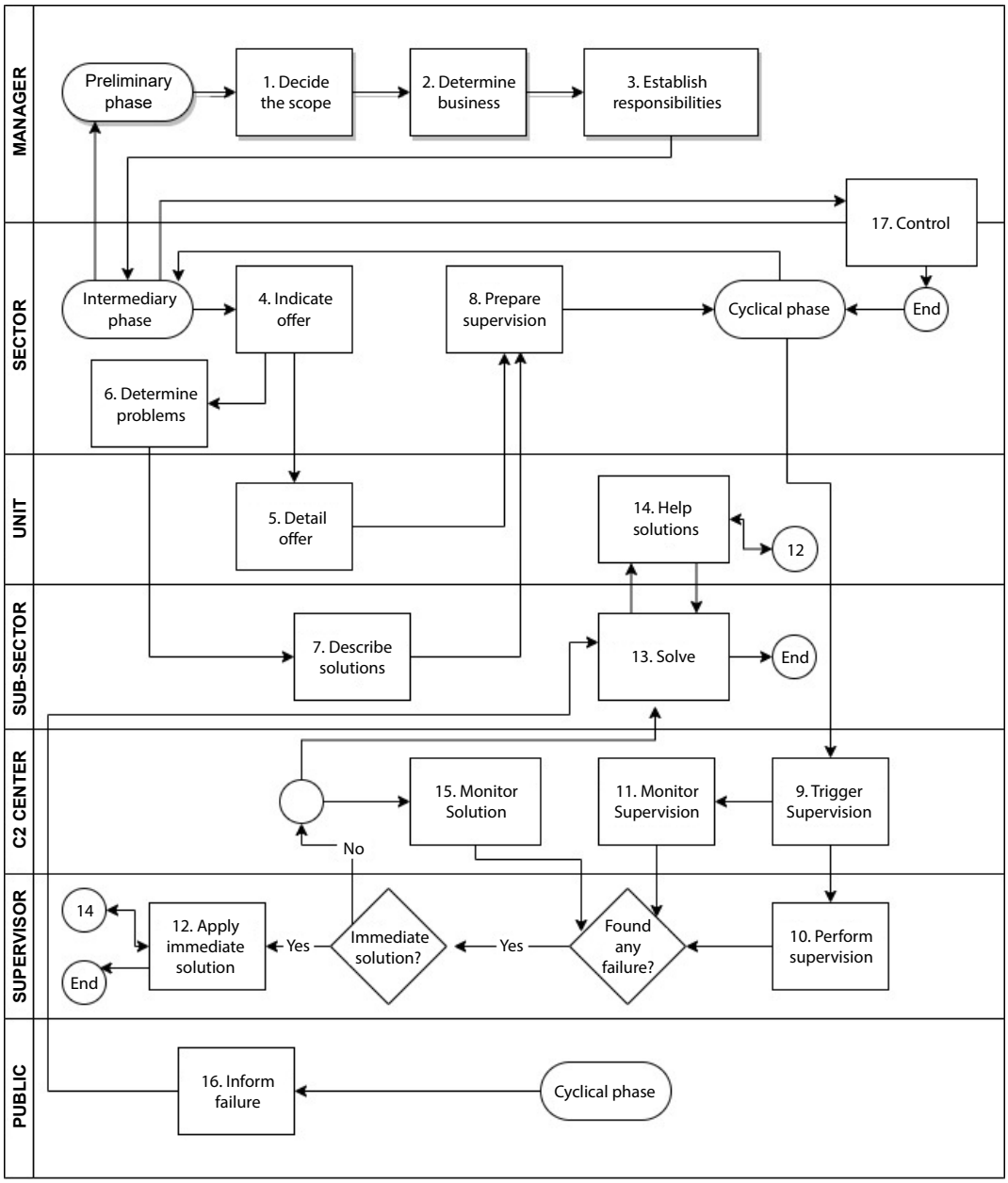
From the above, it is plausible to accept that, by acting to reduce the *tr* necessary for the treatment of a failure, one will be contributing to the improvement of the PSP's performance. It is exactly in this aspect of the PSP that the proposal acts: decreasing the *tr* for the solution of failures, acting on its components *tp*, *td* and *te*.

4.1 Procedural Model

The Business Process Flow Diagram of Figure 6 represents the proposed procedural model. The processes act on the variables related to the treatment of failures, as presented above. Processes nine, ten, eleven, and sixteen reduce the *perception time*; processes

ten, eleven, and twelve reduce the *triggering time*; and processes thirteen, fourteen, and fifteen reduce the *execution time*. Process seventeen may reduce any of the variables.

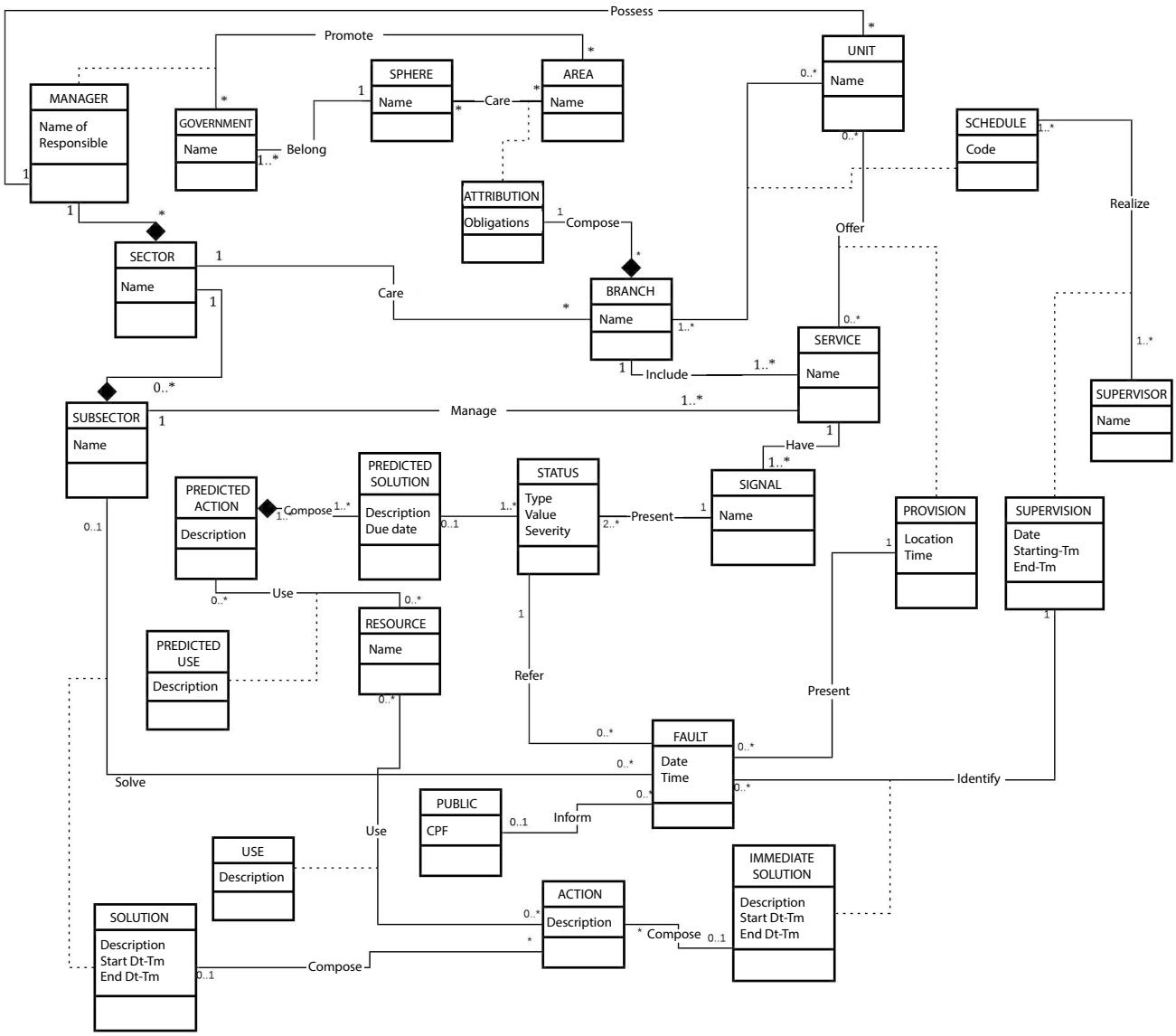
Fig. 6 - Model of C2 and DOCR Processes for the treatment of PSP failure.



4.2 Information Model and Organizational Model

The informational model contains the structure of the information necessary for the processes to be carried out, so that there is no type of ambiguity. This model contains the information elements (Sector, Fault, Solution, etc.), their main characteristics (date of the Fault, name of the Sector, description of the Solution, etc.) and the relationships between the elements (a Subsector acts on a Fault through a Solution, for example), and is presented in FIG. 7.

Fig. 7 - C2 and DOCR Information Model for the treatment of PSP failure.



The organizational model contains the generic functional structures (actors) responsible for carrying out the processes and managing the information. Some actors are pre-existing (Manager, Sector, Sub-sector, Unit, and Public), and others exist only for the treatment of faults (C3 and Supervisors). The actors are also represented in the lanes on the left side of Fig. 6 (actor responsible for the process).

In addition to the definition of the three models mentioned above, the proposal also presents the metamodel and the integration between dimensions, relating the three dimensions to each other, symbolized by the planes and the resulting graph presented in FIG. 4.

A metamodel is a model of the model. Obtaining the metamodel or meta-modeling is the analysis, construction, and development of applicable and useful representations, rules, constraints, models, and theories for modeling a predefined class of problems.

In the particular case of the research, the metamodel offers a view of how the three dimensions considered in the proposed model are organized and related. Thus, while the proposed model informs what

the elements of each of the dimensions addressed are and how the elements of each dimension relate to each other, the metamodel of the proposed model portrays the general characteristics of each dimension and how the dimensions relate to each other.

The class diagram of Fig. 8 contains the meta--model of the proposed model, whose core is formed by classes that represent the dimensions addressed: informational element, procedural element and organizational element.

Each of these classes specializes in the specific classes of each dimension. The procedural element specializes in seventeen classes representative of the processes identified in the procedural model; the organizational element, in the eight classes of the organizational model structure; and the informational element, in the twenty--six information classes of the informational model.

The associations between the meta classes representing the three dimensions provide the foundations for integration. In view of space limitations, the integration is not presented in this article, but the reader can refer to the reference dissertation [1] for further details.

Fig. 8 - Organizational Model of C2 and DOCR for the treatment of PSP failure.

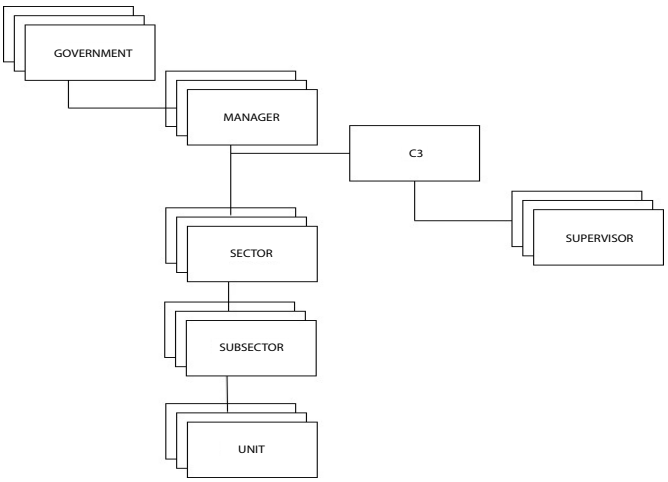
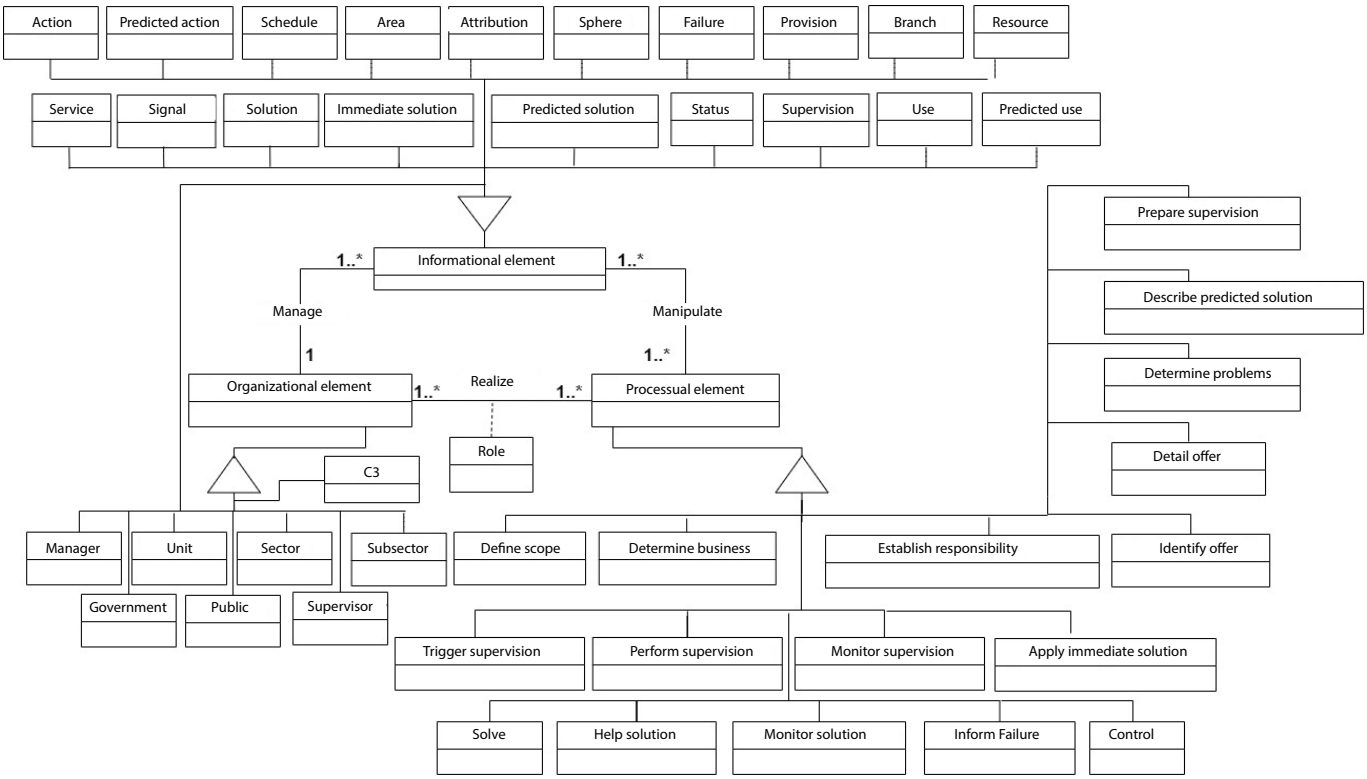


Fig. 8 - Meta-model of C2 and DOCR Information for the treatment of PSP failures.



5. Pilot Project

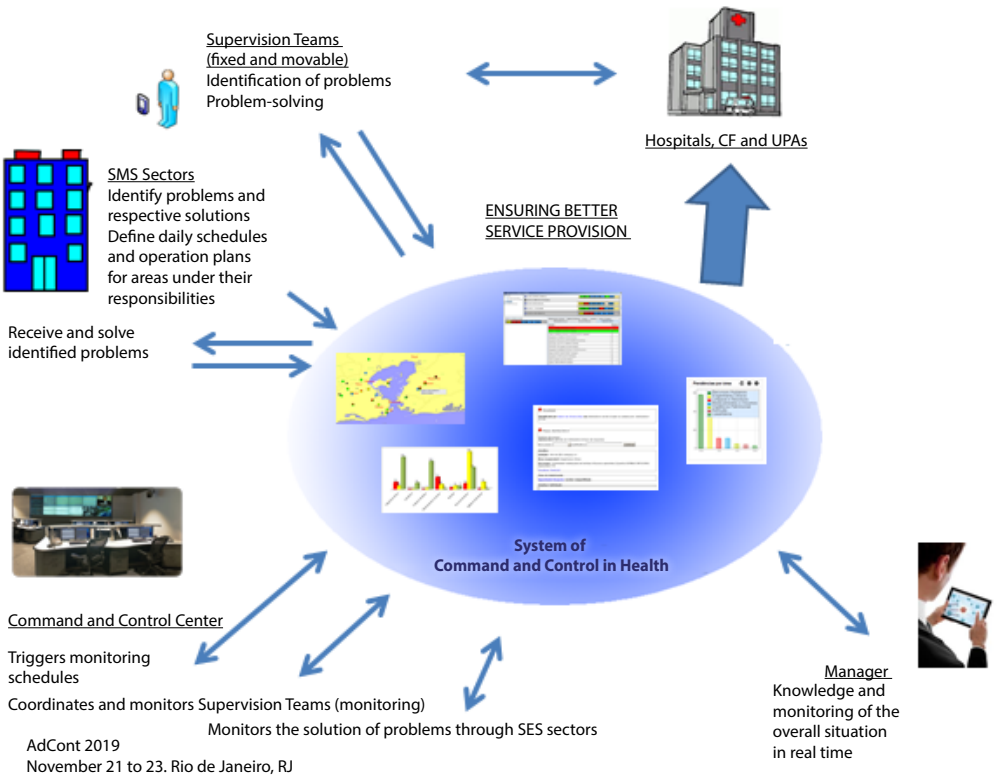
The pilot project was implemented in the period between 10/28/2013 and 03/13/2014, during the experimental and validation stages of the first cycle of action research. It was composed of the C2 Center, the supervision team, an IME team, the team of SES institutional actors involved in the treatment of failures, the monitored units (two hospitals and 28 UPAs) and the C2 computerized system (SC2). The daily operation composed of the processes of the cyclic phase of the procedural model (FIG. 6) is presented schematically in FIG. 9.

784 supervisions were performed, with 37,000 flags analyzed. 736 problems were identified, 632 of which were solved before the next supervision, as a result of the implementation of the proposal (resolution rate of approximately 86.4%).

In addition to decreasing the response time in the treatment of failures, other factors were noticed. For example, almost half of the failures identified throughout the experience (30.77%+17.48%=48.25%) presented reasonable or high severity, with more than 17% being severe. As an example of serious failure, the storage of expired drugs in doctors' offices, which were normally distributed to the public, can be mentioned.

Another finding concerns the previous perception of managers that the 'biggest problem' of SES-RJ would be related to the 'human resources' Branch (absences, delays, and shortages of professionals). It was found that the problems related to this Branch were, for the most part, of medium severity. The 'medicines and supplies' Branch, which was not the focus of greatest concern for managers, received most of the serious problems (almost 40% of all serious problems), as shown in FIG. 10. This finding means that a Branch with serious management problems in the delivery of the service was not perceived as a fragile point, thus not receiving due attention from managers. Particularly in the area of public health, the correct identification of fragile points in the PSP can represent the preservation of many lives.

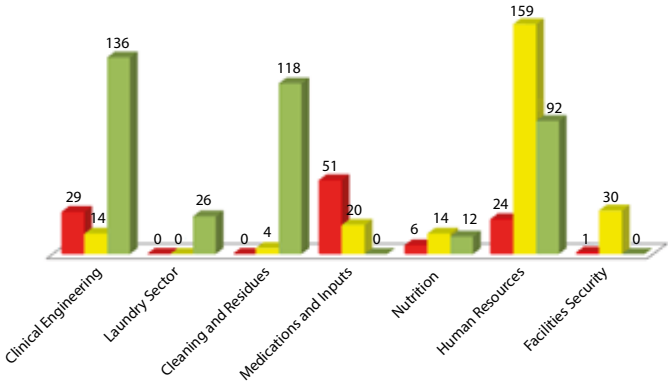
Figure 9 - Cyclical phase of the pilot project.



It was also possible to identify the institutional actors that faced the most problems, showing the need for an internal audit to seek the reasons for this behavior.

After the second cycle of the action research, based on the analysis of the results of the pilot project and a new round of research and development, the initial model was evolved into the proposed final model, which is presented in section 4 (Results). As explained in section 3 (Method), only the first cycle of the action research took place in full, and only one pilot project was implemented.

Figure 10 - Fault distribution according to severity and area. In each group of three bars, the more severe is on the left, and less severe on the right.



A second and third pilot projects are being planned at the time of writing, as part of the completion of the second cycle and the completion of a third action research cycle, which will continue the research produced during the completion of the original doctoral dissertation.

Conclusion

This study presents an open OI proposal for the implementation of C2 and DOCR in the treatment of PSP failures. The objective is to intensely reduce the consequences of failures during the provision of the service to the citizen, by reducing the times related to the solution of failures.

In view of the evolution that occurred in the exploratory stage of the second cycle of action research, the final proposal presents improvements in relation to the model implemented in the pilot project. For example, it is possible to identify how quickly the sectors solve the problems encountered and the public can participate in the treatment of failures, through process 16 (FIG. 2).

As limitations, only one complete cycle of action research can be mentioned; the lack of a mechanism for performance measurement; the impossibility of comparisons between the results of the pilot project and the previous situation, in view of the lack of a prior failure treatment mechanism; and the realization of only one pilot project. Further executions of exploratory and validation stages of the second cycle of action research tend to address these limitations.

Finally, it should be noted that the proposed model is generic, which means that it can be applied to any area of the public service (health, education, etc.), despite having been developed based on the field of public health.

ACKNOWLEDGMENTS

This study received support from FAPERJ (Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Brazil) [Processes E-26/ 210.277/ 2019(248665), E-26/201.409/2021(260810) and E-26/290.027/ 2021(260003)].

References

[1] Souza, Sylvio Jorge Junior. Comando e Controle como Inovação Organizacional na Prestação de Serviços Públicos. 2020. 290 f. Tese (Doutorado em Engenharia de Defesa) – Instituto Militar de Engenharia, Rio de Janeiro, 2020.

[2] Alberts, D. S.; Hayes, R. E. (2006). Understanding Command and Control. DoD Command and Control Research Program. CCRP Publications Series. Recuperado de: http://www.dodccrp.org/files/Alberts_UC2.pdf.

[3] Walker, Guy H.; Stanton, Neville A.; Salmon, Paul M. & Jenkins, Daniel P. (2008) A review of sociotechnical systems theory: a classic concept for new command and control paradigms, Theoretical Issues in Ergonomics Science, (Vol. 9:6, pp. 479-499, DOI: 10.1080/14639220701635470.

[4] QUIPUNGO, Policarpo F.; Rodrigues, E. C. Cavalcante; Peña, C. Rosano; Braga, B. Alexandre. PERCEPÇÃO DA QUALIDADE NO SERVIÇO DE TRANSPORTE PÚBLICO URBANO E DE SEUS IMPACTOS AMBIENTAIS NO DISTRITO FEDERAL - DF - BRASIL. Revista do CEPE. Santa Cruz do Sul, n. 44, p. 100-112, jul. dez. 2016.Ahmad, Zalinah. (2017). Public Service Delivery: Do We Need any Reforms? Religion, State and Society: Exploration of Southeast Asia. (Cap.. 6, pp. 86-91). Political Science Program. Department of Politics and Civics Education. Universitas Negeri Semarang.

[5] PROTASIO, Ane Polline Lacerda et al . Avaliação do sistema de referência e contrarreferência do estado da Paraíba segundo os profissionais da Atenção Básica no contexto do 1º ciclo de Avaliação Externa do PMAQ-AB. Saúde debate, Rio de Janeiro , v. 38, n. spe, p. 209-220, Oct. 2014 . Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-11042014000600209&lng=en&nrm=iso>. Access on 01 Oct. 2019. http://dx.doi.org/10.5935/0103-1104.2014S016.

[6] Serra, Carlos Gonçalves; Rodrigues, Paulo Henrique de Almeida. Avaliação da referência e contrarreferência no Programa Saúde da Família na Região Metropolitana do Rio de Janeiro. Ciênc. saúde coletiva, Rio de Janeiro. (Vol.. 15, supl. 3, pp. 3579-3586). Recuperado de: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232010000900033&lng=en&nrm=iso.

[7] Campos, Rosana Teresa Onocko et al. (2014). Avaliação da qualidade do acesso na atenção primária de uma grande cidade brasileira na perspectiva dos usuários. Saúde em Debate [online] (vol. 38, pp. 252-264). Recuperado de: https://doi.org/10.5935/0103-1104.2014S019.

[8] ANGELINI, Rossana Aparecida Vieira Maia. A qualidade da educação no Brasil: um problema de metodologia?. Rev. psicopedag., São Paulo , v. 23, n. 72, p. 213-220, 2006 . Disponível em <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0103-84862006000300004&lng=pt&nrm=iso>. acessos em 01 out. 2019.

[9] SCHWARTZMAN, S. Os desafios na educação no Brasil. In: BROCK, C.; SCHWARTZMAN, S. (Org.). Os desafios da educação no Brasil. Rio de Janeiro: Nova Fronteira, 2005.

[10] BAND. "MP apura falta de merenda quente em escolas no RJ". 2016. Obtida em "http://noticias.band.uol.com.br/cidades/rio/noticias/100000827894/mp-apura-falta-de-merenda-quente-em-escolas-no-rj.html". Acesso em: 12/12/2017.

