


# Machine Learning Model for Computational Tracking and Forecasting the COVID-19 Dynamic Propagation

Daiana Caroline dos Santos Gomes , *Member, IEEE*,  
and Ginalber Luiz de Oliveira Serra , *Member, IEEE*

**Abstract**—A computational model with intelligent machine learning for analysis of epidemiological data, is proposed. The innovations of adopted methodology consist of an interval type-2 fuzzy clustering algorithm based on adaptive similarity distance mechanism for defining specific operation regions associated to the behavior and uncertainty inherited to epidemiological data, and an interval type-2 fuzzy version of Observer/Kalman Filter Identification (OKID) algorithm for adaptive tracking and real time forecasting according to unobservable components computed by recursive spectral decomposition of experimental epidemiological data. Experimental results and comparative analysis illustrate the efficiency and applicability of proposed methodology for adaptive tracking and real time forecasting the dynamic propagation behavior of novel coronavirus 2019 (COVID-19) outbreak in Brazil.

**Index Terms**—Computational model, covid-19, epidemiological data, interval type-2 fuzzy systems, kalman filtering, machine learning.

## I. INTRODUCTION

IN THE last years, studies involving the integration of fuzzy systems and Kalman filters have been proposed in the literature [1]–[3]. In [4], fuzzy sets are combined with an optimization method based on extended Kalman filter with probabilistic-numerical linguistic information applied for tracking a maneuvering target. In [5], an optimization methodology of adaptive Unscented Kalman Filter (UKF) is presented by an evolutionary fuzzy algorithm named Fuzzy Adaptive Grasshopper Optimization Algorithm, and it is efficiently applied to different benchmark functions.

Recently, with the beginning of the Covid-19 epidemic outbreak, several researchers have proposed model based data analysis approaches applied to novel Coronavirus 2019 [6]–[8].

Manuscript received July 8, 2020; revised October 21, 2020 and December 18, 2020; accepted January 12, 2021. Date of publication January 15, 2021; date of current version March 5, 2021. (*Corresponding author: Ginalber Luiz de Oliveira Serra.*)

Daiana Caroline dos Santos Gomes is with the Master and Doctorate Program in Electrical Engineering, at Federal University of Maranhão, São Luís 70.058-900, Brazil (e-mail: ginalber@ifma.edu.br).

Ginalber Luiz de Oliveira Serra is with the Department of Eletroeletronic, at Federal Institute of Education, Science, and Technology of Maranhão, São Luís 65075-441, Brazil (e-mail: daianagomes159@gmail.com).

Digital Object Identifier 10.1109/JBHI.2021.3052134

The objective of these studies is to characterize the evolution of pandemic in certain regions and, thus, to contribute for the requirements adopted to contain the contamination by virus and allocation of resources. In [9], a mathematical model based on SEIR model (Susceptible - Exposed - Infectious - Recovered) for forecasting the transmission dynamics of Covid-19 in Korea, is proposed. This study is able to predict the final size and the timing of the end of epidemic as well as the maximum number of isolated individuals using daily confirmed cases comparing epidemiological parameters between the national level and the Daegu/Gyeongbuk area. In [10], the role of asymptomatic carriers in transmission poses challenges for control of the Covid-19 pandemic, is addressed.

Differently from aforementioned approaches and others ones found from literature, the scope of this paper outlines a machine learning approach based on integration of Kalman filter and interval type-2 fuzzy systems for adaptive tracking and real time forecasting the COVID-19 dynamic propagation. The design of interval type-2 fuzzy Kalman filter, according to proposed methodology, is based on spectral unobservable components and uncertainty regions extracted from experimental data.

## A. Motivation and Contributions of the Proposed Methodology

The impacts caused by novel coronavirus pandemic has motivated the analysis of epidemiological data, for support of political/health authorities and decision-making [11], [12]. In this context, several modeling methodologies has been proposed in literature for solving epidemiological problems [13]–[15]. However, the uncertainties inherent to experimental epidemiological data (underreporting, lack of information, incubation period of the virus, time to seek care and diagnosis) have open a new research field, in which the proposed methodology belongs to. The originality of the proposed methodology is outlined by following main contributions:

- A new machine learning computational tool based on the successful integration of Kalman filters and type-2 fuzzy systems for adaptive tracking and real time forecasting of experimental epidemiological data, which is useful for analysis of COVID-19 dynamic propagation;
- Formulation of new interval type-2 fuzzy clustering algorithm based on adaptive similarity distance mechanism

ables to define specific operation regions in the epidemiological data associated to the behavior and uncertainty inherent to COVID-19 dynamic propagation;

- Formulation of a new computational model with intelligent machine learning based on interval type-2 fuzzy Kalman filter, for adaptive tracking and real time forecasting the behavior and uncertainty inherent to COVID-19 dynamic propagation, from the specific operation regions in the epidemiological data.

## II. INTERVAL TYPE-2 FUZZY COMPUTATIONAL MODEL

In this section, the proposed methodology for designing the interval type-2 fuzzy Kalman filter computational model from experimental data, is presented.

### A. Pre-Processing by Singular Spectral Analysis

1) *Training Step*: Let the initial experimental dataset referring to  $p$  time series under analysis, with  $N_b$  samples, given by [16]:

$$\mathbf{y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_{N_b}], \quad \mathbf{y} \in \mathbb{R}^{p \times N_b} \quad (1)$$

where  $\mathbf{y}_k \in \mathbb{R}^p$ , with  $k = 1, \dots, N_b$ , is the time series vector at instant of time  $k$ . From this initial dataset, a trajectory matrix  $\mathbf{H}$  is defined, for each of the dimensions of  $\mathbf{y}$ , considering a set of  $\rho$  delayed vectors with dimension  $\delta$ , which is an integer number defined by user with  $2 \leq \delta \leq N_b - 1$  and  $\rho = N_b - \delta + 1$ , given by:

$$\mathbf{H} = \begin{bmatrix} y_1 & y_2 & y_3 & \dots & y_\rho \\ y_2 & y_3 & y_4 & \dots & y_{\rho+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_\delta & y_{\delta+1} & y_{\delta+2} & \dots & y_{N_b} \end{bmatrix}, \quad \mathbf{H} \in \mathbb{R}^{\delta \times \rho} \quad (2)$$

and the covariance matrix  $\mathbf{S}$  is obtained as follows:

$$\mathbf{S} = \mathbf{H}\mathbf{H}^T, \quad \mathbf{S} \in \mathbb{R}^{\delta \times \delta} \quad (3)$$

Applying the Singular Value Decomposition (SVD) procedure to matrix  $\mathbf{S}$ , is obtained a set of eigenvalues in decreasing order such that  $\sigma^1 \geq \sigma^2 \geq \dots \geq \sigma^\delta \geq 0$  with their respective eigenvectors  $\phi^1, \phi^2, \dots, \phi^\delta$ . Considering  $d = \max\{\varsigma, \text{such that } \sigma^\varsigma > 0\}$ , and  $\mathbf{V}^\varsigma = \mathbf{H}^T \phi^\varsigma / \sqrt{\sigma^\varsigma}$  with  $\varsigma = 1, \dots, d$ , the singular value decomposition of the trajectory matrix  $\mathbf{H}$ , can be rewritten as:

$$\mathbf{H} = \mathbf{H}^1 + \mathbf{H}^2 + \dots + \mathbf{H}^d \quad (4)$$

where the matrix  $\mathbf{H}^\varsigma |^{\varsigma=1, \dots, d}$  is elementary (it has rank equal to 1), and is given by:

$$\mathbf{H}^\varsigma = \sqrt{\sigma^\varsigma} \phi^\varsigma \mathbf{V}^{\varsigma T}, \quad \mathbf{H}^\varsigma \in \mathbb{R}^{\delta \times \rho} \quad (5)$$

