

Article

On Comparing Cross-Validated Forecasting Models with a Novel Fuzzy-TOPSIS Metric: A COVID-19 Case Study

Dalton Garcia Borges de Souza ^{1,2,3,*} , Erivelton Antonio dos Santos ^{4,5}  and Francisco Tarcísio Alves Júnior ^{6,7} and Mariá Cristina Vasconcelos Nascimento ^{2,3} 

- ¹ Institute of Science and Technology, Fluminense Federal University, Rio das Ostras 28890-000, Brazil
 - ² Division of Computer Science, Aeronautics Institute of Technology, São José dos Campos 12228-900, Brazil; mariah@ita.br
 - ³ Institute of Science and Technology, Federal University of Sao Paulo, São José dos Campos 12247-014, Brazil
 - ⁴ Institute of Industrial Engineering and Management, Federal University of Itajubá, Itajubá 37500-903, Brazil; d2020101864@unifei.edu.br
 - ⁵ Department of Administration Course, José do Rosário Vellano University, Alfenas 37132-440, Brazil
 - ⁶ Collegiate of Industrial Engineering, University of Amapá State, Macapá 68900-070, Brazil; francisco.junior@ueap.edu.br
 - ⁷ Post-Graduate Program in Intellectual Property and Technology Transfer for Innovation, Federal University of Macapá, Macapá 68903-419, Brazil
- * Correspondence: daltonborges@id.uff.br



Citation: de Souza, D.G.B.; dos Santos, E.A.; Alves Júnior, F.T.; Nascimento, M.C.V. On Comparing Cross-Validated Forecasting Models with a Novel Fuzzy-TOPSIS Metric: A COVID-19 Case Study. *Sustainability* **2021**, *13*, 13599. <https://doi.org/10.3390/su132413599>

Academic Editor: Edmundas Kazimieras Zavadskas

Received: 4 November 2021
Accepted: 7 December 2021
Published: 9 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Time series cross-validation is a technique to select forecasting models. Despite the sophistication of cross-validation over single test/training splits, traditional and independent metrics, such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), are commonly used to assess the model's accuracy. However, what if decision-makers have different models fitting expectations to each moment of a time series? What if the precision of the forecasted values is also important? This is the case of predicting COVID-19 in Amapá, a Brazilian state in the Amazon rainforest. Due to the lack of hospital capacities, a model that promptly and precisely responds to notable ups and downs in the number of cases may be more desired than average models that only have good performances in more frequent and calm circumstances. In line with this, this paper proposes a hybridization of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and fuzzy sets to create a similarity metric, the closeness coefficient (CC), that enables relative comparisons of forecasting models under heterogeneous fitting expectations and also considers volatility in the predictions. We present a case study using three parametric and three machine learning models commonly used to forecast COVID-19 numbers. The results indicate that the introduced fuzzy similarity metric is a more informative performance assessment metric, especially when using time series cross-validation.

Keywords: fuzzy sets; TOPSIS; multicriteria decision making; decision support systems; forecasting

1. Introduction

By 27 October 2021, almost two years after the initial occurrence of SARS-COV-2, the World Health Organization (WHO) announced a total of 219.4 million cases worldwide and 5 million accumulated deaths due to coronavirus disease [1]. Indeed, by 22 November 2021, some countries in Europe have announced a partial or complete lockdown aimed at overcoming the infections spread across Europe, notwithstanding sustainability economy problems [2]. After 80.2 thousand confirmed cases in the world in almost two months, Brazilian authorities stated the SARS-COV-2's primary infection on 25 February 2020 [3]. After a lag of two months, Brazil saw its initial pandemic numbers soar. At the end of October 2021, Brazil had the third-largest number of confirmed cases globally (21.75 million) and the second-highest number of deaths (606 thousand). Moreover, the number of daily

new cases and deaths started decreasing only in June 2021 after mass vaccinations took effect [4].

In Brazil, huge cities, such as São Paulo and Rio de Janeiro, attracted media attention mainly due to their population, economic concentration, and the consequent dimension of SARS-COV-2 numbers. Nevertheless, the pandemic affected even more Brazilian regions, such as the North, which is covered mainly by the Amazon forest and holds about half of the Brazilian area. Although the population density (4.78 inh./km²) and concentration (8.8%) in the North region are low, by the end of September 2021, it was responsible for 30.3% of all Brazilian SARS-COV-2 confirmed cases [5]. Moreover, the death risk by standardized rates in the North region capitals was significantly higher [6,7] than in the rest of the country, mainly due to the poor sanitary and social conditions.

In the North region of Brazil lays a state like an island surrounded and carved out of the Amazon rainforest, as it does not hold land traffic routes with any other state of Brazil (see Figure 1). The Amapá state has just 877,613 residents who live in an area larger than England, but it is 67-times denser in England. As a result, Amapá has already experienced an overload of mortality from transmissible infections, predominantly amidst indigenous groups, such as other parts of the Brazilian Amazon [8]. In addition, despite previous government efforts, several social and health challenges remain for the many people residing in the state, such as public sanitation and minimal access to clean water [9]. This particular scenario makes Amapá receptive to SARS-COV-2 and other disease outbreaks that may occur. Until August 2020, according to the state's official data, the state reported at the second-highest Brazilian infection rate [5]. Consequently, in late May 2021, the capital of Amapá, Macapá, suffered the collapse of its healthcare system due to SARS-COV-2.



Figure 1. Amapá, Brazil [8].

Researchers have been presenting several models to assist local authorities in Amapá and worldwide in many fields [10–15]; some of them assisting in forecasting COVID-19 numbers, such as when the outbreak will peak, how long it will last, how many will be infected or die, and how the hospital demands will evolve [16–19]. The models vary from univariate [16,20,21] to multivariate approaches [18,22] and from ex ante to ex post [23–25]. Other variables such as the number of PCR tests, mobility data, meteorological data, and internet activity are also commonly forecasted or used as exogenous variables while predicting others [18,23,24,26].

Regarding the types of models, compartmental models, such as the SIR model (Susceptible-Infected-Removed) [27,28] and its extensions, are the most used in epidemic outbreaks, especially for medium–long-term forecastings [22,29,30]. Nevertheless, short-term forecastings are also important, especially in supporting operational decisions during the COVID-19 pandemic. Thus, classical parametric and machine learning models have also gained space during the pandemic, such as Autoregressive Integrated Moving Average (ARIMA) [21,31–37], Holt–Winters [35–40], Prophet [20,36,40–42], K-Nearest

Neighbors (KNN) Regressor [37,43–45], Random Forest Regressor (RFR) [11,16,46,47], and Support Vector Regressor (SVR) [16,37,40,47–49]. Researchers may also choose two models [40,43–45,47] or more than three models [16,36,37] to make the forecasts.

There are several alternatives to model and forecast continuous time-dependent variables. Consequently, selecting a proper forecasting model is of essential practical importance. Model performance and evaluation are key to assessing the model quality fit, measured by confronting actual values to the predicted ones [35]. To highlight, in the context of COVID-19, researchers have used many metrics to this end, such as the Root Mean Squared Error (RMSE) [33–38,46], Mean Absolute Percentage Error (MAPE) [34–37,39], Mean Absolute Error (MAE) [16,31,48], Mean Square Error (MSE) [40,46,48], Symmetric Mean Absolute Percentage Error (sMAPE) [16,43], Relative Root Mean Squared Error (RRMSE) [43,48], the Adjusted R-squared (R²) score [33,48], the Improvement Percentage (IP) [16,43], the Akaike Information Criterion (AIC) [32,38], and the Bayesian Information Criterion (BIC) [38].

The RMSE is applicable for evaluating the overall accuracy of the forecasts while punishing significant forecast errors in a square order [50]. The MAE is famous because it is straightforward to calculate and understand. However, it can produce biased results by counting meaningful outliers in datasets; the other measure, MSE, has the same limitation as MAE [51–53]. Another widely used evaluation measure is the MAPE due to its benefits of scale-independency and interpretability. However, MAPE has the notable limitation of resulting in infinite or undefined values for zero or close-to-zero values [51,53,54]. Other metrics used by researchers to assess the performance of machine learning methods are R², R² adjusted, precision, recall, F1-score, and accuracy, or the Matthews Correlation Coefficient (MCC) and the area under the receiver operating characteristic (ROC) curve, also known as AUC [44,48,55]. A forecasting method that has the R², R² adjusted, F1-score, or AUC closest to 1 is the one that should be chosen. Higher values for metrics that carry the word “Mean” on their names indicate poor performance of a given algorithm [52,53,55].

Despite the variety of existing metrics, none of them seem to be specially tailored to cross-validation forecasting approaches or to assess multiple forecasted values to each observation in the testing sets. Commonly, metrics that are based only on averages may not explore other information that the multiplicity of forecasted values may bring [56], such as variability and different fitting expectations to each data point [57]. In fact, in practice, frequently used forecasting metrics are also correlated, meaning that the performance of one model according to a given metric will be linearly correlated to other metrics [55].

The objective of this paper is to present a novel metric that, in addition to averaging errors, also deals with heterogeneous fitting expectations, capturing the volatility of the forecasted values during cross-validation. As a consequence, this measure can relatively assess the performance of COVID-19 forecasting models. To do so, we adopt a similarity metric, the closeness coefficient (CC), which we take from the Fuzzy-TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution), an outranking method generally used in the context of Multi-Criteria Decision Making (MCDM). A responsive metric in fitting expectations is the one that can capture different perspectives or decision criteria over a set of competing forecasting models. In our case, the perspectives are different periods of a time series, such as periods of an increase, stability, or a decrease in the number of COVID-19 cases. Furthermore, by using triangular fuzzy sets, the metric can consider the volatility in the forecasted data related to the output of the models. Finally, it is a relative metric that only makes sense when more than one model is at the pool of potential models to be chosen. We exemplify the usage of this metric with the case of selecting COVID-19 forecasting models in the Amapá state, Brazil.

The main justification for using TOPSIS over other well-consolidated MCDM methods is that it operates with the concept of distance to the ideal solution [58], which is a key aspect to the metric we propose. Other distance to ideal solution methods, such as VIKOR (VIseKriterijumska Optimizacija I Kompromisno Resenje, in English Multicriteria Optimization and Compromise Solution), could also be used to the same end, potentially

producing similar results. However, TOPSIS is still preferred due to its easy usability and level of consolidation among scholars.

To the best of our knowledge, this metric has never been presented before to assess forecasting models, especially in the context of COVID-19 or other epidemics. The main contributions of this paper and the proposed metric are summarized as follows:

- We introduce a novel similarity metric that, in addition to averaging errors, can also capture heterogeneous fitting expectations and volatility in the forecasted values.
- The metric is suitable for comparing forecasting models trained and tested with cross-validation techniques, such as rolling forward forecasting.
- The metric may potentially bring positive implications over automated model selection in a robust and sustainable manner.
- The metric also enables the conjunction of data-driven and expert-driven models, such as TOPSIS and machine learning forecasting models, which are not commonly tight-coupled to produce decision support systems [58].
- To exemplify the usage of the metric, we present a case study that compares three classical and three machine learning forecasting models: Holt–Winters, ARIMA, Prophet, KNN, RFR, and SVR.

The proposed metric also has limitations, especially in the face of more traditional metrics that are widely used to compare forecasting models, such as MAE, RMSE, SMAPE, and others.

