



# Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression

Ricardo Manuel Arias Velásquez\*, Jennifer Vanessa Mejía Lara

Universidad Nacional de San Agustín de Arequipa, Perú

## ARTICLE INFO

### Article history:

Received 12 April 2020

Revised 17 April 2020

Accepted 20 May 2020

Available online 22 May 2020

### Keywords:

COVID-19

Forecast

Gaussian

USA

## ABSTRACT

In this report, we analyze historical and forecast infections for COVID-19 death based on Reduced-Space Gaussian Process Regression associated to chaotic Dynamical Systems with information obtained in 82 days with continuous learning, day by day, from January 21<sup>th</sup>, 2020 to April 12<sup>th</sup>. According last results, COVID-19 could be predicted with Gaussian models mean-field models can be meaning- fully used to gather a quantitative picture of the epidemic spreading, with infections, fatality and recovery rate. The forecast places the peak in USA around July 14<sup>th</sup> 2020, with a peak number of 132,074 death with infected individuals of about 1,157,796 and a number of deaths at the end of the epidemics of about 132,800. Late on January, USA confirmed the first patient with COVID-19, who had recently traveled to China, however, an evaluation of states in USA have demonstrated a fatality rate in China (4%) is lower than New York (4.56%), but lower than Michigan (5.69%). Mean estimates and uncertainty bounds for both USA and his cities and other provinces have increased in the last three months, with focus on New York, New Jersey, Michigan, California, Massachusetts, ... (January e April 12<sup>th</sup>). Besides, we propose a Reduced-Space Gaussian Process Regression model predicts that the epidemic will reach saturation in USA on July 2020. Our findings suggest, new quarantine actions with more restrictions for containment strategies implemented in USA could be successfully, but in a late period, it could generate critical rate infections and death for the next 2 month.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

China and Italy shut down transportation in every single way on January to March 2020; just in China 28 quarantines was implemented [1]. Late January 2020, USA found the first infection with a old woman, who travelled from Wuhan [2], due to infections and suddenly death in New York, the government initiated a quarantine period with strict social distancing measures. The problem associated to fast transmission is associated to asymptomatic proportion in a city, near to 18%, investigated in a Diamond Princess cruise ship [3]. In the last month, some research papers have investigated the evolution of COVID-19 outbreak in Asia (China) [4] and Europe (Italy, France and Spain) [5]. This research have analyzed epidemic data made available on the Center for Systems Science

and Engineering at Johns Hopkins University [6], the available data analyzed is considered between January 21<sup>th</sup> 2020 and April 12<sup>th</sup> 2020, included, with a feedback process in a neural network applied; it allows to examined the information in real time in each state, at Fig. 1, New York is the city with more infection around USA with 7867 deaths and 172,348 total confirmed with COVID-19 with more than 800 new deaths. In this process, the government is implemented new strategies for COVID-19 control, however, it should consider a forecast according USA behavior.

In Fig. 2, the maximum peak of the active cases variation was on March 19<sup>th</sup> with 0.43% until April 12<sup>th</sup> with 0.07%, quarantine effects have reduced 16%.

Besides, many countries have high variability with infections due to quarantine and sanitary decisions taken by the different Governments as Italy, Spain, China, UK. Although the COVID-19 deaths may be affected by many factors, **this study is to explore the effect from forecast in COVID-19 infections, using Reduced-Space Gaussian Process Regression for Data-Driven Probabilistic on infections.**

\* Corresponding author.

E-mail addresses: [ricardoariasvelasquez@hotmail.com](mailto:ricardoariasvelasquez@hotmail.com) (R.M. Arias Velásquez), [jjemejia@gmail.com](mailto:jjemejia@gmail.com) (J.V. Mejía Lara).

