


Proceeding Paper

Estimation of COVID-19 Dynamics in the Different States of the United States during the First Months of the Pandemic [†]

Ignacio Rojas-Valenzuela ^{1,*}, Olga Valenzuela ², Elvira Delgado-Marquez ³ and Fernando Rojas ⁴ 

¹ School of Technology and Telecommunications Engineering, University of Granada, 18071 Granada, Spain

² Department of Applied Mathematics, University of Granada, 18071 Granada, Spain; olgavc@ugr.es

³ Department of Economics and Statistics, University of Leon, 24071 León, Spain; elvira.delgado@unileon.es

⁴ Department of Architecture and Computer Technology, University of Granada, 18071 Granada, Spain; frojas@ugr.es

* Correspondence: e.ignaciorojas@go.ugr.es

[†] Presented at the 7th International conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

Abstract: Estimation of COVID-19 dynamics and its evolution is a multidisciplinary effort, which requires the unification of heterogeneous disciplines (scientific, mathematics, epidemiological, biological/bio-chemical, virologists and health disciplines to mention the most relevant) to work together towards a better understanding of this pandemic. Time series analysis is of great importance to determine both the similarity in the behavior of COVID-19 in certain countries/states and the establishment of models that can analyze and predict the transmission process of this infectious disease. In this contribution, an analysis of the different states of the United States will be carried out to measure the similarity of COVID-19 time series, using dynamic time warping distance (DTW) as a distance metric. A parametric methodology is proposed to jointly analyze infected and deceased persons. This metric allows comparison of time series that have a different time length, making it very appropriate for studying the United States, since the virus did not spread simultaneously in all the states/provinces. After a measure of the similarity between the time series of the states of United States was determined, a hierarchical cluster was created, which makes it possible to analyze the behavioral relationships of the pandemic between different states and to discover interesting patterns and correlations in the underlying data of COVID-19 in the United States. With the proposed methodology, nine different clusters were obtained, showing a different behavior in the eastern zone and western zone of the United States. Finally, to make a prediction of the evolution of COVID-19 in the states, Logistic, Gompertz and SIR models were computed. With these mathematical models, it is possible to have a more precise knowledge of the evolution and forecast of the pandemic.

Keywords: COVID-19; pandemic in the united states; time series; DTW distance; hierarchical clustering; SIR model



check for
updates

Citation: Rojas-Valenzuela, I.; Valenzuela, O.; Delgado-Marquez, E.; Rojas, F. Estimation of COVID-19 Dynamics in the Different States of the United States during the First Months of the Pandemic. *Eng. Proc.* **2021**, *5*, 53. <https://doi.org/10.3390/engproc2021005053>

Academic Editors:
Luis Javier Herrera
and Hector Pomare

Published: 13 July 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/>).

1. Introduction

The COVID-19 epidemic started in Hubei Province, China, around December 2019. Since then, the disease has spread to all continents and countries of the world, being categorized as a pandemic by World Health Organization on 11 March 2020.

In recent months, contributions have been made that analyze the evolution of the pandemic in different countries, implementing mathematical models to predict their evolution. Traditional predictive models for infectious diseases mainly include models for predicting differential equations and models for predicting time series based on statistics and random processes.

For example, in [1] a methodology with the aim of estimating the actual number of people infected with COVID-19 in France is presented, since according to the authors, the number of screening tests carried out and the methodology do not directly calculate

the actual number of cases and infection mortality rate (IFR). A mechanistic–statistical approach was developed that combines an epidemiological SIR model that describes these unobserved epidemiological dynamics, a probabilistic model that describes the data collection process and a method of statistical inference.

The logistic growth model, the generalized logistic growth model, the generalized growth model and the generalized Richards model were used to model the number of infected cases in the 29 provinces of China (and several countries), performing a detailed analysis on the heterogeneous situations by four phases of the outbreak in China [2].

In [3] the Kermack–McKendrick SEIR model (Susceptible, Exposed, Infectious and Recovered) is presented to analyze the effects of behavioral changes on the reduction in community transmission in Mexico. A variable contact rate over time is proposed and the consequences of disease spread in an affected population of non-essential activities is analyzed.