The regrouping of  $\mathbf{H}^\varsigma |^{\varsigma=1, \dots, d}$  into  $\xi$  linearly independent matrices terms  $\mathbf{I}^j |^{j=1, \dots, \xi}$ , such that  $\xi \leq d$ , results in

$$\mathbf{H} = \mathbf{I}^1 + \mathbf{I}^2 + \dots + \mathbf{I}^\xi \quad (6)$$

where  $\xi$  is the number of unobservable components extracted from experimental dataset. The unobservable spectral components  $\alpha^j |^{j=1, \dots, \xi}$  obtained from matrices  $\mathbf{I}^j |^{j=1, \dots, \xi}$ , are given

by:

$$\alpha_k^j = \begin{cases} \frac{1}{k} \sum_{\nu=1}^{k+1} I_{\nu, k-\nu+1}^j & 1 \leq k \leq \delta^* \\ \frac{1}{\delta^*} \sum_{\nu=1}^{\delta^*} I_{\nu, k-\nu+1}^j & \delta^* \leq k \leq \rho^* \\ \frac{1}{N_b - k + 1} \sum_{\nu=k-\rho^*+1}^{N_b - \rho^* + 1} I_{\nu, k-\nu+1}^j & \rho^* < k \leq N_b \end{cases} \quad (7)$$

where  $\delta^* = \min(\delta, \rho)$ ,  $\rho^* = \max(\delta, \rho)$ .

2) *Recursive Step*: The value of  $\rho$  is increased by  $\rho = k - \delta + 1$ , with  $k = N_b + 1, N_b + 2, \dots$ , and the covariance matrix is updated, recursively, as follows:

$$\mathbf{S}_k = \mathbf{S}_{k-1} + \Upsilon_k, \quad \mathbf{S}_k \in \mathbb{R}^{\delta \times \delta} \quad (8)$$

where  $\Upsilon_k = \psi_k \psi_k^T \in \mathbb{R}^{\delta \times \delta}$  with  $\psi_k = [y_\rho, y_{\rho+1}, \dots, y_k]^T \in \mathbb{R}^{\delta \times 1}$ . Applying SVD procedure to covariance matrix  $\mathbf{S}_k$ , the term  $y_k$  can be rewritten by:

$$y_k = h_k^1 + h_k^2 + \dots + h_k^d \quad (9)$$

where  $h_k^\varsigma = \kappa_k^\varsigma \psi_k^T \phi_k^\varsigma$ , with  $\varsigma = 1, \dots, d$ , such that  $\kappa_k^\varsigma$  corresponds to the last element of the eigenvector  $\phi_k^\varsigma$ . Finally, the regrouping of the terms  $h_k^\varsigma |^{\varsigma=1, \dots, d}$  in  $\xi$  disjoint terms  $I_k^j |^{j=1, \dots, \xi}$ , results in

$$y_k = I_k^1 + I_k^2 + \dots + I_k^\xi \quad (10)$$

such that  $I_k^j = \alpha_k^j$ , with  $j = 1, \dots, \xi$  and  $k = N_b + 1, N_b + 2, \dots$ , represents the samples of extracted unobservable components at instant  $k$ .

### B. Parametric Estimation of Interval Type-2 Fuzzy Kalman Filter

The adopted structure of interval type-2 fuzzy Kalman filter presents the  $i |^{i=1, 2, \dots, c}$ -th fuzzy rule, given by:

$$\begin{aligned} R^{(i)} : & \text{ IF } \mathbf{Z}_k \text{ IS } \widetilde{W}^i \\ \text{ THEN } & \begin{cases} \widetilde{\mathbf{x}}_{k+1}^i = \widetilde{\mathbf{A}}_k^i \widetilde{\mathbf{x}}_k^i + \widetilde{\mathbf{B}}_k^i \mathbf{u}_k + \widetilde{\mathbf{K}}_k^i \widetilde{\boldsymbol{\epsilon}}_k^i \\ \widetilde{\mathbf{y}}_k^i = \widetilde{\mathbf{C}}_k^i \widetilde{\mathbf{x}}_k^i + \widetilde{\mathbf{D}}_k^i \mathbf{u}_k \end{cases} \end{aligned} \quad (11)$$

with  $n$ -th order,  $m$  inputs,  $p$  outputs, where  $\mathbf{Z}_k$  is the linguistic variable of the antecedent;  $\widetilde{W}^i$  is the interval type-2 fuzzy set;  $\widetilde{\mathbf{x}}_k^i \in \mathbb{R}^n$  is the estimated interval states vector;  $\widetilde{\mathbf{y}}_k^i \in \mathbb{R}^p$  is the estimated interval output vector and  $\mathbf{u}_k \in \mathbb{R}^m$  is the input signal. The matrices  $\widetilde{\mathbf{A}}_k^i \in \mathbb{R}^{n \times n}$ ,  $\widetilde{\mathbf{B}}_k^i \in \mathbb{R}^{n \times m}$ ,  $\widetilde{\mathbf{C}}_k^i \in \mathbb{R}^{p \times n}$ ,  $\widetilde{\mathbf{D}}_k^i \in \mathbb{R}^{p \times m}$  and  $\widetilde{\mathbf{K}}_k^i \in \mathbb{R}^{n \times p}$  are, respectively, state matrix, input matrix, output matrix, direct transmission matrix and Kalman gain matrix. The residual error  $\widetilde{\boldsymbol{\epsilon}}_k^i$  for  $i$ -th rule is defined as  $\widetilde{\boldsymbol{\epsilon}}_k^i = \mathbf{y}_k - \widetilde{\mathbf{y}}_k^i$ , where  $\mathbf{y}_k \in \mathbb{R}^p$  is the real time series and  $\widetilde{\mathbf{y}}_k^i$  is the interval estimated time series by  $i$ -th interval Kalman filter.

1) *Parametric Estimation of Antecedent*: The interval type-2 fuzzy version of Gustafson-Kessel clustering algorithm, is proposed, as formulated in the sequel. Given the experimental dataset  $\mathbf{Z} \in \mathbb{R}^{p \times N_b}$ , choose the number of clusters  $c$  such that  $1 < c < N_b$ ; the initial partition matrix  $\widetilde{\mathbf{U}}^{(0)} \in \mathbb{R}^{c \times N_b}$ , the termination tolerance  $\mathcal{E} > 0$  and the interval weighting exponent  $\widetilde{m} = [\underline{m}, \overline{m}]$ , where  $\underline{m}$  and  $\overline{m}$  correspond to, respectively,

weighting exponent of upper and lower membership functions of the interval type-2 fuzzy set  $\widetilde{W}^i$ .

**Repeat** for  $l = 1, 2, \dots$

*Step 1:* Compute the centers of the clusters  $\widetilde{\mathbf{v}}^{i(l)}$ :

$$\widetilde{\mathbf{v}}^{i(l)} = \frac{\sum_{k=1}^{N_b} \left( \widetilde{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k)^{i(l-1)} \right)^{\widetilde{m}} \mathbf{Z}_k}{\sum_{k=1}^{N_b} \left( \widetilde{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k)^{i(l-1)} \right)^{\widetilde{m}}}, \quad 1 \leq i \leq c \quad (12)$$

where  $\mathbf{Z}_k$  is the data at sample  $k$  and  $\widetilde{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k)$  is the interval membership degree of  $\mathbf{Z}_k$  in the  $i$ -th cluster.