[11] G1. "Paciente é atendida no chão do IJF, maior hospital de Fortaleza". Disponível em http://g1.globo.com/ceara/noticia/2014/11/paciente-e-atendida-no-chao-do-ijf-maior-hospital-de-fortaleza.html. 2014. Acesso em: 10/08/1016.

[12] G1. Crianças estão sem merenda em escolas públicas de Bela Vista. (2017). Recuperado de: http://g1.globo.com/ma/maranhao/noticia/2017/03/criancas-estao-sem-merenda-em-escolas-publicas-de-bela-vista.html.

[13] G1. Alunos da rede municipal de SP dizem estar proibidos de repetir merenda. (2017). Recuperado de: https://g1.globo.com/sao-paulo/noticia/alunos-da-rede-municipal-de-sp-dizem-estar-proibidos-de-repetir-merenda.ghtml.

[14] G1. Paciente precisa se enrolar em lençol por falta de roupa cirúrgica em hospital. (2017c). Recuperado de: https://g1.globo.com/to/tocantins/noticia/paciente-precisa-se-enrolar-em-lencol-por-falta-de-roupa-cirurgica-em-hospital.ghtml.

[15] PORTAL DA INDUSTRIA. "90% dos brasileiros dizem que qualidade dos serviços públicos deveria ser melhor considerando o valor dos impostos". 2016. Disponível em http://www.portaldaindustria.com.br/agenciacni/noticias/2016/07/90-dos-brasileiros-dizem-que-qualidade-dos-servicos-publicos-deveria-ser-melhor-considerando-o-valor-dos-impostos-2/. Acesso em: 22/09/2017.

[16] R7. "Absurdo: paciente é atendido no chão em hospital no Amapá". Disponível em http://tv.r7.com/record-play/balanco-geral/videos/absurdo-paciente-e-atendido-no-chao-em-hospital-no-amapa-17102015. 2013. Acesso em: 15/05/2017

[17] Bloch, Carter. (2011). Measuring Public Innovation in the Nordic Countries - Final Report. Danish Agency for Science, Technology and Innovation. Recuperado de: https://www.researchgate.net/publication/260793904_Measuring_Public_Innovation_in_the_Nordic_Countries_-_Final_Report.

[18] MERGEL, I.; DESOUZA, K. C. Implementing open innovation in the public sector: The case of Challenge. gov. Public Administration Review, v. 73, n. 6, p. 882-890, 2013.

[19] BOLÍVAR, M. P. R. (2017). Governance Models for the Delivery of Public Services Through the Web 2.0 Technologies: A Political View in Large Spanish Municipalities. Social Science Computer Review, 35(2), 203–225. https://doi.org/10.1177/0894439315609919

[20] Classe, Tadeu Moreira; Araujo, Renata Mendes; Xexéo, Geraldo Bonorino. (2018). Jogos Digitais Baseados em Processos de Prestação de Serviços Públicos: Um Estudo Exploratório. Acta Ludica International Journal of Game Studies. (Vol. 2, nº 1). Recuperado de: https://www.actaludica.com/ojs/index.php/actaludica/article/view/3/1.

[21] EB20-MC-10.205. (2015). Manual de Campanha - Comando e Controle. Ministério da Defesa, Exército Brasileiro, Estado-Maior do Exército. (1a Ed.). Brasil.
[22] MD31-M-03. (2015). Doutrina para o Sistema Militar de Comando e Controle. Ministério da Defesa. Estado-Maior Conjunto das Forças Armadas. (3ª Ed.). Brasil.
[23] Osinga, Frans P. B. (2006). Science, Strategy and War: The Strategic Theory of John Boyd. Routledge Ed. First Edition.
[24] ALBERTS, David S. Information Age Transformation. Command and Control Research Program (CCRP). Department of Defense US. 2002. Obtido em: http://www.dodccrp.org/files/Alberts_IAT.pdf.
[25] DEKKER, A.H. (2001). Applying Social Network Analysis Concepts to Military C4ISR Architectures. Connections, the official journal of the International Network for Social Network Analysis, 24(3), 93–103.
[26] ALSTON, Anthony; Beautement, Patrick; Dodd, Lorraine. Implementing Edge Organizations Exploiting Complexity. 2005 10th International Command and Control Research & Technology Symposium McLean, Virginia. June 2005. Disponível em: http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/107.pdf.
[27] GRANT, T. J.; R. H. P. Janssen and H. Monsuur . Network Topology in Command and Control: Organization, Operation, and Evolution. IGI Global. 2014.
[28] Nunes, João Paulo. (2007). Operações em Rede - Contributos para o seu estudo. Air and Space Power Journal – ASJP (2º Trim). Air University. Força Aérea Portuguesa. Recuperado de: http://www.airpower.au.af.mil/apjinternational/apj-p/2007/2tri07/vicente.html.
[29] Monsuur, Herman & Janssen, Rene & Grant, Tim. (2014). Network Topology in Command and Control. Organization, Operation, and Evolution. IGI Global. Recuperado de: https://www.researchgate.net/publication/269678422_Network_Topology_in_Command_and_Control_Organization_Operation_and_Evolution/citation/doWnload.
[30] AHMAD, Zalinah. Public Service Delivery: Do We Need any Reforms? Religion, State and Society: Exploration of Southeast Asia. Political Science Program. Department of Politics and Civics Education. Universitas Negeri Semarang. Chapter 6. p.86-91. 2017.
[31] Robinson, Mark. From Old Public Administration to the New Public Service - Implications for Public Sector Reform in Developing Countries. UNDP Global Centre for Public Service Excellence. Singapore. 2015. Obtido em: https://www.undp.org/content/dam/undp/library/capacity- development/English/Singapore%20Centre/PS-Reform_Paper.pdf.
[32] MARTINS, H., "Organização Governamental: Demarcação Conceitual do Campo e Análise de Movimentos Recentes em Governos Estaduais", Revista ADM MADE, ano 6 nº 3, 2008;
[33] MARTINS, H. & Marini, C., Governança pública contemporânea – uma tentativa de dissecação conceitual", Revista do TCU, edição nº 130, mai-ago 2014.
[34] MARTINS, H. & Alcoforado, F., "Organização governamental - análise de movimentos recentes em governos estaduais" in: "Avanços e Perspectivas da Gestão Pública nos Estados II". Consad, 2010.
[35] STONER, James A. F. e FREEMAN, R. Edward. (1995). Administração. Rio de Janeiro: Prentice-Hall.
[36] Shah, Anwar. (2005) Public Service Delivery. Public Sector Governance and Accountability Series. Anwar Shah, Series Editor. World Bank.
[37] [37] BRASIL. Constituição da República Federativa do Brasil, 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm/. Acesso em: 10 nov. 2016.
[38] Woolcock, Michael. (2019). World Bank Research and Policy Briefs. No. 117146. World Bank - Development Research Group (DECRG); Harvard University - Kennedy School of Government.
[39] Mbecke, Paulin. (2014). Operations and quality management for public service delivery improvement. Journal of Governance and Regulation (online). (Vol.3, pp. 36-45). Recuperado de: https://www.researchgate.net/publication/280226604_Operations_and_quality_management_for_public_service_delivery_improvement.
[40] Mergel, Ines. (2018). Open innovation in the public sector: drivers and barriers for the adoption of Challenge.gov. Public Management Review. (Vol. 20:5, pp. 726-745). Recuperado de: https://www.tandfonline.com/doi/pdf/10.1080/14719037.2017.1320044?needAccess=true.
[41] OECD. Guidelines for Collecting, Reporting and Using Data on Innovation, The Measurement of Scientific, Technological and Innovation Activities. Oslo manual. 4th Edition. OECD. Publishing Luxembourg. / Eurostat. 2018), Disponível em: https://doi.org/10.1787/9789264304604-en.
[42] Gault, F. (2012), "User innovation and the market", Science and Public Policy, Vol. 39/1, pp. 118-128.
[43] BARBIERI, José Carlos; ÁLVARES, Antônio Carlos Teixeira Sixth Generation Innovation Model Description of a Success Model. Revista de Administ ração e Inovação, São Paulo, v. 13, n.2, p.88-112, abr./jun. 2016.

[44] PAULA, F. D. O., SILVA, J. F. D. Balancing Internal and External R&D Strategies to Improve Innovation and Financial Performance. Brazilian Administration Review, 2018.
[45] Lam, Alice. (2004). Organizational Innovation. The Oxford Handbook of Innovation. ed. Royal Holloway College, University of London.
[46] Sousa, Maria. Inovação Organizacional. (2015). Disponivel em https://www.researchgate.net/publication/270956623.
[47] OECD. The Measurement of Scientific, Technological and Innovation Activities, Proposed Guidelines for collecting and interpreting technological innovation data. OSLO manual. Third edition. OECD. Publishing, Paris/ Eurostat, 2005.
[48] SZYMANSKA, Elzbieta. (2017) User-Driven Innovation – the Concept and Research Results. 7th International Conference on Engineering, Project, and Production Management Procedia Engineering. Volume 182, 2017, Pages 694-700. Volume 182, 2017, Pages 694-700. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877705817313188.
[49] RAUTER, ROMANA; GLOBOCNIK, DIETFRIED, ELKE PERL-VORBACH, RUPERT J. BAUMGARTNER. Open innovation and its effects on economic and sustainability innovation performance. Journal of Innovation & Knowledge, Volume 4, Issue 4, 2019, pp. 226-233.
[50] KOTSEMIR, M. N.; & MEISSNER, D. Conceptualizing the innovation process—trends and outlook. Working papers by NRU Higher School of Economics. Series WP BRP "Science, Technology and Innovation". No. 10/ STI/2013. 2013.
[51] Tripp, David. (2005). Action research: a methodological introduction. Educ. Pesqui., São Paulo. (Vol.. 31, n. 3, pp. 443-466). Recuperado de: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517--97022005000300009&lng=en&nrm=iso.
[52] Mello, C. H. P., Turrioni, J. B., Xavier, A. F., & Campos, D. F. (2012). Pesquisa-ação na engenharia de produção: proposta de estruturação para sua condução. Produção (online), (vol. 22, n. 1, pp. 1-13).
[53] RUP. "Rational Unified Process". (1998). IBM Company. Recuperado de: https://www.ibm.com/developerworks/rational/library/content/03July/1000/1251/1251_bestpractices_TP026B.pdf.
[54] Booch, G.; Rumbaugh, J.; Jacobson, I. "UML – Guia do Usuário, Rio de Janeiro. Ed. Campus. 2000.

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.9115.en

Model-Driven Engineering Applied to Radio Cognition in Military Operations

Marcus Albert Alves da Silva[1]*, Gabriel Bozza[1], Milena Mayara Ruy[1], Cláudia Reis Cavalcanti[1], David Fernandes Cruz Moura[1], David Fernandes Cruz Moura[2]
[1]Military Institute of Engineering (IME)
[2]University of Twente
Praça General Tibúrcio, 80, 22290-270, Praia Vermelha, Rio de Janeiro, RJ, Brazil
* m_albert@ime.eb.br

ABSTRACT: The dynamic nature of the military communications environment makes cognitive radio a promising alternative due to its versatility in sensing the environment and changing its operation mode autonomously. These changes must be grounded in military doctrine and telecommunications standards, which may vary from different military scenarios. In war situations, quick reaction and adaptation to new rules and conditions are desirable to avoid or prevent fratricide. This research applies Model-Driven Engineering (MDE) techniques to achieve fast adaptation, by using Domain-Specific Languages (DSL) and source code model transformations applied to cognitive radios in military operations. In a study case, these concepts were applied, speeding up the update of rules and employing a sliding windows strategy.

KEYWORDS: Cognitive Radio. Military Operations. Model-Driven Engineering. Domain-Specific Language.

RESUMO: O dinamismo do ambiente de comunicações militares faz dos rádios cognitivos uma alternativa promissora, em virtude de sua versatilidade para sensoriar o ambiente e mudar seu modo de operação de forma autônoma. Essas mudanças são baseadas em regras fundamentadas na doutrina militar e nas normas de telecomunicações, que podem mudar em função de alterações no cenário militar. A capacidade de reagir e se adaptar, com rapidez, às novas regras e condições é um diferencial que, em situações de guerra, pode evitar ou prevenir o fratricídio. Buscando atingir essa rapidez, esta pesquisa usou técnicas de engenharia dirigida a modelos, como Linguagens Específicas de Domínio (DSL) e transformações de modelos em código fonte aplicadas a rádios cognitivos, em operações militares. Estes conceitos foram aplicados em um estudo de caso, agilizando a atualização das regras e empregando uma estratégia de janelas deslizantes.

PALAVRAS-CHAVE: Rádio Cognitivo. Operações militares. Engenharia Dirigida a Modelos. Linguagem Específica de Domínio.

1. Introduction

Among the various means of communication used in military communications systems, the electromagnetic spectrum is shared by radios, radars, weapons systems, and other types of devices. During military operations, these systems ensure communicability between command and troop, supporting Command and Control (C2) systems. The diversity of types of operations may require changes in the behavior of communication means, considering technical (signal quality, range, susceptibility to noise, etc.) and tactical requirements of an operation (when the radio must operate at low power to avoid enemy detection). The combination of capabilities of software-defined radio (SDR) and cognitive radio (CR) [1, 2] emerge as an alternative to meet the variations of these requirements, with the agility required by a war environment.

As far as it was possible to investigate, there are studies focused on streamlining the configuration of SDR in military operations, but still without contemplating the definition of rules for autonomous decisions of cognitive radio [3]. On the other hand, other studies in the areas of disaster management [4], medical emergencies involving Internet of Things sensors [5], or even in business, [6] indicate that the use of *Model Driven Engineering* (MDE) could be appropriate in environments that require agility in updating rules that define actions to be performed.

This study proposes the use of an MDE-based approach as an alternative to streamline the construction of rules that will define the behavior of cognitive radios in a military communications system. In this context, a *Domain Specific Language* (DSL) was developed, which facilitates the description of rules, favoring the transformation of models and the automatic generation of codes. In this experimental environment,

the technical and tactical information sensed by the equipment was submitted to the rules that defined the radio's mode of operation.

This study is organized as follows: Section 2 presents a literature review on command and control systems, and communications in military operations, SDR, CR and MDE; Section 3 presents some related work, indicating the potential for the employment of MDE in military communications systems; Section 4 describes a case study; then there are the final considerations and indications of further studies in Sections 5 and 6, respectively.

2. Literature review

In this section, concepts involving military communications in operations are discussed.

2.1 Military operations

Military operation is the name given to the set of actions employing military forces and means, coordinated in time, space and purpose, following a directive, plan or order. It can occur in moments of peace, war, or in crisis situations, under the responsibility of a military authority [7]. In war operations, there is armed conflict and military power manifests itself with the use of violence. In peacetime operations, this power is employed in tasks unrelated to combat, except in special circumstances [8].

Military operations can be classified according to the forces employed and their purpose. As for the use of forces, they can be singular (developed by only one armed force), joint (employs ponderable means of more than one singular force) or combined (composed of ponderable elements of multinational armed forces, under a single command).

Under the finalistic bias, they can be classified as basic or complementary [7]. Basic operations, in situations of war, can be offensive, defensive or in cooperation and coordination with agencies, the latter being more common in situations of peace. Complementary operations, on the other hand, can be divided into seventeen different types of operations aimed at expanding, improving and/or complementing basic operations. The scope of this study does not include detailing the various operations; however, one should note there is a set of rules and restrictions to be adopted by the personnel and communications resources involved for each operation, in line with the context and military doctrine in force. In other words, during a specific operation, a radio may have to operate at a certain frequency, with low signal power, by contextual and doctrinal force.

2.2 Command and Control Systems (C2) and Communications Systems

Command and Control Systems (C2) form an environment where leaders manifest command intentions and perform actions to achieve objectives in a controlled manner. In addition, they can determine roles and responsibilities, as well as establish rules and restrictions to the context and subordinate elements [9]. These actions characterize behaviors defined based on the understanding gained by monitoring and evaluating the context of the environment. Agility in this understanding is a relevant factor for successful decision-making in a timely manner, and may represent an advantage over the enemy.

In the Brazilian Army (EB), through the C2 system supported by communications and information systems, aligned with the communications employment doctrine [10], the commander plans, directs and controls forces and operations [11].

A Communications System is a mesh of devices deployed in an action zone to meet the elements of an echelon. This structure enables the flow of data, voice and images through a network of nodal centers interconnected by various means of communication. Considering radio as a means of communication, different forms of link can be employed, such as satellite, microwave in direct sight or tropo-diffusion, using different waveforms.

Military doctrine [10] defines the means, forms of operation, and the use of communications resources. Such conditions and circumstances may involve tactical, doctrinal and physio-graphic issues, as well as

technical limitations and the role of each equipment in operations.

In military communications systems, range, capacity and mobility are relevant factors [12], which make up the so-called Compensation Triangle, which is a simple way to assess the relevance of using a means of communication.

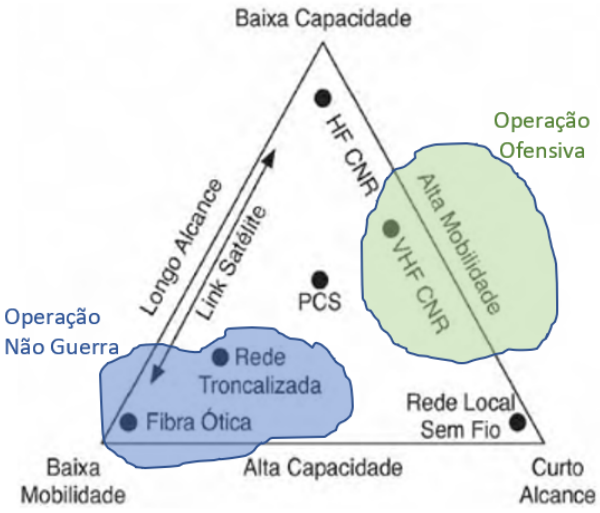
In Table 1, in each example operation (offensive and not war), the technology to be used varies according to the requirements defined, doctrinally, for each of the factors associated with the communications equipment used.

Table 1 - Compensation Triangle Factors applied to Subsystems.

Subsystem	Radio 1	Radio 2
Operation	Offensive	No War
Scope	Average	High
Mobility	High	Low
Capacity	Average	High
Technology	VHF	Trunked Network

Considering the requirements of each operation described in Table 1, they are represented in the compensation triangle as shown in Figure 1.

Figure 1 - Compensation Triangle, Adapted from [12]



Technologies such as software-defined radios and cognitive radios are appropriate to these variations in requirements.

2.3 Software-defined Radio (SDR)

An SDR is a type of radio whose majority of physical components are built using software [1, 2], and which can use virtual machines and intelligent agents [13]. This is considered an evolution of traditional radios, since functions such as modulation, demodulation, and filtering can be implemented by software rather than hardware. Reconfigurability, flexibility, and modularity are hallmarks of SDR, which can be reconfigured without hardware changes [14].

These possibilities make communication systems capable of operating in more than one mode, with different waveforms, with the same hardware [15]. A waveform carries from information describing security mechanisms in data transmission, source coding (voice, image and video compression), to re-transmission mechanisms and modulation and demodulation techniques, among other functionalities [16].

2.4 Cognitive Radio (CR)

A cognitive radio (CR) is an SDR capable of sensing the environment, changing its characteristics and its functioning, being able to adapt its operation to achieve objectives such as performance improvement, energy savings, and adaptation to the operational conditions in the place where it is inserted [2, 15]. The CR is also able to learn from past behavior, employing machine learning techniques to improve its functioning over time, analogous to what a human would do [17].

According to Doyle [18], the CR is a device that perceives inputs or views of the real world and, based on their understanding, makes autonomous decisions, being able to self-configure itself for communication tasks. This type of understanding is divided into four main areas: environment; communication requirements; policies, rules, and doctrines; and the radio's own capabilities.

The understanding of the environment is characterized by the correct detection and decoding of signals that are captured by the antenna, as well as by the knowledge of the operation in which it is inserted and its role in it.

Communication requirements are linked to which capabilities the equipment must have to meet the demands of communications systems (quality of service, tolerated noise level, etc.)

Policies and rules involve both standards for the exploration and use of the electromagnetic spectrum and manuals of military communications doctrine [12].

Knowing one's own capabilities means knowing the limits of equipment perception and operation, such as maximum power or sensed or transmitted waveforms.

Different types of cognitive methods can be applied to cognitive radios; some, such as those based on mathematical models or machine learning algorithms presented by Xu et al [19], could be suitable for definitions of radio behaviors related to technical issues as energy saving, noise level reduction, among others. On the other hand, when the decision is based on rigid doctrinal rules, rule-based models would be appropriate [19], whose knowledge structures, such as tables and decision trees, are desirable.

2.5 Model-Driven Engineering

In the context of software engineering, a model is a form of graphical or textual representation of a system, in which the abstractions and relationships employed are described by a metamodel [20].

Models are useful in a software project because they provide everyone involved, whether in the technical area or not, with a concrete, clear, and common view of the system to be designed [21].

Model-driven engineering (MDE) employs methodologies that conceive models as elementary artifacts in the software development process and not just as documentation of a project [21]. In other words, the model is part of the system and, even undergoing transformations, will always be aligned with its physical implementation.

Comparing MDE with approaches such as structured programming and object-oriented programming, there is greater simplicity in the propagation of changes. This characteristic is justified by the automatic generation of executable code from the system models, through the combined use of meta-modeling techniques and transformations between models (M2M) and from model to text (M2T) [21].

Model-to-Model transformations allow the transformation of one model into another, usually to a lower level of abstraction than the original, or simply so that the new model is more convenient to those involved in the project. On the other hand, *Model-to-Text* transformations generate software artifacts from the models through a technique called code generation. Defining a domain-specific language (DSL) enables the use of M2T transformations.

A DSL is created specifically for a particular application domain, such as languages used to search databases (CSQC) or languages that markup text for Web presentation (HTML). This type of language is close to the universe of understanding of domain users, contributing to the agility in the process of building this level of abstraction.

3. Related works

taking advantage of the characteristics of software-defined radios, in [3] the authors propose a tool that facilitates the configuration of radio equipment that will participate in a military operation. In this case, different software-devised configurations can be distributed to each equipment using a standard configuration file, but without exploring the use of a cognition method on the radio, enabling changes in operation from environment understanding and prior planning definitions.

On the other hand, studies focusing on decision support contexts applied MDE in environments of similar complexity to the communications scenario in military operations, such as bank fraud detection, public health and safety, disaster management, among others. In these environments, the understanding time of the observed information is also an important factor. In addition, the cognitive model is based on rules that can classify and identify perceived situations, which can generate understanding about the environment and support decision-making.

The study by Costa et al. [22] presents a scenario where the system is able to identify situations and react to the environment in which it is inserted. In this context, a DSL applied to the description of real-world situations (*Situation Modeling Language* - SML) serves as the basis for building a set of rules that run on a centralized, rules-based software platform (DROOLS). This platform allows for analyzing data collected from the environment within a time frame of observation. To validate the proposal, a case study was conducted, focusing on the detection of possible bank fraud in a mobile banking system. The research by Moreira et al. [4], aimed at disaster and emergency management, uses MDE in the conversion of texts expressed in SML (the same used by Costa et al. [22]) into rules processable into executable code. In addition, it also addresses the issue of data interoperability, justified by the interaction between the agents involved. The use of ontologies is pointed out as a favorable alternative in conceptual modeling, in applications based on situational awareness.

Brambilla's book [23] presents a DSL used specifically to favor interaction with users in the development of software applications.

Soleymanzadeh et al. [24] created a graphical environment to facilitate the understanding of domain users in the construction of business rules to be met by an application. The study applies MDE techniques in the translation of the rules created into the structure closest to the application domain. In addition, a case study was carried out in the area of collection of attorney's fees.

In this context, the MDE approach has shown good possibilities of use in the conversion of models and in the use and development of domain-specific languages that can help the understanding of the user who will insert the business rules. These initiatives demonstrate that the use of a DSL can favor agility in the elaboration of rules. However, no studies applied to cognitive radios have yet been observed, in favor of agility in the construction and configuration of rules applicable to sensed data. In addition, the approaches evaluated so far also do not address competing rules or establish priorities for them.

4. Modeling rules in cognitive radio systems

this study presents an MDE-based approach to streamline the construction of cognitive radio behavior rules in a military communications system. These rules will be transformed into intelligible code by the equipment and, from them, the radio will be able to change its own behavior (mode of operation) if there are significant variations in the environment that can be perceived and understood by it.

Initially, a small hypothetical scenario will be described, demonstrating cognitive radios acting in military operations, employing varied modes of operation.

In the scenario described by the sequence diagram in Figure 2, we have three radios involved in two operations (A and B). In operation A, there are the radios of the operation commander and of an operator, while in operation B there is only the radio of the operation commander.

