

# The Hybrid Forecasting Method SVR-ESAR for Covid-19

Juan Frausto Solis<sup>1,\*</sup>, José Enrique Olvera Vazquez<sup>1</sup>, Juan J González Barbosa<sup>1</sup>, Guadalupe Castilla Valdez<sup>1</sup>, Juan Paulo Sánchez Hernández<sup>2</sup>, Joaquín Perez-Ortega<sup>3</sup>

Tecnológico Nacional de Mexico/<sup>1</sup>IT Cd Madero/<sup>3</sup>CENIDET,

<sup>2</sup>Universidad Politécnica del Estado de Morelos

<sup>1,\*</sup>Corresponding author: [juan.frausto@gmail.com](mailto:juan.frausto@gmail.com)

## Abstract

*We know that SARS-Cov2 produces the new COVID-19 disease, which is one of the most dangerous pandemics of modern times. This pandemic has critical health and economic consequences, and even the health services of the large, powerful nations may be saturated. Thus, forecasting the number of infected persons in any country is essential for controlling the situation. In the literature, different forecasting methods have been published, attempting to solve the problem. However, a simple and accurate forecasting method is required for its implementation in any part of the world. This paper presents a precise and straightforward forecasting method named SVR-ESAR (Support Vector regression hybridized with the classical Exponential smoothing and ARIMA). We applied this method to the infected time series in four scenarios, which we have taken for the Github repository: the Whole World, China, the US, and Mexico. We compared our results with those of the literature showing the proposed method has the best accuracy.*

## 1. Introduction

In December 2019, Chinese citizens started to suffer a strange respiratory disease; this happened in Wuhan city, Hubei province. This disease was produced by a virus named SARS-CoV-2 or COVID-19, whose origin remains unknown [1], [2]. However, the scientists observed this disease is very contagious. It spread rapidly throughout the world, becoming a pandemic with an increased death rate as it left China. Thus, knowing the behavior of the disease is of utmost importance to take the appropriate sanitary measures. Researchers observed the pandemic rules in different countries; they noticed that when the control measures were relaxed, and a large number of people severely infected, the health services could be saturated. Therefore, it is necessary to establish efficient methods for estimating the number of infected people accurately; thus, this helps the health authorities to take the correct measures at the appropriate time using SIR or other pandemic models. There are two main approaches for describing an epidemic:

- The SIRs models. The mathematical epidemic models, the first of which was proposed by Sir Roland Ross in 1902 and adjusted in 1927 by Kermack and McKendrick [3].

- Forecasting models using time series. This approach is an ancient method for describing an epidemic such as Covid-19. For instance, [4] shows different forecasting epidemic studies using time series.

These two main epidemic approaches are fundamental, and they are related. However, the infected prediction of time series is one of the biggest challenges since the estimations require the lowest error. Currently, the forecasting time series techniques can be classified as:

- Classical forecasting ARIMA and Exponential Smoothing (ES) and they typically obtain good results [4],
- Artificial Intelligence techniques where vector support machines (SVM) and neural networks are the most common.

Nowadays, it is common to use hybrid methods using these two techniques. For instance, [5] combines ES and ARIMA. Roughly speaking, it is a two-phases forecasting technique. Once the first phase obtains a first approximation forecast (or base forecast), a second phase improves the latter by using the residual values of the first step. Koning et al. and Makridakis et al. presented this approach for time series from M3-competition with very good results [6], [7]. Also, among the best methods for M4-competition are hybrid methods such as SVR with ES and ARIMA [8]. Inspired by these ideas, we decided to design a forecasting hybrid method for the Covid-19 time series. This method is named SVR-ESAR (SVR with ES and ARIMA), and it is presented in section three, after a brief background in section two. Then we used infected cases that we have taken from the Github repository ([https://github.com/CSSEGISandData/COVID19/blob/master/archived\\_data/archived\\_time\\_series/time\\_series\\_19-covid-Confirmed\\_archived\\_0325.csv](https://github.com/CSSEGISandData/COVID19/blob/master/archived_data/archived_time_series/time_series_19-covid-Confirmed_archived_0325.csv)) and we prepared four scenarios for and experimentation is described in section four. Finally, we present the conclusions in section five.