*Step 2:* Compute the covariance matrices  $\widetilde{\mathbf{F}}^i$  of the clusters:

$$\widetilde{\mathbf{F}}^i = \frac{\sum_{k=1}^{N_b} \left( \widetilde{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k)^{i(l-1)} \right)^{\widetilde{m}} (\mathbf{Z}_k - \widetilde{\mathbf{v}}^{i(l)}) (\mathbf{Z}_k - \widetilde{\mathbf{v}}^{i(l)})^T}{\sum_{k=1}^{N_b} \left( \widetilde{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k)^{i(l-1)} \right)^{\widetilde{m}}}, \quad (13)$$

$1 \leq i \leq c, \quad 1 \leq k \leq N_b$

*Step 3:* Compute the distances  $\widetilde{D}_{k\widetilde{\mathbf{F}}^i}^i$  between the sample  $\mathbf{Z}_k$  and the center  $\widetilde{\mathbf{v}}^{i(l)}$  of the  $i$ -th cluster:

$$\widetilde{D}_{k\widetilde{\mathbf{F}}^i}^i = \sqrt{(\mathbf{Z}_k - \widetilde{\mathbf{v}}^{i(l)})^T \left[ \det(\widetilde{\mathbf{F}}^i)^{1/n} (\widetilde{\mathbf{F}}^i)^{-1} \right] (\mathbf{Z}_k - \widetilde{\mathbf{v}}^{i(l)})} \quad (14)$$

*Step 4:* Update the interval partition matrix  $\widetilde{\mathbf{U}}^{(l)}$ :

If  $\widetilde{D}_{k\widetilde{\mathbf{F}}^i}^i > 0$  for  $1 \leq i \leq c, 1 \leq k \leq N_b$

$$\widetilde{\mu}_{\widetilde{W}^i}^{i(l)}(\mathbf{Z}_k) = \left[ \underline{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k), \overline{\mu}_{\widetilde{W}^i}^i(\mathbf{Z}_k) \right] \quad (15)$$

where

$$\underline{\mu}_{\widetilde{W}^i}^{i(l)}(\mathbf{Z}_k) = \min \left[ \frac{1}{\sum_{j=1}^c \left( \frac{\widetilde{D}_{k\widetilde{\mathbf{F}}^i}^i}{\widetilde{D}_{k\widetilde{\mathbf{F}}^j}^j} \right)^{2/(\widetilde{m}-1)}}, \frac{1}{\sum_{j=1}^c \left( \frac{\overline{\widetilde{D}}_{k\widetilde{\mathbf{F}}^i}^i}{\overline{\widetilde{D}}_{k\widetilde{\mathbf{F}}^j}^j} \right)^{2/(\widetilde{m}-1)}} \right] \quad (16)$$

is the lower activation degree in  $i$ -th rule and

$$\overline{\mu}_{\widetilde{W}^i}^{i(l)}(\mathbf{Z}_k) = \max \left[ \frac{1}{\sum_{j=1}^c \left( \frac{\widetilde{D}_{k\widetilde{\mathbf{F}}^i}^i}{\widetilde{D}_{k\widetilde{\mathbf{F}}^j}^j} \right)^{2/(\widetilde{m}-1)}}, \frac{1}{\sum_{j=1}^c \left( \frac{\overline{\widetilde{D}}_{k\widetilde{\mathbf{F}}^i}^i}{\overline{\widetilde{D}}_{k\widetilde{\mathbf{F}}^j}^j} \right)^{2/(\widetilde{m}-1)}} \right] \quad (17)$$

is the upper activation degree in  $i$ -th rule. Otherwise,  $\widetilde{\mu}_{\widetilde{W}^i}^{i(l)}(\mathbf{Z}_k) = [0, 0]$  with  $\underline{\mu}_{\widetilde{W}^i}^{i(l)}(\mathbf{Z}_k) \in [0, 1]$  e  $\overline{\mu}_{\widetilde{W}^i}^{i(l)}(\mathbf{Z}_k) \in [0, 1]$

**Until**  $\|\widetilde{\mathbf{U}}^{(l)} - \widetilde{\mathbf{U}}^{(l-1)}\| < \mathcal{E}$

**2) Parametric Estimation of Consequent:** The interval type-2 fuzzy OKID (Observer/Kalman Filter Identification) algorithm, is proposed, as formulated in the sequel. Let the experimental dataset  $\mathbf{Z}$ , such that  $\mathbf{Z}_k = [\mathbf{u}_k \quad \boldsymbol{\alpha}_k^*]^T$ , where  $\boldsymbol{\alpha}_k^*$

corresponds to spectral components extracted from the experimental dataset that presents higher eigenvalue and are more significant to represent the dynamics of experimental dataset. Choose an appropriate number of Markov parameters  $q$ , through the following steps:

*Step 1:* Compute the matrix of regressors  $\boldsymbol{\Lambda}$ , given by:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{u}_q & \mathbf{u}_{q+1} & \cdots & \mathbf{u}_{N_b-1} \\ \mathbf{Z}_{q-1} & \mathbf{Z}_q & \cdots & \mathbf{Z}_{N_b-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_0 & \mathbf{Z}_1 & \cdots & \mathbf{Z}_{N_b-q-1} \end{bmatrix} \quad (18)$$

*Step 2:* Compute the interval Observer Markov Parameters:

$$\widetilde{\mathbf{y}}^T = \sum_{i=1}^c \widetilde{\boldsymbol{\Gamma}}^i \boldsymbol{\Lambda}^T \widetilde{\mathbf{Y}}^i \quad (19)$$

where  $\widetilde{\boldsymbol{\Gamma}}^i$  is the diagonal weighting matrix of the  $i$ -th fuzzy rule obtained from the interval type-2 Gustafson-Kessel fuzzy clustering algorithm and

$$\begin{aligned} \widetilde{\mathbf{Y}}^i &= \left[ \widetilde{\mathbf{D}}_k^i \quad \widetilde{\mathbf{C}}_k^i \widetilde{\mathbf{B}}_k^i \quad \widetilde{\mathbf{C}}_k^i \widetilde{\mathbf{A}}_k^i \widetilde{\mathbf{B}}_k^i \quad \cdots \quad \widetilde{\mathbf{C}}_k^i \widetilde{\mathbf{A}}_k^{i(q-1)} \widetilde{\mathbf{B}}_k^i \right] \\ &= \left[ \widetilde{\mathbf{Y}}_0 \quad \widetilde{\mathbf{Y}}_1 \quad \widetilde{\mathbf{Y}}_2 \quad \cdots \quad \widetilde{\mathbf{Y}}_q \right] \end{aligned} \quad (20)$$

are the interval observer Markov parameters of  $i$ -th rule such that  $\widetilde{\mathbf{A}}_k^i = [\widetilde{\mathbf{A}}_k^i + \widetilde{\mathbf{K}}_k^i \widetilde{\mathbf{C}}_k^i]$ ,  $\widetilde{\mathbf{B}}_k^i = [\widetilde{\mathbf{B}}_k^i + \widetilde{\mathbf{K}}_k^i \widetilde{\mathbf{D}}_k^i, -\widetilde{\mathbf{K}}_k^i]$  and  $C\mathbf{A}^k\mathbf{B} \approx 0$  for  $k \geq q$ , where  $q$  is the number of observer Markov parameters [17]. Manipulating the (19):

$$\boldsymbol{\Lambda} \widetilde{\boldsymbol{\Gamma}}^i \mathbf{y}^T = \boldsymbol{\Lambda} \widetilde{\boldsymbol{\Gamma}}^i \boldsymbol{\Lambda}^T \widetilde{\mathbf{Y}}^i \quad (21)$$