The behavior of the virus in Japan has also been analyzed [4]. By 29 February 2020, in addition to the 619 confirmed cases (passengers and crew members) infected with COVID-19 in a cruise ship (near Tokyo), 215 locally transmitted cases had been also confirmed in Japan. To evaluate the effectiveness of reaction strategies based on avoiding large accumulations or crowded areas and to predict the spread of COVID-19 infections in Japan, in [4] a stochastic transmission model produced by expanding the epidemiological model based on SIR (Susceptible-Infected-Removed) had been presented. The simulation results showed that the number of Infected and Removed patients will increase rapidly if there is no reduction in the time spent in crowded zone.

In [5] using the Maximum-Hasting (MH) parameter estimation method and the SEIR model, the spread of COVID-19 and its prediction in South Africa, Egypt, Nigeria, Senegal, Kenya, and Algeria under three intervention scenarios (suppression, mitigation, mildness) is presented.

In addition to the most relevant epidemiological models used in the literature, models typically based on time series have also been used to analyze the behavior of the pandemic in different countries. The autoregressive integrated moving average (ARIMA) model is a mathematical model widely studied in the context of time series that has been successfully applied in the field of health (estimate the incidence and prevalence of influenza mortality, malaria incidence, hepatitis, and other infectious diseases) as well as in different fields in the past due to its simple structure, fast applicability and ability to explain the data set.

In [6], ten Brazilian states are analyzed using the autoregressive integrated moving average (ARIMA), the cubistic regression (CUBIST), the random forest (RF), ridge regression (RIDGE), the support vector regression (SVR) and the stacking-ensemble learning in the task of time series forecasting of the number of patients infected with COVID-19 with one, three, and six-days ahead. A forecasting model based on ARIMA has also been presented in [7] for Pakistan, presenting the high exponential growth in the number of confirmed cases, deaths and recoveries. In [8], ARIMA time series models were applied to forecast the total confirmed cases of COVID-19 for the next ten days using the model ARIMA (0,2,1), ARIMA (1,2,0) and ARIMA (0,2,1) for Italy, Spain, and France, respectively.

Currently, the analysis of the evolution of COVID-19 in America is of great importance due to the impact of this epidemic on this continent. In this contribution we will focus on the United States. The first patient detected in the United States was a travel-associated case from Washington state on 19 January 2020. The preponderance of initial cases of infected patients with COVID-19 in the United States were correlated with travel to a “high-risk” country or close contacts of previously identified cases corresponding to the testing criteria adopted by the Centers for Disease Control and Prevention (CDC) (<https://www.cdc.gov/>, accessed March 2020). From 1 to 31 March 2020, the number of reported COVID-19 cases in the United States rapidly increased from 30 to 188,172, the number of deaths rising from 1 to 5531, and the virus being detected in all the states. At the end of April, the number of infected reached 1,069,424 and the number of deceased stood at 62,996. At the time of writing this contribution (14 June 2020) the number of infected is more than 2×10^6 and

more than 100.000 deaths, the United States being one of the countries that is suffering the most from COVID-19.

In a recent paper [9], an attempt was made to estimate the actual number of infected people, even if they have not been counted. It was estimated that the true number of COVID-19 cases in the United States is likely in the tens of thousands, suggesting substantial undetected infections and spread within the country [10].

This contribution presents a methodology to analyze the evolution patterns of COVID-19 in the states of United States (including Puerto Rico and the District of Columbia). A parametric similarity measure is presented, based on a robust distance measure between time series, the dynamic time warping distance (DTW), with which the number of infected and dead in each of the states can be compared simultaneously, even though the start of the epidemic originated on different dates in each zone (therefore, the time series that need to be compared have different lengths).

To the best of our knowledge, this contribution is the first study that tries to develop a hierarchical clustering time series algorithm in order to globally compare and classify the behavior of all the states of United State simultaneously in their evolution of infected and deceased patients suffering COVID-19. Carrying out this classification is very useful, since it will allow the establishment of similarities and patterns in the evolution of the pandemic among the states of the United States.

2. Material and Methods

A time series is a sequence of numerical (temporal) data points in successive order, which is naturally high dimensional and large in data size. There are two main operations that could be performed when working with time-series with its sequential data: (a) the analysis of a single time series; (b) the analysis of multiple time series simultaneously. This contribution is concentrated on the analysis of multiple time series for all the states of US suffering COVID-19, with the purpose of finding similarities between multiple time series by performing a clustering time-series methodology.

Clustering such complex objects is particularly advantageous because it may lead to the discovery of interesting patterns in time-series datasets, which contributes to a better understanding of the COVID-19 spread in different regions of the United States.