- Since it is a relative metric, it is not suitable to assess standalone models' performances or compare models that are applied to different time series.
- The metric is more complex and demands more effort to be set than other classical metrics.
- It also requires judgments from decision-makers, making it more time-consuming.

The rest of this paper is organized as follows: Section 2 briefly introduces the TOPSIS, fuzzy-TOPSIS models, and the proposed approach, which is an adaptation of fuzzy-TOPSIS to rank forecasting models. Section 3 presents an actual case study that intends to rank models to the reality of a Brazilian state carved in the Amazon region during the COVID-19 pandemic. In Section 4, we discuss the results and compare the proposed metric to the ones already diffused in the literature. Finally, we provide the conclusions in Section 5.

2. Methodology

This section introduces the models used as the base of the introduced forecasting model performance metric as well as the proposed fuzzy-TOPSIS model to rank forecasting models.

2.1. TOPSIS

The TOPSIS is an MCDM method that calculates the value of a given alternative by distancing it to its ideal and non-ideal concepts (or solutions). Hwang and Yoon [59] presented TOPSIS in 1981, and since then, researchers have been primarily employing it to rank alternatives in different fields of study, such as business [58], construction [60], healthcare [61], and mining [62]. Hwang and Yoon [59] proposed the following step-by-step TOPSIS to select the most suitable alternatives given a list of decision criteria.

Step 1: Let $A = [a_{ij}]_{m \times n}$ be a decision matrix with m rows and n columns. The element in the i -th row and j -th, a_{ij} , indicates a value given to alternative i , regarding a decision criterion j . Consider also an n -dimensional vector C , where each position c_j corresponds to the weight assigned to the decision criterion j , where $\sum_{j=1}^n c_j = 1$.

Step 2: Normalize and weight elements of A by dividing them by the routed summation of the column square values and multiplying them by the corresponding criterion weight, as in Equation (1). Thus, a matrix $A' = [a'_{ij}]_{m \times n}$ with all normalized values is set, as in Equation (2).

$$a'_{ij} = \frac{c_j a_{ij}}{\sqrt{\sum_{i=1}^m a_{ij}^2}} \tag{1}$$

$$A' = \begin{bmatrix} \frac{c_1 a_{11}}{\sqrt{\sum_{i=1}^m a_{i1}^2}} & \frac{c_2 a_{12}}{\sqrt{\sum_{i=1}^m a_{i2}^2}} & \dots & \frac{c_j a_{1j}}{\sqrt{\sum_{i=1}^m a_{ij}^2}} \\ \frac{c_1 a_{21}}{\sqrt{\sum_{i=1}^m a_{i1}^2}} & \frac{c_2 a_{22}}{\sqrt{\sum_{i=1}^m a_{i2}^2}} & \dots & \frac{c_j a_{2j}}{\sqrt{\sum_{i=1}^m a_{ij}^2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{c_1 a_{i1}}{\sqrt{\sum_{i=1}^m a_{i1}^2}} & \frac{c_2 a_{i2}}{\sqrt{\sum_{i=1}^m a_{i2}^2}} & \dots & \frac{c_j a_{ij}}{\sqrt{\sum_{i=1}^m a_{ij}^2}} \end{bmatrix} \tag{2}$$

Step 3: Let V^+ and V^- be m -dimensional vectors corresponding to the Positive and Negative Ideal Solutions, respectively. They are calculated as indicated next.

$$V^+ = \left\{ v_j^+ := \max_i a'_{ij} \mid j = 1, 2, \dots, n \right\} \tag{3}$$

$$V^- = \left\{ v_j^- := \min_i a'_{ij} \mid j = 1, 2, \dots, n \right\} \tag{4}$$

Step 4: Calculate the separation measures d_i^+ and d_i^- between the row vector a'_i and V^+ and V^- , respectively.

$$d_i^+ = \sqrt{\sum_{j=1}^n (a'_{ij} - v_j^+)^2} \tag{5}$$

$$d_i^- = \sqrt{\sum_{j=1}^n (a'_{ij} - v_j^-)^2} \tag{6}$$

Step 5: Calculate the closeness coefficient (CC) as in Equation (7) of each alternative i to the ideal solutions. Then, rank the alternatives by sorting the CC values in their decreasing order.

$$CC_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{7}$$

Under several circumstances, crisp data are inadequate to real-life model situations. Human judgments are often vague and cannot estimate their preference with an accurate numerical value. Thus, we might use linguistic assessments, a more realistic approach, rather than numerical values to produce the classifications and weights of the criteria in the problem [63,64]. Furthermore, crisp averages may lose information carried by a set of samples to the same quantitative observation [57]. To tackle those gaps, researchers have been mainly applying fuzzy set theory to TOPSIS approaches. A positive fuzzy number, \tilde{x} , is characterized by a membership function, $\mu_{\tilde{x}}$, which takes values in the interval $[0, 1]$. It is an extension of classical set theory, and the operations are extensions of the fundamental set theory operations of complement [65,66].

The triangular is the most common membership function in fuzzy-MCDM applications [58]. In contrast to more complex and novel membership functions and types of fuzzy sets, triangular fuzzy numbers are the easiest to learn and understand. Thus, they are traditionally suitable to first fuzzification attempts on novel MCDM methods when uncertainty is primarily tackled [58,66]. For instance, classic MCDM methods, such as AHP [67,68] and TOPSIS [63,69], debut on the fuzzy environment by employing the concept of triangular fuzzy numbers. For a given fuzzy number, \tilde{x} , its associated membership function is defined by a lower limit (x_1), a middle value (x_2), and an upper limit (x_3), with $x_1 \leq x_2 \leq x_3$ (See Figure 2, where:

$$\mu_{\tilde{x}}(z) = \begin{cases} 0, & z < x_1 \\ \frac{z-x_1}{x_2-x_1}, & x_1 \leq z \leq x_2 \\ \frac{x_2-x_3}{z-x_3}, & x_2 \leq z \leq x_3 \\ 0, & z > x_3 \end{cases} \quad (8)$$

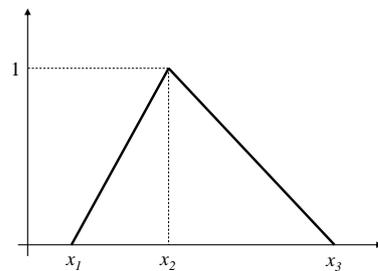


Figure 2. Graphical representation of a triangular membership function.

Chen [63] performed the first extension of TOPSIS with triangular fuzzy numbers. This approach is the baseline to the adaptation of Fuzzy-TOPSIS we propose later on. The procedure of Fuzzy-TOPSIS is similar to classical TOPSIS and is summarized as follows [69].

Step 1: Let $\tilde{A} = [\tilde{a}_{ij}]_{m \times n}$ be the fuzzy decision matrix, with \tilde{a}_{ij} being a triplet $(a_{ij1}, a_{ij2}, a_{ij3})$, where a_{ij1} is the lower limit, a_{ij2} the medium value, and a_{ij3} is the upper value for alternative a_{ij} . Consider also an n -dimensional vector \tilde{C} , where each element \tilde{c}_j is a triplet $(\tilde{c}_{j1}, \tilde{c}_{j2}, \tilde{c}_{j3})$, where \tilde{c}_{j1} , \tilde{c}_{j2} , and \tilde{c}_{j3} are the lower, medium, and upper values for \tilde{c}_j , respectively.

Step 2: Compute the normalized fuzzy decision matrix, $\tilde{R} = [\tilde{r}_{ij}]_{m \times n}$, where each element \tilde{r}_{ij} is calculated according to Equation (9).

$$\tilde{r}_{ij} = \left(\frac{a_{ij1}}{\max_k \{a_{j3}^*\}}, \frac{a_{ij2}}{\max_k \{a_{j3}^*\}}, \frac{a_{ij3}}{\max_k \{a_{j3}^*\}} \right) \quad (9)$$

for $a_{j3}^* = \max_i \{a_{ij3}\}$.

Step 3: Compute the weighted normalized fuzzy decision matrix, $\tilde{V} = [\tilde{v}_{ij}]_{m \times n}$, where each element \tilde{v}_{ij} is calculated as in Equation (10).

$$\tilde{v}_{ij} = \tilde{r}_{ij} \odot \tilde{c}_j \quad (10)$$

Step 4: Compute the Fuzzy Positive (FPIS) and Negative (FNIS) Ideal Solutions.

$$FPIS = (\tilde{v}_1^*, \tilde{v}_2^*, \dots, \tilde{v}_n^*), \text{ where } \tilde{v}_j^* = \max_i \{ \tilde{v}_{ij} \} \quad (11)$$

$$FNIS = (\tilde{v}_1^-, \tilde{v}_2^-, \dots, \tilde{v}_n^-), \text{ where } \tilde{v}_j^- = \min_i \{ \tilde{v}_{ij} \} \quad (12)$$

Step 5: Compute the positive (d_i^*) and negative (d_i^-) distances of each alternative i to its FPIS and FNIS, respectively.

$$d_i^* = \sum_{j=1}^n d(\tilde{v}_{ij}, \tilde{v}_j^*) \quad (13)$$

$$d_i^- = \sum_{j=1}^n d(\tilde{v}_{ij}, \tilde{v}_j^-) \quad (14)$$

where the distance between two generic fuzzy numbers, $\tilde{x}_1 = (l_1, h_1, u_1)$ and $\tilde{x}_2 = (l_2, h_2, m_2)$ (See Figure 3), can be obtained using the vertex method, given as:

$$d(\tilde{x}_1, \tilde{x}_2) = \sqrt{\left(\frac{1}{3}[(l_1 - l_2)^2 + (h_1 - h_2)^2 + (u_1 - u_2)^2]\right)} \tag{15}$$

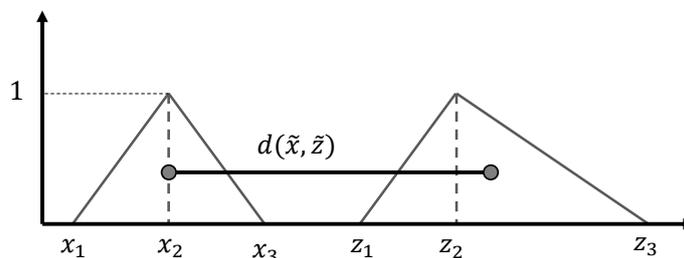


Figure 3. The vertex method: the distance between two TFNs.

Step 6: Compute the closeness coefficients for each alternative i , CC_i , according to Equation (16). Then, sort the alternatives by decreasing value of CC .

$$CC_i = \frac{d_i^-}{d_i^- + d_i^*} \tag{16}$$

2.2. The Proposed Approach

The proposed model is essentially an adaptation of the Fuzzy-TOPSIS model proposed by Chen [63]. In Steps 1–5, it organizes the decision environment and creates data throughout forecasting. To the best of our knowledge, those steps are not found in other TOPSIS-based models. Steps 6–11 directly use the logic and equations proposed by Chen, but we made several adjustments to accommodate the predicted values from the prediction models. We displayed these adjustments following each stage of the proposed model. The step-by-step of the introduced model is given next.

Step 1: Divide the time series observations into two portions: a fixed training set and a fluctuating testing/training set, as represented by Figure 4.