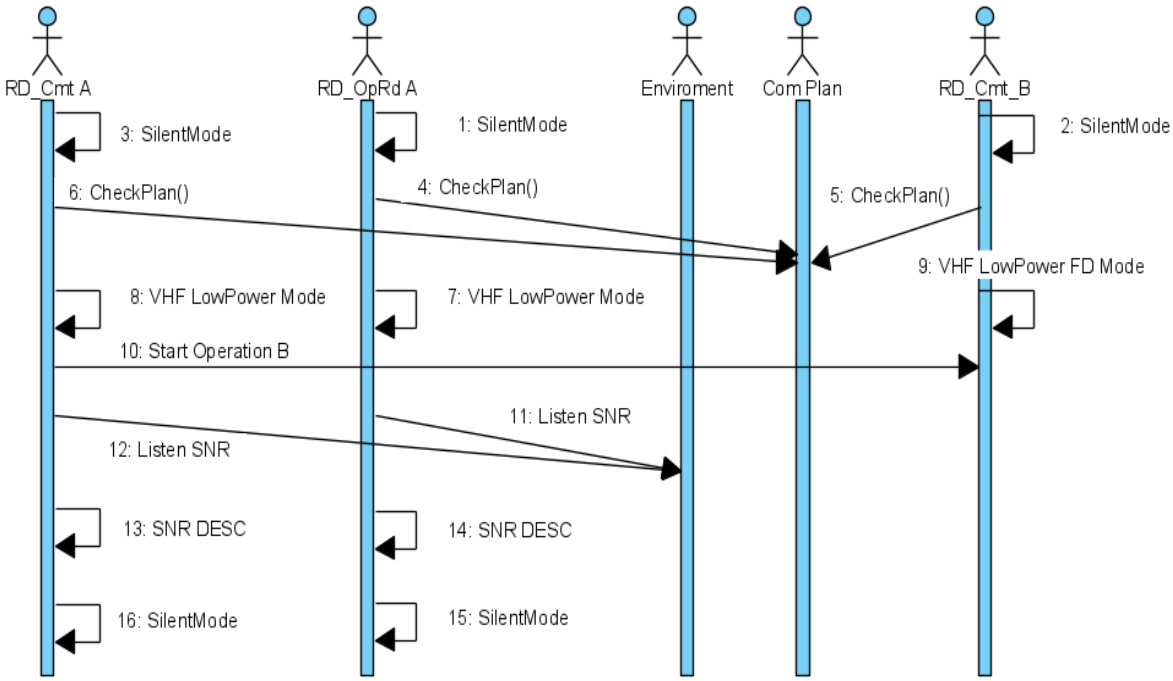
In this environment, three operating modes are used: 1- the Silent mode (*SilentMode*), in which the equipment is turned on, but does not emit any signal; 2 - the low power VHF mode (normal mode); and 3 - the low power VHF mode FD. In the latter, the acronym FD means that the equipment can operate in the Full Duplex In-band mode, in which the radio transmits and receives at the same frequency simultaneously, producing interference on its own signal and increasing safety in the physical layer [8].

The communications plan *(Com Plan)*, accessed by the radios, contains details about the operations in which the equipment participates. It describes the type of operation, the role of the radio, and its possibilities.

From the signal level and noise level readings present in the environment, the *Signal-to-Noise Ratio* (SNR) can be established, i.e., the relationship between the received signal level and the perceived noise level in the electromagnetic environment. A drop in SNR may be caused by a drop in the received signal level or by the increase in the perceived noise level.

The electromagnetic environment (*Environment*) illustrates the environment sensing action performed by the radios. Initially, they all work in silent mode. Then, a communications plan check application is performed and information such as *Security Level*, which depends on the criticality of the operation [25], the role of the equipment in the operation (Operational Role), the frequency of the channel used (Carrier Frequency), among others, can be collected.

Figure 2 - Communications in operations



From this information, the radio will be able to define its mode of operation, together with predefined rules. In this example, the mode set was low power VHF. In the sequence, messages are exchanged between the radio equipment of the Cmt of operations (RD_Cmt_A and RD_Cmt_B). In addition, the equipment of operation A checks the environment and realizes that the SNR ratio shows a decrease (SNR DESC), that is, there was a loss in the quality of the received signal. In this case, the rule defines that there is a change in the operating mode, and it starts to operate in the VHF low power FD mode.

One can classify this type of change as technical, as they aimed to improve the quality of communications. On the other hand, they could have been motivated by tactical issues, such as preparation for an attack, in which silence could be established, characterizing a tactical change.

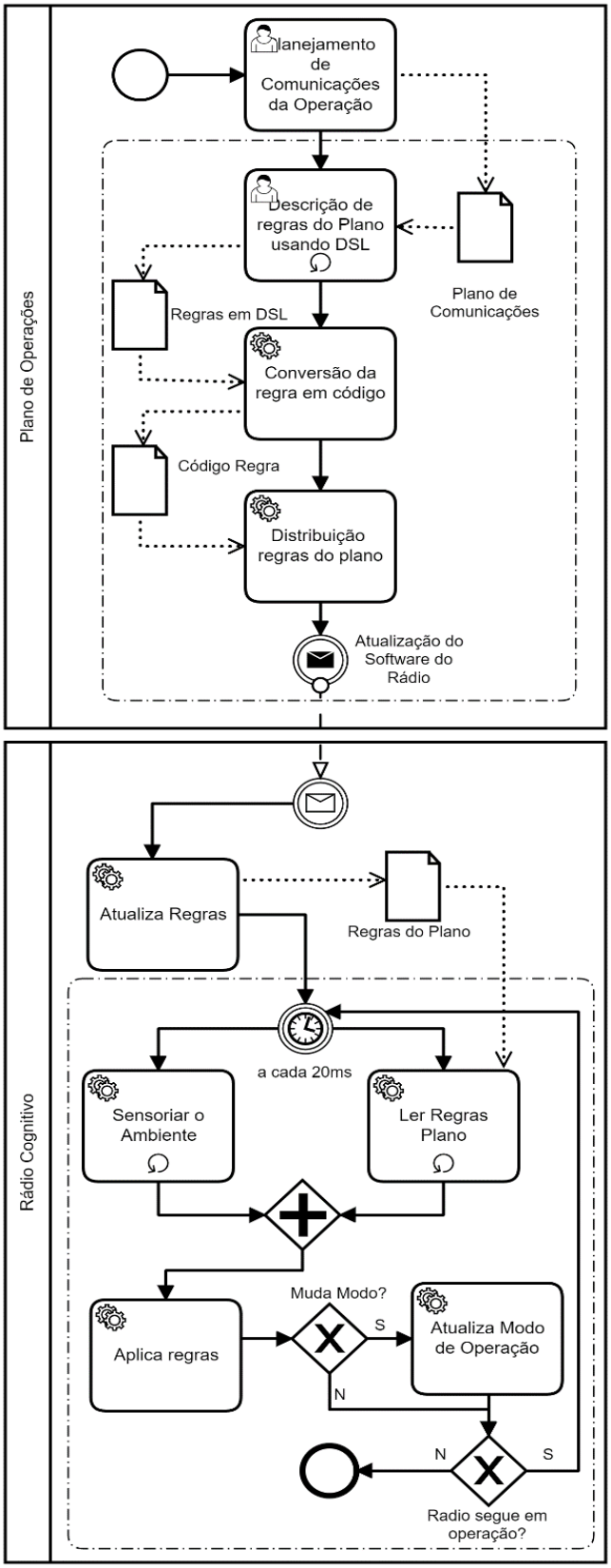
4.1 Research Overview and Scope

Through BPMN notation[1], Figure 3 presents an overview of the approach we are proposing in this work, highlighting tasks and some results produced in each of the steps.

Two *swimlanes* of the BPMN notation are represented, so that one of them represents a system dedicated to communications planning, in which the communications officer prepares the communications plan that will determine the behavior of the cognitive radio.

1 BPMN- from the English: Business Process Model and Notation [26]

Figure 3 - Approach Overview



Once the settings and rules are defined in the Communications Plan, the rules will be described through a DSL language specially designed for this context. The expressed rules are converted into the code of a programming language compatible with the cognitive radio software and hardware platform. At this point, the radios need to be "updated" with the new code, so that the rules take effect and determine their behavior during operation. When entering into operation, the cognitive radio, in a cyclical manner, senses the environment and reads the communications plan, in addition to applying the rules and, as the case may be, changing its mode of operation.

The dashed parts of Figure 3 indicate the scope of what has been implemented in this work. In the planning system, techniques based on MDE were applied; in the cognitive radio, a simulation of environment sensing, plan reading, and the application of rules and definition of the mode of operation was carried out autonomously.

It is worth noting that the set of rules can be changed, and the proposed scheme can make it easier to make changes in an agile way.

In this study, as a validation alternative, a case study simulated the sensing of the environment and the application of rules, defining the radio's mode of operation.

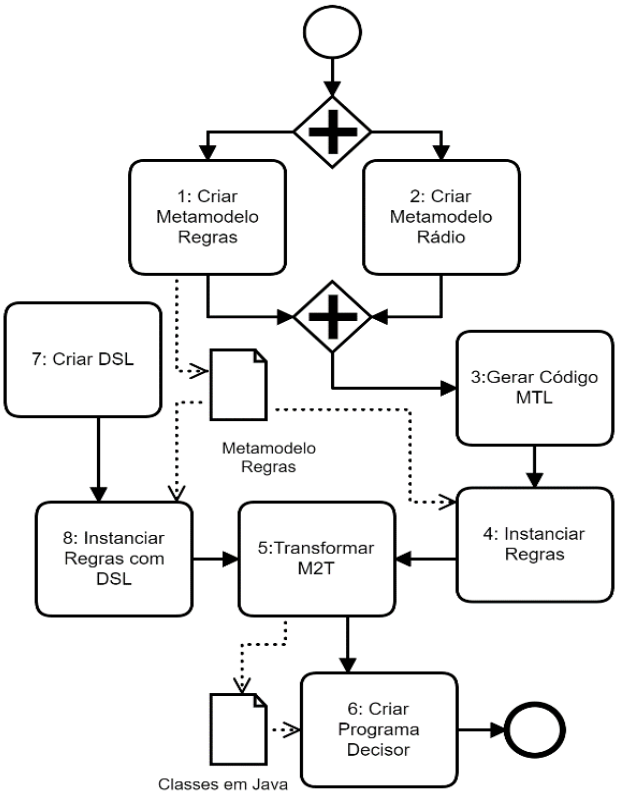
4.2 Applied methodology

Figure 4 describes the methodological procedures of this work, in which a metamodel describing a cognitive radio in an operational environment (step 1) and a metamodel describing the structure of formation and construction of rules that determine the mode of operation of the radio (step 2) were developed.

Then, a code (*template*) in MTL transformation language was developed, based on the rule formation metamodel (step 3).

After this step, rules were instantiated, based on the *template* metamodel (step 4).

Figure 4 - Applied methodology



From the code model and the created rule instances, the M2T transformation tool was used for the automatic generation of a class in Java language for each of the rules (step 5), based on the developed template. In the last step, the classes of all the generated rules are inserted into a Java language decision program. At that moment, the final artifact of the work is ready, that is, the executable decision program (step 6).

To make the process of elaborating the rules simpler and more agile for the user, different from the use of the template generated in step 3, a DSL was developed, i.e., a domain-specific language was created (step 7), which can be used for a more compact representation of the rules (step 8).

4.3 Metamodeling

The metamodel of Figure 5 represents, in a simple way, some concepts of a cognitive radio and a communications plan. In addition, it represents the elements involved in the operation of the cognitive radius. The *Transceiver* class corresponds to the radio equipment. The *Environment* class expresses the environment that the radio can sense, that is, signals and noise, in addition to deriving the signal/noise ratio (SNR). The *Communication Plan* class represents the communications plan of a given operation in which the radio is inserted. The plan informs the frequency of the signal carrier (*Carrier Frequency*), the role of the radio operator (*Operational role*), which can be "*commander*" or "*soldier*", and the security level of the operation in progress (*Security Level*).

In addition, the plan includes a set of rules that determine the operation of the radio. For example, it may contain a rule that says that if the SNR is decreasing, then the mode of operation must be changed. For simplicity, the rules were not represented in this metamodel, but in another, presented in Figure 6.

The *LogEntry* class represents the collection of information from the environment and the communications plan at each instant in time. Each *LogEntry* instance contains the SNR, *Security Level*, *CarrierFrequency* and *OperationalRole*. Finally, the Log class acts as a sliding window, which collects the most recent readings (for example, the last three) from *LogEntry*. Based on these readings, the cognitive radio (*Transceiver*) should refer to the rules of the Communications Plan to see if it should change its mode of operation.

In this composite rule, operators and operands connect, setting a mode of operation for the radio.

The SNR value and the last observed values of the other concepts present in the *Log* (the most recent data perceived) are the basis for the description of the rules and decision of the radio.

Motivated by the rule-based models of Xu et al [19] and inspired by the initiative of Costa et. al. [22] and the work of Horrocks et.al.[27], a set of rules (*ruleSet)* has been devised, and can be composed of a single or many rules. Each of them (*rule*) has an identifier (IDENTIFIER) and a priority. In addition, it consists of a single antecedent (*LeftSide*) and a single consequent (*RightSide*). The left side of the rule is composed of expressions, which can be simple (*SimpleExp*) or compound (*CompoundExp*). Each type of expression has its *operand*(s) and operators, which can be mathematical, logical or comparison. The operands can be attributes or values.

Figure 5 - Cognitive radio metamodel

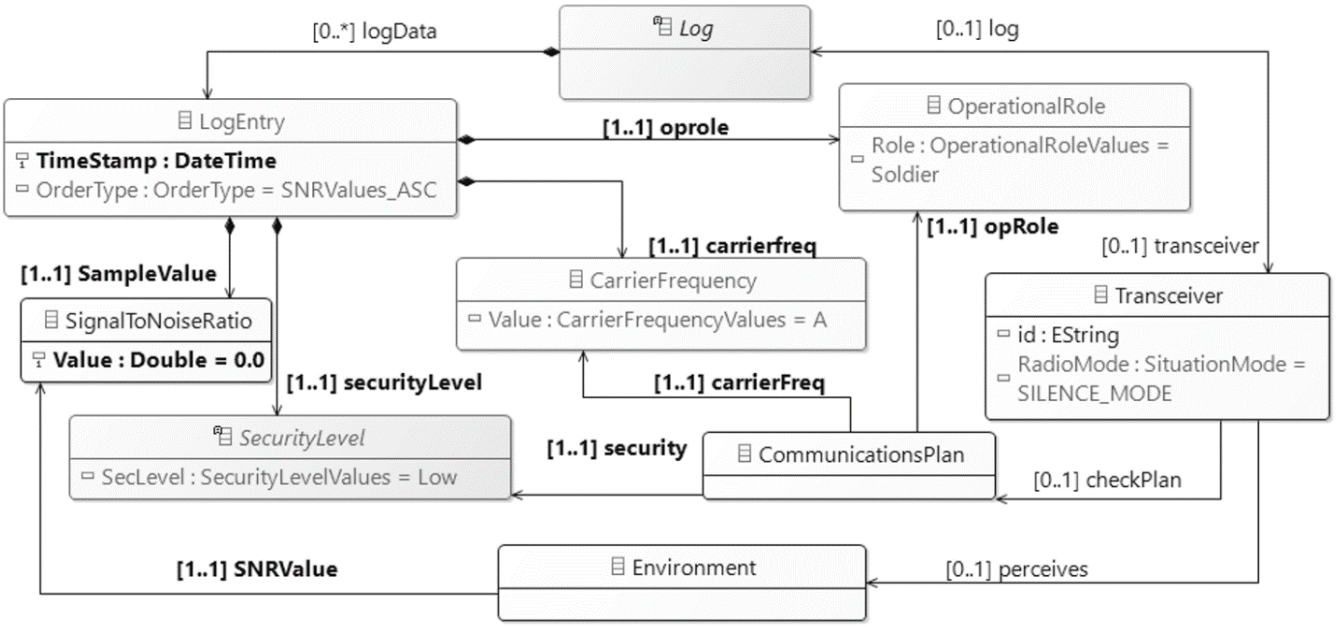
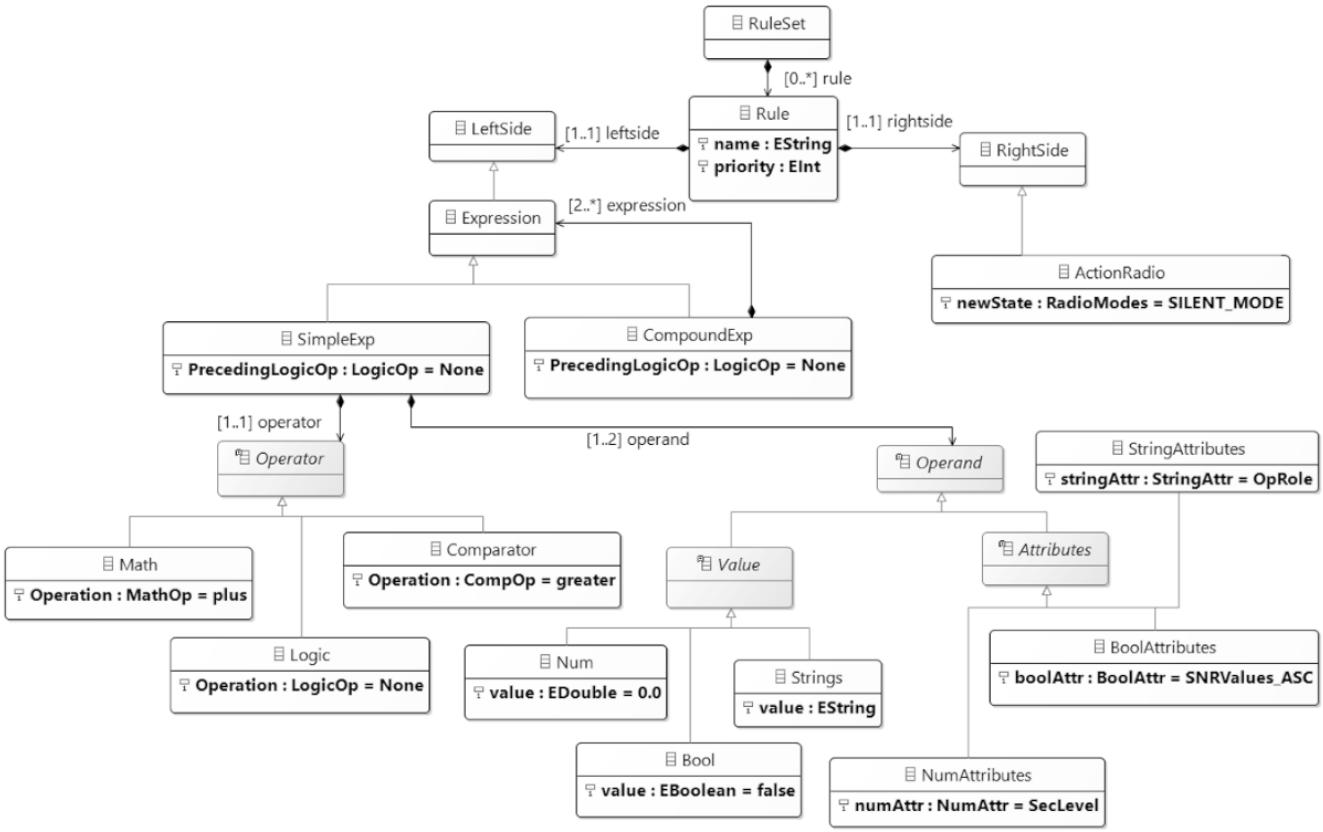


Figure 6 - Rules construction metamodel



The metamodel described in Figure 6 represents the way the rules are structured. Based on this model, a small grammar was developed, represented in Figure 7, employing the formalism of the Backus-Naur Form (BNF) representation.

Figure 7 - Clipping of the BNF grammar

<*rule*> ::= "{" *priority*, <*leftSide*> "→" <*rightSide*> "}"
priority ::= ([*0.0-1.0*]+)
< *rightSide* > ::= "normal" | "default FD" | "silence" | "alert"
<*leftSide*> ::= <*expression*>
<*expression*> ::=< *simpleExp*> | <*compoundExp*>
<*compoundExp*> ::= (< *expression* >, < *expression* >+\,)
<*simpleExp*> ::= <Boolean> <*expr*>
<boolean> ::= *True* | *False* | |ε|
<*expr*> ::= <*operand*> | "(" <*operand*> <*operator*> <*operand*> ")"
<*operand*> ::= <*value*> | <*attribute*>
<*operator*> ::= <*math*> |<*logic*>|<*comparator*>|
<*Math*> ::= "+" | "-" | "*" | "/"
<*logic*> ::= "AND" | "OR"| "NOT" | "NAND"| "NOR"|"NXOR" | "XOR"
<comparator> ::= "=" | "<" | ">" | "!="
<*value*> ::= <*number*> | <*string*> | <*boolean*>
<number>::=(^\d*[0-9](\.\d*[0-9])?$)
<string>::= ([*A-Za-z*][*A-Za-z0-9*]*)
<*attribute*>::= <*numAttribute*>
|<*stringAttribute*> |<*boolAttribute*>
<*numAttribute*> ::= "*securityLevel*" | "*carrierFrequency*"
<*stringAttribute*>::= "*sampleValue*" |"*operationalRole*"

In our example, we represent the universe of attributes that belong to the cognitive radio model presented in Figure 5, which participate in our usage case.

Rule (1) demonstrates a possible use of this grammar in the construction of rules in the domain of cognitive radios.

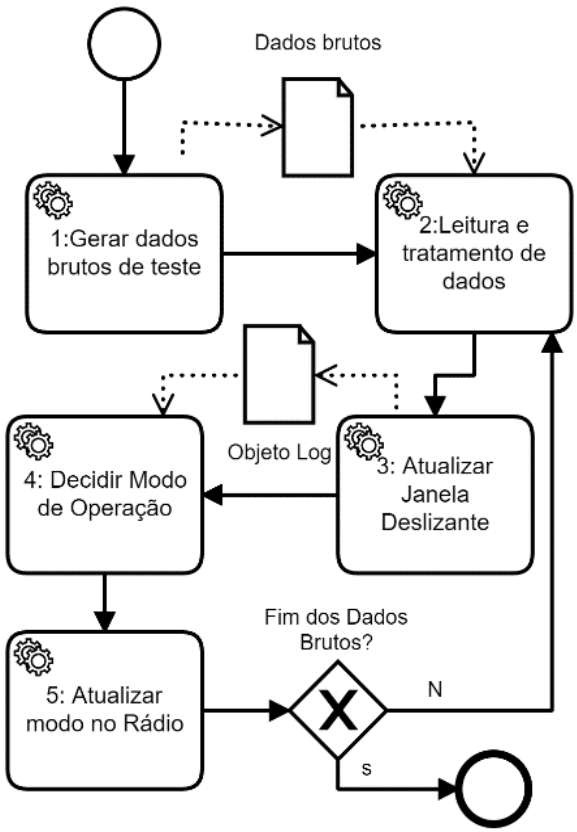
{0,2 True (True(sampleValue ="DESC") AND True (securityLevel < 0.5) AND False(carrierFrequency =100.0) AND (operationalRole ="soldier"))→ normal} (1)

When a rule is created, it receives a single priority value between 0 and 1, with 1 (one) being the highest priority. In this case, the priority is just a simplified strategy of dealing with possible competing rules, where two or more rules could be met simultaneously.

5. Case Study

This case study verifies the feasibility of the approach described in Section 4. According to Figure 8, a raw data file is generated from a test file generator program, which simulates readings of information from the electromagnetic environment and the communications plan, with intervals of 20 milliseconds, simulating a cycle of sensing the environment on a real radio (step 1) [28].

Figure 8 - Stages of the Simulation/Testing Process



From these data, another program reads, validates (step 2), and instantiates them in a class called Log, which represents a set of samples of the environment

sensed or perceived by the device in each simulated reading (step 3). In the next step (step 4), the decision program applies the rules that have been constructed by means of a DSL, resulting in a decision that can determine a new mode of operation for the radio (step 5).

Table 2 is a decision table, which represents the knowledge structure that describes the rules that were simulated in the experiment, representing those related to the doctrine of military communications.

Table 2 - Simulating Rules of military doctrine.

Tipo	Objeto	Atributo	Domínio	ST_1	ST_2	ST_3	ST_4	ST_5	ST_6	ST_7
Entrada	Sample Signal	Sample Value	ASC	true		true		true		false
			DESC		true		true		true	false
		Security level	>=0,5	true			true			
			< 0,5		true	true				
	transceiver	Carrier Frequency	50MHz	true			true			
			100MHz		true	true				
		Operational Role	Cmt	true		true				
			Soldado		true		true			
Saída		Modo Tx	normal							
			padrão FD							
			Silêncio							
			Alerta							
			Sit_Ant							

Taking as an example the rule determined by the context ST_1 of Table 2 (column ST_1), if the trend of the last three values of the received SNR is ascending (i.e., the SNR value of the last three perceived instants is increasing), the channel security level is greater than or equal to a threshold (0.5 in this case), the transmission frequency is 50 MHz and the "operational role" of the radio in question is that of commander ("Cmt"), the mode of operation to be used in the equipment will be the silence mode.

During the experiment, the Eclipse IDE was used, using the Eclipse Modeling Framework (EMF). The Sirius extension of Eclipse was used in the development of metamodels. In step of M2T transformation of the model into a Java program, the Acceleo tool was used.

The rule-building interface is not trivial for a user with no programming experience. To simplify the process of instantiating the rules metamodel, the Xtext tool [29] was used to generate and edit a DSL. This DSL allows the user to instantiate the rules of the radios without having to know any programming language, writing the rules in a simple way. Figure 9 shows a code fragment, demonstrating the description of two rules in the DSL. *Rule_ST1* has priority 1 and describes the conditions for the *Silent Mode*.