Clustering of time-series sequences has received noteworthy attention [11,12], not only as a formidable exploratory method and powerful tool for discovering patterns, but also as a pre-processing step or subroutine for other tasks [13].

In this section, the database used is presented first (Section 2.1). Subsequently, a review of the most popular distance measures for time series is described (Section 2.2) and a new parametric distance is proposed.

2.1. Data Set

The COVID-19 epidemic data set used in this contribution was collected from the Johns Hopkins University [14]. In this platform, the number of confirmed, deaths and recovered cases until 14 June 2020 for different countries are presented. For the United States, two additional .csv files are provided, in with detail of administration and province/state is reported (including Puerto Rico and District of Columbia). In order to compare countries behavior, the time-series data are divided by state population.

2.2. Similarity/Distance Measure in Time Series

In a simplified way, the similarity of two simple time series with the same number of points (denoted by m), and defined by $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_m\}$, can be achieved by simply calculating the Minkowski (or Euclidean) distance (shortest path between two points) between points on both time series that happen at the same time. This

distance is the measure of similarity, denoted as $d(X, Y)$, and it is a function that takes both times series (X, Y) as input and calculates their distance “ d ”, defined as:

$$d(X, Y) = \left(\sum_{i=1}^m |x_i - y_i|^k \right)^{\frac{1}{k}} \quad (1)$$

when $k = 2$, the distance between two series is called Euclidean Distance. Using the Minkowski distance is a good metric to analyze the similarity of two time series, if these time series are synchronized (that is, all similar events in both time series occur at exactly the same time) and have the same length.

The evolution of time series in the different states of the United States present a different start date, both for the number of confirmed and death cases, and therefore its length is also different. Suppose as an analogy the time series of the sound of a mother’s voice when she speaks slowly to her child. If the mother says the same phrase quickly, the child will most likely recognize that she is still his mother. However, if the Euclidean distance between both series were used as a metric, these two time series would have a very low similarity and would not be considered fundamentally equal. This would lead to the conclusion that the two voices did not come from the same person. To solve this problem, the dynamic time warping distance (DTW) method is frequently used in the bibliography [15].

DTW is a technique that can be considered as an extension of the Euclidean Distance between series [16], that calculates an optimal match between two given time series with certain restriction, performing non-linearly in the series (by stretching or shrinking along its time-axis). This distortion (denoted as warping) between two time series is used to find corresponding regions and determine the similarity between them.

The DTW of two series X and Y , defined as $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_m\}$ is computed in the following way. An n -by- m matrix D is computed with the (i, j) th element, defining the local distance of two elements by:

$$d(x_i, y_j) = (x_i - y_j)^2 \quad (2)$$

The point-to-point alignment between series X and Y can be represented by a time warping path W , defined as:

$$W = \begin{pmatrix} w_x(k) \\ w_y(k) \end{pmatrix}, k = 1, 2, \dots, p \quad (3)$$

where p is the length of the warping path W , and $w_x(k)$ and $w_y(k)$ represent the indexes in time series X and Y , respectively. The warping path $\begin{pmatrix} w_x(k) \\ w_y(k) \end{pmatrix}$ indicates that the $w_x(k)$ th element in time series X maps to the $w_y(k)$ th element in time series Y . There are some constraints and rules for the construction of the warping path:

- Every index from the first time series must be matched with one or more indices from the other time series (and vice versa)
- The first (the same for the last index) index from the first time series must be matched (not only this match) with the first (last) index from the other time series. That is, the warping path should start at $W(1) = (1, 1)$ and end up at $W(p) = (n, m)$.
- The mapping of the indices from the first time series to indices from the other time series must be monotonically increasing, and vice versa. The adjacent elements of path W , $W(k)$ and $W(k + 1)$ must be subject to $w_x(k + 1) - w_x(k) \geq 0$ and $w_y(k + 1) - w_y(k) \geq 0$.
- The warping path should also have the property of continuity, mathematically expressed as adjacent elements of path W , $W(k)$ and $W(k + 1)$ must be subject to $w_x(k + 1) - w_x(k) \leq 1$ and $w_y(k + 1) - w_y(k) \leq 1$.