Assuming  $\widetilde{\boldsymbol{\mathbf{U}}}^i = \boldsymbol{\Lambda} \widetilde{\boldsymbol{\Gamma}}^i \boldsymbol{\Lambda}^T$  and  $\widetilde{\boldsymbol{\mathbf{S}}}^i = \boldsymbol{\Lambda} \widetilde{\boldsymbol{\Gamma}}^i \mathbf{y}^T$ , Eq. (21) is rewriting as  $\widetilde{\boldsymbol{\mathbf{U}}}^i \widetilde{\mathbf{Y}}^i = \widetilde{\boldsymbol{\mathbf{S}}}^i$ . Applying QR factorization to the term  $\widetilde{\boldsymbol{\mathbf{U}}}^i$ , it has:

$$\widetilde{\mathbf{Q}}^i \widetilde{\mathbf{R}}^i \widetilde{\mathbf{Y}}^i = \widetilde{\boldsymbol{\mathbf{S}}}^i \quad (22)$$

Because the matrix  $\widetilde{\mathbf{R}}^i$  is upper triangular, Eq. (22) can be solved by backward replacement, obtaining the observer's Markov parameter vector  $\widetilde{\mathbf{Y}}^i$ .

*Step 3:* Compute the observer gain and system Markov parameters:

$$\begin{aligned} \widetilde{\mathbf{Y}}_j^i &= \left[ \widetilde{\mathbf{C}}_k^i \left( \widetilde{\mathbf{A}}_k^i + \widetilde{\mathbf{K}}_k^i \widetilde{\mathbf{C}}_k^i \right)^{(j-1)} \left( \widetilde{\mathbf{B}}_k^i + \widetilde{\mathbf{K}}_k^i \widetilde{\mathbf{D}}_k^i \right), \right. \\ &\quad \left. - \widetilde{\mathbf{C}}_k^i \left( \widetilde{\mathbf{A}}_k^i + \widetilde{\mathbf{K}}_k^i \widetilde{\mathbf{C}}_k^i \right)^{(j-1)} \widetilde{\mathbf{K}}_k^i \right] \end{aligned} \quad (23)$$

$$= \left[ \widetilde{\mathbf{Y}}_j^{i(1)}, -\widetilde{\mathbf{Y}}_j^{i(2)} \right], \quad j = 1, 2, 3, \dots \quad (24)$$

The system Markov parameters  $\widetilde{\mathbf{Y}}_j^i$  are obtained as follows:

$$\widetilde{\mathbf{Y}}_j^i = \widetilde{\mathbf{Y}}_j^{i(1)} - \sum_{\iota=1}^j \widetilde{\mathbf{Y}}_j^{i(2)} \widetilde{\mathbf{Y}}_{j-\iota}^i, \quad \text{for } j = 1, \dots, q \quad (25)$$

$$\tilde{\mathbf{Y}}_j^i = - \sum_{\iota=1}^q \tilde{\mathbf{Y}}_j^{i(2)} \tilde{\mathbf{Y}}_{j-\iota}^i, \text{ for } j = q+1, \dots, \infty \quad (26)$$

and the observer gain Markov parameters  $\tilde{\mathbf{Y}}_j^{i^o}$  are obtained by:

$$\tilde{\mathbf{Y}}_j^{i^o} = \tilde{\mathbf{Y}}_j^{i(2)} - \sum_{\iota=1}^{j-1} \tilde{\mathbf{Y}}_j^{i(2)} \tilde{\mathbf{Y}}_{j-\iota}^{i^o}, \text{ for } j = 2, \dots, q \quad (27)$$

$$\tilde{\mathbf{Y}}_j^{i^o} = - \sum_{\iota=1}^q \tilde{\mathbf{Y}}_j^{i(2)} \tilde{\mathbf{Y}}_{j-\iota}^{i^o}, \text{ for } j = q+1, \dots, \infty \quad (28)$$

Step 4: Construct the Hankel matrix  $\tilde{\mathbf{H}}^i(j-1) \in \mathbb{R}^{\gamma p \times \beta m}$ :

$$\tilde{\mathbf{H}}^i(j-1) = \begin{bmatrix} \tilde{\mathbf{Y}}_j^i & \tilde{\mathbf{Y}}_{j+1}^i & \cdots & \tilde{\mathbf{Y}}_{j+\beta-1}^i \\ \tilde{\mathbf{Y}}_{j+1}^i & \tilde{\mathbf{Y}}_{j+2}^i & \cdots & \tilde{\mathbf{Y}}_{j+\beta}^i \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{Y}}_{j+\gamma-1}^i & \tilde{\mathbf{Y}}_{j+\gamma}^i & \cdots & \tilde{\mathbf{Y}}_{j+\gamma+\beta-2}^i \end{bmatrix} \quad (29)$$

where  $\gamma$  and  $\beta$  are arbitrary integers defined by user.

Step 5: For  $j = 1$ , decompose the Hankel matrix  $\tilde{\mathbf{H}}^i(0)$  using Singular Value Decomposition:

$$\tilde{\mathbf{H}}^i(0) = \tilde{\Xi}^i \tilde{\Sigma}^i \tilde{\Psi}^{iT} \quad (30)$$

where  $\tilde{\Xi}^i \in \mathbb{R}^{\gamma p \times \gamma p}$  and  $\tilde{\Psi}^i \in \mathbb{R}^{\beta m \times \beta m}$  are orthogonal matrices and  $\tilde{\Sigma}^i \in \mathbb{R}^{\gamma p \times \beta m}$  is the diagonal matrix of singular values.

Step 6: Compute the observability matrix  $\tilde{\mathcal{P}}_\gamma^i$  and controllability matrix  $\tilde{\mathcal{Q}}_\beta^i$ :

$$\tilde{\mathcal{P}}_\gamma^i = \tilde{\Xi}_n^i \left( \tilde{\Sigma}_n^i \right)^{1/2} \quad (31)$$

$$\tilde{\mathcal{Q}}_\beta^i = \left( \tilde{\Sigma}_n^i \right)^{1/2} \tilde{\Psi}_n^{iT} \quad (32)$$

Step 7: Compute the matrices that make up the consequent proposition of interval type-2 fuzzy Kalman filter:

$$\tilde{\mathbf{A}}_k^i = \left( \tilde{\Sigma}_n^i \right)^{-1/2} \tilde{\Xi}_n^{iT} \tilde{\mathbf{H}}_n^i(1) \tilde{\Psi}_n^i \left( \tilde{\Sigma}_n^i \right)^{-1/2} \quad (33)$$

$$\tilde{\mathbf{B}}_k^i = \text{first } m \text{ columns of } \tilde{\mathcal{Q}}_\beta^i \quad (34)$$

$$\tilde{\mathbf{C}}_k^i = \text{first } p \text{ rows of } \tilde{\mathcal{P}}_\gamma^i \quad (35)$$

$$\tilde{\mathbf{D}}_k^i = \tilde{\mathbf{Y}}_0^i \quad (36)$$

Step 8: Compute the interval Kalman gain matrix  $\tilde{\mathbf{K}}_k^i$ :

$$\tilde{\mathcal{P}}_\gamma^{iT} \tilde{\Gamma}^i \tilde{\mathbf{Y}}_j^{i^o} = -\tilde{\mathcal{P}}_\gamma^{iT} \tilde{\Gamma}^i \tilde{\mathcal{P}}_\gamma^i \tilde{\mathbf{K}}_k^i \quad (37)$$