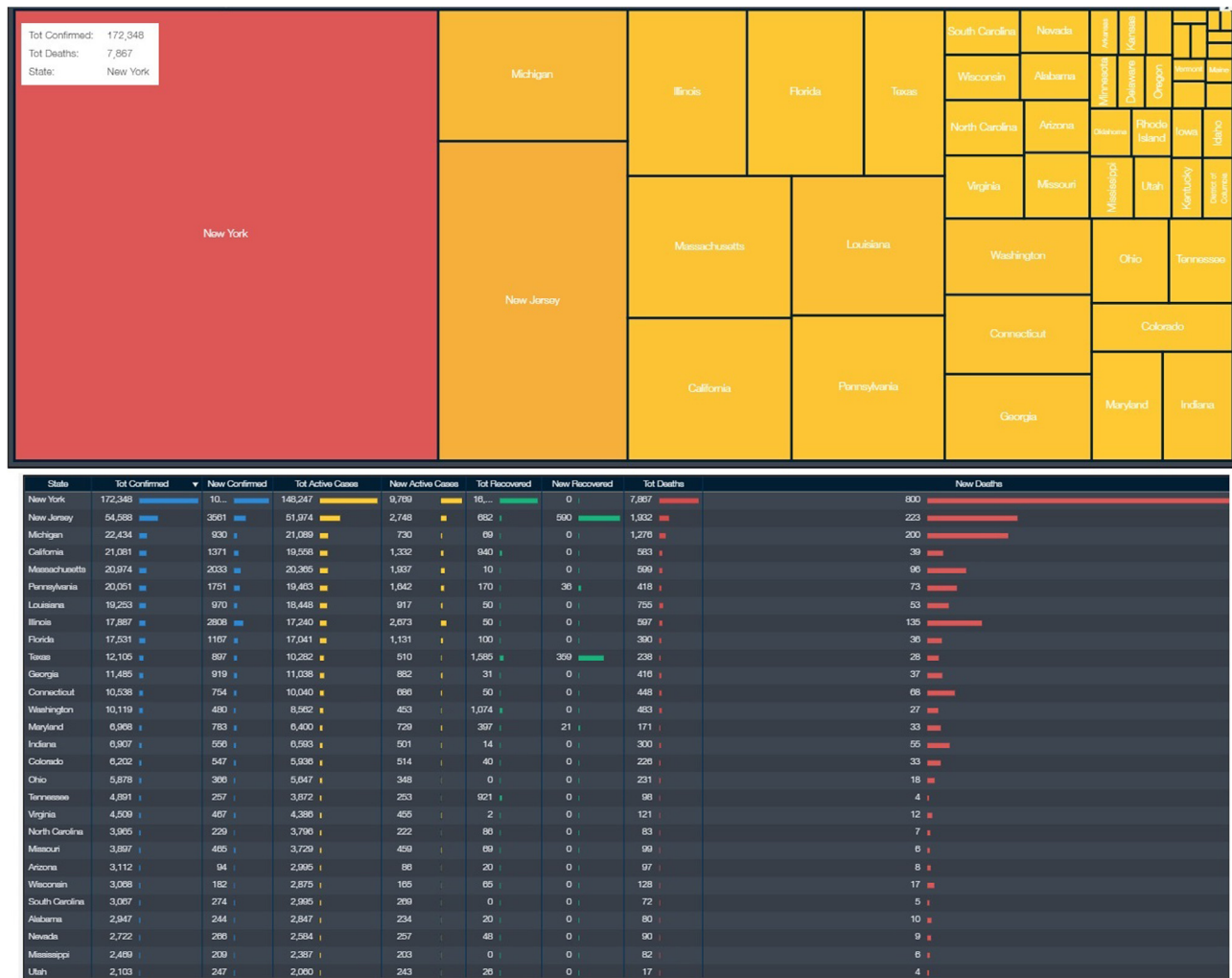


Fig. 1. Distribution infections for USA.

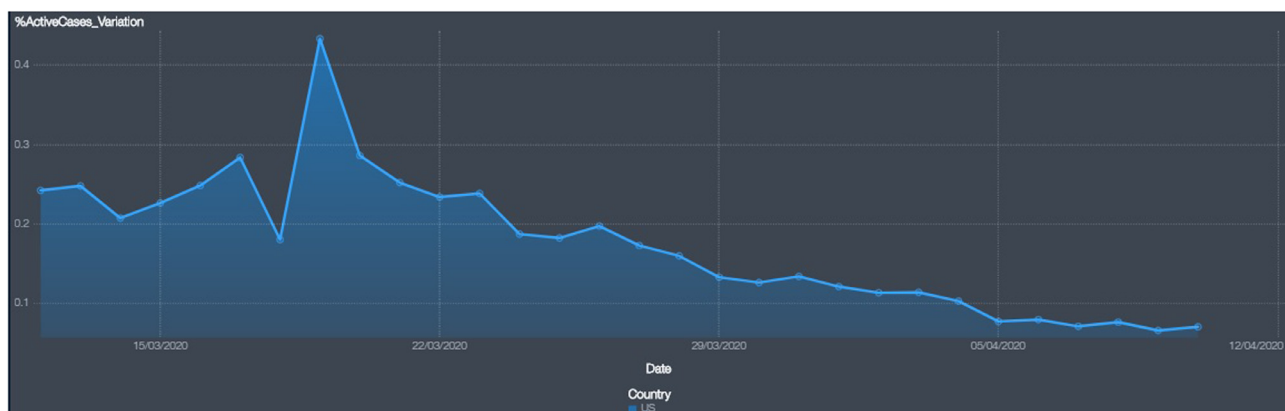


Fig. 2. Active cases variation on percent associated from March to April 2020.

The rest of this paper is organized as follows: Section 2, describe background and Reduced-Space Gaussian Process Regression for Data-Driven Probabilistic for COVID-19. Section 3 develops the results of this functions and main findings. Section 4, concludes.

## 2. Background and methodology

### 2.1. Systematic review

In this field, a new systematic review with Population, Intervention, Comparison, and Outcome (PICO) method, we have reviewed from Science Direct the knowledge contribution based on the following algorithm: ((Forecast)OR(PREDICTION)) AND(COVID-19)NOT (reflections)NOT (Cell)NOT (radiology)NOT (radiation), with 29 research articles available at April 12<sup>th</sup> 2020, the main contribution is described, as follows:

- Patient Information Based Algorithm called PIBA allows to predict case fatality for any infection with asymptomat, however, with new constraints, this model lost accuracy to predict the rate of severe disease [7].
- Imposing control would have imporant impact on COVID-19. Furthermore, “according to the existing data abroad, we also make bold predictions of the epidemic development trends in South Korea, Italy, and Iran, pointing out the possible outbreaks and the corresponding control time, and tracing the earliest transmission dates of countries” [8].
- . With COVID-19, the spatial representation of the disease by using GIS platform allows to verified the “material, population and social psychology at three scales: individual,group and regional” [9]; in this case, with GIS technology is necessary to implement big data techniques, for cross validations and analysis.
- With Bayesian inference and Markov chain Monte Carlo algorithm, they have demonstrated more accuracy in forecast associated to COVID-19 [10].
- Associated to GPR the idea is to “utilize nonlinear diffusion map coordinates and formulate a deterministic dynamical system on the system manifold” [12]. A interesting approach has a reduced-space data-driven dynamical system with an evaluation, with efficiency with low intrinsic dimensionality.
- An “advantage of employing GPR to reconstruct the reduced-order dynamics is the simultaneous estimation of the dynamics and the associated uncertain” [12].

Due to dynamic of this pandemic, a stochastic modeling with Bayesian application should be considered in the forecasting analysis.

**Study design:** We proposed an analytical study associated to correlational study due to the geographical information system (GIS) available.