## 2. Background

Epidemics have always existed in the world. However, their first model (or SIR model) was not published until 1902 [3]. This model and its variants are based on Markov models, where the population is divided into several classes as Susceptible, Infected (or confirmed), and Recovered. This paper is focused on forecasting the confirmed class by using time series. We use a hybrid method that uses Support Vector Regression (SVR), Exponential Smoothing (ES), and ARIMA. These techniques are widespread in several platforms, and they are briefly explained as follows:

- Support Vector Regression (SVR): This is a very effective method for forecasting time series, which is an application of Vector Support Machine (SVM) proposed by Vapnik [9]. SVR and SVM minimize the margin error and use Kernel function for non-separable classes. For obtaining good forecasting results, in the quality and stability of SVR, its parameters and the kernel parameters should be tuned. Typically, they are tuned by a grid search, a genetic or other heuristic selects the best parameters [10].

- **Exponential smoothing (ES).** This forecasting method uses an exponentially weighted moving average with the values of the times series previously observed. ES smooths (averages) the past values of the time series with the last forecast. Exponential smoothing has given good results in different forecasting competitions, especially for short series [11], [12].
- **ARIMA model:** This model is described in terms of parameters currently named p, d, and q, and linked to three types of processes: autoregression, integration, and the moving average [13].

### 3. Description of the SVR-ESAR method

We developed a hybrid Forecasting method named SVR-ESAR, which uses GA-SVR and an adjustment phase. The SVR-ESAR Architecture GA-SVR has two phases and is shown in Figure 1:

❖ **Phase 1:** A Genetic Algorithm adjusts the parameters of an SVR machine and its kernels (linear, rbf, sigmoid) using Grid Search and iteratively improves the  $F(t)$  forecast of SVR [10]. The improved forecast  $F^*_1(t)$  in this phase has a MAPE error dependent on the residuals during the training phase.

❖ **Phase 2:** Three alternative techniques obtain a correction of the  $F^*_1(t)$  forecast previously obtained by GA-SVR in phase 1: SVR, Holt Exponential Smoothing (ES), and ARIMA [13]. These techniques are applied to the residuals between  $F^*_1(t)$  and the actual values in the training stage, and obtaining a better forecasting  $F2_{SVR}$ ,  $F2_{ES}$ , and  $F2_{AR}$  with the application in the second phase of these three techniques SVR, ES, and ARIMA respectively.