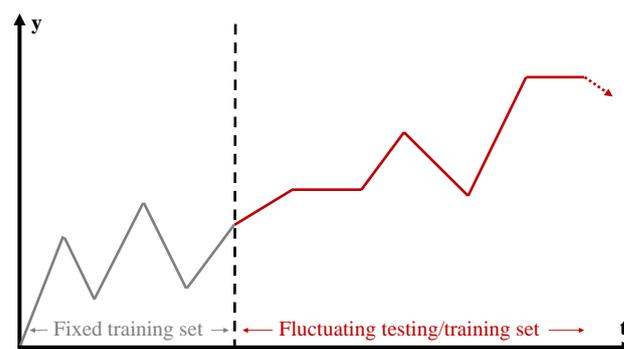


Figure 4. Time series sets.

Step 2: Segment the fluctuating testing/training set in n categories according to decision-makers preferences and/or the time series characteristics. Figure 5 exemplifies $n = 4$. Please notice that non-successive observations may fall into the same category since a time series may display similar behaviors for different periods in time.

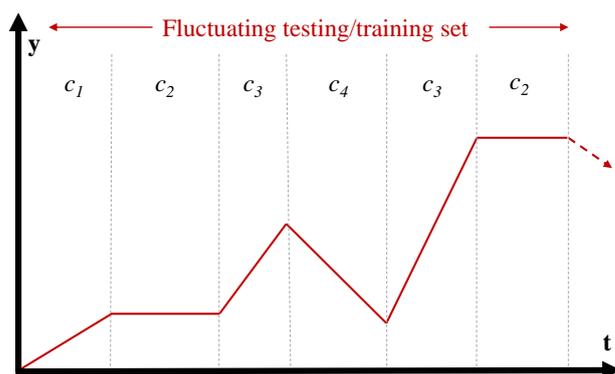


Figure 5. Segmenting and classifying the fluctuating testing/training set.

Step 3: Aggregate the observations according to the m categories and weight the observations inside each category according to the decision-makers preferences. Table 1 shows the weighting scale, containing the conversion from linguistic to crisp and fuzzy weights, used to assess each category. All data points that fall into a category will receive the corresponding weight to that category. Thus, for the p_n observations inside the n categories, their weights are equal to the weight given to the category k , so $c_k = c_1^k, c_2^k, \dots, c_{p_k}^k$. The vector with the weights to all r observations in the fluctuating testing/training set is given in Equation (17):

$$\tilde{C} = [c_i^k]_r = (\tilde{C}_{p_1}^1 \cap \tilde{C}_{p_2}^2 \cap \dots \cap \tilde{C}_{p_n}^n) = [\tilde{c}_1^1, \tilde{c}_2^1, \dots, \tilde{c}_{p_1}^1, \tilde{c}_1^2, \tilde{c}_2^2, \dots, \tilde{c}_{p_2}^2, \tilde{c}_1^n, \tilde{c}_2^n, \dots, \tilde{c}_{p_n}^n] \quad (17)$$

with $r = p_1 + p_2 + \dots + p_n$.

Table 1. Scale to assess the importance of each category.

Linguistic	Crisp	Fuzzy
Low	1	(1,1,2)
Medium Low	2	(1,2,3)
Medium	3	(2,3,4)
Medium High	4	(3,4,5)
High	5	(4,5,5)

In Table 1, the column Linguistic refers to the linguistic weights that the decision-makers may assign to each criterion. The column Crisp refers to the crisp correspondent scale that is commonly used on MCDM methods. The Fuzzy correspondents to the fuzzy scale, which we use to convert the linguistic variable to triangular fuzzy numbers.

Step 4: Select the m forecasting models to be compared and assemble the decision hierarchy (See Figure 6).

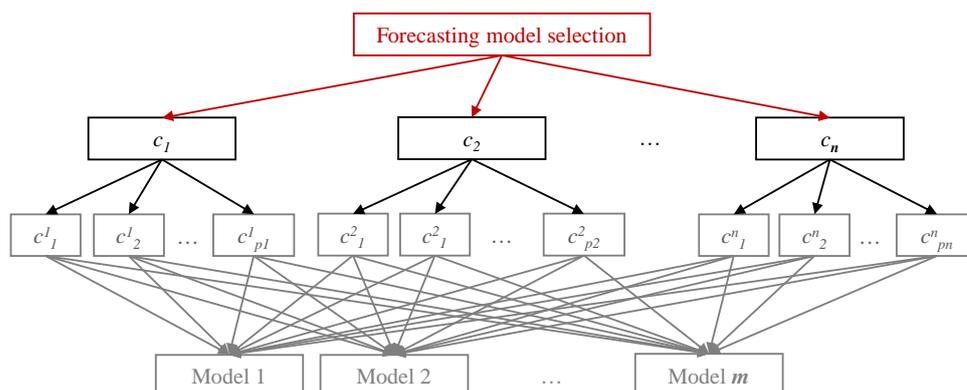


Figure 6. Decision hierarchy—Objective, categories, data points, and models.

Step 5: To the time series defined in Step 1, run the models q time each, where $q = r - w + 1$, with w the forecasting window, which is also the testing set. Thus, to each model run, add one observation to the training set until $r = w$. Figure 7 graphically exemplifies how the training set evolves in time and also how the testing set gives one step ahead in its origin with each new run in the cross-validation.

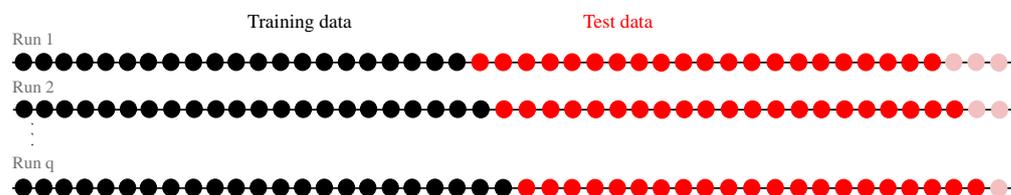


Figure 7. Forecasting window.

Step 6: To observe the fluctuating testing/training set, assemble a triangular fuzzy number, \tilde{e}_{ij}^k , which is also composed by a triplet, with $k = 1, 2, \dots, n$ and where e_{ij1}^k is the smallest residual, e_{ij2}^k is the median of all residuals to a given observation, and e_{ij3}^k is the biggest residual of the forecasting to the r observations and m forecasting models. For the first and last observations of the fluctuating testing/training set, $e_{ij1}^k = e_{ij2}^k = e_{ij3}^k$. Thus, the residual fuzzy matrix, aggregated according to n categories, $\tilde{E}_{r \times m}$, for r observations and m models is given by:

$$\tilde{E} = [\tilde{e}_{ij}^k]_{r \times m} = \begin{bmatrix} \tilde{e}_{11}^1 & \tilde{e}_{12}^1 & \dots & \tilde{e}_{1m}^1 \\ \tilde{e}_{21}^1 & \tilde{e}_{22}^1 & \dots & \tilde{e}_{2m}^1 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{e}_{p_1 1}^1 & \tilde{e}_{p_1 2}^1 & \dots & \tilde{e}_{p_1 m}^1 \\ \tilde{e}_{11}^2 & \tilde{e}_{12}^2 & \dots & \tilde{e}_{1m}^2 \\ \tilde{e}_{21}^2 & \tilde{e}_{22}^2 & \dots & \tilde{e}_{2m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{e}_{p_2 1}^2 & \tilde{e}_{p_2 2}^2 & \dots & \tilde{e}_{p_2 m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{e}_{11}^n & \tilde{e}_{12}^n & \dots & \tilde{e}_{1m}^n \\ \tilde{e}_{21}^n & \tilde{e}_{22}^n & \dots & \tilde{e}_{2m}^n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{e}_{p_n 1}^n & \tilde{e}_{p_n 2}^n & \dots & \tilde{e}_{p_n m}^n \end{bmatrix}_{r \times m} \tag{18}$$

Step 7: Compute the normalized residual fuzzy matrix, $\tilde{G} = [\tilde{g}_{ij}^k]_{r \times m}$, just as presented by Chen [63], where each element \tilde{g}_{ij}^k , for $k = 1, \dots, n; i = 1, \dots, p_k; j = 1, \dots, m$ is given by:

$$\tilde{g}_{ij}^k = \left(\frac{e_{ij1}^k}{e_{j3}^*}, \frac{e_{ij2}^k}{e_{j3}^*}, \frac{e_{ij3}^k}{e_{j3}^*} \right), \tag{19}$$

with the lower, medium, and upper values being, respectively, e_{ij1}^k, e_{ij2}^k , and e_{ij3}^k , coming from matrix \tilde{E} and $e_{j3}^* = \max_{i,k} \{e_{ij3}^k\}$.

Step 8: Construct the weighted normalized fuzzy decision matrix, $\tilde{V} = [\tilde{v}_{ij}^k]_{r \times n}$, where $\tilde{v}_{ij}^k = \tilde{g}_{ij}^k \odot \tilde{c}_i^k$.

Step 9: Compute the Fuzzy Positive Ideal Solution (FPIS) and Fuzzy Negative Ideal Solution (FNIS). We assume that the best forecasting model to a given observation is the one closer in distance to its FPIS and farthest to its FNIS. Thus, the FPIS is the minimal

possible residual, and the FNIS is formed by the maximum residuals among all forecasting models for the given observation.

$$\text{FPIS} = 0 \quad (20)$$

$$\text{FNIS} = (\tilde{v}_1^{1-}, \tilde{v}_2^{1-}, \dots, \tilde{v}_{p_1}^{1-}, \tilde{v}_1^{2-}, \tilde{v}_2^{2-}, \dots, \tilde{v}_{p_2}^{2-}, \dots, \tilde{v}_1^{n-}, \tilde{v}_2^{n-}, \dots, \tilde{v}_{p_n}^{n-}) \quad (21)$$

where $\tilde{v}_i^{k-} = \max_j \{\tilde{v}_{ij}^k\}$.

Step 10: Compute the distance of each model predicted observation from each FPIS, d_j^* , and FNIS, d_j^- , using Equations (22) and (23). We calculate the distances between two Triangular Fuzzy Numbers using the vertex method (See, Figure 3 and Equation (15)).

$$d_j^* = \sum_{k=1}^n \sum_{i=1}^{p_k} d(\tilde{v}_{ij}^k, \text{FPIS}) \quad (22)$$

$$d_j^- = \sum_{k=1}^n \sum_{i=1}^{p_k} d(\tilde{v}_{ij}^k, \tilde{v}_i^{k-}) \quad (23)$$

Step 11: Compute the closeness coefficient for each forecasting model j , CC_j , according to Equation (24) for each forecasting model and rank them from highest closeness coefficient to the lowest.

$$CC_j = \frac{d_j^-}{d_j^- + d_j^*} \quad (24)$$

3. Case Study: Forecasting in the Amazon Region

In this section, we exemplify the usage of the proposed metric with the case study of a Brazilian state, Amapá. We first present the data we use and how we collected them. Then, we briefly introduce the forecasting protocol and the models used to predict future observations of the target variable. Finally, we evaluate the models according to the similarity metric we introduced in Section 2.2.

3.1. Data Acquisition

We performed all modeling to the daily number of confirmed COVID-19 cases in the Amapá state, fully located in the Brazilian Amazonian region. The number of observations/timestamps in this case study since the first official case, in 20 March 2020, up to 28 September 2021, is 558. We gather the data from official reports at the state level. The collected data are also available in an application programming interface provided by Brasil.io repository [5], where the dataset is named "caso" and is presented under the "COVID-19" section.

The data we use may diverge from the Brazilian government's website as the counting protocol may differ from the state of Amapá. Additionally, this paper does not treat cases of sub-notifications, and the target variable is forecasted by the models only considering as predictors lagged values. All data we use, as well as the results from subsequent sections of this paper, can be found in the Supplemental File.