**Methods:** We proposed a novel Bayesian method for chaotic Dynamical Systems, according COVID-19 pandemic behavior, our method is Reduced-Space Gaussian Process Regression.

### 2.2. Gaussian process regression methodology for forecast and infections by COVID-19

This research proposed a Gaussian Process Regression called GPR [11], with application in dynamical and chaotic systems. In this report, GPR is used with probabilistic regression framework, with a training data set with Eq. (1), of  $N$  pairs of vectors with an input  $x_n \in \mathbb{R}$  and “noisy scalar output  $y_n$ ” [11]. For infection population,  $y_n$  should create a model generalized to the distribution of the output at unseen input location associated to confirmed patients and death. Likewise, noise in output models represents ob-

servation error; therefore, Gaussian distribution, generates input-output relationship in Eq. (2).

$$D = (y_n, x_n), n = 1, \dots, N, \quad (1)$$

$$y = f(x) + \epsilon, \epsilon \sim N(0, \sigma_{noise}^2). \quad (2)$$

Where:

$\sigma_{noise}^2$ : “Variance of the noise” [12].  $f(x)$ : Latent variable or function values, non-observed. If Gaussian process is a set of random values, them they must be indexed by some  $x \in X$  as a subset. With Bayes applications as recommended [10], it is possible to “make inferences on function values to unseen inputs conveniently using a finite number of training data” [13]. For this process, we considered mean function  $m(x)$  and a covariance function  $k(x, x')$  as Eqs. (3) and (4).

$$m(x) = E[f(x)], \quad (3)$$

$$k(x, x') = E[(f(x) - m(x)) \times (f(x') - m(x'))]. \quad (4)$$

Where:

$E[.]$ : Expectation.  $m(x)$ : It is zero with data centering, associated to lower and upper value in the forecasting. GPR framework requires functions and dependent on the input  $x$ . It is strictly symmetric and positive semi-definite if they are consolidated at each pair of points, this results are define by prior distribution on  $f(x)$ , in Eq. (4).

$$k(x, x') = \theta_1 \exp\left(-\frac{1}{2} \sum_{l=1}^L \frac{(x_l - x'_l)^2}{\theta_2}\right). \quad (5)$$

Where:

$\theta_1$ : Is a hyper-parameter with maximum covariance in chaotic systems.

$\theta_2$ : Is a strictly positive hyper-parameter, with decrease rate in correlations, it can be used for death, confirmed and recovered patients. In Eq. (5), it calculated the squared exponential covariance function, with  $x_1$  as  $l$ th component of  $x$  and  $x'$ . The function decreases fast if distant pairs of input  $x$  and  $x'$ . A property is the weak correlation associated to  $f(x)$  and  $f(x')$ . Besides,  $\sigma_{noise}^2$  represents uncertainty associated to the order-reduction as a probable observations error.

About forecasting and classifier, the hyper-parameter  $\{\theta_1, \theta_2, \theta_3\}$  should be integrated as  $f$ , as a vector of training latent variables, in Eq. (6), with matrix on Eq. (7). In this case,  $f^*$  is a second subscript, used covariance function  $k(.,.)$  and corresponding hyper-parameters. with a variable change  $y$ , as a conditional probability for the training observation, it should obtain knowledge from data [14], it describes on (8).

$$p(f, f^*) = N(0, X), \quad (6)$$

$$X = \begin{bmatrix} K_{f,f} & K_{*,f} \\ K_{f,*} & K_{*,*} \end{bmatrix}, \quad (7)$$

$$p(y|f) = N(f, \sigma_{noise}^2 I). \quad (8)$$

In Eq. (9), Bayes rule is written with a normalized process to find  $(f, f^*)$  in Eq. (10).

$$p(f, f^*|y) = \frac{p(f, f^*)p(y, f)}{p(y)}, \quad (9)$$

$$p(f^*|y) = \int p(f, f^*|y)df. \quad (10)$$

In Eq. (11), associates to “conditioning the joint Gaussian prior distribution on the observations, resulting in the closed-form Gaussian distribution” [12]. With Eq. (12),  $f^*$ , the mean and covariance should be directly added to obtain Eq. (13).

$$p(f^*|y) = \frac{1}{2} \int p(f, f^*)p(y|f)df \quad (11)$$

$$y^* = f^* + \epsilon^* \quad (12)$$

$$p(f^*|y) = N\left(K_{*,f}(K_{f,f} + \sigma_{noise}^2 I)^{-1}y, K_{**} - K_{*,f}(K_{f,f} + \sigma_{noise}^2 I)^{-1}K_{f,*}\right) \quad (13)$$

$$p(f^*|y) \cong N\left(F', \sum_{f^*} k\right) \quad (14)$$

Finally, in Eq. (14) makes feasible to use up to more than twelve thousands of training data set to make classifiers and forecasting. For cross validation a neural network (NN) is composed of nodes set and synapses, it needs a signal or input of a data set [15], in this research a “performs computation by propagating the signal along the connections”, until, it influences the reply in the inner layer. It is compared to “biological neuron’s spiking action, a non-linear activation function is applied to nodes in any hidden layer and the output layer” [16]. Method has been considered with the fit using Keras, for CNN a nonlinear function is considered in the Eq. (15).

$$P(i, t) = 1 - e^{(-\alpha \times \sum_{j \in I(t)} k(i, j) + \epsilon)} \quad (15)$$

Where:

$\theta$ : Model parameter.