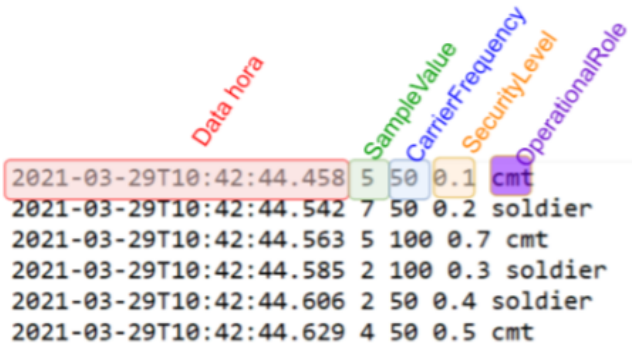
Figure 9 - DSL usage fragment

```
RuleSet{
    Rule 'Rule_ST1' {
        priority = 1
        rule = _[ _(SNRValues_ASC=true) and [_(SecurityLevel>=0.5)
               and(CarrierFrequency=50.0)and(OperationalRole="Cmt")]]
                --> SILENT_MODE
    },
    Rule 'Rule_ST2' {
        priority = 2
        rule = _[ _(SNRValues_DESC=true) and [_(SecurityLevel<0.5) and
               (CarrierFrequency=100.0)and(OperationalRole="Soldier")]]
                --> NORMAL_MODE
    },
```

5.1 Results obtained

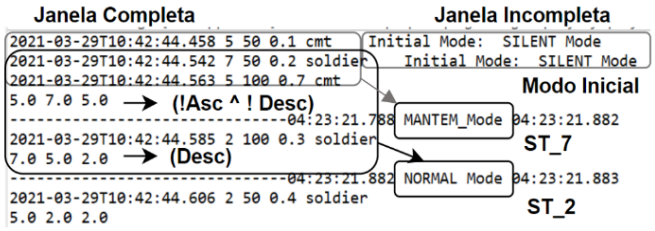
Figure 10 shows a small fragment of the input log, where the data that would be sensed is simulated. Each line represents values of the environment at an instant in time.

Figure 10 - Input log



After the execution of the rules, an output log file is generated, recording all sensed data and each decision made based on the application of the rules on the data contained in the sliding window of the last three observations of the environment. As shown in Figure 11, at the beginning of the execution, while the window is not yet filled, the operating mode of the equipment will always be the silent mode.

Figure 11 - Output log



When the window is filled, the rules are applied, that is, the last three *sampleValue* values are checked to identify whether they are ascending (ASC) or descending (DESC), along with the other values of the most recent reading. It is also possible to observe the changes in the operating mode, at times when the rules ST_7, which maintains the current mode (Silent Mode), and ST_2, which changes to *Normal mode*, are met.

6. Final Considerations

The diversity of types of military operations requires dynamism in changes in media behavior. Cognitive radios are a promising alternative, being able to adapt their operation to the technical and tactical needs of operations.

This study presented an alternative capable of supporting the variations of communications requirements during operations in an agile manner. In this experiment, only one piece of equipment operating in a single operation was simulated, but it was possible to verify its variations in behavior, based on technical and tactical rules. From this initial experiment, the modeling can be extended to a universe compatible with a real operation. The description of rules employing DSL, as well as model transformations and automatic code generation can contribute to agility in updates, in addition to maintaining alignment between the physical implementation and the conceptual model.

As contributions, we can highlight the proposed methodology that enables the replication of this experiment. The use of a sliding observation window can reduce the amount of data to be processed, consuming less computational resources, since the radio has limited computational power. Indirectly, the resource savings in the decision task will make more resources available for tasks that can be performed by machine learning algorithms.

As far as it was possible to observe, many studies have already made use of DSL and automated code transformations in environments that require urgency, however, the particularities of the military communications environment and the domain of cognitive radios were not considered.

Cognitive radios can be part of command and control systems, which operate in an integrated manner. In this context, interoperability between systems is a necessity and ambiguities in the understanding of concepts should be avoided. A failure to understand a concept may represent an error in the construction of rules involving communication systems, causing unwanted operation of equipment during an operation and causing, in some cases, fratricide.

Some studies [4, 22] indicate that the use of ontologies can favor syntactic and semantic interoperability between the various systems. In this sense, ontology-based interoperability will be explored in the next stages of this research, in addition to a possible integration with the mission planner proposed by Souza et al. [3]

The focus of this study is the benefit of the use of the language by the end user, streamlining the decision-making process and reducing the need for knowledge in programming languages. The use of this methodology in systems for performance evaluation and error handling mechanisms is recommended as future work.

Acknowledgments

we thank FINEP (Financier of Studies and Projects) for supporting the S2C2 Project, under agreement No. Ref. 2904/20, contract No. 01.20.0272.00, signed on 12/30/2020 and published in DOU edition No. 16, section 3, page 7, of January 25, 2021.

References

[1] MITOLA, J. The software radio architecture. IEEE Communications Magazine, v. 33, n. 5, p. 26-38, 1995. doi: 10.1109/35.393001.
[2] MITOLA, J.; MAGUIRE, G. Q. Cognitive radio: making software radios more personal. IEEE Personal Communications, v. 6, n. 4, p. 13-18, 1999. doi: 10.1109/98.788210.
[3] OUZA, V.; NAPOLITANO, F.; DIAS, M. Planejador de Missões do Rádio Definido por Software do Ministério da Defesa. In: Simpósio de Aplicações Operacionais em Áreas de Defesa, 21., São José dos Campos: Sige, 2019.
[4] MOREIRA, J. L. R.; PIRES, L. F.; VAN SINDEREN, M.; COSTA, P. D. Towards ontology-driven situation-aware disaster management. Applied Ontology, v. 10, n. 3-4, p. 339-353, 2015. doi: 10.3233/AO-150155.
[5] MOREIRA, J. L. R.; PIRES, L. F.; VAN SINDEREN, M.; DANIELE, L., SAREF4health: IoT Standard-Based Ontology-Driven Healthcare Systems. In: Formal Ontology in Information Systems – Proceedings of the International Conference, 10. Cape Town: FOIS, 2018. p. 239-252. doi: 10.3233/978-1-61499-910-2-239.
[6] TEIXEIRA, S.; AGRIZZI, B. A.; PEREIRA FILHO, J. G.; ROSSETTO, S.; PEREIRA, I. S. A.; COSTA, P. D. *et al.*, LAURA architecture: Towards a simpler way of building situation-aware and business-aware IoT applications, Journal of Systems and Software, v. 161, 2020. doi: 10.1016/j.jss.2019.110494.
[7] BRASIL. Operações: Manual de campanha. 5. Ed. Brasília, DF: Ministério da Defesa, 2017. Disponível em: http://bdex.eb.mil.br/jspui/handle/1/848. Acesso: 8 jun. 2017.
[8] CAMILO, M. J.; MOURA, D. F.; SALLES, R. M. Redes de comunicações militares: desafios tecnológicos e propostas para atendimento dos requisitos operacionais do Exército Brasileiro. Revista Militar de Ciência e Tecnologia, v. 37, n. 3, p. 5-25, 2020. doi: 10.2307/2215650.
[9] ALBERTS, D.; HAYES, R. Understanding Command and Control. CCRP, 2006.
[10] BRASIL. Manual de Campanha: As comunicações nas operações. Brasília, DF: Ministério da Defesa, 2020. Disponível em: http://bdex.eb.mil.br/jspui/handle/123456789/7073. Acesso: 28 out. 2020.
[11] BRASIL. Manual de Campanha: Comando e Controle. Brasília, DF: Ministério da Defesa, 2015.
[12] RYAN, M. J.; FRATER, M. R.. *Tactical Communications for the Digitized Battlefield*. Boston: Artech House, 2002.
[13] KRISHNAN, R.; BABU, R. G.; KAVIYA, S.; KUMAR, N. P.; RAHUL, C.; RAMAN, S. S. Software defined radio (SDR) foundations, technology tradeoffs: A survey. In: IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, 2017. p. 2677-2682. doi: 10.1109/ICPCSI.2017.8392204.
[14] BARROS, L. G. O Rádio Definido por Software. Trabalho de Graduação (Faculdade de Tecnologia) – Universidade de Brasília, Brasília, DF, 2007.
[15] MITOLA, J. An Integrated Agent Architecture for Software Defined Radio. Dissertation (Ph.D.) – Royal Institute of Technology, Stockholm, 2000.
[16] GALDINO, J.; MOURA, D.; MORAES, R.; SILVA, F.; MARQUES, E., PAIVA JUNIOR, N. Introdução ao desenvolvimento de rádios definidos por software para aplicações de defesa. In: Simpósio Brasileiro de Telecomunicações, 30., Brasília, DF, 2012. doi: 10.14209/sbrt.2012.211.
[17] SILVA, W.; CORDEIRO, J. R. S.; MACEDO, D. F.; VIEIRA, M. A. M.; VIEIRA, L. F. Introdução a rádios definidos por software com aplicações em GNU Radio. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 35., Vitória: UFES, 2015. p. 315.
[18] DOYLE, L. E. Essentials of Cognitive Radio. Cambridge: Cambridge University Press, 2009. doi: 10.1017/CBO9780511576577.

[19] XU, W.; XU, Y.; LEE, C.; FENG, Z.; ZHANG, P.; LIN, J. Data-Cognition-Empowered Intelligent Wireless Networks: Data, Utilities, Cognition Brain, and Architecture. IEEE Wireless Communications, v. 25, n. 1, p. 56-63, 2018. doi: 10.1109/MWC.2018.1700200.

[20] ALVES VIEIRA, M.; CARVALHO, S. Model-driven Engineering in the Development of Ubiquitous Applications: Technologies, Tools and Languages. In: Proceedings of the Brazillian Symposium on Multimedia and the Web, 23., 2017, p. 29–32. doi: 10.1145/3126858.3131633.

[21] SILVA, A. Model-driven engineering: A survey supported by the unified conceptual model. Computer Languages, Systems & Structures, v. 43, p. 139-155, 2015. doi: 10.1016/j.cl.2015.06.001.

[22] COSTA, P. D.; MIELKE, I. T.; PEREIRA, I.; ALMEIDA, J. P. A. A Model-Driven Approach to Situations: Situation Modeling and Rule-Based Situation Detection. IEEE International Enterprise Distributed Object Computing Conference,16., Beijing, 2012. p. 154-163.

[23] BRAMBILLA, M.; FRATERNALI, P. Interaction Flow Modeling Language: Model-Driven UI Engineering of Web and Mobile Apps with IFML. Burlington: Morgan Kaufmann, 2014.

[24] SOLEYMANZADEH, K.; BUL, Y.; BAĞCI, S.; KARDAS, G. A Tool for Modeling JsonLogic based Business Process Rules. In: International Informatics and Software Engineering Conference, 1., Ankara, 2019. p. 1–5. doi: 10.1109/UBMYK48245.2019.8965462.

[25] CAMILO, M. J.; MOURA, D. F. C.; SALLES, R. M. Combined Interference and Communications Strategy as a Defense Mechanism in Cognitive Radio Military Networks. IEEE Military Communications Conference, Norfolk, 2019. p. 113-118. doi: 10.1109/MILCOM47813.2019.9020787.

[26] CHINOSI, M.; TROMBETTA, A. BPMN: An introduction to the standard. Computer Standards & Interfaces, v. 34, n. 1, p. 124-134, 2012. doi: 10.1016/j.csi.2011.06.002.

[27] HORROCKS, I.; PATEL-SCHNEIDER, P. F.; BOLEY, H.; TABET, S.; GROSOF, B.; DEAN, M. SWRL: A Semantic Web rule language combining OWL and RuleML, W3C Member Submission, v. 21, 2004.

[28] FONTÁN. F. P.; ESPIÑEIRA, P. M. Modelling the wireless propagation channel: a simulation approach with Matlab. Hoboken: John Wiley & Sons, 2008.

[29] JONATHAN, B.; AVETYAN, R.; ABELN, S. Create Domain-Specific Language and Syntax Checker Using Xtext. International Journal of Industrial Research and Applied Engineering, v. 4, n. 1, 2020. doi: 10.9744/jirae.4.1.26-32.

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.8902.en

Analysis of glass substituted hydroxyapatite under different sintering routes

Andrey Mariano dos Santos[1], Iana Costa Carvalho[1], Ygor Pereira de Lima[1], Amal Elzubair[1]*, Suzana Noronha Ferreira Ribeiro[1], André Luís de Vasconcelos Cardoso[1], Marcelo Henrique Prado da Silva[1]
[1]Instituto Militar de Engenharia.
Praça General Tibúrcio, 80 - Urca, Rio de Janeiro - RJ, 22290-270
***amal@ime.eb.br**

ABSTRACT: Hydroxyapatite is a calcium phosphate that is classified as a biomaterial. It is used in bone restorations as it has characteristics similar to the mineral phase of bone tissue. Substitutional apatites have been especially researched to obtain better parameters for their syntheses and improve their impact on final applications. In view of the importance of these compounds as a biomaterial, this research aimed to study sintered substituted hydroxyapatite compounds to identify the formed phases, their microstructure, morphology, and shrinkage after sintering. HA and a composite of HA with 4% calcium phosphate and magnesium-based glass were produced from precursors in solutions and then sintered at different temperature. Results were obtained regarding the phases present and their morphology for the different conditions of chemical composition and sintering route by XRD analysis and scanning electron microscopy. Density and diameter shrinkage were analyzed after green body sintering at 1350 ºC. The HA samples showed a hexagonal phase, greater density, greater shrinkag, and uniform morphology with fewer pores than the HA/glass composite samples. The addition of magnesium glass partially transformed HA into tricalcium phosphate.

KEYWORDS: Biomaterial, Hydroxyapatite, Glass, XRD, Microstructure.

RESUMO: A hidroxiapatita é um fosfato de cálcio classificado como biomaterial utilizado em restaurações ósseas, dado que apresenta características similares à fase mineral do tecido ósseo. Em particular, as apatitas substitucionais vêm sendo pesquisadas a fim de se obter melhores parâmetros de suas sínteses e, consequentemente, de seu impacto na aplicação final. Tendo em vista a importância desses compostos como biomateriais, a presente pesquisa teve o objetivo de estudar compostos sinterizados de hidroxiapatita substituída, visando identificar as fases formadas, a morfologia da microestrutura e a retração do material após sinterização. HA e um compósito de HA e 4% de vidro a base de fosfato com cálcio e magnésio, produzidos a partir de precursores em soluções e depois sinterizadas em diferentes patamares de temperatura. Resultados foram obtidos quanto às fases presentes e sua morfologia para as diferentes condições de composição química e rota de sinterização aplicada a partir da análise de DRX, microscopia eletrônica de varredura. A densidade e a retração de diâmetro foram analisadas após a sinterização dos corpos verdes à 1350 ºC. As amostras da HA apresentaram uma fase, maior densidade, maior retração e morfologia uniforme com poucos poros, comparada às amostras do compósito HA/vidro. A adição de vidro de magnésio resultou na transformação parcial da HA em fosfato tricálcico.

PALAVRAS-CHAVE: Biomaterial, Hidroxiapatita, Vidro, DRX, Microestrutura

1. Introduction

Increasingly interest has been devoted to calcium phosphate materials because of a chemical composition similar to that of the bone mineral. Thus, calcium phosphate ceramics with expected superior biological properties began to be investigated. They are called bioactive ceramics, a family that mainly includes calcium phosphate, so-called hydroxyapatite (HA), tricalcium phosphate (TCP), and biphasic calcium phosphates (BCP, the mixture of HA and TCP). Hydroxyapatite is considered the most used biomaterial in bone restorations [1-3]. Given that there are several methods of obtaining this compound in the literature, the synthesis of hydroxyapatite powder is of commercial and scientific interest, as the processing conditions have a direct influence on the structure, morphology and application of the desired final compound [4].

In addition, hydroxyapatite may act as a coating for orthopedic implants in physiological media. It is also noted that its application in dental implants is feasible, as well as calcium and phosphorus reserves, considering the ease of anion and cation substitutions, providing the release or storage of ions in body fluids [5].The porosity of hydroxyapatite is responsible for the adsorption capacity of molecules, favoring the incorporation of drug-conducting drugs [3], as HA

easily substitutes anions and cations. Thus, bone growth around the implant is also promoted [6].

This work aimed to process HA in different routes, using samples of pure hydroxyapatite, as well as hydroxyapatite with a glass based on calcium and magnesium phosphate, in order to identify phases formed and morphology of the microstructure.

2. Materials and Methods

2.1 Glass Synthesis

For the production of glass, individual solutions of the precursors $MgCl_2.6H_2O$ 99.99%, $CaCO_3$ 99.99%, $H_3PO_4$85% were prepared to meet the proportion of 0.01 mol Mg: 0.03 mol, Ca: 0.06 mol, P.

The amount of 3.033 g of $CaCO_3$, 2.033 g of $MgCl_2.6H_2O$ were weighed on a Gehaka analytical balance, model BK 300, and 8.08 ml of H_3PO_4 were measured in a beaker, all reagents measured. Then they were diluted in distilled water separately, totaling 50 ml of solution for each. The solutions were then mixed in a 500 ml Becker with magnetic stirrer with heater marca 5L-HS9T.

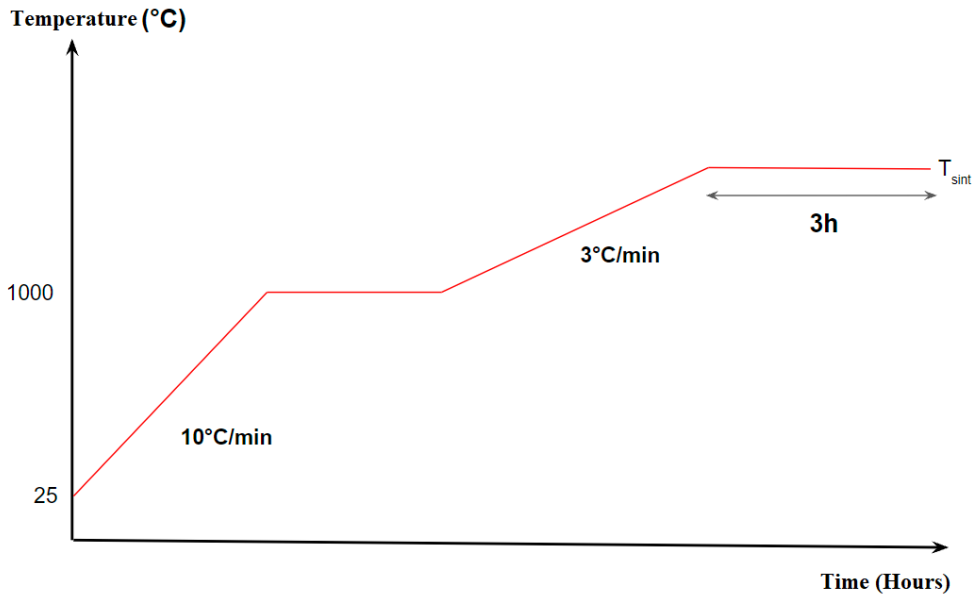
2.2 Sintering of samples

Four samples were prepared, two with pure hydroxyapatite (HA) and two composite samples with 96% hydroxyapatite and 4% magnesium glass (HA/glass composite). For this, 11.95 g of hydroxyapatite was weighed for the pure tablets, 11.87 g of hydroxyapatite and 0.49 g of magnesium phosphate glass for the composite tablets. The hydroxyapatite used was previously synthesized in the ceramic laboratory. The composite powders were ground in a MARCONI horizontal ball mill, model MA 500.

Tablets of 2 g were produced in a 20 cm diameter matrix, initially pressed in a manual hydraulic press with a load of 2 tons applied for 30 s, totaling a load of 62.38 MPa. Subsequently, the pellets were sintered in the Jung oven, and the samples were sintered in pairs consisting of an HA pellet and one of the HA/glass composite, at two different temperatures: 1100°C and 1350°C.

Figure 1 shows the sintering route in which the furnace was first heated to 1000°C with a heating rate of 10°C/min. When stabilized at this temperature, heating was continued until the sintering temperature with a heating rate of 3°C/min, and then remained at this temperature for 3 hours.

Figure 1 - Schematic representation of the sintering route.



3. Characterization of samples

3.1 Physical Properties

The linear contraction, bulk density and porosity of the samples were measured. The measurement of linear contraction was performed with the aid of a manual caliper, of the diameters of the samples in the green body condition and after sintering. The density and porosity were calculated by the Archimedes method applied to the sintered samples:

$$\text{Bulk density} = (M_{dry}).\rho_{water} / (M_{wet} - M_{immersed}) \quad (1)$$

$$\text{Porosity \%} = (M_{wet} - M_{dry}) \times 100 / (M_{wet} - M_{immersed}) \quad (2)$$

Where: M is the mass of the sintered sample, measured dry, immersed and wet.

ρ_{water} is the density of water ~ 1 g/cm^3.

3.2 Scanning Electron Microscopy (SEM)

The morphologies and microstructure analyses were performed with the Scanning Electron Microscope, QUANTA model FEG 250 (SEM-FEG). The following parameters were used: 10 to 15 kV acceleration voltage, 30 µs scan time, 1000 to 16,000 X magnification. In addition, the technique was applied with secondary electrons.

3.3 X-Ray Diffraction

Crystallographic phase analyses were performed using X-ray diffraction (XRD) with a Panalytical diffractometer, model X'Pert PRO MRD. Cobalt source (Co-Kα = 1.789010 Å), 40 kV voltage, 40 mA current, 2θ scan from 20° to 80° under 0.029° step was used.

4. Results and discussion

4.1 Physical Properties

Table 1 shows the contraction content of the samples after sintering. It was not possible to obtain the contraction value of the sintered samples at 1100 °C, as they fractured after the process. When HA and HA/glass densification above 1000 °C is initiated, grain contours form [7], total pore volume decreases, and samples contract to reach their maximum bulk density. Then, the linear contraction occurs due to the compaction of the grains that make up the ceramic material during the sintering process, increasing the density of the material and consequently decreasing its dimensions.

It is seen that the percentage of contraction for the pure HA samples was much higher than that of HA/glass, indicating that the presence of the glass promoted a drop to more than half contraction of the sintered sample.

Density measurements (Table-1) revealed that samples of the pure HA sintered at 1350ºC have higher density than that of the HA/sintered glass composite at the same temperature. This resulted in a higher porosity in the composite when compared to that of pure HA.

Samples sintered at 1100ºC had lower density and higher porosity than those sintered at 1350ºC.

Table 1 - Linear contraction, density and porosity of sintered samples.

Sample	Contraction ± SD (%)	Density ± SD (g/cm^3)	Porosity ± SD (%)
HA (1350ºC)	19.10 ± 0.34	2.7769 ± 0.0291	2.96 ± 0.48
HA/Glass (1350ºC)	8.8 ± 0.94	2.5472 ± 0.0034	9.47 ± 1.42
HA (1100ºC)	-	2.434	15.56
HA/Glass (1100ºC)	-	1.7982	37.69

4.2 Scanning Electron Microscopy (SEM)

Figures 2 to 5 show the micrographs of the sintered samples obtained by SEM. Figures 2 and 3 show HA and HA/glass samples sintered at 1100 °C. In Figures 2a and 3a, it is possible to see crack formation in the samples, but with greater intensity in the HA/glass composite. It is observed in Figures 2b and 3b that there was a growth and coalition of the grains in the process, thus decreasing the total interfacial energy of the grain contours. It can be noted that the HA/glass composite grains have greater growth and have a significant presence of pores.

Figure 2 - Microstructure of the pure HA sample sintered at 1100°C.