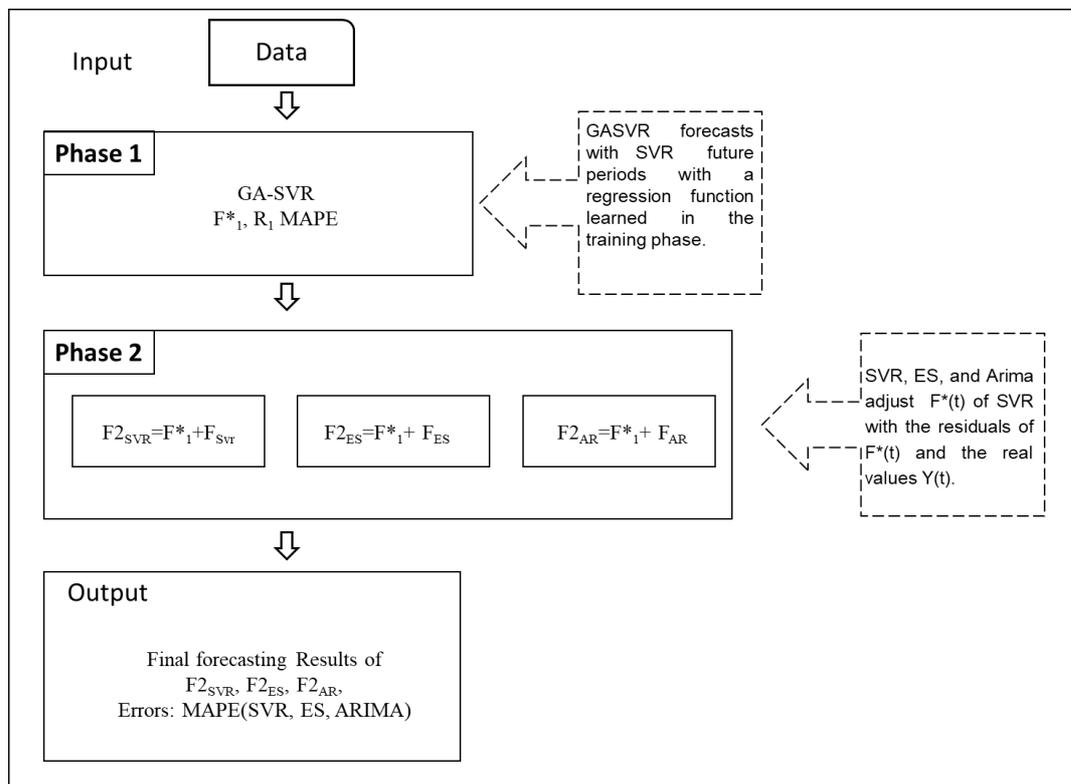


Figure 1. SVR-ESAR Architecture.

## 4. Experimental Results

The presented forecasting method was applied to people infected with Coronavirus 19. The period for this time series is from 22/January until 25/April/ 2020 [14]; we have taken the data from the GitHub repository. In the references [4] and [13], the forecast is for the last ten days. Thus, we used the same parameters to compare their results with our proposed forecasting method. We have four scenarios:

- **The Whole World:** The forecasting results obtained for infected people [15]. We compare the results obtained for our proposed forecasting method using the infected cases.
- **China:** This country was the origin of the pandemic. There are published results for this case [4], which we use in this section for validating the proposed method.
- **United States (US):** This country has the highest number of incidences in the American continent; also, it is significant because it has a lot of communication with the rest of the world.
- **Mexico** implemented some particular policies and has a lot of relationships with many countries, particularly with the US, and Canada.

Table 1 presents the results for the Whole World and China, obtained with the application of the proposed method for confirmed forecasting case. In this table, we show the results of SVR-ESAR using the adjustment techniques SVR, ES, and ARIMA. We notice that all these techniques have a small value of MAPE error. We also present the results obtained for this case in [4] and [15]. We observe that SVR-ARIMA achieved the best results for these cases. Furthermore, Figure 2 shows the results obtained by SVR-ESAR for these scenarios.

Table 1. Forecasting for the infected in the Whole World and China using the SVR-ESAR method with three adjustment techniques (SVR, Exponential Smoothing, and ARIMA)

Scenario 1: Infected in the Whole Word.				
Forecasting method	Exponential Smoothing, [15]	SVR-ESAR		
	Adjusted Technique:	Adjusted Technique		
	Multiplicative Trend model	SVR	ES	Arima
MAPE error (%)	6.9	2.46	0.77	0.56
Scenario 2: Infected in China				
Forecasting method	Neural Networks [4]	SVR-ESAR		
	Adjusted Technique	Adjusted Technique		
	Unknown	SVR	ES	Arima
MAPE error (%)	4.79	1.90	3.53	1.41