The Eq. (15) has the results of “weights on the connections” among nodes of the neural network. This model starts with the input layer, later, four hidden layers and ends at the output layer [17]. Covariance are calculating linear combination of the input nodes in the Eq. (16).

$$x \in R \quad (16)$$

In the Eq. (16), individually node in the hidden layer has “linear combination” [18] with a chain events. The hidden layers are composed as following:

- Convolutional layers: For abstract local features at different locations [15].
- Pooling layers: It uses the average value from each subarea of previous layer.
- Fully-connected layers: It has a function similar to regular neural network. This type is powerful for seizing local geometric features, spatial patterns and detects larger-scale features in deeper layers [19].

“Linear transformation of the covariates  $x$ ” [20] or inputs are described in the Eqs. (17) and (18).

$$h^{in} = W^{(1)} \times x + b^{(1)} \quad (17)$$

$$W^{(1)} \in R^{M \times M} \quad (18)$$

In Eq. (18) describes a matrix of weights, it is analogous to regression coefficient in the Eq. (19).

$$b^{(1)} \in R^M \quad (19)$$

$$h^{in} \in R^M \quad (20)$$

In the Eq. (20), the hidden layer is described. Besides, a sigmoid non-linearity is indicated in the Eq. (21), it is a hyperbolic tangent function.

$$h^{out} = \sigma \times (h^{in}) = \frac{1}{1 + e^{-h^{in}}} \quad (21)$$

Where:  $h$ : it is a description of the hidden layer, due to restriction of the “linearly transformed and passed to the output layer”.

$$y^{in} = W^{(2)} \times h^{out} + b^{(2)} \quad (22)$$

In Eq. (22), it is the output of the process, the hidden layer is linearly transformer through the output layer. With Eq. (22) the multinomial logistic regression (MLR) function is applied in the Eq. (23), it provides the input vector probability, the results belong “to class  $k$ ” [21], with the limits indicated in the Eq. (24), for cross-validation.

$$y^{out_k} = P(Y = k|x) = \frac{e^{y^{in_k}}}{\sum_{l=1}^K e^{y^{in_l}}} \quad (23)$$

$$k \in 1, \dots, K \quad (24)$$

Where:  $y$ : It is the maximum output vector; it has described the classifier as to the class of  $x$ .  $k$ : Class.

About the Fig. 3, we provide a basic view of training input and neural network size and shape along with other training hyper-parameters. From GIS platform, we obtain information from USA, with probabilistic regression framework, with a training data of  $N$  pairs of vectors with an input  $x_n$  and noisy scalar output  $y_n$  [21], therefore, Gaussian distribution, generates input location associated to confirmed patients and death. In this Fig. 3, the function decreases fast if distant pairs of input  $x'$ , a weak correlation associated to  $f(x)$  and  $f(x')$  with  $\sigma_{noise}^2$  associated to the order-reduction.

This process, uses a Convolutional Neural Network (CNN) [22], instead of other neural networks, since considers five types of layers: Convolution Layer, Activation Layer, Pooling Layer, Fully Connected Layer and Soft-Max Classifier, and from this group, it is selected the predictions that have better precision and lower entropy, which leads to a shorter computational time.

### 3. Results, case study USA

If we considered a baseline China results until April 2020, the fatality rate is 4%, then USA has a mean fatality rate in 2.9% on April 12<sup>th</sup>, with a recovery rate on 12%. As Fig. 4, New York with 172,348 confirmed and fatality rate of 4.56% and total deaths of 7867 has an especially consideration for density population. Besides, New Jersey has confirmed with 54,588 with a fatality rate of 3.54% and total deaths of 1932. The last one, is under China reference. On the other hand, Kentucky and Michigan have more fatality rate with 5.32% and 5.69% respectively.

However, China has implemented robust strategies for distancing control and quarantine around this country, therefore, the recovery rate is 90%. Far away from New York, with recovery rate of 9.42% and on the other hand, cities as Texas is 13.09%. Besides, Hawaii has different rate of 56.79% for the restrictions initiated with more influence in their population.

Combining a mathematical model with multiple datasets, Figs. 6 and 7, we found that the median daily  $R_t$  of COVID-19 in USA probably varied between 2.6 and 5.6 in April, 2020, before 3 weeks of travel restrictions were introduced. We also estimated on May 11<sup>th</sup>, the estimated active patients will be 829,110 (peak), with 95,670 deaths and estimated contagion 1,066,549. Based on our estimates of  $R_t$ , assuming COVID-19 variation, we calculated that in locations with similar transmission potential to New York in early April, once there are at least four independently introduced