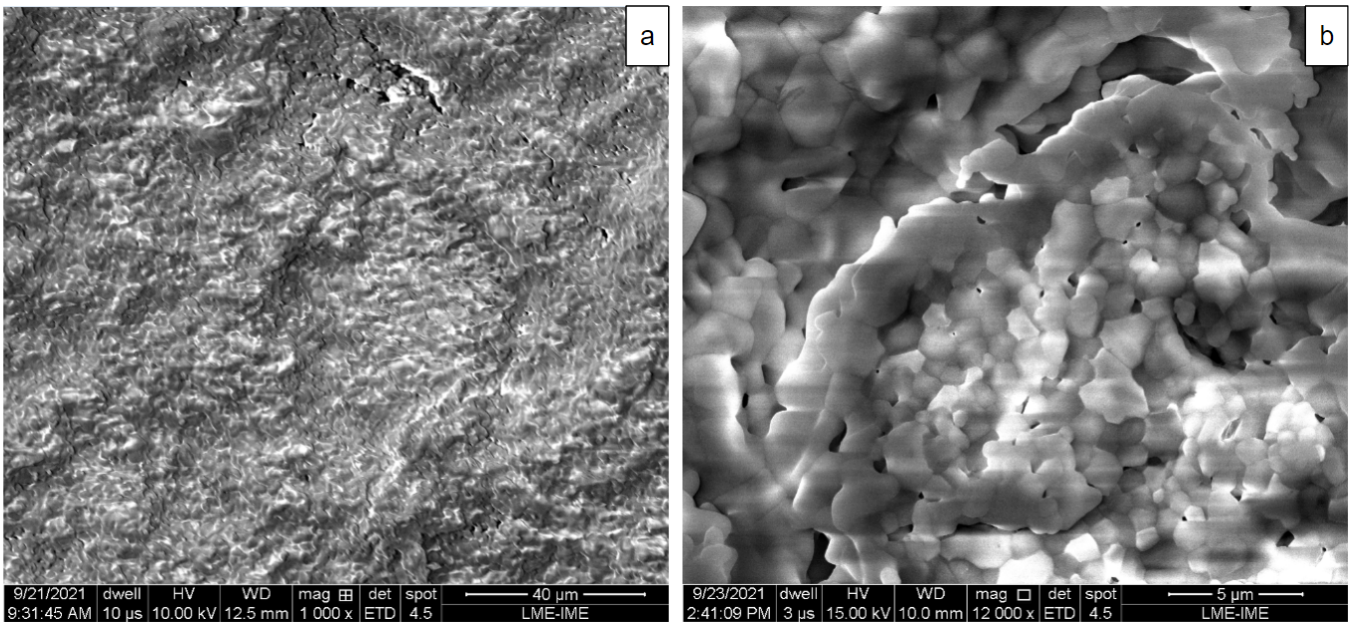
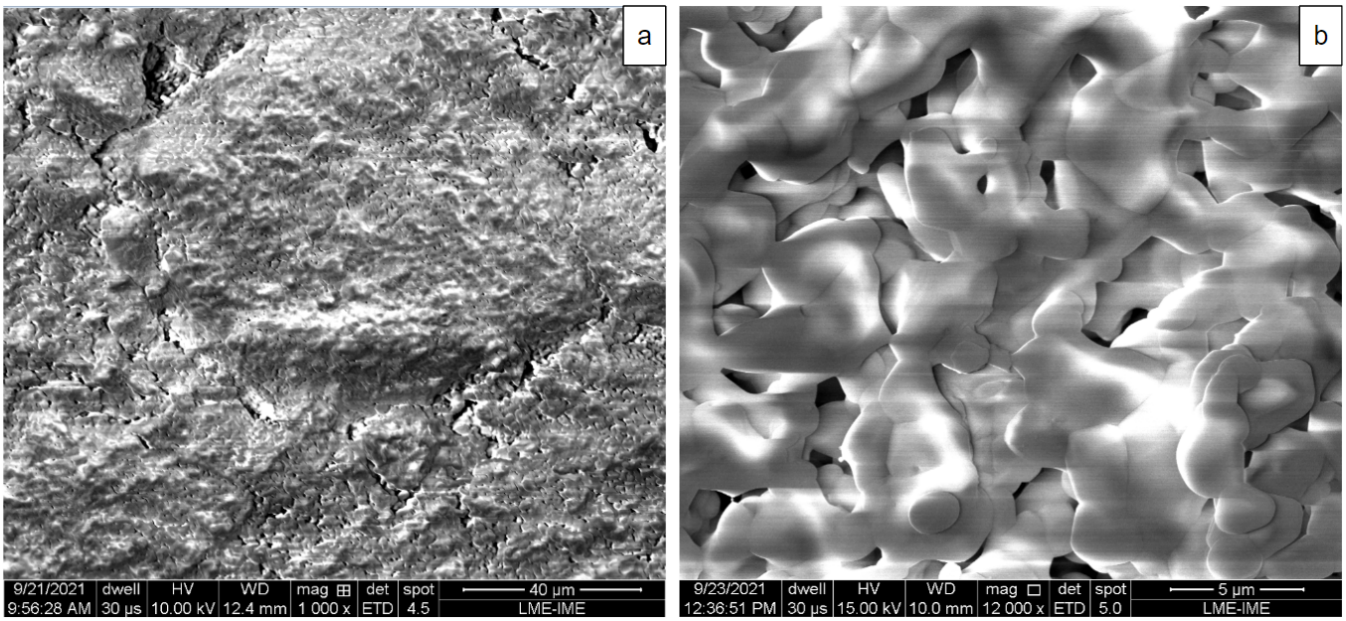


Figure 3 - Microstructure of the HA/Sintered glass sample at 1100°C.



In the case of the samples sintered at 1350°C, a less rough surface appearance morphology was observed in Figures 4a and 5a than in the samples sintered at 1100°C. There is also a lower occurrence of cracks and greater densification. There are still some intergranular pores and channels, most notably in the HA/glass samples (Figure 5a). Figure 4a shows that HA sintering produced ceramic, almost ideally dense (bulk density ~ 97% as shown in Table-1), without significant grain growth. The final microstructure in Figures 4a and 5a consists of hexagonal grains. Figure 4b showed regions with residual closed pores. An interesting structure can be seen in Figure 5b, with increase magnification, there is presence of pores or small circular cavities distributed on the surface of the grains.

Figure 4 - Microstructure of the pure HA sample sintered at 1350°C.

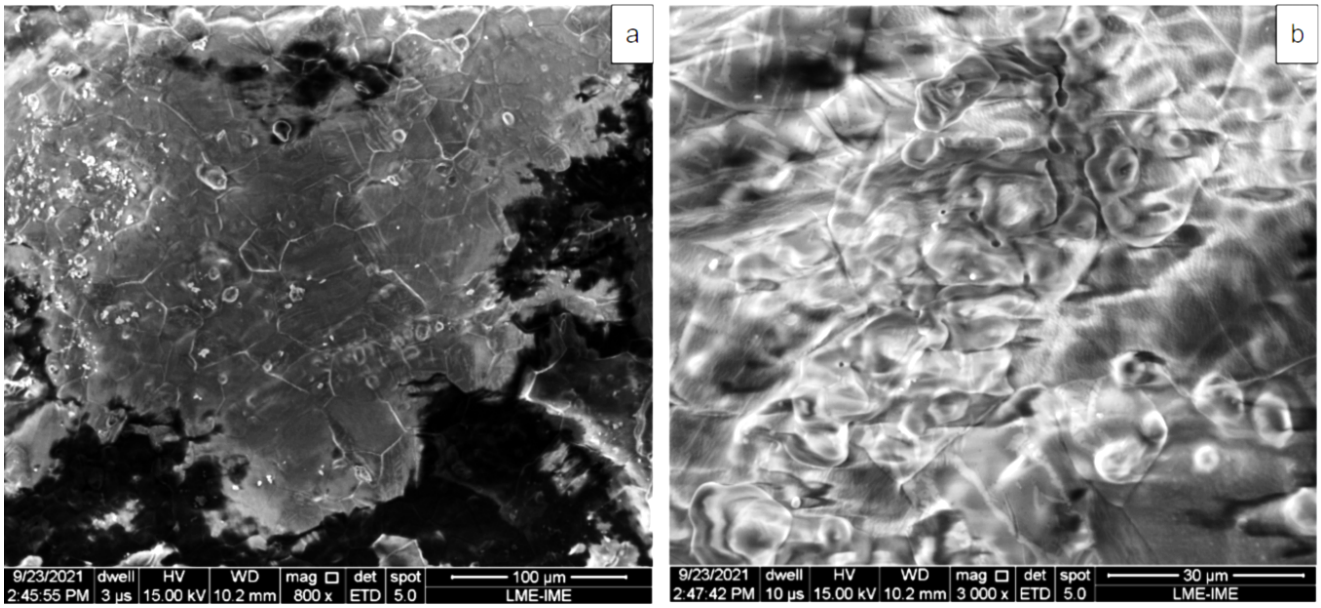
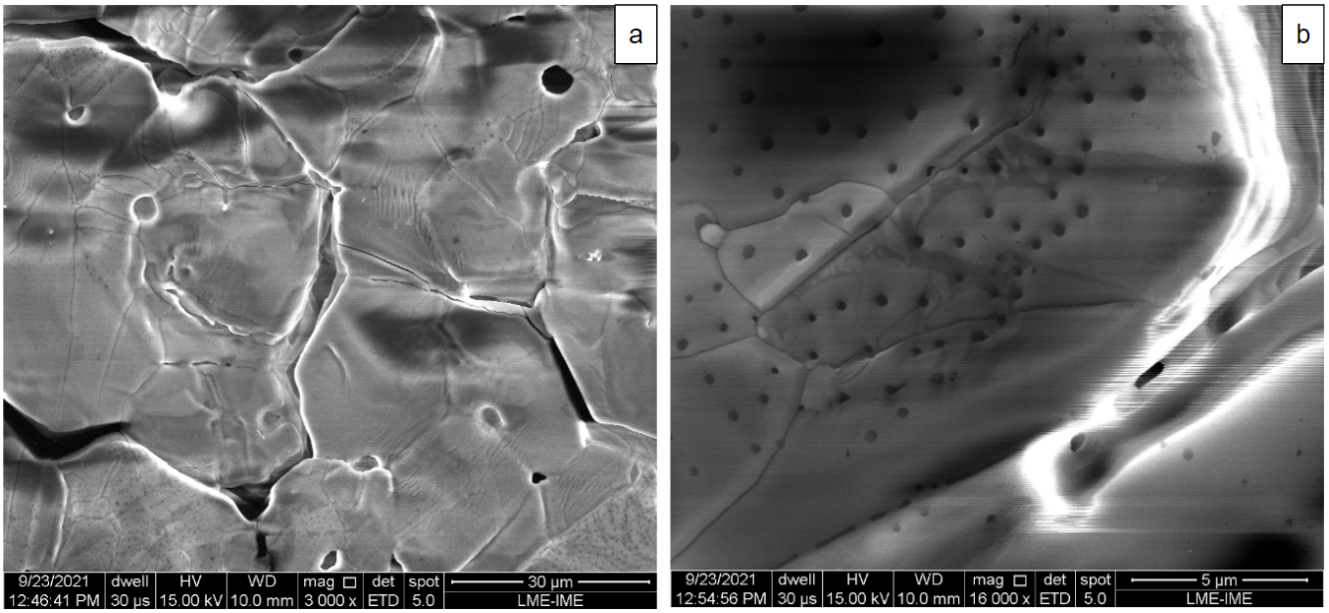


Figure 5 - Microstructure of the HA/glass sample sintered at 1350 °C.

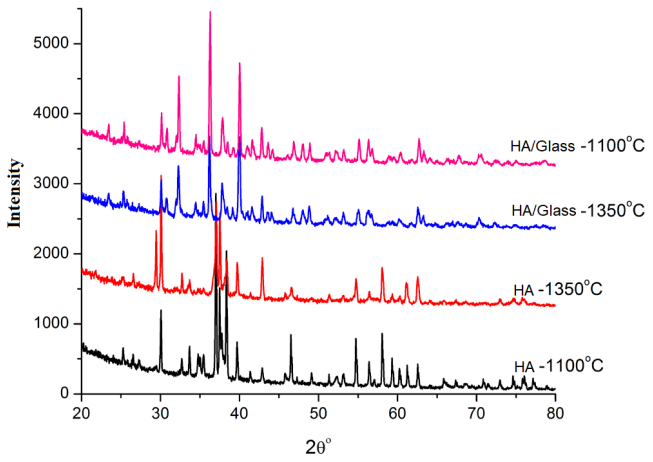


4.3 X-Ray Diffraction

The X-Ray Diffraction measurements resulted in the diffractograms of Figure 6. In the case of the pure HA sample sintered at 1100 °C, a majority hydroxyapatite phase and another unidentified minority phase were identified. In the pure HA sample sintered at 1350°C, the formation of a single crystalline phase of hydroxyapatite was observed (JCPDS 09-0432).

In both cases, the indexed HA phase has hexagonal symmetry and lattice parameters of a = b = 9.4180 Å and c = 6.8840 Å. This result is compatible with the hexagonal microstructure of Figure 4a.

Figure 6 - XRD diffractograms of sintered samples at 1100ºC and 1350ºC.



The HA/glass samples, on the other hand, presented the same crystalline structure at the two sintering temperatures, a pattern different from pure HA, being better indexed by the rhombohedral phase (JCPDS 09-0169) of the β-TCP, with lattice parameters of a = b = 10.4290 Å and c = 37.3800 Å. This indicates that HA has partially transformed into β-TCP (β-tricalcium phosphate $Ca_3(PO_4)_2$). Thus, the substitution of Ca^{2+} in the crystal lattice of the HA by Mg^{2+} favors the thermal decomposition of the HA into a biphasic mixture of HA and TCP [8, 9]. Magnesium also stabilized this transformation at both sintering temperatures.

It is noted that the lattice parameter c of the samples with glass addition is much higher than for the pure HA phase. This difference is consistent with the result of linear contraction, which may be the cause of the observed effect of greater retraction of pure HA samples in relation to that of the composite.

Conclusions

From these analyses, the following conclusions can be cited:

- The production of HA and glass based on calcium and magnesium phosphate under the use of precursors in liquid medium ensured the homogeneity of the compounds;
- The contraction of the surface area after sintering indicates the coalescence of the grains of the green body, elimination of pores and consequently densification. The pure HA samples had higher bulk density and lower porosity than the HA/glass composite at the same temperature, meaning that the addition of glass slows the densification of the HA;
- There was a growth, coalition of the grains and formation of grain contours in the sintering process at both temperatures. The HA/glass composite had higher grain growth, presence of pores and intergranular channels assimilate cracks;
- The microstructure of the HA samples at 1350ºC showed the grains in hexagonal shape, which was confirmed by XRD, with hexagonal crystal structure. While the HA/glass composite had rhombohedral crystal structure which indicates that HA partially transformed into β-TCP due to partial replacement of calcium with magnesium.

Acknowledgments

This work was carried out entirely on the premises of the IME. The authors would like to thank the funding agencies, CAPES and CNPq, for their financial support.

Authorship and Collaborations

All authors participated equivalently in the article preparation.

References

[1] Hench, L. L.; Polak, J. M. Third-generation biomedical materials. *Science*, [*s. l.*], v. 295, p. 1014– 1017, 2002.
[2] Jarcho, M. Calcium phosphate ceramics as hard tissue prosthetics. *Clinical Orthopaedics and Related Research*, [*s. l.*], v. 157, p. 259–78, 1981.
[3] Groot, K de. *Bioceramics of calcium phosphate*. Boca Raton: CRC Press, 1983.
[4] Viana, J. R. *et al*. Análise comparativa da síntese de hidroxiapatita via estado sólido. *Matéria*, Rio de Janeiro, v. 25, n. 1, 2020. DOI: https://doi.org/10.1590/S1517-707620200001.0914
[5] Sygnatowicz, M.; Keyshar, K.; Tiwari, A. Antimicrobial Properties of Silver-doped Hydroxyapatite nano-powders and Thin Films. *Biological Biomedical Materials*, [*s. l.*], v. 62, p. 65-70, 2010.
[6] Santos, M. V. B.; Osajima, J. A.; Silva-Filho, E. C. Hidroxiapatita: suporte para liberação de fármacos e propriedades antimicrobianas. *Cerâmica*, [*s. l.*], v. 62, p 256-265, 2016.
[7] Rootare, H. M.; Craig, R. G. Characterization of hydroxyapatite powders and compacts at room temperature and after sintering at 1200ºC. *Journal of Oral Rehabilitation*, [*s. l.*], v. 5, n. 3, p. 293–307, 1978.
[8] Cacciotti, I.; Bianco, A.; Lombardi, M.; Montanaro, L. Mg-substituted hydroxyapatite nanopowders: Synthesis, thermal stability and sintering behaviour. *Journal of the European Ceramic Society*, [*s. l.*], v. 29, n.14, p. 2969–2978, 2009.
[9] Gibson, I.R.; Bonfield, W. Preparation and characterization of magnesium/carbonate co-substituted hydroxyapatites. *Journal of Materials Science: Materials in Medicine*, [*s. l.*], v. 13, n. 7, p. 685–93, 2002.

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.9267.en

Recommendations for filling geospatial data abstracts

Carolina Coutinho Salustiano Silva[1], Ivanildo Barbosa[1]
[1]Military Institute of Engenharia, Praça General Tibúrcio, 80, 22290-270, Praia Vermelha, Rio de Janeiro, Brazil
carolsalustiano@gmail.com
ivanildo@ime.eb.br

ABSTRACT: As spatial data infrastructures evolves, geospatial data producers became able to provide them to a wide scope of potential users. This user, either human or a search engine, decides to adopt that dataset based on the analysis of the correspondent metadata by comparing data characteristics and their expectations. The Brazilian Geospatial Metadata Profile (in Portuguese, PMGB) offers guidelines to support agents in charge to fill geospatial metadata. However, they often meet the expectations of web search engines. This study aims to propose guidelines to fill the metadata element Abstract to make data more attractive to both human and machine users. This study created alternative versions of Abstracts of geospatial data available at the web based on search engines optimization techniques and the PMGB guidelines. A group of expert users assessed the alternatives by considering their preferences regarding their perception of gain of information between the proposed alternatives. In total, 84.6% of respondents approved the proposed guidelines for filling the Abstract metadata element.

***KEYWORDS**: Geospatial Metadata; Search Engine Optimization; Free-Text; Abstracts*

RESUMO: Com o avanço das Infraestruturas de Dados Espaciais, os produtores de dados geoespaciais podem disponibilizá-los para um amplo número de potenciais usuários. A decisão do usuário, humano ou motor de busca na Web, por acessar esses dados se baseia, prioritariamente, na análise dos seus metadados, ponderando as características do dado disponibilizado e as suas expectativas. As recomendações de preenchimento de metadados do Perfil de Metadados Geospaciais Brasileiro (PMGB) servem como referência aos agentes responsáveis pelo preenchimento de metadados, porém nem sempre atendem aos critérios adotados pelos motores de busca na Web. O objetivo deste trabalho é apresentar propostas de diretrizes de preenchimento do elemento de metadados Resumo, com o intuito de tornar a descrição do dado geoespacial mais atrativa para usuários humanos e máquinas. Neste trabalho, foram criadas versões alternativas de resumos de dados já disponibilizados na Web, aplicando técnicas de otimização para motores de busca e as recomendações de preenchimento indicadas no PMGB. Em seguida, um grupo de usuários avaliaram a sua percepção de ganho de informação dentre as opções apresentadas. Observou-se que 84,6% dos respondentes aprovaram a sistemática de preenchimento proposta.

***PALAVRAS-CHAVE**: Metadados Geoespaciais; Otimização de Motores de Busca; Texto-Livre; Resumos*

1. Introduction

The implementation of Spatial Data Infrastructures (SDI), regardless of their hierarchical level, allowed for the optimization of processes of dissemination and user access to Geospatial Data Sets (CDG) [1]. In this context, *Catalog Services for Web* (CSW) stand out for facilitating the search for CDG based on metadata elements such as title, abstract, and keywords.

Broadening the scope to data search through web search engines, it becomes imperative that one selects words that increase the chances a search engine will present the CDG as an answer to a related query.

Logical and semantic inconsistencies may arise during metadata filling, due to a mistaken understanding of the meaning of metadata elements, concomitantly with a lack of knowledge of the dataset to be documented [1].

An example of this misunderstanding is the description of data attributes rather than the information pertaining to the data itself. In addition, the author has the possibility of perceiving the relevance of abstract contents as subjective, whereas this is an important parameter for search engine indexing and classification.

Several CDGs arouse interest for academic research, governments, companies or related activities, and the importance of an adequate filling of metadata is evident, since inadequate filling prevents the search engine from correctly locating and indexing spatial data on the Web.

A ranking was elaborated through a study by Benjelloun *et al*. [2] to list the notability of metadata elements in scientific texts published on the internet.

According to the authors, the abstract influences 100% in the discovery of a data set, occupying the first position in the list of most relevant elements.

Consequently, it makes clearer the need for proper completion of the abstract. To mitigate the problem of metadata filling, the following steps are advisable: proper documentation, including the characteristics and information of geographical data and those made available on the Web by the producers; associated with knowledge on document and data indexing strategies, in addition to understanding of search engine operation.

This work aims to propose filling guidelines for the Abstract metadata element, in the context of an IDE, so to make the CDG description more attractive to human and machine users.

After this contextualization, Section 2 presents the concepts that support the guidelines herein. Section 3 describes the methodology used to verify the gains obtained with the existing recommendations and Section 4 presents and discusses the results obtained. The fifth section presents the final considerations of the study.

2. Conceptual review

2.1 Search Engine Optimization

Metadata search engines are typically implemented on Web catalog service platforms or other specific local repositories for spatial data. In these cases, textual search occurs only between repository records, including fields such as keywords, title, and abstract, considering that querying in web search engines has become almost instinctive, since not all geospatial data of interest are part of some IDE and its respective metadata catalog.

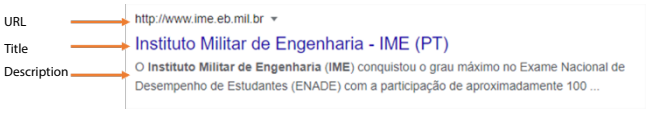
The operation of a search engine is divided into four main aspects and algorithms. First are the crawlers, which look for new content on the Web. Then comes indexing, which registers important information in the engine's search index, such as keywords. Third, a search engine that ranks and organizes hundreds of billions of web pages, analyzing factors such as query words and usability and knowledge of sources and settings. The weight applied to each factor varies according to the type of search. Fourth is the display of results, presenting the user with pages related to the search. Some preponderant factors help to determine the result, such as: understanding of natural language, which involves interpreting typos; finding commonalities frequently appearing in titles, headings, or text body; interpreting contexts, as the user's location, language preference, and search history, for example. [3] [4].

The search results that appear to the user are listed according to what is considered most relevant from the words used in the search for titles and abstracts. Inadequate filling of metadata can generate a non-coherent indexing of keywords in search engines, causing a difficulty of access.

In this context, techniques or tricks were identified to improve a system or Web page, to optimize its indexing by search engines. This process is known as "search engine optimization" (*SEO*) [5, 6, 7, 8, 9, 10].

Among the good optimization practices, known as *white hats*, we can mention the use of meta descriptors, used by search engines to display the text results of the second and third lines of search results, just below the title of the sites (Figure 1). Meta descriptors have a character limit, around 120 to 153, depending on the search engine used. They themselves do not increase the reputation of a page, but if users find what they are looking for in the text of the meta description, the chance of accessing the site increases [11].

Figure 1 - Example of a Google search using meta description *tags*



Source: prepared by the authors and adapted from Google.

The second best practice to highlight is to avoid identical or similar descriptions in all metadata

when individual pages appear in web results. This goes against the practice of creating long *templates* for dynamic content generation, reinforcing the need to emphasize the uniqueness of each CDG in the abstract.

The development of a fluid text is also recommended, to the detriment of long sequences of keywords. First, it makes the abstract human-readable; then, it allows natural language processing algorithms to process the context and eliminate any ambiguities.

Datasets are easier to find when they provide supporting information, such as name, description, creator, and distribution formats like structured data. Based on the approach that Google applies to the discovery of datasets, the use of 'schema.org' is recommended, as well as other metadata patterns that can be added to the pages describing the datasets.

The purpose of this information is to improve discovery of datasets from fields such as geosciences, life sciences, social sciences, machine learning, civic and government data, and more. Thus, abstracts are more attractive not only to human users, but also to search algorithms that refine search results based on the interpretation of the available text [12].

Finally, the first two sentences are often displayed in search engine results. Therefore, making them attractive, with relevant keywords, encourages people to click through to the page to read the content in its entirety. Ideally, one should try to repeat these keywords three to six times in the abstract, maintaining the readability of the text [13]. The naturalness of the textual production must be maintained, in addition to being composed of clear and concise central points, respecting the limit between 50,000 and 5,000 characters [12].

This function is intuitive for queries by text documents, image captions and video descriptions. In the case of CDG, the initiatives for indexing CSW services and map servers aim to make such content visible to search engine indexers. An example is the implementation of the *Geosearch* module on *Geoserver* map servers [14]. Another solution is the generation of web pages with content based on the title, abstract, and keyword elements, for publication and indexing by search engines.

2.2 Geospatial metadata

The most simplistic concept of metadata is that it consists of the description of a given data [15]. The purpose of metadata is to document and organize, in a systematic and structured way, the data of organizations to facilitate their sharing and maintenance, discipline the production and storage of data, and guide the use of data in different applications.

The concept extends to bibliographic and objects cataloging in digital format. Different metadata schemes have been proposed to suit the characteristics of the described objects, such as the *International Standard Bibliographic Description* (ISBD) [16] and the *Dublin Core Metadata Initiative* (DCMI) [17].

In the context of geospatial data, different metadata profiles have been proposed: *Content Standard for Digital Geospatial Metadata* (CSDGM) [18], ISO 19115 [19], and several national profiles drawn from the latter. The Brazilian Geospatial Metadata Profile (MGB Profile) [20] is the standard adopted in Brazil, based on [19], and recently adapted to suit the updates made by ISO 19115 in 2014 [21].

Each profile presents dozens of metadata elements that aim to describe technical, legal, and identification aspects, so that users can query the repository content to discover the CDGs related to their expectations. Some of these elements have values that can be automatically filled — the spatial extension and the reference coordinate system of a vector file, for example. A second set of elements have a controlled domain, i.e., they can only be filled with values defined in a pre-established list. Finally, another set of elements are filled in free text format, which may give rise to subjectivity on the part of the agent who fills in the metadata.

In section 2.1, two metadata elements with great influence on the search for resources on the Web were mentioned: keywords and abstract.

The keyword element is designated to describe a feature, its aspect, or the source. The selection of

terms can be facilitated by using the controlled list, MD_TopicCategoryCode [19, 20], which has a category code, as they contain defined themes and taxonomy. Another way to favor the choice of keywords is the use of a lexicon or *thesaurus* [20].