We divided the time series into two parts, respecting Step 1 of the methodology we proposed. Let the fixed training set be denoted by A and the fluctuating testing/training set be denoted by B . From the 558 observations/days, A encompasses 238 observations, while B encompasses 320. Figure 8 shows A and B , as well as their centered moving average, for 42 days.

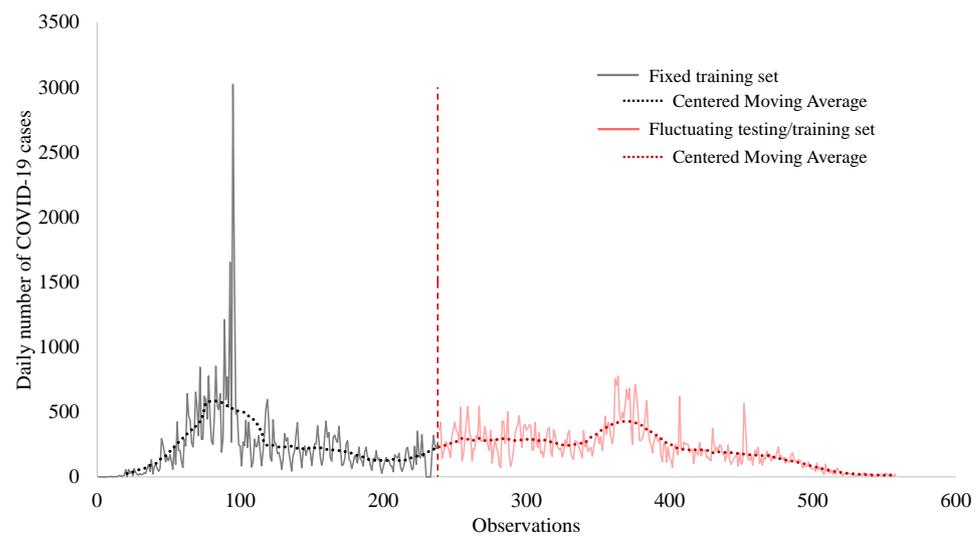
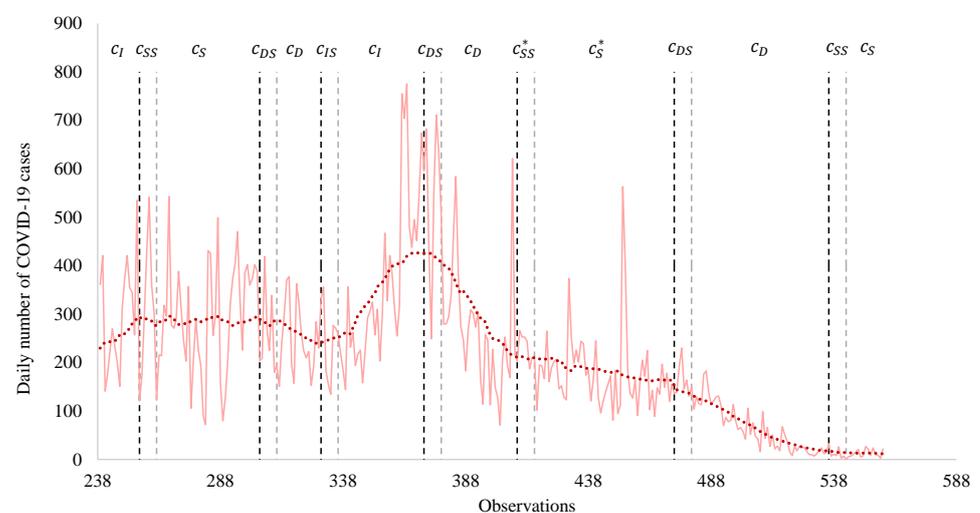


Figure 8. Observations for the case study.

Furthermore, in respect to the proposed methodology's Step 2, we classify different segments of B , according to six categories:

- Stability Start (C_{SS}): the first week (7 days) of a stability period;
- Stability (C_S): a period with no significant increase or decrease in the tendency of the time series;
- Increasing Start (C_{IS}): the first week of an increasing period;
- Increasing (C_I): a period of a significant increase in the tendency of the time series;
- Decreasing Start (C_{DS}): the first week of a decreasing period;
- Decreasing (C_D): a period of a significant decrease in the tendency of the time series.

These categories reflect local decision-makers' concerns, which also weigh the importance of each category according to their needs. For instance, better predictions during periods with an increase in the number of daily COVID-19 confirmed cases are the most appreciated since those times may require more infrastructure investments or negotiation with other regions to transfer the excess of new patients. Similarly, but less critically, are the predictions during decreasing periods since during these times, the state may be open to receiving patients from other regions or temporarily close temporary facilities, such as campaign hospitals. Figure 9 shows the splitting of B into categories.



* Periods are in stability categories due to their small angle in tendency

Figure 9. Observation categories defined by the decision-makers.

Then, following Step 3, Table 2 presents the linguistic grades given by the decision-makers to each of the categories. The column Category refers to each category defined recognized by the decision-makers, Code refers to its abbreviation, and Scale shows the verbal scale assigned, according to Table 1.

Table 2. Scale to access the importance of each category

Category	Code	Scale
Stability Start	C_{SS}	Low
Stability	C_S	Medium
Increasing Start	C_{IS}	Medium High
Increasing	C_I	High
Decreasing Start	C_{DS}	Medium High
Decreasing	C_D	Medium

3.2. Forecasting Protocol

Thus, the first forecasting run uses all the fixed training set A as the training set and predicts the first 21 observations of B . Then, we perform the walking forward over the fluctuating testing/training B , now using 239 observations (238 from A and 1 from B), and attempt to predict from the 2nd to the 22th observation of B , as shown in Figure 7. Finally, we repeat this procedure until the training set encloses 537 observations (238 from A and 299 from B), and the last 21 observations of B are the testing set. Therefore, we have only 1 prediction to the first and last observations of B , 2 predictions to the 2nd first and 2nd last observations of B , ..., 20 predictions to the 20th and 301th observations of B , and 21 predictions to all observations of B between positions 21th and 300th. Figure 10 shows a flowchart describing the adopted forecasting protocol.



Figure 10. Forecasting protocol.

3.3. Forecasting Models

We run six forecasting models according to the forecasting protocol defined in Section 3.2 and take the data presented in Section 3.1 as input. First, we select the six forecasting models commonly used for predicting the number of COVID-19 cases. Once the models are defined, we complete the decision hierarchy, as prescribed by Step 4 in the proposed methodology subsection (see Figure 11).

For the forecasting in Figure 9, we split and started the time series from points 1 (equal to the point 238 original) to 320 (equal to 558) included in the forecasting window, as shown in Table 3, which shows the defined categories, their correspondent codes, and the data points included in the fluctuating testing/training set, that fall inside each one of the categories.

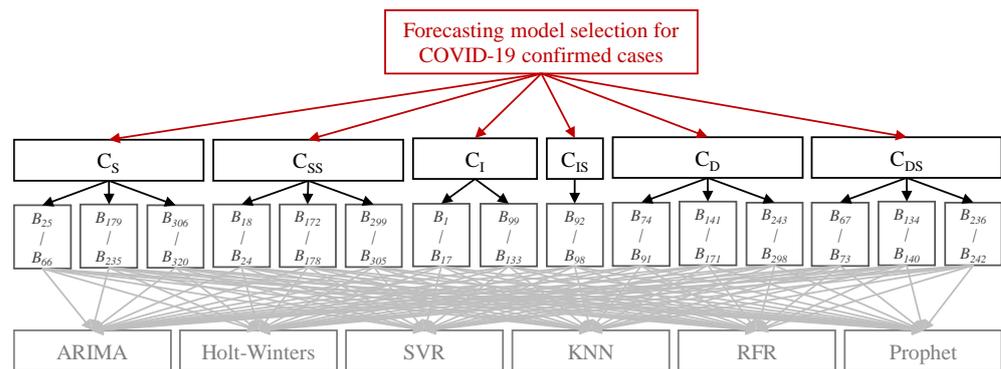


Figure 11. The case decision hierarchy.

Table 3. Categories and forecasting widows.

Categories	Code	Points Included in the Forecasting Widow
Stability Start	C_{SS}	18–24; 172–178; 299–305
Stability	C_S	25–66; 179–235; 306–320
Increasing Start	C_{IS}	92–98
Increasing	C_I	1–17; 99–133
Decreasing Start	C_{DS}	67–73; 134–140; 236–242
Decreasing	C_D	74–91; 141–171; 243–298

As proposed in Step 5 of the proposed methodology, for each model, we perform a total of 320 runs. Figure 12 presents a graphical summary of the mean predicted value for each observation in the fluctuating testing/training set.

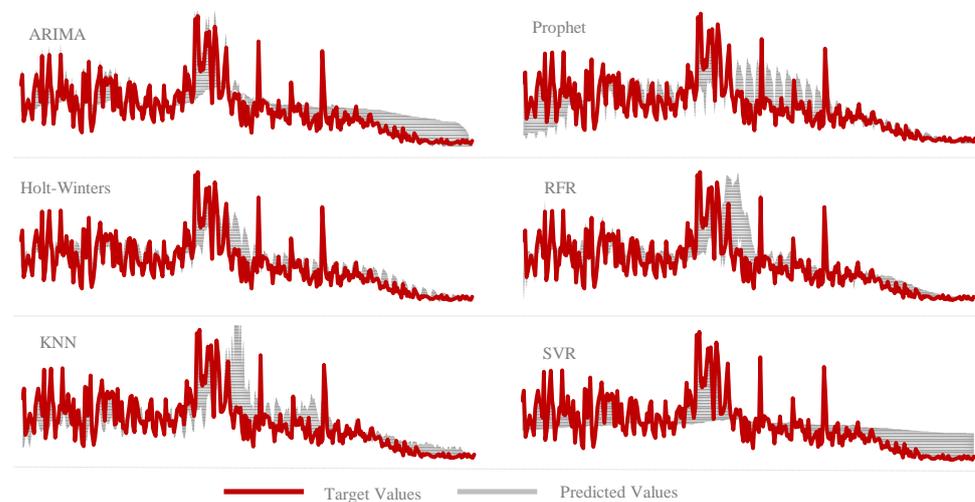


Figure 12. Prediction achieved by the forecasting models.

The parameters and hyperparameters are tuned to each new run of the models. Thus, we do not list the optimal combination of parameters to each one of the models for each new run during the cross-validation. The following section briefly presents the six forecast models used in this research.

3.3.1. ARIMA

ARIMA, also known as the Box–Jenkins model [70], is a statistical approach commonly used for time series analysis and forecasting. The model’s composition of integration (I), autoregressive (AR), and moving average (MA) comprises the ARIMA model. Time series components consist of trend, seasonal, cyclic, and random or irregular movement

categories [35]. ARIMA can additionally be set to recognize seasonality, the optimal value of which can be found after running a Canova–Hansen test [71].