Assuming  $\tilde{\mathbf{A}}^i = -\tilde{\mathcal{P}}_\gamma^{iT} \tilde{\Gamma}^i \tilde{\mathcal{P}}_\gamma^i$  and  $\tilde{\mathbf{N}}^i = \tilde{\mathcal{P}}_\gamma^{iT} \tilde{\Gamma}^i \tilde{\mathbf{Y}}_j^{i^o}$ , Eq. (37) is rewriting  $\tilde{\mathbf{A}}^i \tilde{\mathbf{K}}_k^i = \tilde{\mathbf{N}}^i$ , which is solved by QR factorization method being applied to  $\tilde{\mathbf{A}}^i$  and obtaining the interval Kalman gain matrix  $\tilde{\mathbf{K}}_k^i$ .

*Recursive Updating of Interval Type-2 Fuzzy Kalman Filter Inference System:* Considering the regressors vector  $\lambda_k$ , at instant  $k = N_b + 1, k = N_b + 2, \dots$ , given by

$$\lambda_k = [\mathbf{u}_{k+1} \quad \mathbf{Z}_k \quad \mathbf{Z}_{k-1} \quad \cdots \quad \mathbf{Z}_{k-q}]^T \quad (38)$$

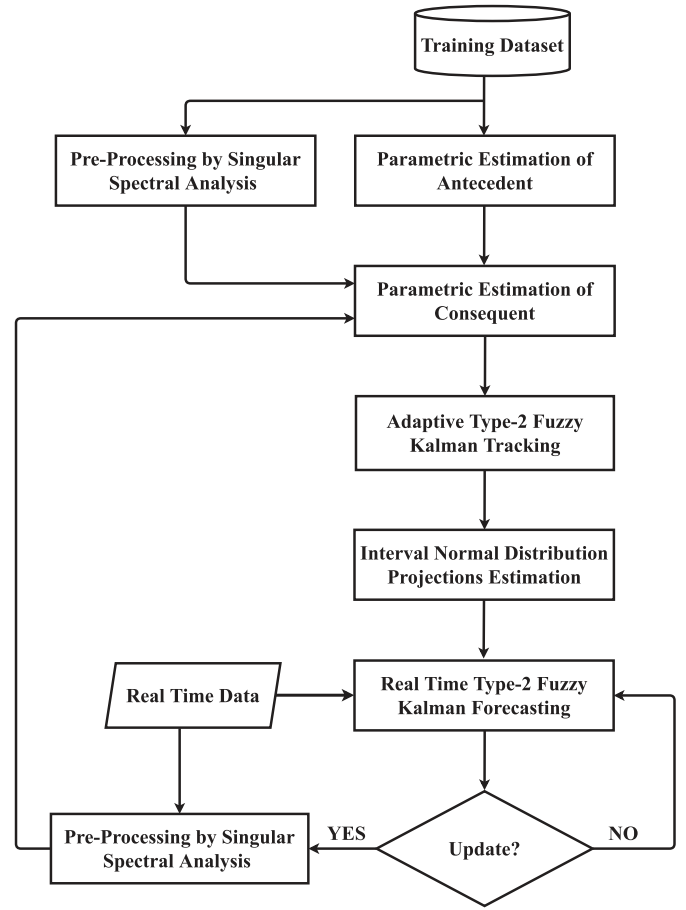


Fig. 1. The flowchart of the proposed methodology corresponding to computational aspects for designing the interval type-2 fuzzy Kalman filter.

the interval observer Markov parameters  $\tilde{\mathbf{Y}}_k^i$  are obtained by recursive updating of  $\tilde{\mathbf{U}}^i \tilde{\mathbf{Y}}^i = \tilde{\mathbf{N}}^i$ , as follows:

$$\tilde{\mathbf{U}}_k^i = \tilde{\mathbf{U}}_{k-1}^i + \tilde{\mu}_{W^i}^i(\mathbf{Z}_k) \lambda_k \lambda_k^T \quad (39)$$

$$\tilde{\mathbf{N}}_k^i = \tilde{\mathbf{N}}_{k-1}^i + \tilde{\mu}_{W^i}^i(\mathbf{Z}_k) \lambda_k \mathbf{y}_k^T \quad (40)$$

The consequent proposition of the type-2 fuzzy Kalman filter is updated recursively by repeating the Step 3 to Step 7. Similarly, the interval type-2 fuzzy Kalman gain matrix  $\tilde{\mathbf{K}}_k^i$  is obtained by recursive updating of  $\tilde{\mathbf{A}}^i \tilde{\mathbf{K}}_k^i = \tilde{\mathbf{N}}^i$ , as follows:

$$\tilde{\mathbf{A}}_k^i = \tilde{\mathbf{A}}_{k-1}^i + \tilde{\mu}_{W^i}^i(\mathbf{Z}_k) \lambda_k \lambda_k^T \quad (41)$$

$$\tilde{\mathbf{N}}_k^i = \tilde{\mathbf{N}}_{k-1}^i + \tilde{\mu}_{W^i}^i(\mathbf{Z}_k) \lambda_k \lambda_k^T \quad (42)$$

In the sense to illustrate the sequential steps of the computational aspects for interval type-2 fuzzy Kalman filter design, for better understanding from readers, a flowchart of the proposed methodology is shown in Fig. 1. The code of interval type-2 fuzzy Kalman filter algorithm,

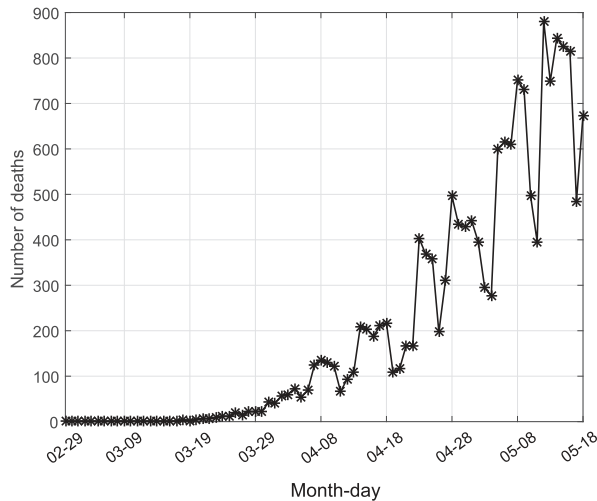


Fig. 2. The experimental dataset of daily deaths reports within period from 29 of February 2020 to 18 of May 2020, in Brazil.

based on the proposed methodology, is of open access from the link <<https://drive.google.com/drive/folders/1BvzMnaZZhtleJ1dVggJJGWISpNfwdvqD?usp=sharing>>.

### III. EXPERIMENTAL RESULTS

In this section, experimental results for forecasting analysis the COVID-19 dynamic propagation, including comparative analysis with the approaches in [18], [19] and with the machine learning models Least Absolute Shrinkage and Selection Operator (LASSO), Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) recurrent neural network, taking into account the experimental dataset of daily deaths reports caused by coronavirus disease in Brazil, are presented.