The optimal match is denoted by the match that satisfies all the restrictions and the rules and that has the minimal cost, where the cost is computed as the sum of absolute differences, for each matched pair of indices, between their values. The DTW (minimal distance and optimal warping path) could be found using a dynamic programming algorithm:

$$RD(x_i, y_j) = d(x_i, y_j) + \min \begin{cases} RD(x_{i-1}, y_{j-1}) \\ RD(x_{i-1}, y_j) \\ RD(x_i, y_{j-1}) \end{cases} \quad (4)$$

$$DTW(X, Y) = \min\{RD(x_n, y_m)\}$$

where $RD(x_i, y_j)$ is the minimal cumulative distance from $(0, 0)$ to (i, j) in matrix D . In the methodology proposed in this paper, for each of the states analyzed, both the time series of the number of infected and the time series of deaths will be simultaneously taken into account.

If each of these time series needs to be weighted differently, the following parametric metric, $DTW_\alpha(S_A, S_B)$ is defined:

$$DTW_\alpha(S_A, S_B) = \alpha DTW(TSD_A, TSD_B) + (1 - \alpha)DTW(TSC_A, TSC_B) \quad (5)$$

that measures the similarity in the evolution of the COVID time series for two states of the United States (S_A y S_B), TSC_A and TSC_B represent the time series of the number of infected, TSD_A and TSD_B represent the time series of the number of deaths for the states S_A and S_B , respectively. The parameter α (with $0 \leq \alpha \leq 1$) indicates the relative relevance given to the similarity measure, taking into account the time series of infected or deaths.

2.3. Clustering Method for Time Series

Clustering is a data mining technique in which similar data are divided into related or homogeneous groups, in an unsupervised way, that is, without a priori advanced knowledge of the data. For the problem presented in this contribution, working with time series of states of the United States suffering COVID-19, given a set of individual time series data, the objective is to group similar time series into the same cluster.

The problem of grouping time series data is formally defined by, given a dataset of N time series data $Q = \{X_1, X_2 \dots X_N\}$, finding in an unsupervised way a partition of Q into K cluster, denoted as $C = \{C_1, C_2, \dots C_k\}$, taking into account that homogeneous or similar series are grouped together based on a certain similarity/distance measure. In this paper, the parametric metric $DTW_\alpha(S_A, S_B)$ is used and there is no intersection between clusters, therefore:

$$Q = \cup_{i=1}^K C_i; \text{ with } C_i \cap C_j = \emptyset \text{ (} i \neq j \text{)} \quad (6)$$

The methods used in the area of time series clustering [11,17] are usually based in a conventional clustering algorithm by substituting standard distance measurements with a more suitable distance to compare time series (raw methods) or converting series into normal data and using directly classical algorithms (feature-based methods and models).

Among the most popular clustering algorithms, the hierarchical clustering and the k-means algorithm are widely used in time series clustering. In this contribution the hierarchical clustering is used, mainly due to its great visualization power and its simple and intuitive interpretation.

Hierarchical clustering creates a nested hierarchy of similar time series, according to a pair-wise distance matrix of the time series analyzed. The similarity measure $DTW_\alpha(S_A, S_B)$ used is therefore essential in this time-series clustering process.

The most widely used linkage criteria, such as single, average and complete linkage variants [18], were analyzed. Hierarchical clustering can be converted into a partitional clustering, with k cluster, by cutting the first k links.

3. Results and Discussion

To evaluate the performance of the proposed method, several experiments are conducted in this section for three values of the parameter α in the distance metric $DTW_\alpha(S_A, S_B)$. The time series of the states of the United States was taken from John Hopkins database. Data range from 22 January 2020 to 14 June 2020, covering all the states (including Puerto Rico and District of Columbia). For the computation of the distance metric, a threshold I_{min} was defined, defining the minimum number of infected people to start the time series, being for this study $I_{min} = 5$ (the number of confirmed was greater than 5). Therefore, the length of the time series is different for each state, being on average 101 days (Figure 1). The index of each of the states is presented in Table 1.