The ARIMA model is commonly referred to as an ARIMA (p, d, q) , where p is the order of autoregression, d is the degree of difference, and q is the order of the moving average [31]. The optimum values of autoregressive (p), degree of their differences (d), and moving average (q) may also be found by search-grid. Generally, the chosen parameter values are those that minimize the Information Criterion (AIC). Benvenuto et al. [21], Ceylan [31], and Singh et al. [32] present examples of the ARIMA applicability to forecast the number of COVID-19 cases. The general equations for AR and MA models are [31]:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (25)$$

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (26)$$

where Y_t , ε_t , ϕ , and θ are the observed values at time t , the value of the random shock at time t , AR, and MA parameters, respectively. Thus, an ARIMA model is given by:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (27)$$

where α is a constant. When dealing with non-stationarity, the data may be differentiated first, and the ARIMA model is then performed.

3.3.2. Holt–Winters

Holt [72] and Winters [73] are the forerunners of the Holt–Winters method, likewise acknowledged as triple exponential smoothing. The Holt–Winters method is an improved version of the simple exponential smoothing model to recognize trend and seasonality in a time series. The method has three parameters: α , the smoothing factor, β , a trend smoothing parameter, and γ , which relates to seasonality. Numerous authors have used this model to forecast the number of COVID-19 cases [38,39].

The two literature Holt–Winters models use additive or multiplicative settings based on the seasonal component. The additive model can be applied with a linear trend and an exponential trend. Moreover, the Holt–Winters additive model is suitable for data with trends and seasonality that do not grow over time [35,36]. The equations of the additive model are as follows:

$$S_t = \alpha \frac{y_t}{I_t - L} + (1 - \alpha)(S_{t-1} + b_{t-1}), \quad (28)$$

where t indicates a given period, S_t is the smoothed observation (level) at period t , L the cycle length, α is the smoothing parameter of level, and y_t the value in t for a target variable. The trend factor (b_t), the seasonal index (I_t), and the forecast at m steps ($F_t + m$) are given by Equations (29)–(31), respectively.

$$b_t = \beta(S_t - S_{t-1}) + (1 - \beta)b_{t-1} \quad (29)$$

$$I_t = \gamma \frac{y_t}{S_t} + (1 - \gamma)I_{t-L+m} \quad (30)$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (31)$$

where β and γ are the smoothing parameters of trend and season, sequentially.

3.3.3. Support Vector Regression (SVR)

The Support Vector Machine (SVM) is a supervised machine learning technique adopted for regression, classification proposals, and time series data forecasts [37]. Vapnik [74] is the vanguard of this technique, as well as its regression variant, the support vector regression (SVR), which was vastly broadcasted by the work of Drucker et al. [75]. We can find some employment of SVR [16,48,49] in the COVID-19 forecasting context. The common logic of an SVR is moderately simple. For example, assume a linear regression, which aims to minimize the sum of square errors:

$$\text{Minimize } f(x) = \sum_{i=1}^n (y_i - w_i x_i)^2 \quad (32)$$

where y_i is the target, w_i the coefficient, and x_i the feature. Then, the SVR training intends to minimize the following system.

$$\text{Minimize } f(x) = \frac{1}{2} \|w\|^2 \quad (33)$$

$$\text{Subject to } g(x) = |y_i - w_i x_i + b_i| \leq \varepsilon \quad (34)$$

where b_i is a linear coefficient and ε is the error. Cost and Kernel are two examples of hyperparameters usually tuned in this algorithm.

3.3.4. K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) technique is a nonparametric and lazy learning classification method [37,44]. Initially, the KNN was designed to accomplish classification problems. Notwithstanding, decades after the KNN's original conceptualizations, around the early 1990s, researchers began investigating it for regression goals [76]. Instead of learning the training dataset, KNN does not need a training phase and holds the training dataset [44].

The KNN algorithm explores the k nearest past comparable values (nearest neighbors) by minimizing a similarity measure in a time series context [37]. Later, the forecasting is an average of these K -nearest neighbors. Moreover, the KNN gives the smallest similarity measure within the past and new cases [37]. Although it sounds straightforward, it requires a significant computational cost [43]. In the COVID-19 environment, several researchers have employed this approach in classification problems. Nevertheless, only some have applied it to forecast the number of COVID-19 cases [43]. For this algorithm, the number of neighbors is the most common hyperparameter to be tuned. The central distance functions employed for continuous variables are:

$$\text{Minkowski, } d(x,y) = \left(\sum_{i=1}^k (|x_i - y_i|^q) \right)^{1/q} \quad (35)$$

where k refers to the number of samples. When $q = 1$, the metric produces the Manhattan distance, whereas when $q = 2$, one has the Euclidean distance (both frequently used distance metrics for this end).

3.3.5. Random Forest Regression (RFR)

The Random Forest (RF) approach is a machine learning algorithm with several decision trees proposed by Breiman [77], which is a compound of bagging and random subspaces methods. Currently, practitioners of machine learning and researchers apply the RF approach in regression and classification assignments. For example, the authors [11,16,46,47] have employed the RF approach to deal with COVID-19 forecasting.

In the RF algorithm, initially, data are randomly split into two parts: training data (the in-Bag) for learning and validation (the out-of-Bag data) for the testing learning levels.

Next, the algorithm randomly creates many decision trees with “boot-strap samples” from the data [11,44,46]. Subsequently, the randomly selected predictors define the branching of each tree at node points. Lastly, the results from each tree average are the final RF estimation [44,46]. It is remarkable when applied with the randomness of the time series [16]. As for dividing criteria in each tree’s branch in regression applications, we use the Mean Squared Error (MSE). For RFR, the number of estimators and the maximum tree depth are the most common hyperparameters tuned.

3.3.6. Prophet

The Facebook team released a decomposed model called Prophet for forecasting, which is an open-source library [40]. It applies a decomposable times series model, with three central model elements: the trend function ($g(t)$), the periodical function ($s(t)$), and the holidays ($h(t)$). It also appropriates an error ϵ_t if the model does not predict any abnormal changes.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (36)$$

where

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)\gamma) \quad (37)$$

In Equation (37), k is the increase rate, δ is the rate arrangements, m is the offset parameter, and γ_j is set to $s_j \delta_j$ to make a continuous function. An extra important feature is that the model does automated changepoint election, setting a sparse prior on δ .

On the other hand, it relies on the Fourier series to incorporate daily, weekly, and annual seasonalities. In the case of COVID-19, we are more concerned about weekly seasonality [8].

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right) \quad (38)$$

For instance, Prophet has few occurrences in forecasting the death and the accumulated cases confirmed in the COVID-19 context [20,41].

3.4. Model Evaluation and Comparison

After running each model 320 times, each time, one step ahead over the fluctuating testing/training set (see Figure 10, we reorganize the observations into the categories and then build triangular fuzzy numbers, $\tilde{e}_{ij}^k = (e_{ij1}^k, e_{ij2}^k, e_{ij3}^k)$, for each predicted observations, all in accordance with Step 6 of the proposed methodology. For instance, after running the Holt–Winters method, represented by $j = 2$, the observation $t = 274$, which occupies position $i = 13$ inside category $k = 2$, has the following rounded predicted values: [331, 406, 396, 308, 249, 390, 383, 355, 370, 331, 301, 284, 280, 250, 343, 342, 361, 338, 346, 335, 305] for the actual value $y = 203$. Thus, the lower limit $\tilde{e}_{13,1,1}^2 = 250$ is the predicted value with minimum deviation from the actual value, $\tilde{e}_{13,1,3}^2 = 406$ is the value with maximum deviation, and $\tilde{e}_{13,1,2}^2 = 338$ is the median prediction considering all 21 predicted values, as collected from runs 17 to 37. This way $\tilde{e}_{13,1,1}^2 = (250, 338, 406)$.

In possession of all fuzzy numbers for each observation of B and for the six models, we then build the Residual Fuzzy Matrix, $\tilde{E}_{320 \times 6}$, whose three first and last rows we represent in a tabular form in Table 4. Please notice that each model represents one column of the matrix defined by Equation (18), and the observations inside the fluctuation testing/training set are represented by the rows. Tables 6–9 are also tabular forms of the matrices defined by the proposed model step-by-step, and l , m , and u in the sub-heading of the tables denote the lower, middle, and upper values of the triangular fuzzy numbers.

Table 4. Residual Fuzzy Matrix for the case with the first and last three rows.

HW			RFR			KNN			SVR			Prophet			ARIMA		
l	m	u	l	m	u	l	m	u	l	m	u	l	m	u	l	m	u
167	167	167	161	161	161	285	285	285	179	179	179	320	320	320	169	169	169
241	220	199	289	194	99.1	280	271	263	238	237	236	381	376	372	232	208	184
21.5	28.5	89.4	13.3	154	242	40.4	46	136	42.1	44	46.1	126	123	120	50.6	111	168
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
53.8	49.3	40.6	6.23	7.27	8.53	20.9	25.9	28.7	194	195	195	22.6	21.3	20.1	13.8	14.9	16.7
2.63	8.69	14.8	14.8	15	15.1	23.7	29.3	34.9	200	200	201	28.2	26.5	24.9	19.9	21.3	22.7
15.3	15.3	15.3	0.07	0.07	0.07	14	14	14	181	181	181	54.5	54.5	54.5	0.82	0.82	0.82

We compute them in the normalized fuzzy matrix, $\tilde{R}_{320 \times 6}$, according to Step 7. Table 5 shows the first and last three rows of the tabular form of \tilde{R} .

Table 5. Normalized Fuzzy Matrix for the case with the first and last three rows.

HW			RFR			KNN			SVR			Prophet			ARIMA		
l	m	u	l	m	u	l	m	u	l	m	u	l	m	u	l	m	u
0.52	0.52	0.52	0.50	0.50	0.50	0.89	0.89	0.89	0.56	0.56	0.56	1.00	1.00	1.00	0.53	0.53	0.53
0.52	0.58	0.63	0.26	0.51	0.76	0.69	0.69	0.74	0.62	0.62	0.63	0.98	0.98	1.00	0.48	0.52	0.61
0.09	0.09	0.37	0.05	0.17	1.00	0.17	0.17	0.56	0.17	0.18	0.19	0.50	0.50	0.52	0.21	0.21	0.69
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.21	0.21	0.28	0.03	0.04	0.04	0.11	0.13	0.15	0.99	0.99	1.00	0.10	0.10	0.12	0.07	0.08	0.09
0.01	0.04	0.07	0.07	0.07	0.08	0.12	0.15	0.17	1.00	1.00	1.00	0.12	0.12	0.14	0.10	0.10	0.11
0.08	0.08	0.08	0.00	0.00	0.00	0.08	0.08	0.08	1.00	1.00	1.00	0.30	0.30	0.30	0.00	0.00	0.00

We transform them into linguistic weights assigned by the decision-makers to the criteria in a fuzzy scale, according to Table 1, so we construct the Weighted Fuzzy Matrix, $\tilde{V} = [\tilde{v}_{ij}]_{320 \times 6}$, as requested by Step 8. Table 6 shows the first and last three rows of the tabular form of \tilde{V} .

Table 6. Weighted Fuzzy Matrix for the case with the first and last three rows.