#### A. Interval Type-2 Fuzzy Kalman Filtering and Forecasting Analysis of the COVID-19 Dynamic Propagation in Brazil

The experimental dataset corresponding to daily deaths reports within the period ranging from 29 of February 2020 to 18 of May 2020, in Brazil, is shown in Fig. 2, which were extracted from official report by Ministry of Health of Brazil.<sup>1</sup> The Variance Accounted For (VAF) was considered for evaluating the appropriate number of unobservable components, within a range from 2 to 15 ones for best representation of experimental dataset. Considering the cost-benefit balance for computational practical application of proposed methodology, the appropriated number of unobservable components was  $\xi = 10$ , with VAF value of 99.98%. For implementing the proposed type-2 fuzzy clustering algorithm, the following parameters were adopted: number of clusters  $c = 3$ , interval weighting exponent  $\tilde{m} = [1.5, 2.3]$  and termination tolerance  $\mathcal{E} = 10^{-5}$ . The implementation of interval type-2 fuzzy OKID algorithm took into account the parameters

<sup>1</sup>Available at: <https://covid.saude.gov.br/>

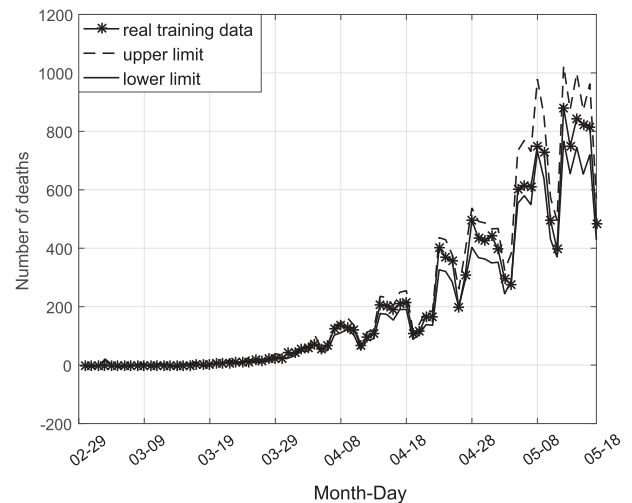


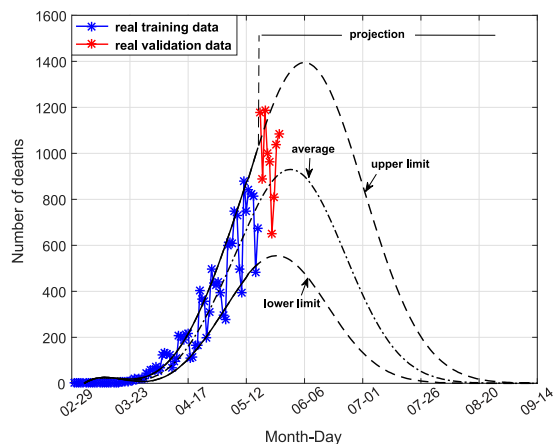
Fig. 3. The confidence region generated by interval type-2 fuzzy Kalman filter for tracking the experimental dataset of daily deaths reports, from 29 of February 2020 to 18 of May 2020, in Brazil.

values  $q = 1$ ,  $\gamma = 15$  and  $\beta = 15$ . The confidence region, as shown in Fig. 3, created by initial estimation of interval type-2 fuzzy Kalman filter, illustrates its efficiency for tracking the experimental dataset of daily deaths reports in Brazil. From this confidence region, an interval normal distribution projections were estimated, delimiting upper and lower limits for forecasting the further daily deaths reports in Brazil. The efficiency of interval type-2 fuzzy Kalman filter based on its initial estimation by training step from experimental dataset of daily deaths reports, for forecasting the further (validation) experimental dataset of daily deaths reports, is shown in Fig. 4. The results of updating of interval type-2 fuzzy Kalman filter for tracking and forecasting the COVID-19 dynamic propagation related to the daily deaths reports, are shown in Figs. 4(b)–4(f). The efficiency of interval type-2 fuzzy Kalman filter, during its recursive updating for tracking and forecasting the COVID-19 dynamic propagation related to daily deaths reports in Brazil, was validated through Variance Accounted For (VAF) criterion, as shown in Fig. 5.

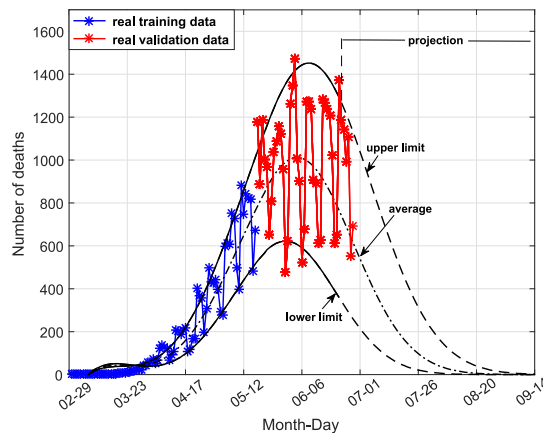
#### B. Comparative Analysis and Discussions

In this section, a more detailed discussion on the results shown in Section III-A, according to comparative analysis of proposed methodology with the approaches in [18], [19] as well as with the machine learning models LASSO, ARIMA and LSTM recurrent neural network, considering the metrics RMSE (Root Mean Square Error), MAE (Mean Absolute Error), RMSPE (Root Mean Square Percentage Error),  $R^2$  (coefficient of determination), MAD (Median Absolute Deviation) and MAPE (Mean Absolute Percentage Error), is presented.

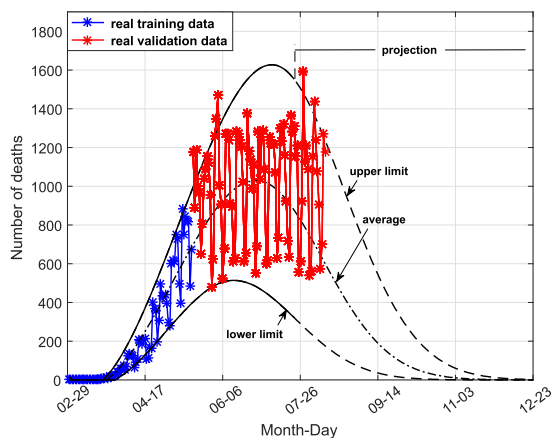
The approach in [18] is based on Wavelet-Coupled Random Vector Functional Link (WCRVFL) network for forecasting the COVID-19 dynamic propagation in Brazil, using normalized data. The efficiency of interval type-2 fuzzy Kalman filter, compared to approach in [18], is shown in Table I. As it can be seen,



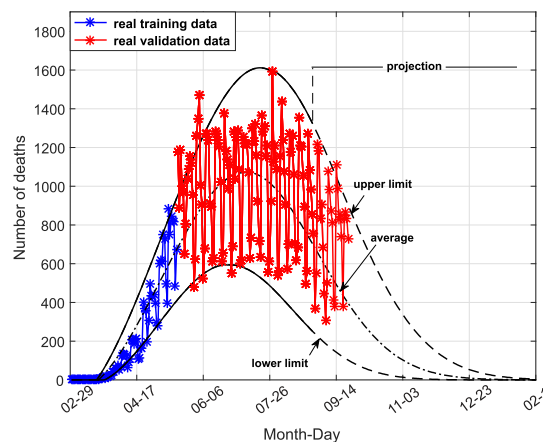
(a) The performance of interval type-2 fuzzy Kalman filter based on its initial estimation by training step for forecasting the daily deaths reports within the period ranging from 19 of May 2020 to 27 of May 2020.