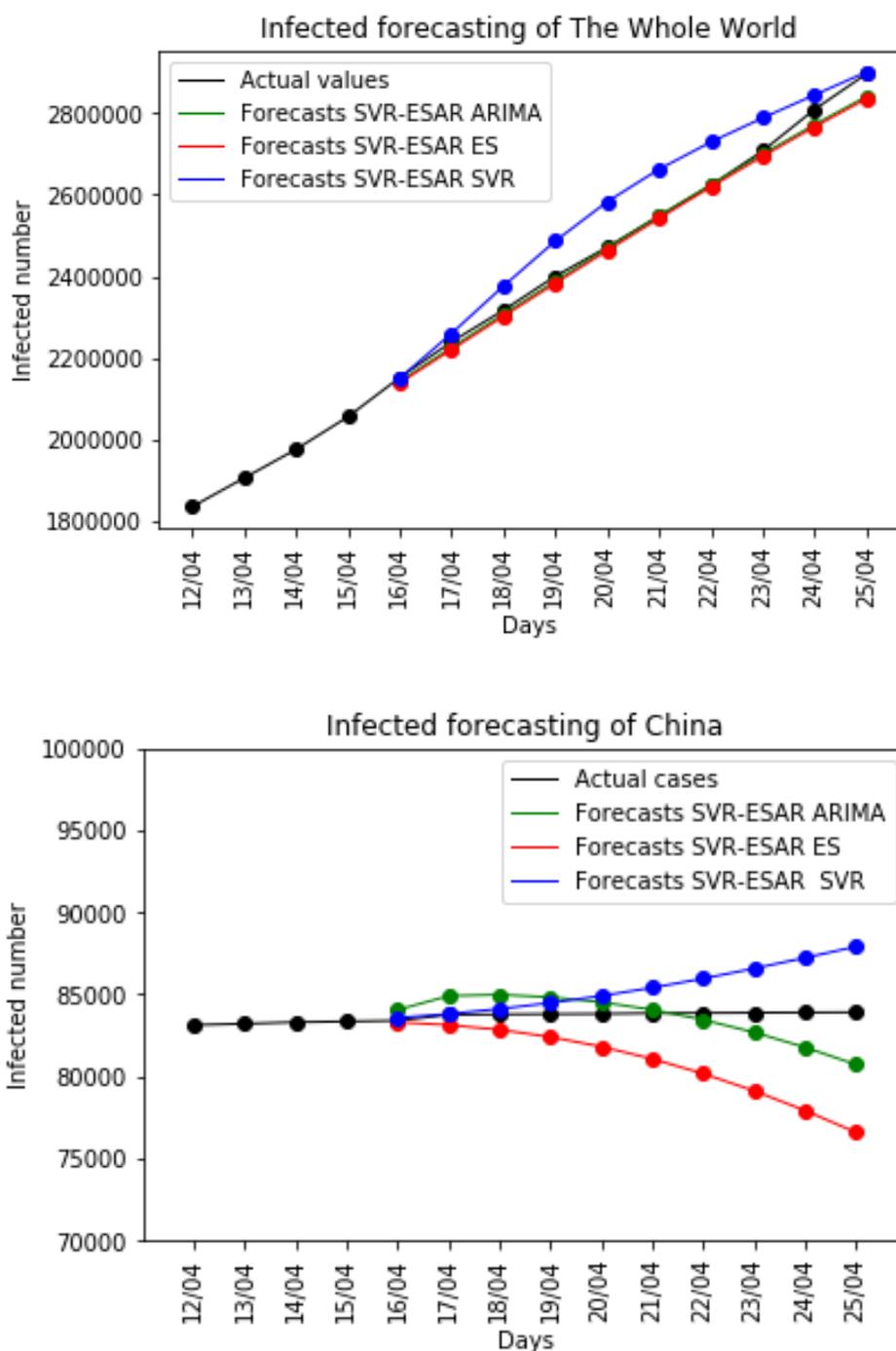


Figure 2: Forecast of SVR-ESAR for the Whole World and China

We present in Table 2 the forecasting confirmed results for the US and Mexico obtained with SVR-ESAR. We show the results of the proposed method using the adjustment obtained by SVR, ES, and ARIMA (Figure 3). We point out that in the case of the US, the ARIMA adjustment technique achieves the best result. This table also shows the results obtained by SVR-ESAR for Mexico. This time the proposed method achieves a modest forecasting result with all three adjustment techniques.