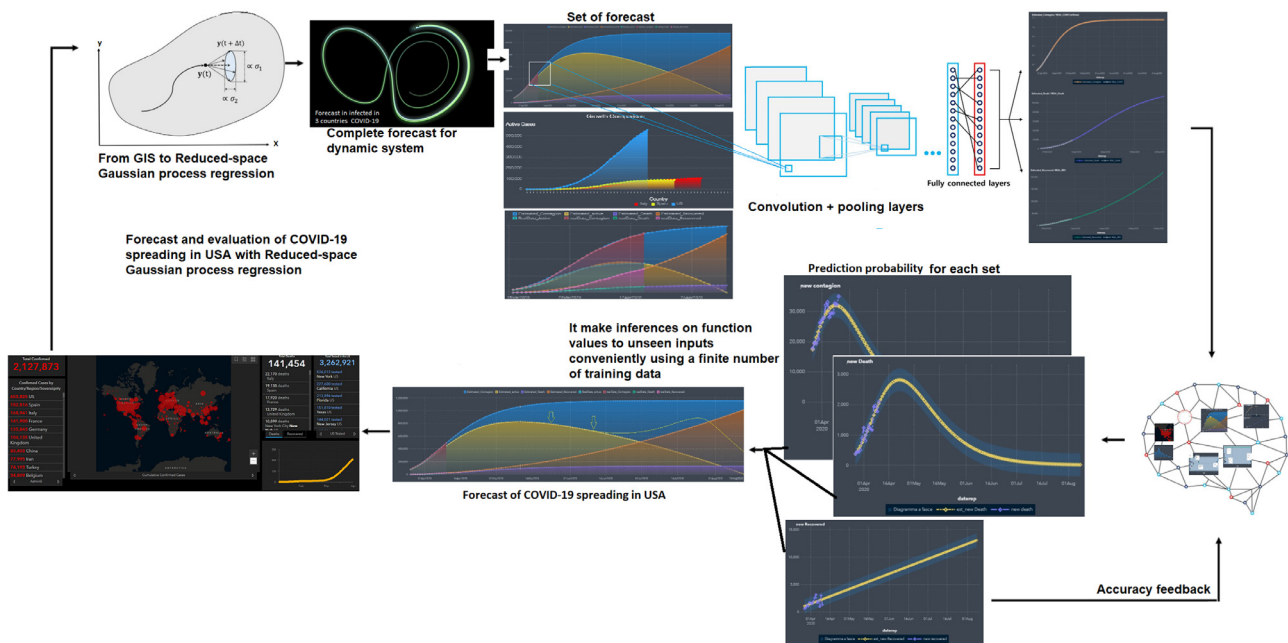


Fig. 3. Methodology diagram.



Fig. 4. Fatality rate (%) in function of confirmed patients for COVID-19 in USA.

cases, there is a more than 50% chance the infection will establish within that population in Michigan, New Jersey, California, Massachusetts, Illinois, Florida, Louisiana and Pennsylvania with higher than 1.12% in each state.

Besides, on Fig. 7, the forecast places the peak in USA around July 14<sup>th</sup> 21 2020, with a peak number of 132,074 death with infected individuals of about 1,157,796 and a number of deaths at the end of the epidemics of about 132,800.

In the GPR analysis, forecast has detected three clusters, in the early days of March 2020, before quarantine, after quarantine actions during 2 weeks later and the last cluster during the second week on April 2020.

Associated to the median  $R_t$  estimated during April for travel restrictions were introduced, we estimated that a single introduction of COVID-19 transmission would have a 30% to 34% probability of causing a large outbreak, Fig. 8.

In Fig. 8, the new contagion curve demonstrated the maximum value on April 9<sup>th</sup> with 35,049 new contagion, after this day, the values have been lower. However, new death peak will occur on 23<sup>th</sup> with 3202 deaths. Finally new recoveries will occur the peak on August 19<sup>th</sup>.

Some states as Wyoming, Utah, South Dakota and West Virginia have lower COVID-19 infections than other states. It requires verified if test has been done or if population has abided by all regulations associated to quarantine in a local level, as Fig. 4.

The results of GPR has an average Target Value of 0.9688312 with an inverted Covariance Matrix with highest value of 0.9936848383. At the same time, the inverted covariance matrix \* Target-value Vector has highest value of 0.496957.

About the results of the model, they have a correlation coefficient of 98.91% and relative absolute error 99.19 %. With the model associated to infections are described in Fig. 7.

### 3.1. Discussion about the process

About the deterministic models, where the future is considered already fixed, and knowing that in the past, many infections studies and physics phenomenon were performed under the model mentioned, we agree that this is the common viewpoint to come away with. However, we think that in previous big scale studies as pandemic physics, these equations have not demonstrated accuracy in



Fig. 5. Recovery rate (%) in function of total recovered patients for COVID-19 in USA.