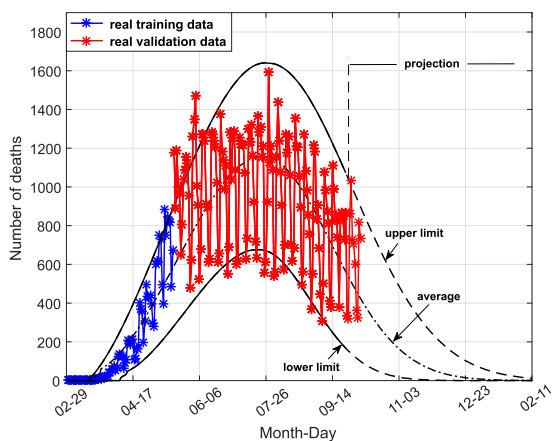
(b) The performance of interval type-2 fuzzy Kalman filter based on its recursive updating on 24 of June 2020 for forecasting the daily deaths reports within the period ranging from 25 of June 2020 to 30 of June 2020.



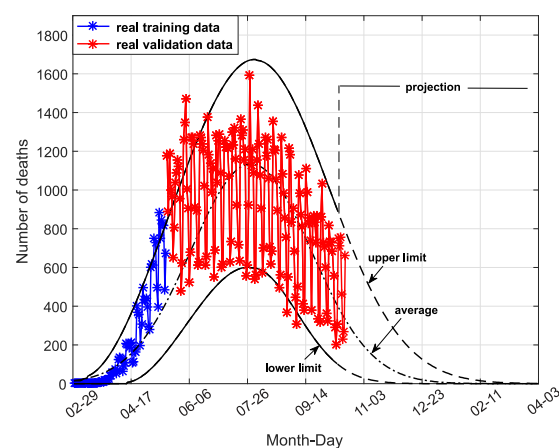
(c) The performance of interval type-2 fuzzy Kalman filter based on its recursive updating on 23 of July 2020 for forecasting the daily deaths reports within the period ranging from 24 of July 2020 to 14 of August 2020.



(d) The performance of interval type-2 fuzzy Kalman filter based on its recursive updating on 28 of August 2020 for forecasting the daily deaths reports within the period ranging from 29 of August 2020 to 25 of September 2020.



(e) The performance of interval type-2 fuzzy Kalman filter based on its recursive updating on 25 of September 2020 for forecasting the daily deaths reports within the period ranging from 26 of September 2020 to 13 of October 2020.



(f) The performance of interval type-2 fuzzy Kalman filter based on its recursive updating on 13 of October 2020 for forecasting the daily deaths reports within the period ranging from 14 of October 2020 to ahead.

**Fig. 4.** Performance of the interval type-2 fuzzy Kalman filter for adaptive tracking and real time forecasting the COVID-19 dynamic propagation related to daily deaths reports: (a) updating based on training data from 29 of February 2020 to 18 of May 2020; (b) recursive updating on 24 of June 2020; (c) recursive updating on 23 of July 2020; (d) recursive updating on 28 of August 2020; (e) recursive updating on 25 of September 2020; (f) recursive updating on 13 of October 2020.

TABLE I

COMPARATIVE ANALYSIS BETWEEN THE INTERVAL TYPE-2 FUZZY KALMAN FILTER AND APPROACH IN [18] FOR FORECASTING THE COVID-19 DYNAMIC PROPAGATION IN BRAZIL

Methodology	RMSE	MAE	RMSPE	R <sup>2</sup>	MAD	MAPE(%)
approach in [18]	0.006190	0.004880	0.2359	0.999450	0.176	0.00745
interval type-2 fuzzy Kalman filter	0.003388	0.000701	0.00339	0.999677	0.1655	0.00701

TABLE II

COMPARATIVE ANALYSIS BETWEEN THE INTERVAL TYPE-2 FUZZY KALMAN FILTER AND APPROACH IN [19] FOR FORECASTING THE COVID-19 DYNAMIC PROPAGATION IN BRAZIL

Methodology	RMSE	MAE	RMSPE	R <sup>2</sup>	MAD	MAPE(%)
approach in [19]	922.83	170.77	0.00407	0.609	33614	3.701
interval type-2 fuzzy Kalman filter	563.15	104.21	0.002486	0.998	97	0.0025494

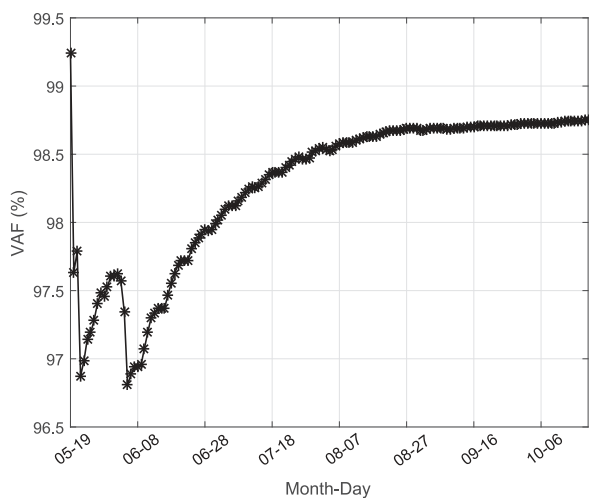


Fig. 5. Efficiency of interval type-2 fuzzy Kalman filter, in tracking and forecasting the COVID-19 dynamic propagation within period ranging from 18 of May to 20 of October 2020.

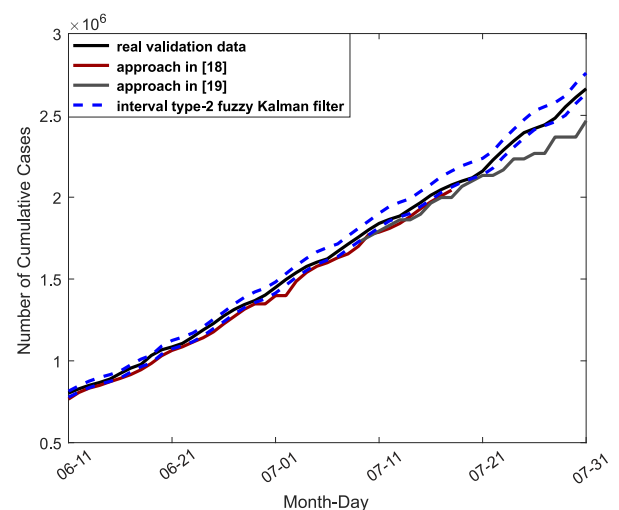


Fig. 6. Comparative analysis of the prediction results between approaches [18], [19] and proposed interval type-2 fuzzy Kalman filter, according to the number of cumulative cases of COVID-19 in Brazil.

once that the approach in [18] uses different types of wavelets to process non-stationarity of experimental dataset, it presents competitive results compared to interval type-2 fuzzy Kalman filter, but the performance is slightly inferior due to its computing limitation from determination of the optimal number of nodes in the hidden layer of the WCRVFL network, tuning the scaling of the uniform randomization range for wavelet estimator and accurate data availability. The approach in [19] is based on ARIMA model for forecasting the COVID-19 dynamic propagation in Brazil. The efficiency of interval type-2 fuzzy Kalman filter, compared to approach in [19], is shown in Table II. The approach in [19] is fast in processing speed but presents performance more inferior than interval type-2 fuzzy Kalman filter due to consider only linear characteristics for modeling the COVID-19 dynamic propagation, which tends to increase forecasting errors in the time varying epidemiological data [20]. Considering the prediction results available in approaches [18], [19], in the sense of clearly and intuitively illustrates the prediction performance of each method as compared to performance of proposed interval

type-2 fuzzy Kalman filter, according to the number of cumulative cases of COVID-19 in Brazil, a comparative analysis is shown in Fig. 6.