Table 2. Forecasting for the infected in the Whole World and China using the SVR-ESAR method with three adjustment techniques (SVR, Exponential Smoothing, and ARIMA)

Scenario 3 and 4: Infected in United States and Mexico			
Forecasting Method	SVR-ESAR		
	Adjusted Technique		
	SVR	ES	Arima
MAPE US (%)	5.09	5.39	0.90
MAPE Mexico (%)	12.44	13.25	12.57

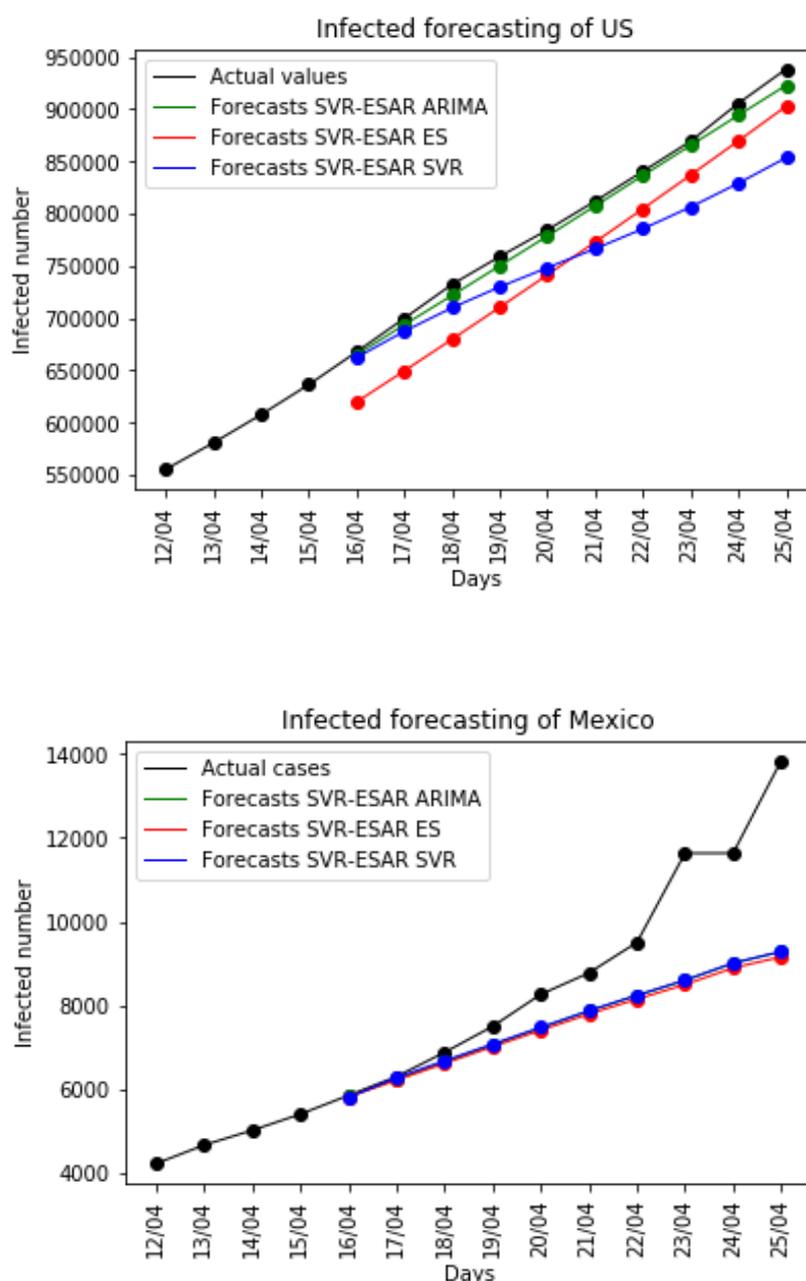


Figure 3. The forecasting of infected cases for the US and Mexico

## 5. Discussion and conclusion

This paper presents a forecasting method named SVR-ESAR (Support Vector Regression with Exponential Smoothing and ARIMA), which was applied for the estimation of confirmed (infected) cases of Covid-19 in four scenarios (the Whole World, China, the US, and Mexico). This method is straightforward and uses SVR with an iterative adjustment phase. In the adjustment phase, we applied three algorithms to the SVR residuals: ES, SVR, and ARIMA. We found SVR-ESAR with an ARIMA adjustment obtained the best or equivalent results for all these scenarios. We applied this method for data of Mexico with modest quality results. However, the results of SVR-ESAR were compared with those published in the literature for the US and China, and we found the proposed method achieved the best predictions.