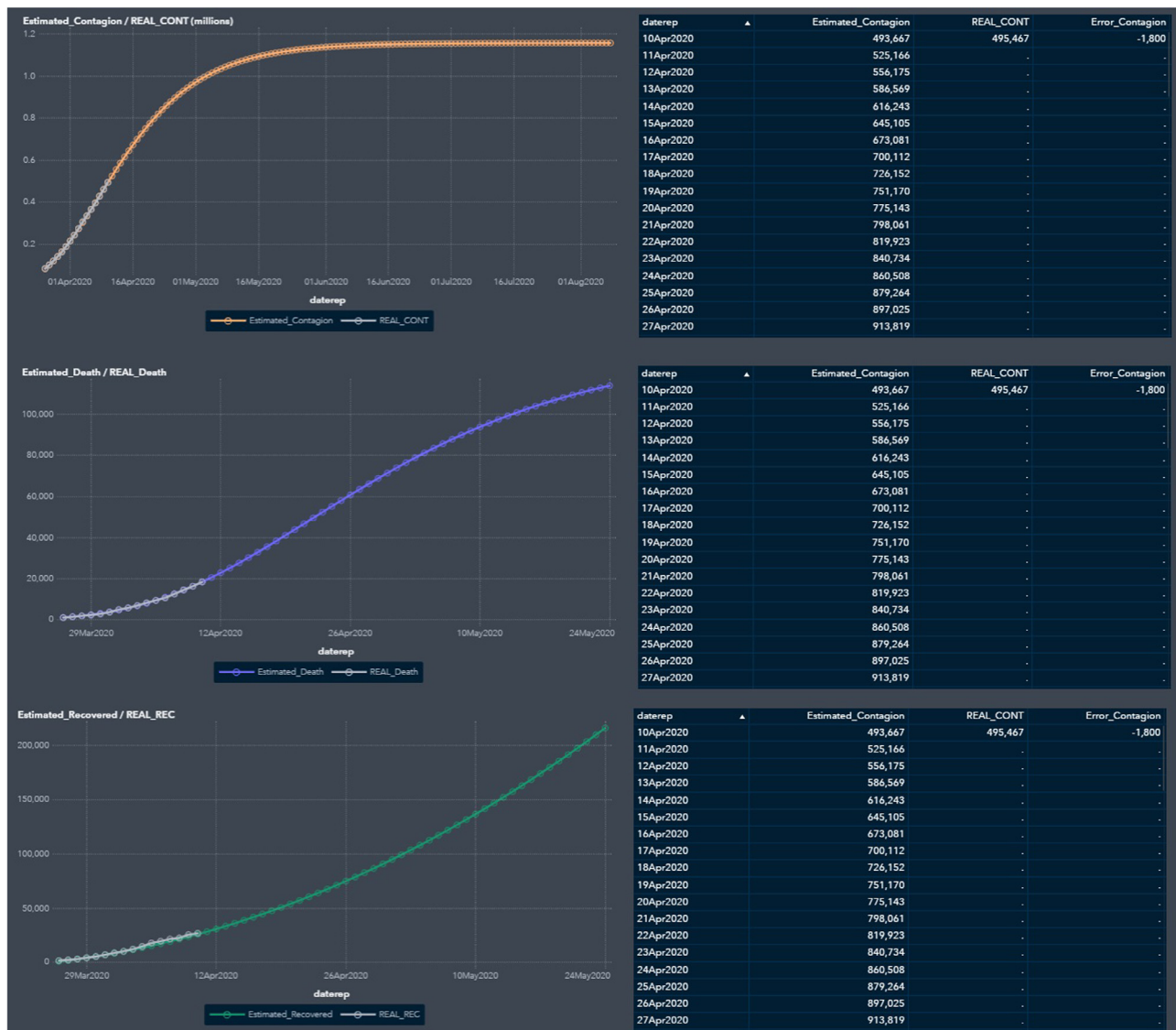


Fig. 6. Forecast for contagion, death and recovery associated to COVID-19 in USA.

the forecast curves, since their behavior, never match in the space and this is because, for a long range of initial conditions, the system evolves in one state. Therefore, the traced paths never intersect, since the pandemic responds to external stimulation in the way that managed and, that is why establishes new starting points that do not intersect.

Then, in Fig. 9, the curves never form loops and consequently are not predictable, and this is where we see that small actions (for example government quarantines), could dramatically change the behavior of the pandemic, and that is precisely what has been observed in all countries, given that in each of them, the pandemic has behaved in a different way, unpredictable and changing, ac-



Fig. 7. Forecast contagion, active, death, recovered until last estimate active patients.