The comparative analysis between the interval type-2 fuzzy Kalman filter and the machine learning models LASSO, ARIMA and LSTM recurrent neural network, for forecasting the COVID-19 dynamic propagation in Brazil, within the horizon of 10 days, is shown in the Table III. A possible limitation of the interval type-2 fuzzy Kalman filter is the determination of some parameters ( $\gamma$ ,  $\beta$  and  $q$ ), which requires some intuition by expert and depends on the experimental dataset. The parameters  $\gamma$  and  $\beta$  are related to dimension and rank of Hankel matrix in Eq. (29) so that a good conditioning can be guaranteed for parametric estimation of consequent proposition in interval type-2 fuzzy Kalman filter, whose most typical values are into the interval  $\frac{N_b}{10} < \gamma, \beta < \frac{N_b}{2}$ , where  $N_b$  is the length of experimental dataset [17]. The numerical value of  $q$  is related to the most representative factors of impulse response from experimental dataset and most typical values are into the interval  $1 \leq q \leq 10$  [21].

TABLE III

COMPARATIVE ANALYSIS BETWEEN THE INTERVAL TYPE-2 FUZZY KALMAN FILTER AND MACHINE LEARNING MODELS LASSO, ARIMA AND LSTM RECURRENT NEURAL NETWORK FOR FORECASTING THE COVID-19 DYNAMIC PROPAGATION IN BRAZIL

Model	RMSE	MAE	RMSPE	R <sup>2</sup>	MAD	MAPE(%)
LASSO	191.5600	163.4643	2.5081	0.7652	234.1	0.1283
ARIMA	203.0832	140.9561	2.6938	0.3680	245.7	0.1106
LSTM	116.0065	94.9875	1.3265	0.8290	193.4	0.0746
interval type-2 fuzzy Kalman filter	10.3586	3.9000	1.0542	0.9984	108.3	0.0427

#### IV. CONCLUSION

The results shown the applicability of machine learning approach based on interval type-2 fuzzy Kalman filter due to its recursive updating mechanism, for adaptive tracking and real time forecasting the COVID-19 dynamic propagation. For further works, the formulation and applicability of proposed methodology in the context of evolving interval type-2 fuzzy systems, is of particular interest.

#### ACKNOWLEDGMENT

The authors are grateful to Coordination for the Improvement of Higher Education Personnel (CAPES) and to the Master and Doctorate Program in Electrical Engineering at the Federal University of Maranhão (PPGEE-UFMA).

#### REFERENCES

- [1] D. S. Pires and G. L. O. Serra, "Methodology for evolving fuzzy kalman filter identification," *Int. J. Control Automat. Syst.*, vol. 17, no. 3, pp. 793–800, Feb. 2019.
- [2] I. Eyoh, R. John, G. D. Maere, and E. Kayacan, "Hybrid learning for interval type-2 intuitionistic fuzzy logic systems as applied to identification and prediction problems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 2672–2685, Oct. 2018.
- [3] P. Gil, T. Oliveira, and L. Palma, "Adaptive neurofuzzy control for discrete-time nonaffine nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 8, pp. 1602–1615, Aug. 2019.
- [4] X. Wang, Z. Xu, X. Gou, and L. Trajkovic, "Tracking a maneuvering target by multiple sensors using extended kalman filter with nested probabilistic-numerical linguistic information," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 2, pp. 346–360, Feb. 2020.
- [5] R. M. Asl, R. Palm, H. Wu, and H. Handroos, "Fuzzy-based parameter optimization of adaptive unscented kalman filter: Methodology and experimental validation," *IEEE Access*, vol. 8, no. 1, pp. 54887–54904, 2020.
- [6] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, and D. Liu, "Early prediction of the 2019 novel Coronavirus outbreak in the mainland China based on simple mathematical model," *IEEE Access*, vol. 8, no. 1, pp. 51761–51769, 2020.
- [7] Q. Lin *et al.*, "A conceptual model for the Coronavirus disease 2019 (COVID-19) outbreak in wuhan, china with individual reaction and governmental action," *Int. J. Infect. Dis.*, vol. 93, no. 1, pp. 211–216, Apr. 2020.
- [8] N. Chintalapudi, G. Battineni, and F. Amenta, "COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in italy: A. data driven model approach," *J. Microbiol. Immunol. Infect.*, vol. 53, no. 3, pp. 396–403, Jun. 2020.
- [9] S. Kim, Y. B. Seo, and E. Jung, "Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes," *Epidemiol. Health*, Apr. 2020, Art. no. e2020026.
- [10] S. W. Park, D. M. Cornforth, J. Dushoff, and J. S. Weitz, "The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak," *Epidemics*, vol. 31, Jun. 2020, Art. no. 100392.
- [11] Y. Huang *et al.*, "SARS-CoV-2 viral load in clinical samples from critically ill patients," *Amer. J. Respir. Crit. Care Med.*, vol. 201, no. 11, pp. 1435–1438, Jun. 2020.
- [12] K. Kanagarathinam and K. Sekar, "Estimation of reproduction number and early prediction of 2019 Novel Coronavirus Disease (COVID-19) outbreak in India using statistical computing approach," *Epidemiol. Health*, May 2020, Art. no. e2020028.
- [13] N. Piovella, "Analytical solution of SEIR model describing the free spread of the COVID-19 pandemic," *Chaos Solitons Fractals*, vol. 140, Nov. 2020, Art. no. 110243.
- [14] R. Takele, "Stochastic modelling for predicting COVID-19 prevalence in east africa countries," *Infect. Dis. Modelling*, vol. 5, no. 1 pp. 598–607, 2020.
- [15] C. A. Varotsos and V. F. Krapivin, "A new model for the spread of COVID-19 and the improvement of safety," *Saf. Sci.*, vol. 132, Dec. 2020, Art. no. 104962.
- [16] J. B. Elsner, "Analysis of time series structure: SSA and related techniques," *J. Amer. Statist. Assoc.*, vol. 97, no. 460, pp. 1207–1208, Dec. 2002.
- [17] J. N. Juang, *Applied System Identification*. Hoboken, NJ, USA: Prentice Hall, 1994.
- [18] B. B. Hazarika and D. Gupta, "Modelling and forecasting of COVID-19 spread using wavelet-coupled random vector functional link networks," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106626.
- [19] A. K. Sahai, N. Rath, V. Sood, and M. P. Singh, "ARIMA modelling & forecasting of COVID-19 in top five affected countries," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 1419–1427, Sep. 2020.
- [20] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, no. 1 pp. 159–175, Jan. 2003.
- [21] F. M. Callier and C. A. Desoer, "Realization theory," in *Springer Texts in Electrical Engineering*. Berlin, Germany: Springer, 1991.