HW			RFR			KNN			SVR			Prophet			ARIMA		
l	m	u	l	m	u	l	m	u	l	m	u	l	m	u	l	m	u
2.09	2.61	2.61	2.01	2.51	2.51	3.57	4.46	4.46	2.23	2.79	2.79	4.00	5.00	5.00	2.12	2.64	2.64
2.09	2.89	3.17	1.04	2.55	3.80	2.76	3.46	3.68	2.48	3.12	3.13	3.91	4.89	5.00	1.93	2.61	3.05
0.35	0.44	1.84	0.22	0.83	5.00	0.67	0.87	2.81	0.69	0.91	0.95	1.98	2.48	2.60	0.84	1.05	3.46
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.21	0.21	0.55	0.03	0.04	0.09	0.11	0.13	0.29	0.99	0.99	2.00	0.10	0.10	0.23	0.07	0.08	0.17
0.01	0.04	0.15	0.07	0.07	0.15	0.12	0.15	0.35	1.00	1.00	2.00	0.12	0.12	0.28	0.10	0.10	0.23
0.08	0.08	0.17	0.00	0.00	0.00	0.08	0.08	0.15	1.00	1.00	2.00	0.30	0.30	0.60	0.00	0.00	0.01

Following Step 9, we compute the Fuzzy Positive and Negative Ideal Solutions, FPIS and FNIS. Table 7 shows the first and last three FPIS and FNIS rows of the tabular form of the Fuzzy Ideal Solutions.

Table 7. Fuzzy Ideal Solutions for the case with the first and last three rows.

Obs.	Fuzzy-PIS			Fuzzy-NIS		
	l	m	u	l	m	u
239	0.00	0.00	0.00	4.00	5.00	5.00
240	0.00	0.00	0.00	3.91	4.89	5.00
241	0.00	0.00	0.00	1.98	2.48	5.00
	⋮	⋮	⋮	⋮	⋮	⋮
556	0.00	0.00	0.00	0.99	0.99	2.00
557	0.00	0.00	0.00	1.00	1.00	2.00
558	0.00	0.00	0.00	1.00	1.00	2.00

According to Step 10, we calculate the Euclidian distances of each model's predicted values from each FPIS and FNIS by using the vertex method, which also defuzzifies the fuzzy numbers, obtaining once again crisp values. Table 8 shows the first and last three rows for the Positive and Negatives to each model.

Table 8. First and last three and Positive and Negative Distances for each forecasting model.

Positive Distances						Negative Distances					
HW	RFR	KNN	SVR	Prophet	ARIMA	HW	RFR	KNN	SVR	Prophet	ARIMA
2.45	2.36	4.18	2.62	4.69	2.48	2.24	2.33	0.51	2.07	0.00	2.21
2.76	2.71	3.32	2.93	4.63	2.57	1.89	2.25	1.31	1.70	0.00	2.07
1.11	2.93	1.74	0.86	2.37	2.14	2.36	1.39	1.74	2.61	1.38	1.38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.36	0.06	0.20	1.41	0.16	0.12	1.05	1.35	1.22	0.00	1.25	1.30
0.09	0.11	0.23	1.41	0.19	0.15	1.33	1.31	1.19	0.00	1.22	1.26
0.12	0.00	0.11	1.41	0.43	0.01	1.30	1.41	1.30	0.00	0.99	1.41

Finally, the total distances, d_j^* and d_j^- (see Step 10), as well as the closeness coefficient, CC_j (see Step 11), are calculated to each model (see Table 9). According to the model we propose, forecasting models with a higher CC are preferred, which results from the association of great negative distances d_j^- to small positive distances. In Table 9, the positive distances contribute more to the turn of Prophet over Holt–Winters than the negative distances. Nevertheless, ARIMA has a significant distance to the negative ideal solution, its distance to the ideal solution is not sufficient to overcome Holt–Winters and Prophet.

Table 9. Total distances and closeness coefficient for each forecasting model.

	HW	RFR	KNN	SVR	Prophet	ARIMA
d+	450.6284	509.4003	466.2830	476.2906	405.4136	461.7832
d-	392.9327	340.9680	359.4671	321.0382	387.8993	392.4330
CCj	0.4658	0.4010	0.4353	0.4026	0.4890	0.4594

The application of the metric and the input data can be found in the Supplemental File.

4. Results and Discussion

To evaluate the proposed similarity metric, we first calculate the Mean Absolute Error (MAE) for each model according to the six categories illustrated in Sections 3.1 and 3.3. We also calculate the overall MAE for all models, as it is commonly found in the literature.

As we may observe in Figure 13, Holt–Winters is the model with better overall performance according to MAE. It is the model with the best performance after changes since it has the smallest MAE for the categories Stability Start and Decreasing Start and

holds the second-best performance for Increasing Start. Moreover, according to MAE, Holt–Winters is also the best model during Increasing, the second-best model during Stability and Decreasing. In the radio chart of Figure 13, Holt–Winters seems to have the smallest area, with an advantage over the second and third best models, RGR and KNN. These results are not surprising since Holt–Winters has shown good forecasting performance, especially for short-term forecastings and even against fashioned models, such as machine learning models [39,78–80]. In an overall manner, Prophet and ARIMA are the worst models, displaying higher MAE values.

We can also calculate the Weighted Mean Absolute Error (WMAE) by taking as weights the crisp values correspondent to the linguistic variable assigned by the decision-makers in Table 2. As observed in Figure 14, Holt–Winters still is the best overall model, but it loses performance, especially during periods of Decreasing, where Prophet has the best performance. In fact, with this approach, Prophet improves its overall performance since periods of decreasing accounts for 33% of all observations in the fluctuating testing/training set. Notice how the polar graph goes almost to zero for the effect of some chunks of data related to some categories over the overall result, such as c_{SS} , c_{IS} , and c_{DS} . It happens due to the smaller weights given to starting periods and the number of observations that fall into those categories.

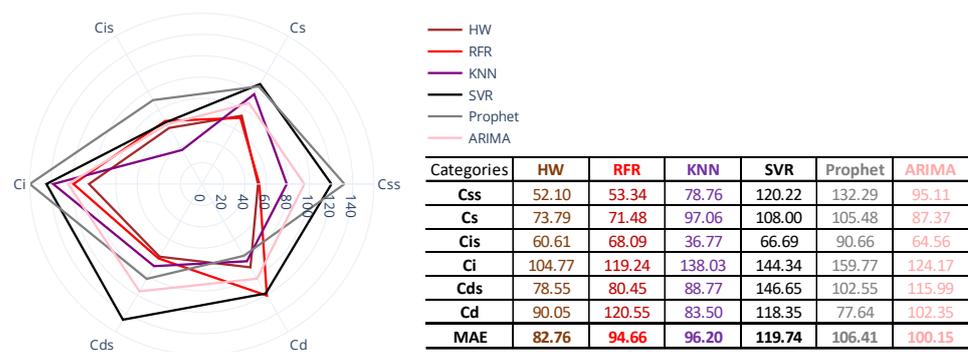


Figure 13. Mean Average Error for each category.

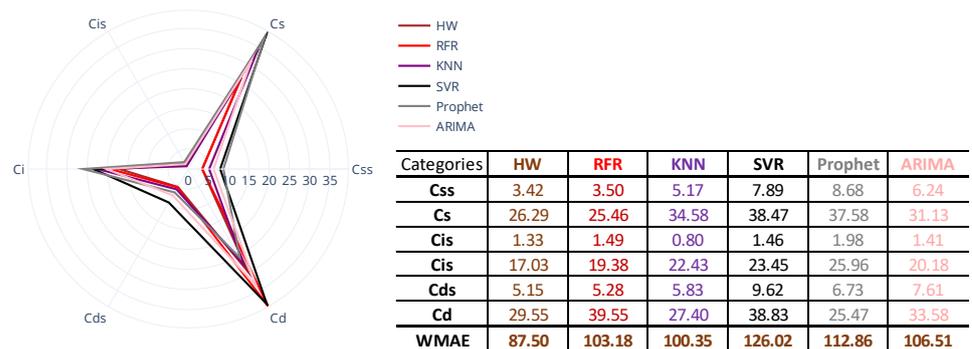


Figure 14. Weighted Mean Average Error for each category.

However, both MAE and WMAE consider only averages, and high volatilities in the data are not considered. For example, by looking at Figure 15, we notice how volatile Holt–Winters is when compared to Prophet. The first presents a much more significant difference between the lower and upper values computed to the fuzzy numbers than the latter; however, Holt–Winters medium values are closer to the target values.

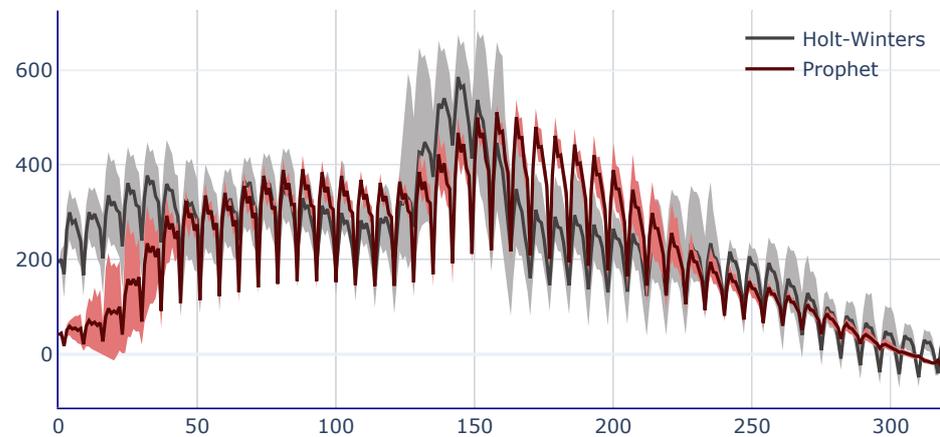


Figure 15. Variability in Holt–Winters and Prophet.

This variation can be better perceived by looking at the difference between the fuzzy residuals from Holt–Winters and Prophet and the positive ideal solutions, $FPIS = (0, 0, 0)$. When Holt–Winters has more significant residuals than Prophet, the difference is positive. On the other hand, when Prophet has greater residuals than Holt–Winters, the difference is negative. In Figure 16, we calculate the differences between the residuals of Holt–Winters and Prophet. The residuals are calculated between the target variable and the medium values (which is also correspondent to the MAE values, See Figure 13) for both models. Since we compute the difference between the residuals in Figure 16, for each cross-validation run, only one model will present non-null values: positive values if Holt–Winters has greater residuals, or negative values if Prophet is worst. As it can be noticed, Holt–Winters displays a smaller area in general, but it already loses performance during c_D and in most recent observations.

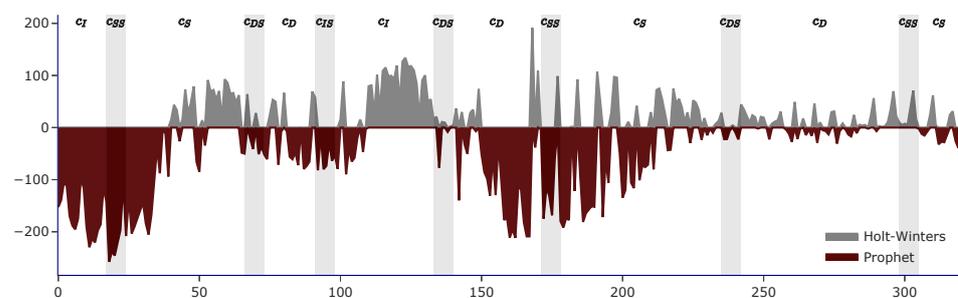


Figure 16. Comparison of residuals to the medium fuzzy number.

However, when we use the upper residual values, the results are far different (see Figure 17). Thus, the extremes in the predicted values are more evident. In this scenario, Prophet displays better results than Holt–Winters, not only in the most important categories, such as c_D , but also where it lacked accuracy before, such as c_S .