The Abstract element is defined as a brief summary of the resource [19, 20, 21, 22] and "must synthesize the fundamental aspects of the resource in terms of content, geographical extent, date, scale, series name, producer or responsible entity, sources used, etc." [20]. The subjectivity inherent in the term "brief" can induce overly simplified filling, failing to register content that can be found in keyword searches. On the other hand, long-winded summaries may omit relevant information to the detriment of aspects that do not add value to the search, in addition to unnecessarily taking up more storage space. Below, respectively, are examples of overly simplified and wordy summaries:

• The Digital Elevation Model, which is part of the RJ-25 project, represents the numerical model of the surface altimetric characteristics, articulated by sheets according to the framework of the Brazilian systematic mapping. It covers a geographical square of 07'30' 'latitude by 07'30" longitude [23];

• "URBAN WATER SERVICE INDEX – IN023. Indicator of the National Sanitation Information System SNIS. Calculation formula. IN023 = AG026_R / G06a * 100. Percentage unit. Reference Year 2011". [24].

"Several states today monitor the quality of surface water in their territory and pass on the data to ANA. However, from a national perspective, it is not always possible to compare the data generated, since the states adopt different criteria, methodologies and parameters, and there is no standardization on a Country scale. The National Water Quality Monitoring Network (RNQA) is the main component of the National Water Quality Assessment Program (PNQA), and its main objective is to standardize and expand monitoring in the country, eliminating existing temporal and geographical gaps. The points of RNQA were determined based on a point allocation methodology developed by ANA and were later analyzed together with all states and the Federal District to seek to take advantage of the monitoring points of existing networks. In addition, the ANA is responsible for the operation of the National Hydrometereological Network, which contains fluviometric stations and generates river flow information throughout the country. In part of these stations, approximately 1600, there is also the monitoring of four water quality parameters measured with multi-parametric probes (Dissolved Oxygen, Turbidity, Temperature and pH)" [25].

2.3 Related works

Studies were developed describing the SEO line.

Cahill and Chalut [5] examined techniques used by marketing for optimization: the different optimization tactics between the terms "white hat" and "black hat" were observed, and why it was important for librarians to understand these techniques and the impact on search engine results pages. They also looked at ways library staff could help their users develop awareness of the factors influencing search results and better assess quality and relevance on the results page.

Shih, Chen, and Chen [6] developed a search engine optimization that could be used by a company. Social networking sites were included in the Internet marketing strategy. The proposed technique was applied in the operations of an online e-bookstore. Website rankings were monitored in two search engines: Google and Yahoo. The results revealed that a well-designed SEO, with the incorporation of social networks, can effectively increase website visibility and exposure.

Zilincan [7] looked at the most important factors that can help improve placement in search results. He points out that no technique can guarantee high ranking, because search engines have sophisticated algorithms that measure the quality of web pages and derive their position in search results. Zilincan also developed a website for the purpose of implementing and testing key SEO techniques. Then, the relevant optimization factors that influenced the search engine increased the ranking of your site, in addition to subsequently verifying higher traffic.

Katumba and Coetzee [10] identified and categorized the search terms typically employed by users when searching for geospatial resources on the Web. Guided by these terms, metadata on geospatial sources was published "directly" on the Web and empirical tests were performed with search engine optimization (SEO) techniques. Two sets of HTML pages were prepared and registered in Google and Bing, respectively. The metadata in one set was tagged with Dublin Core, the other with Schema.org.

3. Methodology

The methodology used was divided into four parts: extraction of metadata from the INDE repository, analysis of the filling structure, analysis of the filling structure used in the INDE, and compilation of recommendations. The validation of results obtained consisted of the evaluation performed by users, choosing the option with the highest semantic representativeness.

3.1 Extraction of metadata

The process of extracting metadata from the INDE catalog was based on the script developed and documented in [26]. The identifiers, titles, and abstracts of 5,808 metadata stored in the INDE repository on 11/05/2020 were extracted. CSW services allow the preparation of HTTP requests, receiving responses in XML format, which can be interpreted and stored in a structured way. In this work, data were stored in comma-separated value format.

3.2 Analysis of the filling structure

The MGB Profile is the Brazilian normative reference for filling in geospatial metadata. As mentioned in Section 2.3, it specifies the information on some aspects of the data that should be included in the summary. In addition, this Profile, as well as dozens of other initiatives around the world, is based on the specifications of ISO 19115 [19]. The same occurs in international standards such as INSPIRE [27, 28] and IDE Espanha [29], which were chosen by this study, due to greater maturity and, consequently, availability of documentation with a higher level of detail and greater adherence by member countries.

3.3 Analysis of the filling structure used in INDE

This step aims to understand what content the agents responsible for filling in geospatial metadata use in the preparation of summaries.

One hundred and four (104) metadata were selected from the population of 5,808 metadata extracted from the catalog in Section 3.1. The selection took place based on the diversity of types and themes of the data; varied producing institutions and, mainly, an attempt to avoid duplication of abstracts.

This sample size implies a 95% confidence level with a margin of error of 9.4%. However, it was observed that some producers followed filling templates for dozens of products, which could influence the statistics of model identification.

The standardization of metadata filling in the abstracts was analyzed based on the requirements suggested by the PMGB. The results of this stage were obtained from the presence of the following factors in abstracts: geographical area, date, scale, series name, producer, and sources used. Each aspect of this filling structure was analyzed and classified with the following criteria: a) fully meets the requirement, or b) partially meets the requirement, or c) does not meet or not found.

3.4 Compilation of recommendations

The recommendations were divided into two groups: in terms of form and content.

As for the form, the concatenation of the items for the preparation of the abstract text followed the SEO recommendations whenever possible, by using the most relevant keywords with a repetition of 3 to 6 times naturally in the abstract, including the main keywords in the first two sentences of the abstract, avoiding oblique and wordy texts and avoiding repetition of abstract templates for different geospatial data in an IDE.

As for the content, Table 1 compares the filling contents in the abstracts. Items presented in it are assumed to be the basis of the recommendations.

Table 1 - Incidence of requirements on published recommendations

Item	PMGB	INSPIRE	IDE Espanha
Geographical area	X	X	X
Date	X	X	
Scale	X	X	
Sources Used:	X	X	X
Work Importance			X
Key Attributes			X
Producer	X	X	
Legal references			X
Grade	X	X	

Source: Prepared by the authors.

Some adaptations were necessary to meet the recommendations. In short, the content of the prepared abstract should bring together nine elements:

• Theme – subject to which it refers, or seeks to develop, or the proposition to be addressed. Main theme of the CDG;

• Product – result of the survey, i.e., what was produced;

• Spatial framework – result of the survey, i.e., what was produced;

• Time frame – main locations according to the scale: less than 1:10,000,000 (country); less than 1:5,000,000 (states and capitals); less than 1,000,000 (cities with more than 1,000,000 inhabitants);

• Scale – denominator compatible with the dimension of the smallest detail representable in the data. Use the scale for vector CDGs. For matrix data, use the scale compatible with the smallest detail representable on the ground or the spatial resolution, expressed in meters;

• Series name (applicable to a series or data collection) – name of a cartographic series, sheet-by-sheet documents or fieldwork documents, for the constitution of a given resource. The name is applicable to a series or collection of data;

• Producer and/or entity responsible for the institution and/or responsible for the geospatial data;

• Sources used – origin of spatial data, for example aerophotogrammetric survey, charts, maps, images, mosaics, cartographic bases, etc.;

Contextualization (which helps to understand the data produced) – describes the purpose of the data, presents a context for the creation of the data.

3.5 Validatio

Validation aims to evaluate the gain obtained with the use of the proposed completion guidelines, i.e., the perception of completeness of the information contained in the abstract when compared to the original texts.

To this end, a questionnaire was developed according to the self-explanation model, in which the form is delivered to the respondents to be completed without the intervention of the researcher [19, 30].

The questions within each section were designed following complexity and respondents' reflection levels [30]: the initial questions in each section asked for quick answers, while the final ones were more reflective, complex, and abstract. The form was released during a scientific event and published through emails and publications in groups of geotechnologies users on *Facebook* and *Whatsapp*. Responses were received for approximately two months (between November 2020 and January 2021).

The form was developed in *Google Forms* and divided into two fundamental sections: characterization of respondents and validation of the abstract filling systematization.

Respondent characterization aimed to quantify the expertise of the respondents regarding academic training and time of experience in the use of geospatial data. The following are the main questions of the form for the characterization:

1. To which of the categories below do you belong?
• Undergraduate Student;
• Graduate Student;
• Technical Course Student;
• Professor/Researcher;
• Professional/Producer;
• Professional/User;
• Others.

2. How long have you been in contact with geospatial data, either in your training or professional experience (years)?
• 0 – 1;
• 2- 5;
• 6 – 10;
• More than 10.

In the second section, respondents compared the abstracts of three products originally made available in the INDE metadata catalog with other versions, written in accordance with existing guidelines in the PMGB (see section 2.2) and the ones compiled in this study. The aim of the analysis of responses obtained in the form is to verify whether the abstracts prepared based on the proposed structure are clearer and more representative, compared to the abstracts based on the PMGB guidelines or those originally filed in the INDE.

If the original abstract did not contain all the items necessary for the preparation of the alternative abstracts, it would be necessary to manually locate the complete metadata in the INDE repository to consult the content of other elements.

The abstracts were rewritten, keeping the original texts of the INDE and adding the missing information from the PMGB guidelines in one version, and adding information according to the recommendations suggested in this study in the other.

This measure was taken to minimize the influence of the form of writing from one author to another, so that the respondent could exclusively evaluate the filling structure of content presented in the abstract. The following are the main form questions for the abstract comparison section.

3. Which Abstract option do you consider the most representative and clear?
• "The Digital Elevation Model, which is part of the RJ-25 project, represents the numerical model of the surface altimetric characteristics, articulated by sheets according to the framework of the Brazilian systematic mapping. It covers a geographical square of 07'30' 'latitude by 07'30" longitude [23];
• "The Digital Elevation Model, which is part of the RJ-25 project, represents the numerical model of the surface altimetric characteristics, articulated by sheets according to the framework of the Brazilian systematic mapping. It covers a geographical square of 07'30' 'latitude by 07'30" longitude in the locality of Folha de Guaçuí. This feature was created in 2008/08/08 with a scale of 1:25,000 and the name of the series is 1:25,000 Digital Elevation Model, developed at IBGE by the Cartography Coordination. Aerial photographs obtained from an aerial survey carried out by the company Base Aerofotogrametria e Projetos S.A. were used as a data source;
• The Digital Elevation Model, which is part of the RJ-25 project, covers a geographical square of 07'30' 'latitude by 07'30" longitude, covering the area of Folha Guaçuí. This feature was created in 2008/08/08 with a scale of 1:25,000 and the name of the series is 1:25,000 Digital Elevation Model, developed at IBGE by the Cartography Coordination. Aerial photographs obtained from an aerial survey carried out by the Base Aerofotogrametria e Projetos S.A. company were used as a data source. This survey aims to represent, through a numerical model, the altimetric characteristics of the surface.

4. Which Abstract option do you consider the most representative and clear?
• Several states today monitor the quality of surface water in their territory and pass on the data to ANA. However, from a national perspective, it is not always possible to compare the data generated, since the states adopt different criteria, methodologies and parameters, and there is no standardization on a Country scale. The National Water Quality Monitoring Network (RNQA) is the main component of the National Water Quality Assessment Program (PNQA), and its main objective is to standardize and expand monitoring in the country, eliminating existing temporal

and geographical gaps. The points of RNQA were determined based on a point allocation methodology developed by ANA and were later analyzed together with all states and the Federal District to seek to take advantage of the monitoring points of existing networks. In addition, the ANA is responsible for the operation of the National Hydrometereological Network, which contains fluviometric stations and generates river flow information throughout the country. In part of these stations, approximately 1600, there is also the monitoring of four water quality parameters measured with multi-parametric probes (Dissolved Oxygen, Turbidity, Temperature, and pH)

• Several states today monitor the quality of surface water in their territory and pass on the data to ANA. However, from a national perspective, it is not always possible to compare the data generated, since the states adopt different criteria, methodologies and parameters, and there is no standardization on a Country scale. The National Water Quality Monitoring Network (RNQA) is the main component of the National Water Quality Assessment Program (PNQA), and its main objective is to standardize and expand monitoring in the country, eliminating existing temporal and geographical gaps. The points of RNQA were determined based on a point allocation methodology developed by ANA and were later analyzed together with all states and the Federal District to seek to take advantage of the monitoring points of existing networks. In addition, the ANA is responsible for the operation of the National Hydrometereological Network, which contains fluviometric stations and generates river flow information throughout the country. In part of these stations, approximately 1600, there is also the monitoring of four water quality parameters measured with multi-parametric probes (Dissolved Oxygen, Turbidity, Temperature and pH). These data comprise the Brazilian territory. The date identifies when the appeal was issued on 2016-03-22, with the scale 1:1000000, developed by the National Water Agency;

• The points of the National Water Quality Monitoring Network (RNQA) were determined based on a point allocation methodology developed by ANA and were later analyzed together with all states and the Federal District to seek to take advantage of the monitoring points of existing networks. The appeal was issued on 2016-03-22, with the scale 1:1000000, developed by the National Water Agency. This survey aims to provide geospatial information on the panorama of water resources in Brazil.

5. Which Abstract option do you consider the most representative and clear?

• URBAN WATER SERVICE INDEX – IN023. Indicator of the National Sanitation Information System SNIS. Calculation formula. IN023 = AG026_R / G06a * 100. Percentage unit. Reference Year 2011. The research covered 4864 municipalities, making up 87% of the possible sample universe. The scale of these data is 1:2500000, developed by the Ministry of Cities. This survey aimed to generate information necessary for actions related to future land use;

• URBAN WATER SERVICE INDEX – IN023. Indicator of the National Sanitation Information System SNIS. Calculation formula. IN023 = AG026_R / G06a * 100. Percentage unit. Reference Year 2011. The research covered 4864 municipalities, making up 87% of the possible sample universe. The scale of these data is 1:2500000, developed by the Ministry of Cities.

• URBAN WATER SERVICE INDEX – IN023. Indicator of the National Sanitation Information System SNIS. Calculation formula. IN023 = AG026_R / G06a * 100. Percentage unit. Reference year 2011.

6. What items do you think are important in the abstract? (There is no limit of options for the answer)

• Theme;
• Spatial framework;
• Time frame;
• Date;
• Series name;
• Scale;
• Product;
• Producer and responsible entity;
• Sources used;
• Context;
• Status;
• Reference system;

- Distribution format;
- Type of representation;
- Language.

The respondent was encouraged to contribute to the construction of abstract recommendations, opening the opportunity for interaction based on a subjective question, which allowed the inclusion or removal of information different from those mentioned in previous questions.

7. Would you like to add any additional information regarding the completion of the Abstract?

To quantify the users' acceptance of the proposal to systematize the filling structures, a scoring metric was established: each question or alternative accepted by the user was assigned a value of 1 (one) point; in the answers that generated doubts about the user's opinion, the value of 0.5 (half) point was assigned; finally, 0 (zero) for those that were not accepted by the respondents. The validated abstracts were considered and the systematization of the abstract filling by the users was approved, when the score was equal to or greater than 70%, this percentage represents how much the proposals made in this study were accepted by the respondents.

Complementing the validation of the proposed recommendations, the respondent was also consulted on the relevant information to include in an abstract. No option limits were established to be chosen. The nine recommended items were presented as options, as well as five items unrelated to the recommendation: status, reference system, distribution format, type of representation, and language. These items were selected among elements of the summarized MGB profile [13].

The metadata element Data is present in Table 1 and, despite its temporal character, has a more limited meaning than the time frame, designed to characterize periods represented in the geographical data instead of a single reference date. The question was presented so that respondents did not know how to distinguish the recommended items from the additional items.

4. RESULTS OBTAINED

From the main recommendations for the Abstract (Table 1) in the PMGB, the following requirements were extracted: geographical area, date, scale, series name, producer, and sources used. This information was evaluated and it was verified if these elements appeared in the abstracts, presented in Table 2.

Table 2 - Incidence in percentage of the requirements for summary in the PMGB in the INDE metadata analyzed

Element	Requirement	INDE
Abstract	Geographical area	49%
	Date	22%
	Scale	33%
	Series name	17%
	Producer	23%
	Sources Used	31%

Source: Prepared by the authors.

One can see that geographical area, scale, and sources used are the most found, however with percentages below 50% of incidence. Based on this result, it is worth mentioning the importance of a greater dissemination of good metadata filling practices among geospatial data producers.

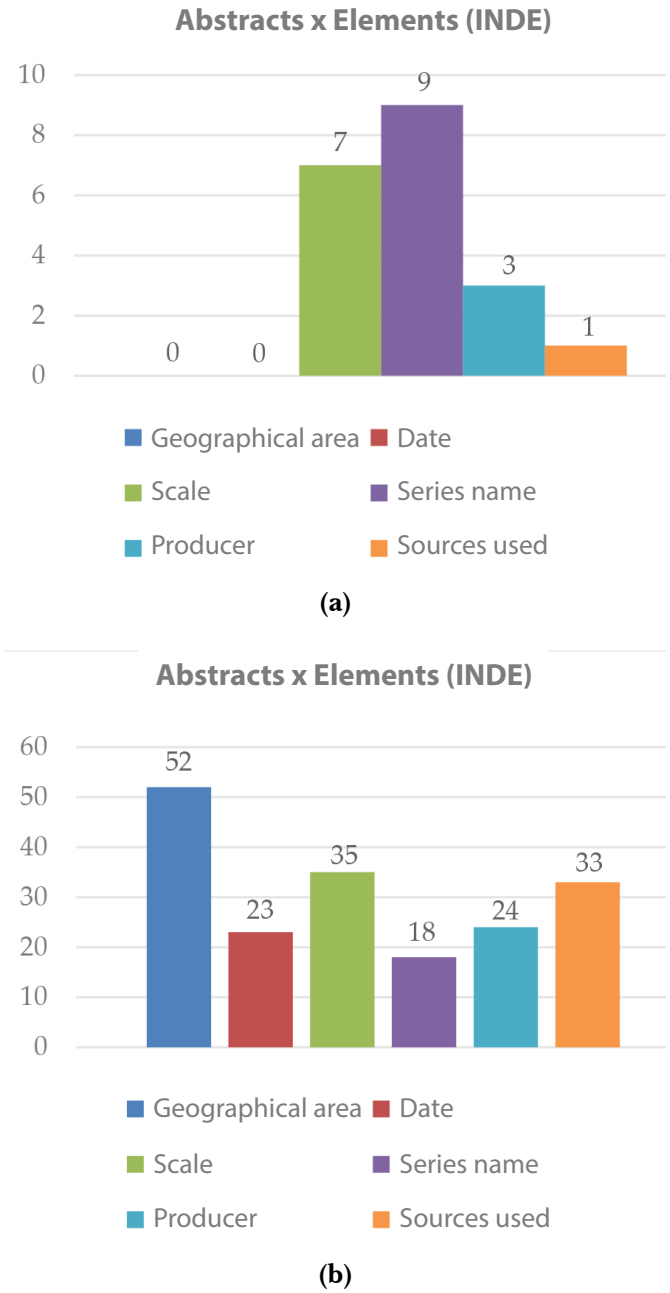
The analysis of abstracts of the metadata sample analyzed indicates that not all recommended items were met (Figure 2), but there were also unsuggested items such as the periodicity of data updating, the calculations and descriptions of the methodology used in the data attributes, the names of projects mentioned instead of the series name, in addition to explaining the operation of these projects, among others.

Seventy-five responses to the questionnaire were obtained. Despite being a quantity that makes the representativeness of the results obtained questionable, the profile of the respondents is composed of 66% of teachers, researchers and professionals, and about 60% of people with 10 or more years of experience with geospatial data.

The respondents were 29 teachers/researchers, 16 graduate students, 13 professionals/producers, 9 undergraduate students, 7 professionals/users, 1 high school teacher, and no technical course students.

Regarding the time of contact with geospatial data by training or professional experience, 44 reported having more than 10 years, 18 reported having between 6 and 10 years, 11 reported having between 2 and 5 years and 2 reported having between 0 and 1 year.

Figure 2 - Comparison between the items suggested by the PMGB in the abstracts referring to (a) systematic and (b) thematic data



Source: Prepared by the authors.

The characterization of respondents included a group with different levels of training and experience. However, most were research professors, with a level of experience of more than 10 years. The qualification of respondents and the time of interaction with geospatial data demonstrate reliability in the responses received.

Results of the comparison between the evaluated abstracts is shown in Table 1. Each row corresponds to the respective abstract, and the columns indicate the number of times each alternative was selected. One can observe that the abstract prepared following the recommendations was selected more frequently, in all cases. However, the abstract prepared only with the recommendations of the MGB profile already shows a significant gain in two of the three abstracts evaluated. The third summary stands out from the other two by originally describing the context in detail, being the only one to have alternatives of reduced size.

Chart 1 - Summary of responses regarding users' preference for the abstracts presented

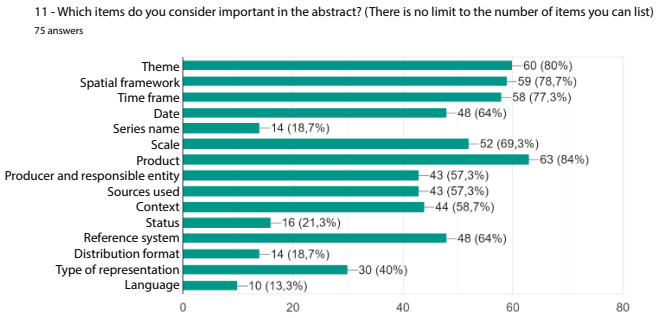
Responses / Abstract	INDE (responses)	PMGB (responses)	Alternative (responses)
1	15	29	31
2	11	30	34
3	19	17	39

Source: Prepared by the authors.

According to the score established to quantify the acceptance of the proposals, abstracts 1 and 2 received 0.5 points, as the difference between the values of the second and third columns are almost identical. However, in abstract 3, the alternative abstract based on the recommendations was selected as preferred by more than half of the respondents. This means assigning 1 point to this abstract.

The second stage of validation consisted of consulting the opinion of users regarding the items deemed important in the elaboration of a clear and comprehensive abstract. The graph illustrated in Figure 3 contains the compilation of the responses.

Figure 3 - Representation of the choice of the most important requirements for the abstract



Source: Prepared by the authors.

In the fourth quartile of responses, with the items most selected by the respondents, there are the items product, theme, spatial framework, and time frame (more than 77% of selections), all included in the recommendations. In the third quartile of responses, there are the items scale, date, and reference system (more than 59% of selections). With the exception of the reference system, these items are among the recommendations specified in the MGB Profile. It should be noted that the users treated the time frame item (proposed as the period represented in the geographic data) differently from the date (proposed as a metadata element indicated in Table 1, referring to a single date, which may be the date of publication of data or the oldest input used). The second quartile of responses (more than 31% of selections) includes the elements of producer, sources and contextualization (recommended in the PMGB specifications) and the representation format (not mentioned in those recommendations). Items such as series name, distribution format, status, and language were the least selected.

In analyzing this result, one may notice that not all recommended elements were the most voted by the respondents, so that the items Date and Reference System are among the most indicated items, to the detriment of the items Sources and Contextualization.

The validation of the abstracts accumulated 10 out of 13 points and an acceptance with the respondent of 84.6%.

5. Conclusions

The movement of geospatial open data has increasingly motivated data sharing. Therefore, the way metadata is described has become paramount, as it is directly related to the search engine's ability to locate the material made available on the internet.

Currently, the metadata filling instructions presented in the MGB Profile are not sufficient to efficiently describe the interesting characteristics of the product to its users. In addition, some factors are considered harmful in this process:–producers who use the PMGB recommendations based on their own criteria, i.e., subjectively; the recommendations need to be reviewed periodically to monitor technological developments, which are increasingly updated; the diverse environment of INDE, in which the multiplicity of consumers and data producers was observed, with different training and expertise; among other factors.

This study aims to propose guidelines for filling in the Abstract metadata element, in the context of an IDE, to make the description of the CDG more attractive to human and machine users.

To this end, the recommendations for filling in the abstracts in the PMGB were raised, criteria for analyzing the abstracts made available in the INDE catalog were established, and guidelines for filling in the abstracts based on SEO techniques in the PMGB and in the abstracts analyzed by INDE were developed. The analysis indicated the need for producers to insert in the abstract content that mentions theme, product, spatial framework, time frame, scale, series name, producer, sources used, and contextualization.

Considering that search engines are in constant change, in which they evoke a continuous refinement for SEO techniques, the main focus was directed at the quality of the content useful to the audience to be reached. In concrete terms, sharing information relevant to one who generated the data and one who will use it. In the responses obtained from the consumers and producers consulted, there was a gain in representativeness in the proposed alternative abstracts, i.e., the respondents validated the proposed recommendations with the preference of the product and theme requirements.

Based on this knowledge, metadata filling guidelines were developed for geospatial data summaries with a simple and natural language for users and aimed at meeting the demands of search engines. Such recommendations can be employed in developing solutions for suggesting abstracts for geospatial resources published in an IDE or in a catalog of geospatial products on the Web.

Finally, it should be noted that this topic requires constant updating, since the volume of data made available increases daily and the profile of producing and consuming agents changes continuously. Some improvements in the method presented in this work include the use of complementary analysis techniques, as well as the experimentation of other databases, in the deepening and addition of geospatial metadata elements, in the development of natural language processing techniques, machine readability compared to natural language, in the automation of abstracts, among others.