### Authors contribution

**Conceptualization:** Juan Frausto Solís.

**Data curation:** José Enrique Olvera Vázquez and Javier Gonzalez Barbosa.

**Investigation:** Juan Frausto Solís.

**Methodology:** Guadalupe Castillo Valdez.

**Project Supervisor:** Juan Frausto Solís.

**Software:** Juan Paulo Sánchez, Guadalupe Castillo, José Enrique Olvera Vázquez.

**Validation:** Javier González Barbosa.

**Writing – original draft:** Juan Frausto Solís, Javier Gonzalez Barbosa.

**Writing – review & editing:** Joquin Perez Ortega.

### References

- [1] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, and R. Siddique, “COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses,” *J. Adv. Res.*, vol. 24, pp. 91–98, 2020, DOI: 10.1016/j.jare.2020.03.005.
- [2] F. Di Gennaro *et al.*, “Coronavirus diseases (COVID-19) current status and future perspectives: A narrative review,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 8, 2020, DOI: 10.3390/ijerph17082690.
- [3] W. O. Kermack and A. G. McKendrick, “A Contribution to the Mathematical Theory of Epidemics,” vol. 115, no. 772. 2018, DOI: 10.1002/mma.5067.
- [4] M. A. A. Al-Qaness, A. A. Ewees, H. Fan, and M. A. El Aziz, “Optimization method for forecasting confirmed cases of COVID-19 in China,” *Appl. Sci.*, vol. 9, no. 3, 2020, DOI: 10.3390/JCM9030674.
- [5] E. Spiliotis, F. Petropoulos, and V. Assimakopoulos, “Improving the forecasting performance of temporal hierarchies,” *PLoS One*, vol. 14, no. 10, pp. 1–21, 2019, DOI: 10.1371/journal.pone.0223422.
- [6] A. J. Koning, P. H. Franses, M. Hibon, and H. O. Stekler, “The M3 competition: Statistical tests of the results,” *Int. J. Forecast.*, vol. 21, no. 3, pp. 397–409, 2005, DOI: 10.1016/j.ijforecast.2004.10.003.
- [7] S. Makridakis and M. Hibon, “The M3-competition: Results, conclusions and implications,” *Int. J. Forecast.*, vol. 16, no. 4, pp. 451–476, 2000, DOI: 10.1016/S0169-2070(00)00057-1.
- [8] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The M4 Competition: Results, findings, conclusion, and way forward,” *Int. J. Forecast.*, vol. 34, no. 4, pp. 802–808, 2018, DOI: 10.1016/j.ijforecast.2018.06.001.
- [9] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector

- regression machines,” *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 155–161, 1997.
- [10] G. Santamaría-Bonfil, J. Frausto-Solís, and I. Vázquez-Rodarte, “Volatility Forecasting Using Support Vector Regression and a Hybrid Genetic Algorithm,” *Comput. Econ.*, vol. 45, no. 1, pp. 111–133, 2013, DOI: 10.1007/s10614-013-9411-x.
- [11] S. Makridakis *et al.*, “The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition,” *J. Forecast.*, vol. 1, no. June 1981, pp. 111–153, 1982.
- [12] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The M4 Competition: 100,000 time series and 61 forecasting methods,” *Int. J. Forecast.*, vol. 36, no. 1, pp. 54–74, 2020, DOI: 10.1016/j.ijforecast.2019.04.014.
- [13] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. New Jersey, 2008.
- [14] Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (<https://github.com/CSSEGISandData/COVID-19> accessed 26/04/2020)
- [15] Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. *PLoS ONE* 15(3): e0231236. <https://doi.org/10.1371/journal.pone.0231236>