This new variability perspective, added to the perspective of the mean, explains the results of Table 9 and the dominance of Prophet over the other models. Thus, despite presenting the worst MAE compared to other models, it performs better in more critical categories for the decision-makers and displays lesser variability in the predicted data when we perform the moving forward validation.

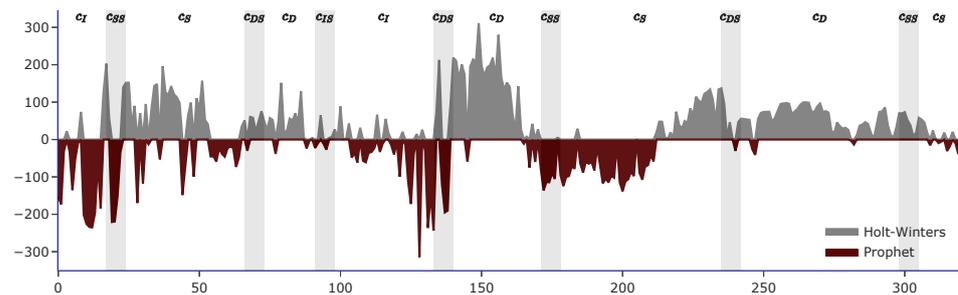


Figure 17. Comparison of residuals to the upper fuzzy number.

If classical TOPSIS were used instead of fuzzy-TOPSIS, the variability would not be captured, and some weighted means would only form the final metric. With classical TOPSIS, the ranking would be HW, KNN, RFR, ARIMA, Prophet, and SVR, which is the same ranking provided by the WMAE in Figure 14. Table 10 summarizes the proposed rankings from MAE, WMAE, TOPSIS similarity metric, and Fuzzy-TOPSIS similarity metric. To MAE and WMAE, the smaller the ranking, the better the results. To TOPSIS and Fuzzy-TOPSIS, the greater the results, the better.

Table 10. Rankings from MAE, WMAE, TOPSIS, and Fuzzy-TOPSIS.

Forecasting Model	MAE	WMAE	Fuzzy-TOPSIS	TOPSIS
HW	82.76	87.50	0.465	0.593
RFR	94.66	103.18	0.401	0.546
KNN	96.20	100.35	0.435	0.567
SVR	119.74	126.02	0.402	0.367
Prophet	106.41	112.86	0.489	0.452
ARIMA	100.15	106.51	0.459	0.471

For the sake of comprehension, the results of Table 10 are normalized in Table 11. The greater the value, the better for all ranking approaches. According to the four ranking approaches, the normalized values also allow us to better notice the distance between the models.

Table 11. Normalized rankings from MAE, WMAE, TOPSIS, and Fuzzy-TOPSIS.

Forecasting Model	MAE	WMAE	Fuzzy-TOPSIS	TOPSIS
HW	0.1988	0.1996	0.1754	0.1979
RFR	0.1738	0.1692	0.1513	0.1822
KNN	0.1710	0.1740	0.1641	0.1893
SVR	0.1374	0.1386	0.1516	0.1225
Prophet	0.1546	0.1547	0.1845	0.1509
ARIMA	0.1643	0.1639	0.1731	0.1572

As noted, while WMAE and classic TOPSIS are able to capture heterogeneous fitting expectations, they do not penalize models with greater volatility. Thus, their ranks are similar to those obtained with MAE. This way, the volatility treatment brought by the usage of fuzzy numbers is the biggest game-changer of the model we propose, at least for the data we explored.

5. Conclusions

This paper proposed a novel forecasting metric, a fuzzy similarity metric, that, in addition to averaging errors, can capture heterogeneous fitting expectations and volatility in the forecasted values, especially when using cross-validation approaches. We tested the metric to select the best model to predict future values for the number of COVID-19

confirmed cases in Amapá, Brazil. Models that quickly respond to increases in the series are preferred over others that tend to be more conservative under those situations. Besides having a small mean error associated with each predicted observation, low volatility during cross-validation is also desired since more steady models may lead to more expected scenarios. The similarity metric proposed in this paper ranks Holt–Winters as a second option to Prophet. Despite presenting a modest performance according to MAE, Prophet presents low volatility in the forecasted data, meaning that its worst predictions are still closer to the target variable than the worst predictions of Holt–Winters, for example. Furthermore, when weighting the time series periods according to the decision-maker's preferences, Prophet has a good performance during periods of decreasing, which accounts for almost 33% of the fluctuating testing/training set. Nevertheless, other metrics that may capture heterogeneous fitting expectations also penalize Holt–Winters, such as WMAE or even classical TOPSIS, but they still rank similarly to MAE. Thus, capturing volatility in the data is the biggest game-changer to the case we presented. Since it is the first introduction to a novel metric, future explorations must be performed to test it in other datasets and under different time series perspectives. Our work is also limited to univariate time series, and its response to multiple target variables is still unknown.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/su132413599/s1>. Availability of data, code, and other materials File S1.

Author Contributions: In this research, all authors contributed in some way. D.G.B.d.S. took part in the project administration, investigation, conceptualization, methodology, formal analysis, validation, writing (original draft), and writing (review and editing). E.A.d.S. took part in the conceptualization, writing (original draft), and writing (review and editing). F.T.A.J. took part in funding acquisition, data curation and writing (review and editing). M.C.V.N. did the formal analysis, writing (original draft), writing (review and editing), and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Dean of Research and Graduate Studies (PROPESP) of the State University of Amapá (UEAP) and in part by Coordination for the Improvement of Higher Education Personnel (CAPES), together with the Graduate Support Program (PROAP) of the Federal University of São Paulo (UNIFESP).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are contained within the article.

Conflicts of Interest: All authors have approved the manuscript and agree with its submission. Furthermore, we confirm that this manuscript has not been previously published and is not considered for publication in any other journal. Therefore, no conflict of interest needs to be disclosed.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike Information Criterion
ARIMA	Integration (I) between autoregressive (AR) and moving average (MA) models
AUC	the area under the receiver operating characteristic (ROC)
BIC	The Bayesian Information Criterion
CC	closeness coefficient
FNIS	Fuzzy Negative Ideal Solution
FPIS	Fuzzy Positive Ideal Solution
HW	Holt–Winters
KNN	K-Nearest Neighbors
MAE	Mean Absolute Error
MCC	Matthews Correlation Coefficient

MCDM	Multi-Criteria Decision Making
Prophet	Forecasting approach developed by Facebook
R2	R-squared
RFR	Random Forest
RMSE	Root Mean Squared Error
RRMSE	Relative Root Mean Squared Error
SMAPE	Symmetric Mean Absolute Percentage Error
SVR	Support Vector Regression
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
WMAE	Weighted Mean Absolute Error

References

- World Health Organization. Coronavirus Disease 2019 (COVID-19): Situation Report—64. 2020. Available online: <https://www.who.int/publications/m/item/situation-report---64> (accessed on 21 November 2021).
- BBC. Covid: Austria Back in Lockdown Despite Protests. 2021. Available online: <https://www.bbc.com/news/world-europe-59369488> (accessed on 22 November 2021).
- Ministry of Health of Brazil Painel Coronavírus. Coronavírus Brasil. 2020. Available online: <https://covid.saude.gov.br/> (accessed on 21 November 2021).
- Roser, M.; Ritchie, H.; Ortiz-Ospina, E.; Hasell, J. Coronavirus Pandemic (COVID-19). *Our World in Data*. 2020. Available online: <https://ourworldindata.org/coronavirus> (accessed on 10 October 2021).
- Justen, A. Brazil.io: COVID-19: Coronavirus Newsletters and Cases by Municipality per Day. 2021. Available online: <https://brasil.io/dataset/covid19/caso> (accessed on 10 October 2021).
- Silva, G.A.; Jardim, B.C.; Lotufo, P.A. Mortalidade por COVID-19 padronizada por idade nas capitais das diferentes regiões do Brasil. *Cad. De Saúde Pública* **2021**, *37*, e00039221. [[CrossRef](#)]
- Brasil, A. Mortalidade por COVID-19 na Região Norte é Mais Alta. 2021. Available online: <https://agenciabrasil.ebc.com.br/saude/noticia/2021-07/mortalidade-por-covid-19-na-regiao-norte-e-mais-alta-diz-pesquisa> (accessed on 11 October 2021).
- de Souza, D.G.B.; Júnior, F.T.A.; Soma, N.Y. Forecasting COVID-19 cases at the Amazon region: A comparison of classical and machine learning models. *bioRxiv* **2020**. [[CrossRef](#)]
- Kaplan, H.S.; Trumble, B.C.; Stieglitz, J.; Mamany, R.M.; Cayuba, M.G.; Moye, L.M.; Alami, S.; Kraft, T.; Gutierrez, R.Q.; Adrian, J.C.; et al. Voluntary collective isolation as a best response to COVID-19 for indigenous populations? A case study and protocol from the Bolivian Amazon. *Lancet* **2020**, *395*, 1727–1734. [[CrossRef](#)]
- Yang, Z.; Li, X.; Garg, H.; Qi, M. Decision support algorithm for selecting an antivirus mask over COVID-19 pandemic under spherical normal fuzzy environment. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3407. [[CrossRef](#)]
- Dabbah, M.A.; Reed, A.B.; Booth, A.T.; Yassaee, A.; Despotovic, A.; Klasmer, B.; Binning, E.; Aral, M.; Plans, D.; Labrique, A.B.; et al. Machine learning approach to dynamic risk modeling of mortality in COVID-19: A UK Biobank study. *arXiv* **2021**, arXiv:2104.09226.
- Zangmeister, C.D.; Radney, J.G.; Vicenzi, E.P.; Weaver, J.L. Filtration efficiencies of nanoscale aerosol by cloth mask materials used to slow the spread of SARS-CoV-2. *ACS Nano* **2020**, *14*, 9188–9200. [[CrossRef](#)] [[PubMed](#)]
- Ranjbari, M.; Esfandabadi, Z.S.; Zanetti, M.C.; Scagnelli, S.D.; Siebers, P.O.; Aghbashlo, M.; Peng, W.; Quatraro, F.; Tabatabaei, M. Three pillars of sustainability in the wake of COVID-19: A systematic review and future research agenda for sustainable development. *J. Clean. Prod.* **2021**, *297*, 126660. [[CrossRef](#)] [[PubMed](#)]
- Zheng, N.; Du, S.; Wang, J.; Zhang, H.; Cui, W.; Kang, Z.; Yang, T.; Lou, B.; Chi, Y.; Long, H.; et al. Predicting COVID-19 in China using hybrid AI model. *IEEE Trans. Cybern.* **2020**, *50*, 2891–2904. [[CrossRef](#)]
- Garg, H.; Kaur, G. Algorithms For Screening Travelers During Covid-19 Outbreak Using Probabilistic Dual Hesitant Values Based On Bipartite Graph Theory. *Appl. Comput. Math.* **2021**, *20*, 22–48.
- Ribeiro, M.H.D.M.; da Silva, R.G.; Mariani, V.C.; dos Santos Coelho, L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos Solitons Fractals* **2020**, *135*, 109853. [[CrossRef](#)]
- Zhang, X.; Ma, R.; Wang, L. Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. *Chaos Solitons Fractals* **2020**, *135*, 109829. [[CrossRef](#)]
- Goic, M.; Bozanic-Leal, M.S.; Badal, M.; Basso, L.J. COVID-19: Short-term forecast of ICU beds in times of crisis. *PLoS ONE* **2021**, *16*, e0245272. [[CrossRef](#)] [[PubMed](#)]
- Koç, E.; Türkoğlu, M. Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: The COVID-19 pandemic in Turkey. *Signal Image Video Process.* **2021**, *1*, 1–9. [[CrossRef](#)]
- Wang, P.; Zheng, X.; Li, J.; Zhu, B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos Solitons Fractals* **2020**, *139*, 110058. [[CrossRef](#)] [[PubMed](#)]
- Benvenuto, D.; Giovanetti, M.; Vassallo, L.; Angeletti, S.; Ciccozzi, M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data Brief* **2020**, *29*, 105340. [[CrossRef](#)] [[PubMed](#)]
- Capistran, M.A.; Capella, A.; Christen, J.A. Forecasting hospital demand in metropolitan areas during the current COVID-19 pandemic and estimates of lockdown-induced 2nd waves. *PLoS ONE* **2021**, *16*, e0245669. [[CrossRef](#)]