References

[1] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Acesso e uso de dados geoespaciais*, *Manuais técnicos em Geociências*. n. 14. Rio de Janeiro: IBGE, 2019. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101675.pdf Acesso em: 20 out. 2021.

[2] BENJELLOUN, O.; CHEN, S.; NOY, N. Google Dataset Search by the Numbers. *arXiv, Cornell* University, Nova York, 2020. DOI: https://doi.org/10.48550/arXiv.2006.06894

[3] CENDÓN, B. V. Ferramentas de busca na Web. *Ciência da Informação*, [*s. l.*], v. 30, n. 1, p. 39-49, 2001.

[4] GOOGLE. Home Search Work. *Google*, [*s. l.*], 2019. Disponível em: https://www.google.com/search/howsearchworks/crawling-indexing/ Acesso em: 18 dez. 2019

[5] CAHILL, K.; CHALUT, R. Optimal results: what libraries need to know about google and search engine optimization. *The Reference Librarian*, [*s. l.*], v. 50, n. 3, p. 234–247, 2009. DOI: https://doi.org/10.1080/02763870902961969

[6] SHIH, B.-Y.; CHEN, C.-Y.; CHEN, Z.-S. Retracted: An empirical study of an Internet Marketing Strategy for Search Engine Optimization. *Human Factors and Ergonomics in Manufacturing & Service Industries*, [*s. l.*], v. 23, n. 6, p. 528–540, 2012. DOI: https://doi.org/10.1002/hfm.20348

[7] ZILINCAN, J. Search engine optimization. *CBU International Conference Proceedings*, [*s. l.*], v. 3, p. 506–510, 2015. DOI: https://doi.org/10.12955/cbup.v3.645

[8] TAYLOR & FRANCIS. *Writing your paper*. Milton Park: Taylor & Francis, 2021. Disponível em: https://authorservices.taylorandfrancis.com/resources/writing-paper-ebook/. Acesso em: 2 mar. 2021.

[9] GABRIEL, M. *Marketing na era digital*: conceitos, plataformas e estratégias. São Paulo: Novatec, 2010.

[10] Katumba, S.; Coetzee, S. Empregando técnicas de otimização de mecanismos de busca (SEO) para melhorar a descoberta de recursos geoespaciais na Web. *ISPRS International Journal of Geo-Information*, [*s. l.*],v 6, n. 9, p. 284, 2017. DOI: https://doi.org/10.3390/ijgi6090284

[11] ROCK CONTENT. SEO 2.0: o guia definitivo e atualizado para conquistar a primeira página do Google. Ebook. Disponível em: https://rockcontent.com/materiais-educativos/seo-o-guia-definitivo-da-rock-content/. Acesso em: 10 ago. 2019.

[12] CENTRAL DA PESQUISA GOOGLE. Conjunto de dados. *Google*, [*s. l.*], 2021 Disponível em: https://developers.google.com/search/docs/advanced/structured-data/dataset. Acesso em: 27 abr. 2021.

[13] TAYLOR & FRANCIS. *Writing your paper*. Abingdon: Taylor & Francis, 2021. Disponível em: https://authorservices.taylorandfrancis.com/resources/writing-paper-ebook/. Acesso em: 27 abr. 2021.

[14] OSGeo – OPEN SOURCE GEOSPATIAL FOUNDATION. Layer. 2024. Disponível em: https://docs.geoserver.org/stable/en/user/data/webadmin/layers.html. Acesso em: 30 ago. 2024.

[15] PRADO, B. R.; HAYAKAWA, E. H.; BERTANI, T. C.; SILVA, G. B. S.; PEREIRA, G.; SHIMABUKURO, Y. E. Padrões para metadados geográficos digitais: modelo ISO 19115:2003 e modelo FGDC. *Revista Brasileira de Cartografia*, Uberlândia, v. 62, n. 1, p. 33-41, 2010. DOI: https://doi.org/10.14393/rbcv62n1-43665

[16] IFLA – International Federation of Library Associations and Institutions. ISBD International Standard Bibliographic Description. *IFLA*, [*s. l.*], 2011. Disponível em: https://repository.ifla.org/bitstream/123456789/786/1/ifla-isbd-international-standard-bibliographic-description-2011.pdf. Acesso em: 14 out. 2019.

[17] DCMI – Dublin Core Metadata Initiative. 2022. Disponível em: https://dublincore.org/specifications/dublin-core/. Acesso em: 16 mar. 2022.

[18] FGCD – Federal Geographic Data Committee. Content Standard for Digital Geospatial Metadata. *FGCD*, Washington DC, 1998. Disponível em: https://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdf. Acesso em: 30 jun. 2019.
[19] ISO 19115. Geographic Information – Metadata. 2003.
[20] CONCAR – Comissão Nacional de Cartografia. *Perfil de Metadados Geoespaciais do Brasil*. Perfil MGB. Brasília, DF: Ministério do Planejamento, 2009.
[21] ISO 19115. Geographic Information -Metadata. 2014. Disponível em: https://www.iso.org/standard/53798.html. Acesso em: 23 mai. 2019.
[22] DCMI – Dublin Core Metadata Initiative. 2022. Disponível em: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/abstract/https://www.dublincore.org/specifications/dublin-core/dcmi-terms/terms/abstract/. Acesso em: 12 jul. 2022.
[23] IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA *Modelo Digital de Elevação 1:25.000 - GUAÇUÍ SF-24-V-A-IV-4-NO 2613-4-NO*. 2008. Disponível em: https://metadados.inde.gov.br/geonetwork/srv/por/catalog.search#/metadata/ea5778a6-e4db-495a-9b1c-f7a22976921f. Acesso em: 30 ago. 2024.
[24] MINISTÉRIO DO PLANEJAMENTO, ORÇAMENTO E GESTÃO. *Índice de atendimento urbano de água com rede de abastecimento - Valor realizado*. 2011. Disponível em: https://metadados.inde.gov.br/geonetwork/srv/por/catalog.search#/metadata/e466791c-0a49-4a8f-b8fe-e31f8a16bf15. Acesso em: 30 ago. 2024.
[25] AGÊNCIA NACIONAL DE ÁGUAS. *Evolução da Rede de Monitoramento de Qualidade da Água*. 2015. Disponível em: https://metadados.inde.gov.br/geonetwork/srv/por/catalog.search#/metadata/5f5da94d-f61b-4706-857f-63f7152618eb. Acesso em: 30 ago. 2024.
[26] GOTTARDO, T. V. *Proposta de metodologia para avaliação de ide através de indicadores sobre a disseminação de dados especiais*. Rio de Janeiro: Instituto Militar de Engenharia, 2018.
[27] SILVA, H.; SERRONHA, A. *Metadados Inspire*. Portugal: Direção-Geral do Território, CCDR-LVT, CCDR Algarve, CCDR Norte, CCDR Centro, CCDR Alentejo, 2015.
[28] INSPIRE – Infrastructure for Spatial Information in Europe. *Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007*. [*S. l.*]: INSPIRE, 2017.
[29] GOBIERNO DE ESPAÑA, Catálogo de Datos y Servicios IDEE. Disponível em: https://www.idee.es/csw-inspire-idee/srv/por/catalog.search;jsessionid=9F3F01DDBFAD57D7FDB569FCD810EA47#/home. Acesso em: 30 ago. 2024.
[30] VIEIRA, S. *Como elaborar questionários*. São Paulo: Atlas, 2009.

VOL.40 Nº1 2023
https://doi.org/10.22491/rmct.v40i1.9273.en

Pralidoxime: a review of its synthesis and antidotal properties against warfare nerve agents

Pedro Augusto Gomes Buitrago*; Samir Frontino de Almeida Cavalcante[1]; Valdir Florêncio da Veiga Júnior[1].
[1]Instituto de Defesa Química, Biológica, Radiológica e Nuclear; Centro Tecnológico do Exército; Instituto Militar de Engenharia
Avenida das Américas, 28705, CEP 23020-470, Rio de Janeiro, RJ, Brasil
Praça General Tibúrcio, 80, CEP 222290-270, Rio de Janeiro, RJ, Brasil
***pedro@ime.eb.br**

ABSTRACT: *Acetylcholinesterase (AChE) is an enzyme in the central and peripheral nervous systems that has been studied in fields of research such as that for Alzheimer's and Parkinson's diseases. AChE inhibitors may be either natural or synthetic, which is the case of the organophosphorus compounds, developed as chemical weapons or pesticides, the latter of which is less toxic. The inhibition of organophosphorus is irreversible and is carried out by binding the phosphorus atom to the hydroxyl group of the serine residue within the active site of the AChE, thus preventing AChE from fulfilling its physiological task in cholinergic transmissions, possibly leading to respiratory failure and death. Due to their strong nucleophilic character, AChE reactivators can cleave the bond between the serine residue and the adduct, reestablishing enzymatic activity. This study describes the biological properties and the diverse synthetic methods for pralidoxime, the first AChE reactivator clinically applied.*

KEYWORDS: *Acetylcholinesterase. Antidote. Oxime. Organophosphorus. Pralidoxime. Quaternary reactivator.*

RESUMO: *A acetilcolinesterase (AChE), enzima presente nos sistemas nervosos central (SNC) e periférico, é estudada em pesquisas relacionadas à doença de Alzheimer e à doença de Parkinson. Os inibidores de AChE podem ser naturais ou sintéticos, como os organofosforados desenvolvidos para o uso como armas químicas ou pesticidas, sendo estes menos tóxicos. A inibição por organofosforados ocorre irreversivelmente através da formação de uma ligação entre o átomo de fósforo e a hidroxila do resíduo de serina presente no sítio ativo da AChE. Com isso, a AChE perde sua capacidade de cumprir sua função fisiológica nas transmissões colinérgicas, podendo levar à parada respiratória e morte. Os reativadores de AChE, devido ao seu forte caráter nucleofílico, conseguem romper a ligação entre o resíduo de serina e o aduto, restabelecendo a atividade enzimática. Este trabalho aborda diferentes metodologias sintéticas e propriedades biológicas da pralidoxima, o primeiro reativador de AChE empregado clinicamente.*

PALAVRAS-CHAVE: *Acetilcolinesterase. Antídoto. Oxima. Organofosforado. Pralidoxima. Reativador Quaternário.*

1. Introduction

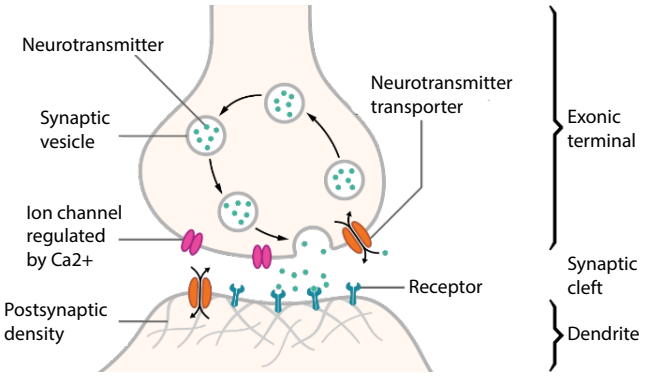
1.1 Structure and function of acetylcholinesterase and acetylcholine

Acetylcholinesterase (AChE; Enzyme Commission Number [EC] 3.1.1.7) is an enzyme of the hydrolase class, responsible for the hydrolysis of carboxylic acid esters [1], whose active site is composed of a catalytic triad containing serine, histidine, and glutamate residues, with the serine residue being responsible for the attack on the carboxylic ester [2]. Its action occurs in the central and peripheral nervous systems, as well as in the neuromuscular junctions where, together with muscarinic and nicotinic acetylcholine (ACh) receptors, AChE regulates the transmission of electrical impulses (action potentials) along neuromuscular synapses. The physiological function of AChE is the hydrolysis of the neurotransmitter ACh, which ends the action potentials generated by the stimulation of cholinergic receptors. The enzyme acts when ACh is released by the presynaptic neuron in response to an action potential, preventing the accumulation of the neurotransmitter in the synaptic cleft (Fig. 1) [3]–[5].

ACh is transported along the synapse and, when it binds to its receptors, it leads, among other responses, to an influx of K-ions^{+} in the postsynaptic nervous process or in a muscle cell. This process initiates

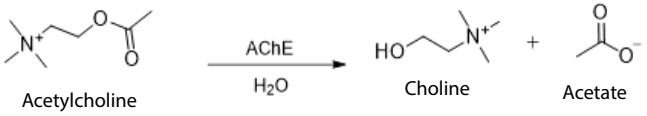
action potentials in the postsynaptic cell, which is quickly stopped by the action of AChE by hydrolyzing ACh into its breakdown products, choline and acetate (Fig. 2), which are used to regenerate ACh in the peripheral nerve [3, 4].

Fig. 1 - Structure of a synapse, in which the release and capture of a neurotransmitter occurs.



Source: [6]

Fig. 2 - Enzymatic hydrolysis of acetylcholine into its precursors: acetate and choline.



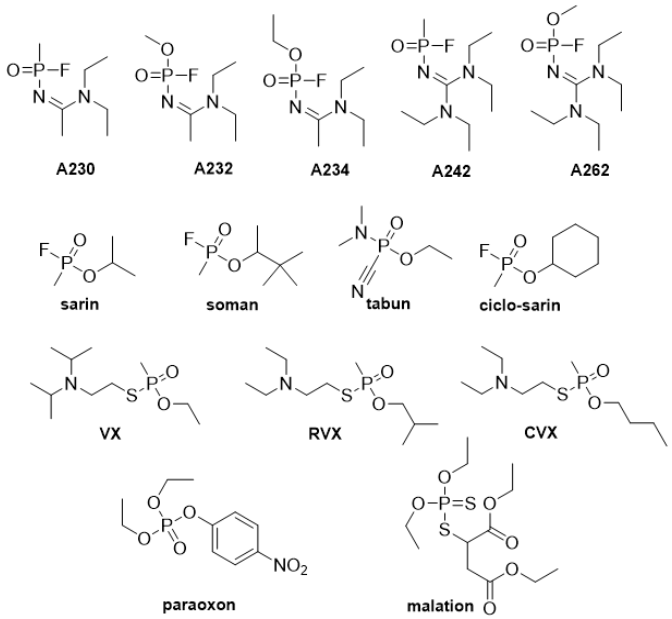
Source: prepared by the authors

1.2 Inhibition of acetylcholinesterase

Inhibition of AChE causes an accumulation of the neurotransmitter ACh in the postsynaptic cleft, leading to hyperstimulation of cholinergic receptors (muscarinic and nicotinic). This inhibition can be reversible and thus temporary, as is the case with therapies for the treatment of Alzheimer's disease (AD), in which reversible inhibitors such as donepezil, galantamine, and rivastigmine are employed, an approach known as the "cholinergic hypothesis" [7]mainly affecting older people. The unclear root cause and involvement of various enzymes in the pathological conditions confirm the complexity of the disease. Quantitative structure-activity relationship (QSAR. However, in the case of poisoning by pesticides (paraoxon, malathion; Fig. 3) or nerve agents, inhibition occurs irreversibly. Although pesticides are less toxic than neurotoxic agents, both poisonings can lead to malfunction of the central nervous system (CNS) and neuromuscular junctions, which might be lethal [8], [9].

Nerve agents are organophosphorus compounds that can be divided into three series: the oldest, known as the G series (sarin, soman, tabun, and cyclosarin, among others), which are volatile under normal temperature and pressure conditions; the V series (VX, RVX, and CVX, for instance) (Fig. 3), which are more persistent in the environment [10] e.g. in Alzheimer's disease, Parkinson's disease, or in eco-toxicology as a biological marker. Many inhibitors of AChE have been identified in nature as well as prepared in chemical labs as a result of systematic synthetic efforts. The organophosphorus (OP; and the A series, whose substances are known as Novichoks (Fig. 3) [11], which were added to Schedule 1 of the Chemical Weapons Convention (CWC) in June 2020 [12].

Fig. 3 - Structures of G, V, and A series nerve agents and pesticide examples

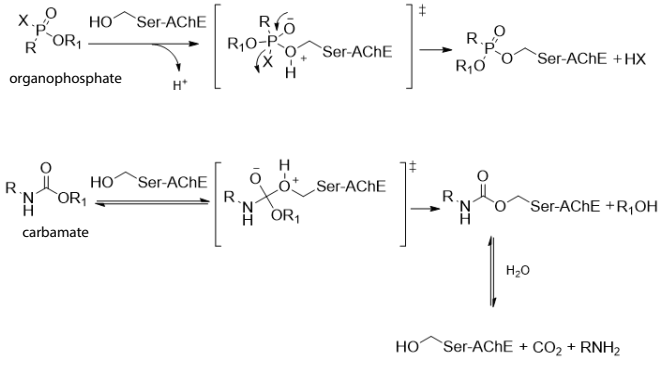


Source: prepared by the authors.

The inhibition of AChE occurs by the formation of a covalent bond between the organophosphate and the hydroxyl of the serine residue at the active site of the enzyme, which makes it impossible for the enzyme to interact with the ACh molecule (Fig. 4). The accumulation of this neurotransmitter results in hyperstimulation of cholinergic receptors, failure of cholinergic synaptic transmission, muscle paralysis, and CNS impairment. These effects constitute a "cholinergic crisis," characterized by symptoms such as miosis (pupil constriction), excessive salivation, bradycardia, diarrhea, emesis, and bronchoconstriction, caused by overstimulation of muscarinic receptors. They also lead to seizures, paralysis, and muscle dysfunction, resulting from the overstimulation of nicotinic receptors. The action on the neuromuscular junctions of the diaphragm's smooth muscle can lead to death by respiratory arrest [13], [14].

The AChE inhibition reaction follows an addition and elimination mechanism (Fig. 4), in which the hydroxyl of the AChE serine residue binds to the electrophilic site of the inhibitor, followed by the elimination of a leaving group. In the case of irreversible inhibitors, such as organophosphates, the leaving group can be, for example, a halogen (fluoride in the case of sarin and soman, Fig. 3) or a cyanide (tabun, Fig. 3). In the case of reversible inhibitors, such as carbamates, the leaving group is an alkoxide [8], [15], [16].

Fig. 4 - AChE reactions with irreversible and reversible inhibitors.



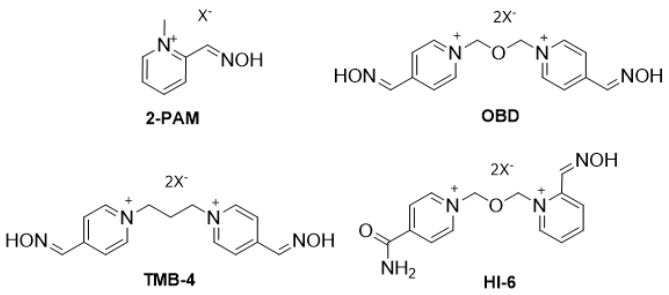
Source: prepared by the authors.

1.3 Reactivation of acetylcholinesterase

To avoid the lethality of organophosphate poisoning, medicines should be administered as soon as possible, especially in cases in which AChE may undergo the "aging" process, in which the antidotes available in the clinic lose their capacity to act, with soman being one of the agents most likely to cause this phenomenon. The rapid drug action also enables the reduction of risks related to neurological issues that impact victims of nerve agents [17], [18]. In addition to an antimuscarinic agent (atropine), which acts by antagonizing the effects of excess neurotransmitter in the synaptic cleft, and an anticonvulsant (diazepam), it is essential to administer an AChE reactivator antidote to reverse the effects of poisoning [19].

AChE reactivators must possess a strong nucleophilic character to break the strong P-O bond between the organophosphate and the serine residue of the AChE catalytic triad. This characteristic is found in cationic oximes derived from pyridine aldehydes, which are the only class of substances used clinically to treat poisoning by nerve agents. Among these, pralidoxime (2-PAM), obidoxime (OBD), trimedoxime (TMB-4), and asoxime (HI-6) are available on the market (Fig. 5) [20]–[23].

Fig. 5 - Clinical acetylcholinesterase reactivators (X^-: Cl^-, I^-, MsO^-).

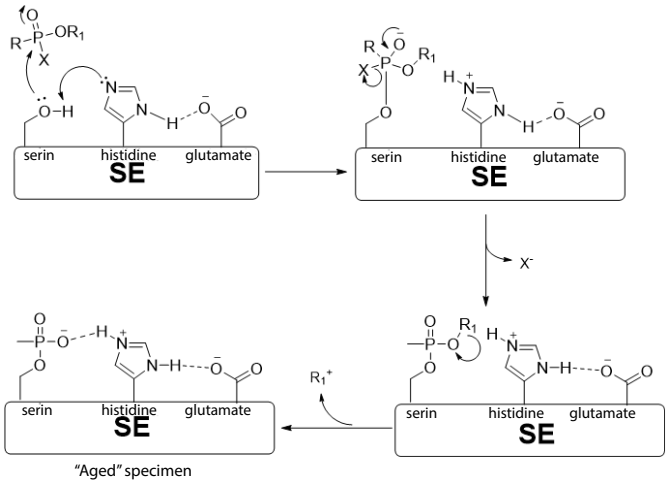


Source: prepared by the authors.

Although these oximes are used in clinical practice, they show limitations, including high toxicity, which restricts their dosage, and a limited spectrum of action against structurally distinct organophosphates, without broad-spectrum reactivation oxime

currently available [13]. Other limitations include their low penetration of the blood-brain barrier due to their cationic nature and their inability to reactivate the "aged" form of AChE, caused by the disproportion of the formed enzyme adduct. The "aging" process consists of the dealkylation of the phosphorus adduct of the inhibited AChE (Fig. 6). To improve the reactivation spectrum of current clinical oximes, one or more of them can be combined when AChE is inhibited by different nerve agents [24], [25].

Fig. 6 - Mechanism of inhibition of AChE by an organophosphate followed by its aging.



Source: prepared by the authors.

2. Pralidoxime – properties and synthesis

2.1 Biological properties

Synthesized in the United States in 1955 [20], 2-PAM was the first molecule capable of reactivating AChE inhibited by organophosphates to be used in clinical practice. As it is a cationic oxime, it is found in the form of a salt, and can be associated with chloride, iodide, methyl sulfate, or mesylate anions. It is used not only by the Brazilian Army, but also by the armies of the United States, France, and the United Kingdom, in addition to being listed in the Brazilian Ministry of Health's RENAME (Brazilian National List of Essential Medicines) [10], [26], [27]e.g. in Alzheimer's disease, Parkinson's disease, or in eco-toxicology as a biological marker. Many inhibitors of AChE have been identified in nature as well as prepared in chemical labs as a result of systematic synthetic efforts. The organophosphorus (OP.

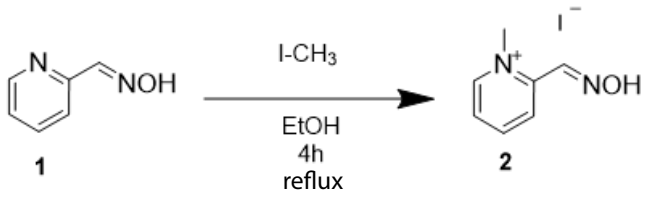
Pralidoxime has been found to demonstrate efficacy in reactivating sarin- or VX-inhibited AChE, especially when combined with atropine [28], [29], but not in reactivating the enzyme inhibited by tabun or soman [30], which reiterates the absence of a "universal antidote" [13]. Another limitation of this reactivator lies in its low rate of penetration in the blood-brain barrier due to the presence of a positively charged nitrogen. Sakurada and collaborators [31] determined that this rate is approximately 10% but later studies suggest that this value is overestimated [32]. This pharmacokinetic limitation is common to all AChE reactivators available in the clinic, which drives the search for new compounds that are increasingly active and efficient for CNS reactivation.

Administration of 2-PAM in humans at a dose of 10 mg/kg resulted in concentrations greater than 4 µg/ml in blood plasma in less than 10 minutes, which was maintained for the subsequent 50–55 minutes due to its high stability in water [33], [34]. The use of this reactivator can include side effects such as dizziness, blurred vision, diplopia (double vision), nausea, and headaches [33], [35].

2.2 Synthetic methodologies for pralidoxime

In 1956, Green and collaborators [36] described a synthesis of pralidoxime iodide (2), in which 2-pyridine aldoxime (1) was reacted with methyl iodide in ethanol under reflux for 4 hours (Fig. 7). The research group also presented a possible interaction between oximes and the nerve agent sarin. The yield of the reaction was not reported in the article.

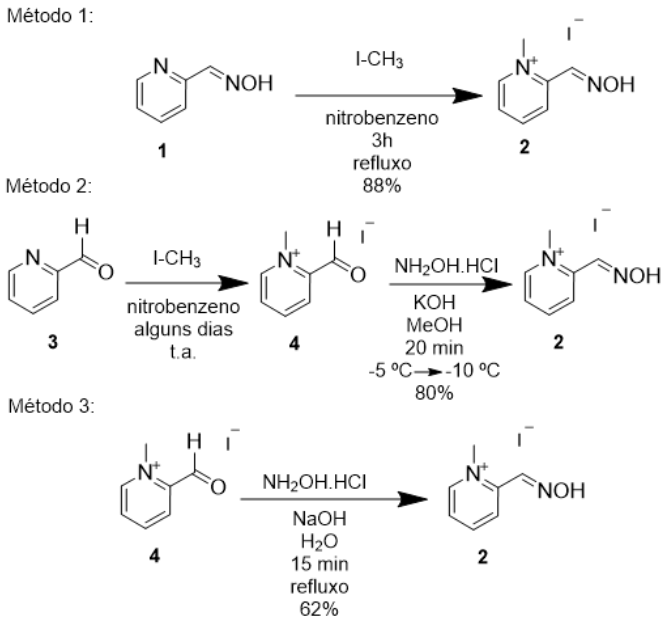
Fig. 7 - Synthesis of pralidoxime iodide proposed by Green.