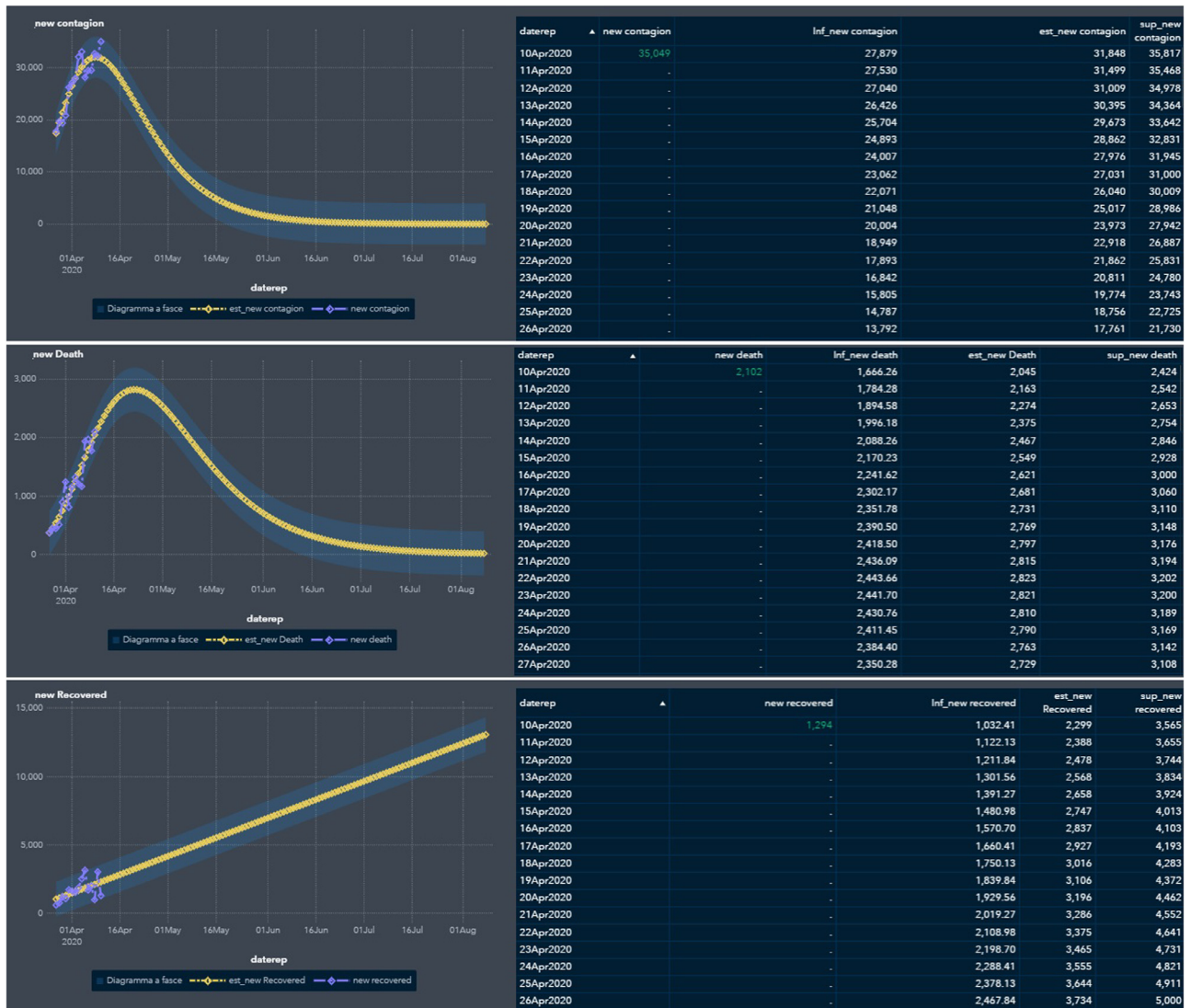
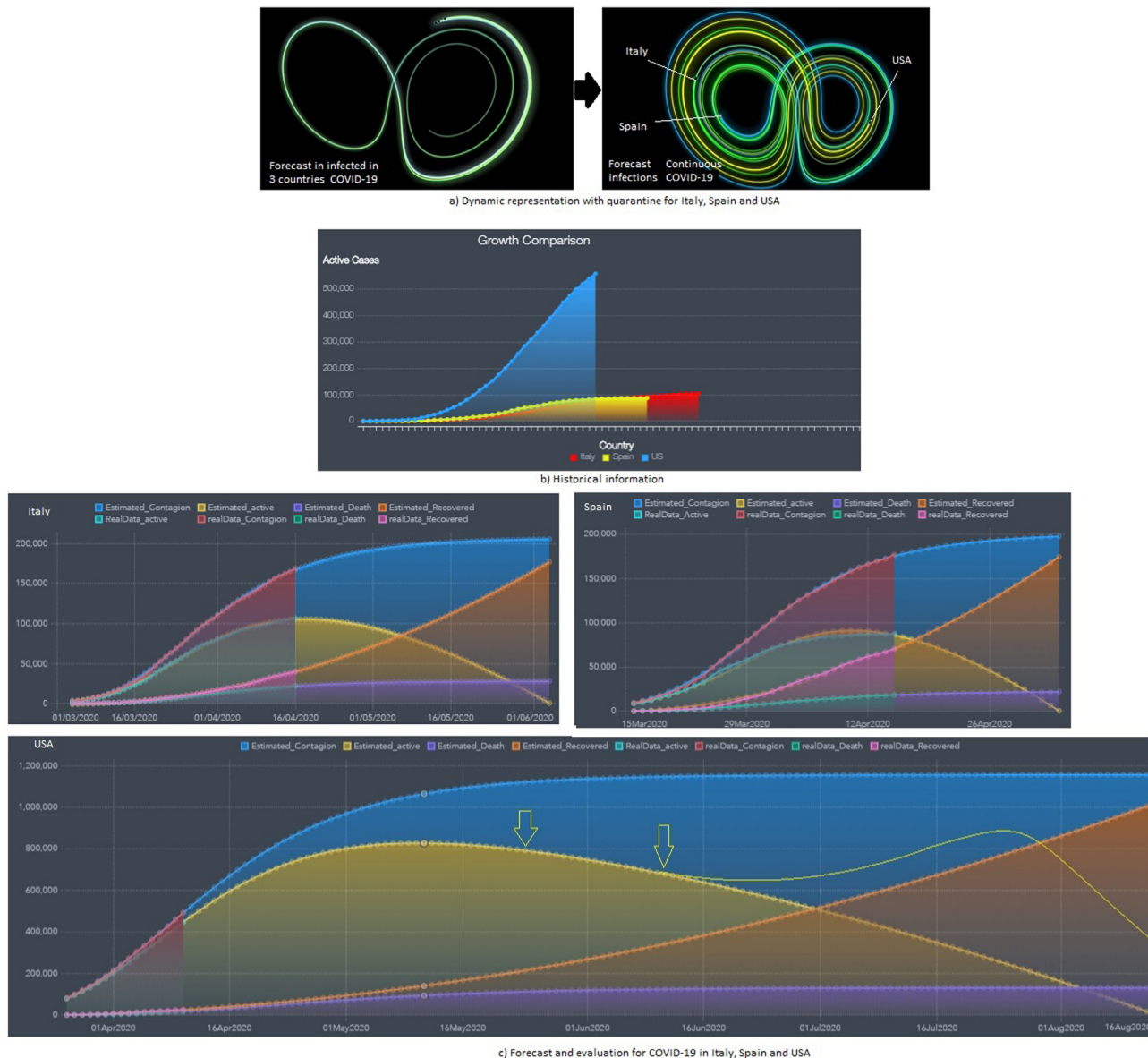


Fig. 8. Reported cases by date of onset (blue points) and estimated USA cases from each state by date of onset (Yellow line) for new contagion, deaths and recovered patients. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)





**Fig. 9.** a) Dynamic representation, b) historical information c) forecast and evaluation in USA, Spain and Italy.

According to the factors that have been inferred in its treatment. Finally, the COVID-19 forecast, performed as chaotic-dynamic system, has a good result than traditional models, because of the sensitive dependence on initial conditions. It uses the Lorenz's system displayed, so it seems like a paradox, where the system is both deterministic and unpredictable, because, you could never know the initial conditions with perfect accuracy, for the contagious pattern.

In traditional methods, we detect mistakes in the forecast associated to old pandemic researches, and as we can see, the COVID-19 has behaved in unpredictable and changing ways, according to the factors used in its treatment. Therefore, we think that the pandemic behaves as a dynamic-chaotic system, which sets new starting points that do not intersect and changes its behavior according to the way it is treated. Consequently, this study could be a real contribution for understanding this problem with a good tool of processing a systematic review. As we can see in the Fig. 9, the example started with three countries, with three initial states. Hence they have been evolving together until some days, but after a month, they diverge (e.g. Spain and Italy), due to the strategies implemented by the government as quarantine, closed fron-

tiers and more constraints. From being arbitrarily close together, they end up on different trajectories; then, it demonstrates that depends on initial conditions and the actions taken. In the equations described in the Section 2, there is nothing random at all about this system of equations, even if you could input exactly the same initial conditions, you would get exactly the same result. Though the case presented behaves setting different starting points, so any difference in initial conditions, no matter how tiny, will variates the final state.