23. da Silva, T.T.; Francisquini, R.; Nascimento, M.C. Meteorological and human mobility data on predicting COVID-19 cases by a novel hybrid decomposition method with anomaly detection analysis: A case study in the capitals of Brazil. *Expert Syst. Appl.* **2021**, *182*, 115190. [[CrossRef](#)] [[PubMed](#)]
24. Liu, D.; Clemente, L.; Poirier, C.; Ding, X.; Chinazzi, M.; Davis, J.T.; Vespignani, A.; Santillana, M. A machine learning methodology for real-time forecasting of the 2019–2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. *arXiv* **2020**, arXiv:2004.04019.
25. Pandey, G.; Chaudhary, P.; Gupta, R.; Pal, S. SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv* **2020**, arXiv:2004.00958.
26. Sears, J.; Villas-Boas, J.M.; Villas-Boas, V.; Villas-Boas, S.B. *Are We# Stayinghome to Flatten the Curve?* 1st ed.; Department of Agricultural and Resource Economics: Storrs, CT, USA, 2020; Volume 5. Available online: <https://escholarship.org/uc/item/5h97n884> (accessed on 28 October 2021).
27. Garg, H.; Nasir, A.; Jan, N.; Khan, S.U. Mathematical analysis of COVID-19 pandemic by using the concept of SIR model. *Soft Comput.* **2021**, *23*, 103970. [[CrossRef](#)] [[PubMed](#)]
28. He, S.; Peng, Y.; Sun, K. SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dyn.* **2020**, *101*, 1667–1680. [[CrossRef](#)]
29. Rivera-Rodriguez, C.; Urdinola, B.P. Predicting hospital demand during the COVID-19 outbreak in Bogota, Colombia. *Front. Public Health* **2020**, *8*, 710. [[CrossRef](#)] [[PubMed](#)]
30. Massonnaud, C.; Roux, J.; Crépey, P. COVID-19: Forecasting short term hospital needs in France. *medRxiv* **2020**. [[CrossRef](#)]
31. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* **2020**, *729*, 138817. [[CrossRef](#)] [[PubMed](#)]
32. Singh, R.K.; Rani, M.; Bhagavathula, A.S.; Sah, R.; Rodriguez-Morales, A.J.; Kalita, H.; Nanda, C.; Sharma, S.; Sharma, Y.D.; Rabaan, A.A.; et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR Public Health Surveill.* **2020**, *6*, e19115. [[CrossRef](#)] [[PubMed](#)]
33. Chakraborty, T.; Ghosh, I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos Solitons Fractals* **2020**, *135*, 109850. [[CrossRef](#)]
34. Wang, Q.; Li, S.; Li, R.; Jiang, F. Underestimated impact of the COVID-19 on carbon emission reduction in developing countries—A novel assessment based on scenario analysis. *Environ. Res.* **2022**, *204*, 111990. [[CrossRef](#)]
35. Cihan, P. Impact of the COVID-19 lockdowns on electricity and natural gas consumption in the different industrial zones and forecasting consumption amounts: Turkey case study. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107369. [[CrossRef](#)]
36. Talkhi, N.; Fatemi, N.A.; Ataei, Z.; Nooghabi, M.J. Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods. *Biomed. Signal Process. Control* **2021**, *66*, 102494. [[CrossRef](#)]
37. Saba, T.; Abunadi, I.; Shahzad, M.N.; Khan, A.R. Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. *Microsc. Res. Tech.* **2021**, *84*, 1462–1474. [[CrossRef](#)] [[PubMed](#)]
38. Panda, M. Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states. *medRxiv* **2020**. [[CrossRef](#)]
39. Petropoulos, F.; Makridakis, S. Forecasting the novel coronavirus COVID-19. *PLoS ONE* **2020**, *15*, e0231236. [[CrossRef](#)]
40. Guo, L.; Fang, W.; Zhao, Q.; Wang, X. The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. *Comput. Ind. Eng.* **2021**, *161*, 107598. [[CrossRef](#)]
41. Yadav, D.; Maheshwari, H.; Chandra, U. Outbreak prediction of covid-19 in most susceptible countries. *Glob. J. Environ. Sci. Manag.* **2020**, *6*, 11–20.
42. Dash, S.; Chakraborty, C.; Giri, S.K.; Pani, S.K. Intelligent computing on time-series data analysis and prediction of COVID-19 pandemics. *Pattern Recognit. Lett.* **2021**, *151*, 69–75. [[CrossRef](#)]
43. da Silva, R.G.; Ribeiro, M.H.D.M.; Mariani, V.C.; dos Santos Coelho, L. Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. *Chaos Solitons Fractals* **2020**, *139*, 110027. [[CrossRef](#)]
44. Arslan, H. COVID-19 prediction based on genome similarity of human SARS-CoV-2 and bat SARS-CoV-like coronavirus. *Comput. Ind. Eng.* **2021**, *161*, 107666. [[CrossRef](#)] [[PubMed](#)]
45. Atsa'am, D.D.; Wario, R. Classifier Selection for the Prediction of Dominant Transmission Mode of Coronavirus Within Localities: Predicting COVID-19 Transmission Mode. *Int. J. E-Health Med. Commun. (IJEHMC)* **2021**, *12*, 1–12. [[CrossRef](#)]
46. Yeşilkanat, C.M. Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos Solitons Fractals* **2020**, *140*, 110210. [[CrossRef](#)]
47. Tena, A.; Clarià Sancho, F.; Solsona Tehàs, F. Automated detection of COVID-19 cough. *Biomed. Signal Process. Control* **2022**, *71*, 103175. [[CrossRef](#)]
48. Rustam, F.; Reshi, A.A.; Mehmood, A.; Ullah, S.; On, B.; Aslam, W.; Choi, G.S. COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access* **2020**, *8*, 101489–101499.
49. Herlawati, H. COVID-19 Spread Pattern Using Support Vector Regression. *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.* **2020**, *8*, 67–74. [[CrossRef](#)]
50. Zhang, J.; Florita, A.; Hodge, B.M.; Lu, S.; Hamann, H.F.; Banunarayanan, V.; Brockway, A.M. A suite of metrics for assessing the performance of solar power forecasting. *Sol. Energy* **2015**, *111*, 157–175. [[CrossRef](#)]
51. Chen, C.; Twycross, J.; Garibaldi, J.M. A new accuracy measure based on bounded relative error for time series forecasting. *PLoS ONE* **2017**, *12*, e0174202. [[CrossRef](#)]

52. Sahoo, B.B.; Jha, R.; Singh, A.; Kumar, D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys.* **2019**, *67*, 1471–1481. [[CrossRef](#)]
53. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021. Available online: <https://otexts.com/fpp3/> (accessed on 18 October 2021).
54. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 669–679. [[CrossRef](#)]
55. Yang, E.; Park, H.W.; Choi, Y.H.; Kim, J.; Munkhdalai, L.; Musa, I.; Ryu, K.H. A simulation-based study on the comparison of statistical and time series forecasting methods for early detection of infectious disease outbreaks. *Int. J. Environ. Res. Public Health* **2018**, *15*, 966. [[CrossRef](#)]
56. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
57. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [[CrossRef](#)]
58. de Souza, D.G.B.; dos Santos, E.A.; Soma, N.Y.; da Silva, C.E.S. MCDM-Based R&D Project Selection: A Systematic Literature Review. *Sustainability* **2021**, *13*, 11626. [[CrossRef](#)]
59. Hwang, C.L.; Yoon, K. Methods for multiple attribute decision making. In *Multiple Attribute Decision Making*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 58–191. [[CrossRef](#)]
60. Dehdasht, G.; Ferwati, M.S.; Zin, R.M.; Abidin, N.Z. A hybrid approach using entropy and TOPSIS to select key drivers for a successful and sustainable lean construction implementation. *PLoS ONE* **2020**, *15*, e0228746. [[CrossRef](#)]
61. Zhao, Q.; Chen, J.; Li, F.; Li, A.; Li, Q. An integrated model for evaluation of maternal health care in China. *PLoS ONE* **2021**, *16*, e0245300. [[CrossRef](#)] [[PubMed](#)]
62. Bi, Q.P.; Li, Y.C.; Shen, C. Screening of Evaluation Index and Construction of Evaluation Index System for Mine Ventilation System. *Sustainability* **2021**, *13*, 11810. [[CrossRef](#)]
63. Chen, C.T. Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets Syst.* **2000**, *114*, 1–9. [[CrossRef](#)]
64. Chen, C.T.; Lin, C.T.; Huang, S.F. A fuzzy approach for supplier evaluation and selection in supply chain management. *Int. J. Prod. Econ.* **2006**, *102*, 289–301. [[CrossRef](#)]
65. Mizumoto, M.; Tanaka, K. Fuzzy sets and their operations. *Inf. Control* **1981**, *48*, 30–48. [[CrossRef](#)]
66. Souza, D.G.; Silva, C.E.; Soma, N.Y. Selecting projects on the Brazilian R&D energy sector: A fuzzy-based approach for criteria selection. *IEEE Access* **2020**, *8*, 50209–50226.
67. Van Laarhoven, P.J.; Pedrycz, W. A fuzzy extension of Saaty’s priority theory. *Fuzzy Sets Syst.* **1983**, *11*, 229–241. [[CrossRef](#)]
68. Liu, Y.; Eckert, C.M.; Earl, C. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Syst. Appl.* **2020**, *161*, 113738. [[CrossRef](#)]
69. Nădăban, S.; Dzitac, S.; Dzitac, I. Fuzzy TOPSIS: A general view. *Procedia Comput. Sci.* **2016**, *91*, 823–831. [[CrossRef](#)]
70. Box, G.E.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 734.
71. Canova, F.; Hansen, B.E. Are seasonal patterns constant over time? A test for seasonal stability. *J. Bus. Econ. Stat.* **1995**, *13*, 237–252. [[CrossRef](#)]
72. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages. *Int. J. Forecast.* **2004**, *20*, 5–10. [[CrossRef](#)]
73. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Manag. Sci.* **1960**, *6*, 324–342. [[CrossRef](#)]
74. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 1995. [[CrossRef](#)]
75. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1997; pp. 155–161.
76. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [[CrossRef](#)]
77. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
78. Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R.; Newton, J.; Parzen, E.; Winkler, R. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecast.* **1982**, *1*, 111–153. [[CrossRef](#)]
79. Makridakis, S.; Hibon, M. The M3-Competition: Results, conclusions and implications. *Int. J. Forecast.* **2000**, *16*, 451–476. [[CrossRef](#)]
80. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. *Int. J. Forecast.* **2020**, *36*, 54–74. [[CrossRef](#)]