Source: prepared by the authors.

In 1957, Ginsburg and collaborators [37] described the synthesis of pralidoxime and several other derivatives of it. For the synthesis of pralidoxime iodide, three different synthetic routes were presented (Fig.8).

Fig. 8 - The three different synthetic routes for pralidoxime iodide presented by Ginsburg.



Source: prepared by the authors.

In the first method, **1** reacts with excess methyl iodide in nitrobenzene for 3h with reflux, forming **2** with a yield of 88% (Fig.8) [37].

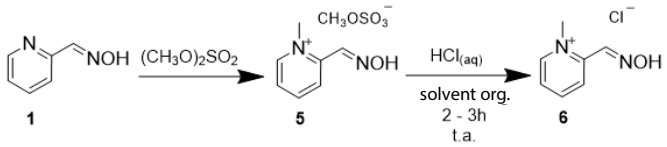
Then, in the second method, 2-pyridinecarboxaldehyde (3) was reacted with excess methyl iodide in nitrobenzene, stirred for a few days at room temperature The product obtained (2-formyl-1-methyl pyridinium iodide, **4**) was precipitated using acetone. In the next step, **4** was added to a methanolic solution of hydroxylamine hydrochloride and potassium hydroxide at −5°C, then stirred for 20 minutes as the temperature dropped to −10°C. The product was precipitated in diethyl ether, obtaining 80% yield (Fig. 8).

In the third method, **4** was reacted with excess hydroxylamine hydrochloride in water under reflux for 15 minutes, followed by pH adjustment to a range of 6–7 using a sodium hydroxide solution. The oxime was recrystallized by methanol or ethanol, reaching a yield of 62% (Fig.8).

In 1964, an innovative route was proposed by Bloch for the synthesis of pralidoxime chloride (**6**) [38]. Previously, **6** was synthesized from the reaction of a solution of **2** with solid silver chloride. Then the silver iodide formed was filtered and the aqueous solution evaporated to dryness at low temperatures. The disadvantage of this method was the residual traces of silver in the product, which were difficult to remove, as well as the use of an expensive reagent, silver chloride. Another method for converting pralidoxime iodide to chloride involved using an anion exchange resin; however, this method was expensive and impractical, as regenerating the iodide-saturated resin was difficult and, similar to the previous method, it required the evaporation of large volumes of water at low temperatures. Direct quaternization of **1** with methyl chloride in a pressure reactor was also performed but the yield was low and successive recrystallizations were required to achieve acceptable purity [38]. Aware of these disadvantages, Bloch proposed a synthetic methodology that involved an intermediate step, forming pralidoxime methyl sulfate (**5**) by reacting **1** with dimethyl sulfate. Molecule **5** was then converted to the chloride by reacting it with concentrated hydrochloric acid and a water-miscible organic solvent (Fig. 9). The solvents tested were isopropanol (85% yield), methanol (30%), absolute ethanol (70%), isobutanol (84%), propylene glycol (27%), dioxane (29%), and acetone (75%). Since **5** proved to be significantly more soluble than

6 in the solvents mentioned, the product could be easily separated by filtration at the end of the reaction and then washed with acetone, achieving a high degree of purity [38].

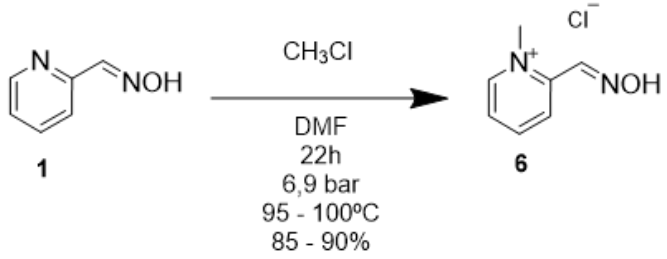
Fig. 9 - Synthesis of pralidoxime chloride proposed by Bloch.



Source: prepared by the authors.

In the work of Ellin and collaborators [39], molecule **6** was synthesized by reacting **1** with methyl chloride in *N*,*N*-dimethylformamide (DMF) (Fig. 10). At atmospheric pressure, both DMF and other solvents such as acetone, ethanol, tetrahydrofuran, and benzene showed low yields for the method presented. However, using DMF at pressures around 7 bar resulted in a yield approximately four times greater than those obtained with the other solvents. This increase is due to DMF being a polar aprotic solvent, and the reaction follows an SN_2 mechanism, which presents, as slow step, a dipolar transition state formation, whose energy is reduced by the solvation effect of DMF, leading to greater stability [39].

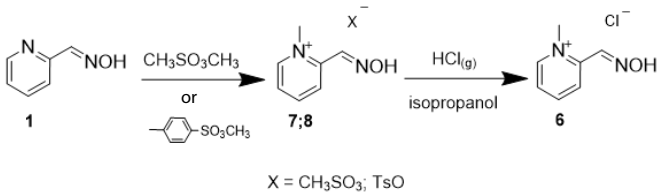
Fig. 10 - Synthesis of pralidoxime chloride proposed by Ellin.



Source: prepared by the authors.

In the study by Rao and colleagues [40], aimed at synthesizing pralidoxime chloride via the methylation of **1**, various methylating agents were tested, including methyl methanesulfonate and p-toluenesulfonate (Fig. 11). For methylation with methyl methanesulfonate, the tested solvents were toluene (70% yield), tert-butyl methyl ether (52%), dimethoxyethane (60%), acetonitrile (90%), and 1,4-dioxane (55%). For methylation with p-toluenesulfonate, the solvents tested were toluene (91%) and acetonitrile (70%). In all reactions, with the exception of methyl methanesulfonate methylation in acetonitrile, the crude product crystallized in an ethanol solution with ethyl acetate. In the final step, to convert to pralidoxime chloride, a solution of the methylation product in isopropanol was bubbled with anhydrous hydrogen chloride gas to form **6** (Fig. 11).

Fig. 11 - Synthesis of pralidoxime chloride proposed by Rao.



Source: prepared by the authors.

3. Discussion

The 2-PAM is part of a broad family of compounds with strong nucleophilic characteristics, capable of reactivating AChE inhibited by neurotoxic agents. From a structural standpoint, it is used as a quaternary ammonium salt and shows only one oxime group, whereas the other three available reactivators show two. The quaternization of oximes aims to increase their affinity for the anionic catalytic site of AChE, enhance water solubility, and adjust the pKa values (from 7.0 to 8.35) to facilitate the reactivation process [41]. A major limitation of cationic oximes is their poor penetration of the blood-brain barrier due to their low lipophilicity, which causes them to act predominantly in the peripheral nervous system [42]. However, 2-PAM showed a 10% penetration of the blood-brain barrier in rats, compared to 1 to 3% for the bipyridinium oximes [31],

[43]. Among the available antidotes, asoxime is the least toxic, followed by 2-PAM, whereas obidoxime and trimedoxime are the most toxic [44].

Although clinical antidotes are cationic and have one or two pyridinium rings, recent studies have explored AChE reactivators with different structural characteristics. In search of greater lipophilicity, neutral oximes and oxime derivatives containing nitrogen heterocycles in their structure have been evaluated as potential new classes of AChE reactivators [45]–[47]. Other approaches with the same objective include structural modification of cationic oximes by adding fluorine atoms to the pyridinium rings and transforming these oximes into prodrugs—drugs administered in an inactive form and activated by biotransformation within metabolic pathways of the body [48]. In addition to chemical approaches, research have been exploring new methods of administering antidotes, developing techniques that facilitate the entry of reactivators into the central nervous system, such as intranasal administration [49].

Regarding the synthesis methodologies presented, considering the number of steps and yield, the most efficient synthetic methods were those reported by Ginsburg (one step; 88% yield) and Ellin (one step; 85–90% yield).

4. Conclusion

Several synthetic methodologies for 2-PAM have been described, making it one of the most relevant pyridinium oximes available for combating poisoning by nerve agents. We searched for articles published since 1955, the year 2-PAM was first reported. In the discussion, the synthetic routes were compared in terms of efficiency, considering the number of steps, yield, and feasibility of the syntheses. A comparison was also made between the properties and limitations of 2-PAM relative to other AChE reactivators. Approaches were also cited in the search for new classes of AChE reactivators.

List of abbreviations and acronyms

2-PAM	=	Pralidoxime
ACh	=	Acetylcholine
AChE	=	Acetylcholinesterase
CWC	=	Chemical Weapons Convention
AD	=	Alzheimer's disease
DMF	=	*N*,*N*-dimethylformamide
EC	=	Enzyme Commission Number
HI-6	=	Asoxime
OBD	=	Obidoxime
RENAME	=	Brazilian National List of Essential Medicines
SNC	=	Central Nervous System
TMB-4	=	Trimedoxime

References

[1] SCHOMBURG, I.; CHANG, A.; SCHOMBURG, D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, Oxford, v. 30, n. 1, p. 47–49, 2002. DOI: 10.1093/nar/30.1.47.

[2] SOREQ, H.; SEIDMAN, S. Acetylcholinesterase: new roles for an old actor. *Nature Reviews Neuroscience*, [s.l], v. 2, n. 4, p. 294–302, 2001. DOI: 10.1038/35067589.

[3] FRANJESEVIC, A. J.; SILLART, S. B.; BECK, J. M.; VYAS, S.; CALLAM, C. S.; HADAD, C. M. Resurrection and reactivation of Acetylcholinesterase and Butyrylcholinesterase. *Chemistry - A European Journal*, Weinheim, v. 25, n. 21, p. 5337–5371, 2019. DOI: 10.1002/chem.201805075.

[4] QUINN, D. M. Acetylcholinesterase: enzyme structure, reaction dynamics, and virtual transition states. *Chemical Reviews*, Washington, D.C., v. 87, n. 5, p. 955–979, 1987. DOI: 10.1021/cr00081a005.

[5] TAYLOR, P. The cholinesterases. *Journal of Biological Chemistry*, Washington, D.C., v. 266, n. 7, p. 4025–4028, 1991. DOI: 10.1016/S0021-9258(20)64277-6.
[6] SPLETTSTOESSER, T. Schematic of a synapse. *In:* Wikimedia Commons, 2015. Disponível em: https://commons.wikimedia.org/wiki/File:SynapseSchematic_en.svg. Acesso em: 5 dez. 2020.
[7] AMBURE, P.; ROY, K. Advances in quantitative structure-activity relationship models of anti-Alzheimer's agents. *Expert Opin on Drug Discovery*, Londres, v. 9, n. 6, p. 697–723, 2014. DOI: 10.1517/17460441.2014.909404.
[8] DELFINO, R. T.; RIBEIRO, T. S.; FIGUEROA-VILLAR, J. D. Organophosphorus compounds as chemical warfare agents: a review. *Journal of the Brazilian Chemical Society*, Campinas, v. 20, n. 3, p. 407–428, 2009. DOI: 10.1590/S0103-50532009000300003.
[9] POHANKA, M. Cholinesterases, a target of pharmacology and toxicology. *Biomedical Papers*, [s.l.], v. 155, n. 3, p. 219–223, 2011. DOI: 10.5507/bp.2011.036.
[10] MALINAK, D.; KORABECNY, J.; SOUKUP, O.; GORECKI, L., NEPOVIMOVA, E. A review of the synthesis of quaternary acetylcholinesterase reactivators. *Current Organic Chemistry*, [s.l.], v. 22, n. 6, p. 1619–1648, 2018. DOI: 10.2174/1385272822666180711123529.
[11] KLOSKE, M.; WITKIEWICZ, Z. Novichoks – The a group of organophosphorus chemical warfare agents. *Chemosphere*, [s.l.], v. 221, p. 672–682, 2019. DOI: 10.1016/j.chemosphere.2019.01.054.
[12] COSTANZI, S.; SLAVICK, C. K.; HUTCHESON, B. O.; KOBLENTZ, G. D.; CUPITT, R. T. Lists of chemical warfare agents and precursors from international nonproliferation frameworks: structural annotation and chemical fingerprint analysis. *Journal Chemical Information Modeling*, Washington, D.C., v. 60, n. 10, p. 4804–4816, 2020. DOI: 10.1021/acs.jcim.0c00896.
[13] KONING, M. C. de; HORN, G.; WOREK, F.; VAN GROL, M. Discovery of a potent non-oxime reactivator of nerve agent inhibited human acetylcholinesterase. *European Journal Medicinal Chemistry*, [s.l.], v. 157, p. 151–160, 2018, doi: 10.1016/j.ejmech.2018.08.016.
[14] VITOROVIĆ-TODOROVIĆ, M. D.; WOREK, F.; PERDIH, A.; BAUK, S. D.; VUJATOVIĆ, T. B.; CVIJETIĆ, I. N. The in vitro protective effects of the three novel nanomolar reversible inhibitors of human cholinesterases against irreversible inhibition by organophosphorous chemical warfare agents. *Chemico-Biological Interactions*, [s.l.], v. 309, 108714, 2019. DOI: 10.1016/j.cbi.2019.06.027.
[15] GREENFIELD, S.; VAUX, D. Parkinson's disease, Alzheimer's disease and motor neurone disease: identifying a common mechanism. *Neuroscience*, [s.l.], v. 113, n. 3, p. 485–492, 2002. DOI: 10.1016/S0306-4522(02)00194-X.
[16] TOUGU, V. Acetylcholinesterase: mechanism of catalysis and inhibition. *Current Medicinal Chemistry – Central Nervous System Agents*, [s.l.], v. 1, n. 2, p. 155–170, 2005. DOI: 10.2174/1568015013358536.
[17] YANAGISAWA, N.; MORITA, H.; NAKAJIMA, T. Sarin experiences in Japan: acute toxicity and long-term effects. *Journal of the Neurological Science*, [s.l.], v. 249, n. 1, p. 76–85, 2006. DOI: 10.1016/j.jns.2006.06.007.
[18] YOKOYAMA, K. Our recent experiences with sarin poisoning cases in Japan and pesticide users with references to some selected chemicals. *Neurotoxicology*, [s.l.], v. 28, n. 2, p. 364–373, 2007. DOI: 10.1016/j.neuro.2006.04.006.
[19] CANNARD, K. The acute treatment of nerve agent exposure. *Journal of the Neurological Science*, [s.l.], v. 249, n. 1, p. 86–94, 2006. DOI: 10.1016/j.jns.2006.06.008.
[20] WILSON, I. B.; GINSBURG, B. A powerful reactivator of alkylphosphate-inhibited acetylcholinesterase. *BBA - Biochimica et Biophysica Acta*, [s.l.], v. 18, n. 1, p. 168–170, 1955. DOI: 10.1016/0006-3002(55)90040-8.
[21] POZIOMEK, E. J.; HACKLEY Jr., B. E.; STEINBERG, G. M. Pyridinium Aldoximes. *Journal of Organic Chemistry*, v. 23, n. 5, p. 714–717, 1958. DOI: 10.1021/jo01099a019.
[22] LUTTRINGHAUS, A.; HAGEDORN, I.; ENGELHARD, H.; ENGELHARD, N. *Preparation of bis-quaternary pyridinium salts*. 3,187,702. Concessão: 16 jun. 1964.
[23] OLDIGES, H.; SCHOENE, K. Pyridinium- und Imidazoliumsalze als Antidote gegenüber Soman- und Paraoxonvergiftungen bei Mäusen. *Fühner-Weilands Sammlung von Vergiftungsfällen Archives für Toxikologie*, Welling, v. 26, p. 293–305, 1970. DOI: 10.1007/BF00577721.
[24] AMEND, N.; NIESSEN, K. V.; SEEGER, T.; WILLE, T.; WOREK, F.; THIERMANN, H. Diagnostics and treatment of nerve agent poisoning –current status and future developments. *Annals of the New York Academy Sciences*, Nova Iorque, v. 1479, n. 1, p. 1–16, 2020. DOI: 10.1111/nyas.14336.
[25] WOREK, F.; THIERMANN, H. The value of novel oximes for treatment of poisoning by organophosphorus compounds. *Pharmacology & Therapeutics*, [s.l.], v. 139, n. 2, p. 249–259, 2013. DOI: 10.1016/j.pharmthera.2013.04.009.

[26] JOKANOVIĆ, M.; PROSTRAN, M. Pyridinium oximes as cholinesterase reactivators: structure-Activity relationship and efficacy in the treatment of poisoning with organophosphorus compounds. *Current Medicinal Chemistry*, [s.l.], v. 16, n. 17, p. 2177–2188, 2009. DOI: 10.2174/092986709788612729.

[27] BRASIL. Ministério da Saúde. *Relação Nacional de Medicamentos Essenciais*. Brasília, DF: Ministério da Saude, 2020. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/relacao_medicamentos_rename_2020.pdf. Acesso em: jan. 15, 2022.

[28] SIDELL, F. R.; GROFF, W. A. The reactivatibility of cholinesterase inhibited by VX and sarin in man. *Toxicology and Applied Pharmacology*, [s.l.], v. 27, n. 2, p. 241–252, 1974. DOI: 10.1016/0041-008X(74)90195-1.

[29] HARRIS, L. W.; STITCHER, D. L. Reactivation of VX-Inhibited Cholinesterase by 2-Pam and Hs-6 in Rats. *Drug and Chemical Toxicology*, Londres, v. 6, n. 3, p. 235–240, 1983. DOI: 10.3109/01480548309017814.

[30] KOPLOVITZ, I.; STEWART, J. R. A comparison of the efficacy of HI6 and 2-PAM against soman, tabun, sarin, and VX in the rabbit. *Toxicology Letters*, [s.l.], v. 70, n. 3, p. 269–279, 1994. DOI: 10.1016/0378-4274(94)90121-X.

[31] SAKURADA, K.; MATSUBARA, K.; SHIMIZU, K.; SHIONO, H.; SETO, Y.; TSUGE, K.; YOSHINO, M.; SAKAI, I.; MUKOYAMA, H.; TAKATORI, T. Pralidoxime iodide (2-PAM) penetrates across the blood-brain barrier. *Neurochemical Research*, [s.l.], v. 28, n.9, p. 1401–1407, 2003.

[32] LORKE, D.; KALASZ, H.; PETROIANU, G.; TEKES, K. Entry of oximes into the Brain: a review. *Current Medicinal Chemistry*, [s.l.], v. 15, n. 8, p. 743–753, 2008. DOI: 10.2174/092986708783955563.

[33] SIDELL, F. R.; GROFF, W. A. Intramuscular and intravenous administration of small doses of 2-pyridinium aldoxime methochloride to man. *Journal of Pharmaceutical Sciences*, [s.l.], v. 60, n. 8, p. 1224–1228, 1971. DOI: 10.1002/jps.2600600823.

[34] DAWSON, R. M. Review of oximes available for treatment of nerve agent poisoning. *Journal Applied Toxicology*, [s.l.], v. 14, n. 5, p. 317–331, 1994. DOI: 10.1002/jat.2550140502.

[35] JAGER, B. V.; STAGG, G. N. Toxicity of diacetyl monoxime and of pyridine-2-aldoxime methiodide in man. *Bulletins of the John Hopkins Hospital*, Baltimore, v. 102, n. 4, p. 203–11, 1958. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/13523283. Acesso em: 7 mar. 2022.

[36] GREEN, A. L.; SAVILLE, B. The reaction of oximes with isopropyl methylphosphono-fluoridate (Sarin). *Journal of the Chemical Society*, n. 0, p. 3887-3892, 1956. DOI: 10.1039/jr9560003887.

[37] GINSBURG, S.; WILSON, I. B. Oximes of the Pyridine Series. *Journal American Chemical Society*, v. 79, n. 2, p. 481–485, 1957. DOI: 10.1021/ja01559a067.

[38] BLOCH, L. P. *Method for producing 2-pyridinealdoxime methochloride*. 3,123,613. Concessão: 3 mar. 1964.

[39] ELLIN, R. I.; EASTERDAY, D. E.; KONDRITZER, A. A. *Process for preparing methyl pyridinium 2 aldoxime chloride*. 3,140,289. Concessão: 7 jul. 1964.

[40] UNNISA, L.; SUMAKANTH, M.; RAO, B. L.; DIVI, M. K.; RAO, M. A. A simple process for the preparation of pralidoxime chloride. *Indian Journal of Chemistry*, v. 53B, n. 4, p. 431–435, 2014.

[41] GORECKI, L.; KORABECNY, J.; MUSILEK, K.; DAVID, M.; NEPOVIMOVA, E.; DOLEZAL, R.; JUN, D.; SOUKUP, O.; KUCA, K. SAR study to find optimal cholinesterase reactivator against organophosphorous nerve agents and pesticides. *Archives of Toxicology*, v. 90, n. 12, p. 2831–2859, 2016. DOI: 10.1007/s00204-016-1827-3.

[42] CHAMBERS, J. E.; MEEK, E. C.; CHAMBERS, H. W. Novel brain-penetrating oximes for reactivation of cholinesterase inhibited by sarin and VX surrogates. *Annals New York Academy of Sciences*, Nova Iorque, v. 1374, n. 1, p. 52–58, 2016. DOI: 10.1111/nyas.13053.

[43] KARASOVA, J. Z.; ZEMEK, F.; BAJGAR, J.; VASATOVA, M.; PROCHAZKA, P.; NOVOTNY, L.; KUCA, K. Partition of bispyridinium oximes (trimedoxime and K074) administered in therapeutic doses into different parts of the rat brain. *Journal of Pharmaceutical and Biomedical Analysis*, [s.l], v. 54, n. 5, p. 1082–1087, 2011. DOI: 10.1016/j.jpba.2010.11.024.

[44] SOUKUP, O.; TOBIN, G.; KUMAR, U. K.; BINDER, J.; PROSKA, J.; JUN, D.; FUSEK, J; KUCA, K. Interaction of nerve agent antidotes with cholinergic systems. *Current Medicinal Chemistry*, [s.l.], v. 17, n. 16, p. 1708–1718, 2010. DOI: 10.2174/092986710791111260.

[45] SIT, R. K.; RADIĆ, Z.; GERARDI, V.; ZHANG, L.; GARCIA, E.; KATALINIĆ, M; AMITAI, G.; KOVARIK, Z.; FOKIN, V. V.; SHARPLESS, K. B.; TAYLOR, P. New structural scaffolds for centrally acting oxime reactivators of phosphylated cholinesterases. *Journal of Biological Chemistry*, [s.l.], v. 286, n. 22, p. 19422–19430, 2011. DOI: 10.1074/jbc.M111.230656.

[46] RADIĆ, Z.; SIT, R. K.; KOVARIK, Z; BEREND, S.; GARCIA, E.; ZHANG, L.; AMITAI, G; GREEN, C.; RADIĆ, B.; FOKIN, V. V.; SHARPLESS, K. B.; PALMER, T. Refinement of structural leads for centrally acting oxime

reactivators of phosphylated cholinesterases. *Journal of Biological Chemistry*, [s.l.], v. 287, n. 15, p. 11798–11809, 2012. DOI: 10.1074/jbc.M111.333732.

[47] KOVARIK, Z.; NIKOLINA, M.; SIT, R. K.; RADIĆ, Z.; FOKIN, V. V.; SHARPLESS, K. B.; TAYLOR, P. Centrally acting oximes in reactivation of tabun-phosphoramidated AChE. *Chemical Biological Interaction*, [s.l.], v. 203, n. 1, p. 77–80, 2013. DOI: 10.1016/j.cbi.2012.08.019.

[48] MERCEY, G.; VERDELET, T.; RENOU, J.; KLIACHYNA, M.; BAATI, R.; NACHON, F.; JEAN, L.; RENARD, P. Y. Reactivators of acetylcholinesterase inhibited by organophosphorus nerve agents. *Accounts of Chemical Research*, Washington, D.C., v. 45, n. 5, p. 756–766, 2012. DOI: 10.1021/ar2002864.

[49] Krishnan, J. K. S.; ARUN, P.; APPU, A. P.; VIJAYAKUMAR, N.; FIGUEIREDO, T. H.; BRAGA, M. F. M.; BASKOTA, S.; OLSEN, C. H.; FARKAS, N.; DAGATA, J.; FREY, W. H.; MOFFETT, J. R.; NAMBOODIRI, A. M. A. Intranasal delivery of obidoxime to the brain prevents mortality and CNS damage from organophosphate poisoning. *Neurotoxicology*, [s.l.], v. 53, p. 64–73, 2016. DOI: 10.1016/j.neuro.2015.12.020.

PÓS-GRADUAÇÃO NO IME

Bolsas: CAPES, CNPq e FAPERJ

Mestrado

- Sistemas e Computação
- Engenharia Mecânica

Mestrado e Doutorado

- Química
- Ciência dos Materiais
- Engenharia de Defesa
- Engenharia Nuclear
- Engenharia Elétrica
- Engenharia de Transportes

Matrículas: Semestral

Tel.: (21) 2546-7114 – Fax: (21) 2546-7089 – www.ime.eb.br – sd1@ime.eb.br