Finally, with almost the same initial conditions, the forecasts never release the same evolution of the COVID-19 and it makes changes in the behavior. Hence, the infection will forever be unpredictable.

#### 4. Conclusions

While it is very difficult to evaluate if some state has implemented all the test required, an assessment tieh fatality rate, recovery rate (Figs. 4 and 5), our data evaluates that unreported or undetected cases are likely, especially in countries with low HAQ



index. In this research, we proposed timely short-term forecasts of the cumulative number of reported cases of the COVID-19 epidemic in USA per states as of January 21<sup>th</sup> to April 12<sup>th</sup>, 2020. As the epidemic continues, we have created a new Bayesian system with GIS information available [6], the selected algorithms for chaotic dynamic system is Reduced-Space Gaussian Process Regression with information obtained in 82 days with continuous learning, day by day, the correlation coefficient is 98.91%. This new model will recognize the new changes in the constraints associated to COVID-19. Our analyses identified: May 11<sup>th</sup>, the estimated active patients will be 829,110 (peak), with 95,670 deaths and estimated contagion 1,066,549; likewise, July 14<sup>th</sup> 2020, with a peak number of 132,074 death with infected individuals of about 1,157,796 and a number of deaths at the end of the epidemics of about 132,800 is a key date for new measurements associated to control of this infections. Further studies are needed to investigate new government restrictions or economic activation and more associated issues.

### Grants/financial support

None.

### Declaration of Competing Interest

There is no conflict of interest in this work.

### CRediT authorship contribution statement

**Ricardo Manuel Arias Velásquez:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **Jennifer Vanessa Mejía Lara:** Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing.

### Acknowledgments

Recognition to Universidad Nacional de San Agustín de Arequipa, for the knowledge contribution in this research.

### References

- [1] Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus 2019-ncov outbreak. *J Travel Med* 2020; taaa020.
- [2] Chowell G, Mizumoto K. The COVID-19 pandemic in the USA: what might we expect? *Lancet* 2020;395(102304–10):1093–4.
- [3] Neher RA, Dyrda R, Druelle V, Hodcroft EB, Albert J. Potential impact of seasonal forcing on a SARS-cov-2 pandemic. *medRxiv* 2020. Published online March 3.
- [4] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, Yan P, Chowell G. Real-time forecasts of the COVID-19 epidemic in china from february 5th to february 24th. *Infectious Disease Modelling* 2020;5:256–63. 2020.

- [5] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in china. Italy and France. *Chaos, Solitons & Fractals* 2020;134. Article 109761.
- [6] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;3099(20):19–20.
- [7] Wang L, Li J, Guo S, Xie N, Yao L, Cao Y, et al. Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. *Sci Total Environ* 2020:138394.
- [8] Li L, Yang Z, Dang Z, Meng C, Huang J, Meng H, et al. Propagation analysis and prediction of the COVID-19. *Infect Dis Model* 2020;5:282–92.
- [9] Zhou C, Su F, Pei T, Zhang A, Du Y, Luo B, et al. COVID-19: Challenges to GIS with big data. *Geogr Sustainability* 2020:1–13.
- [10] Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model* 2020;5:271–81.
- [11] Rasmussen C, Williams C. Gaussian processes in machine learning. Cambridge, MA: The MIT Press; 2005.
- [12] Wan ZY. Themistoklis p. sapsis, reduced-space gaussian process regression for data-driven probabilistic forecast of chaotic dynamical systems. MIT Ind Liaison Program 2016:1–31.
- [13] Berry T, Giannakis D, Harlim J. Nonparametric forecasting of low-dimensional dynamical systems. *Phys Rev E* 2015;91(3):32915.
- [14] Velásquez RMA, Jennifer vanessa mejía lara, andres melgar, converting data into knowledge for preventing failures in power transformers. *Eng Fail Anal* 2019;101:215–29.
- [15] Velásquez RMA, Lara JVM. Corrosive sulphur effect in power and distribution transformers failures and treatments. *Eng Fail Anal* 2018;92:240–67.
- [16] Keras C.F. 2015. <https://github.com/fchollet/keras>, 1–2.
- [17] Pattern recognition and image preprocessing, 1–5. Bow S., editor. CRC Press; 2002.
- [18] Zhang W, Jin L, Song E, Xu E. Removal of impulse noise in color images based on convolutional neural network. *Appl Soft Comput* 2019;82(1–11):10558.
- [19] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–324.
- [20] Selvikvag A, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* 2019;29(2):102–27.
- [21] Velásquez RMA, Lara JVM. Harmonic failure in the filter of static var compensator. *Eng Fail Anal* 2020;107:104207.
- [22] Shanthi T, Sabeenian RS, Anand R. Automatic diagnosis of skin diseases using convolution neural network. *Microprocess Microsyst* 2020;76:103074.



**Ricardo Manuel Arias Velásquez.** IEEE senior member, he studied his PhD in engineering by PUCP, MSC concentration in project management, project engineer and electrical engineer degree by UNSA. Scientific researcher certificated by CONCYTEC Perú. He is expert in energy, artificial intelligence, non-linear system and robotics. Currently, he works in ENEL associated to operational efficiency for renewable energy in Latin America. He is vice chairman in IEEE PES Perú. He has published with affiliation to UTEC, MIT, UNSA, PUCP.



**Jennifer Vanessa Mejía Lara.** IEEE senior member, she studied his PhD in engineering by PUCP, MSC concentration in project management, electrical engineer degree by Universidad Nacional de Colombia. Currently, she works as project manager of renewable energies in Latin America. Scientific researcher certificated by CONCYTEC Perú. She has published with affiliation to MIT, UNSA, PUCP.