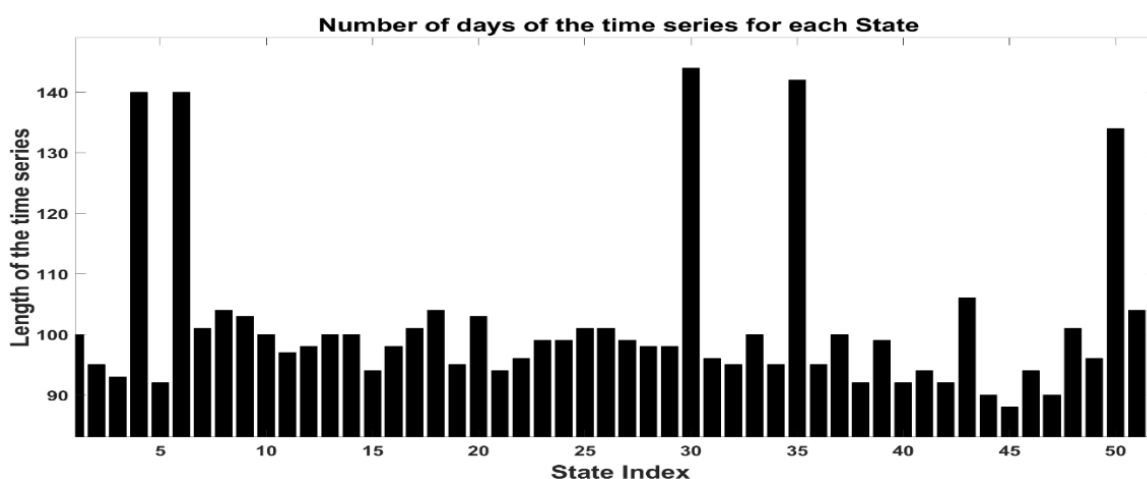


Figure 1. Length of the different time series, corresponding to the states analyzed.

Table 1. Distribution of the states obtained by means of hierarchical clustering with 9 clusters ($\alpha = 0.5$ and, in bold, the state for which the SIR model is calculated). $D_{Cluster}$ is the distance between the elements that make up a cluster (its value is zero in the case that there is only one element in a cluster).

Cluster Number (C_N)	$D_{Cluster}$	States ($\alpha = 0.5$)
1	0.015	(1) Nebraska , (2) South Dakota
2	0.042	(3) Alabama, (4) Arizona, (5) Arkansas, (6) California , (7) Colorado, (8) Florida, (9) Georgia, (10) Indiana, (11) Iowa, (12) Kansas, (13) Kentucky, (14) Minnesota, (15) Mississippi, (16) Missouri, (17) Nevada, (18) New Hampshire, (19) New Mexico, (20) North Carolina, (21) North Dakota, (22) Ohio, (23) Oklahoma, (24) South Carolina, (25) Tennessee, (26) Texas, (27) Utah, (28) Vermont, (29) Virginia, (30) Washington, (31) Wisconsin
3	0.019	(32) Michigan, (33) Pennsylvania
4	0.031	(34) Delaware, (35) Illinois , (36) Louisiana, (37) Maryland
5	0.024	(38) Alaska, (39) Hawaii, (40) Idaho, (41) Maine, (42) Montana , (43) Oregon, (44) Puerto Rico, (45) West Virginia, (46) Wyoming
6	0	(47) District of Columbia
7	0	(48) New Jersey
8	0.033	(49) Connecticut, (50) Massachusetts, (51) New York
9	0	(52) Rhode Island

The values of the parameter α analyzed will be $\{0, 0.5, 1\}$. This section of results begins with the value of $\alpha = 0.5$, that is, the information of the confirmed patients time series with the same relevance as the time series of deaths for the final computation of the distance

$DTW_\alpha(S_A, S_B)$. The distance matrix between the different states is presented in Figure 2 (for a better visual representation, the distance matrix was multiplied by a constant and the states are ordered according to the cluster to which each one belongs).

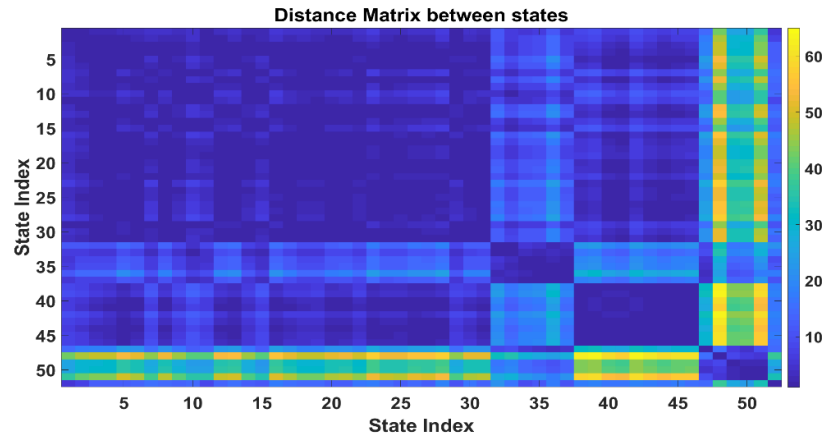


Figure 2. Distance or similarity symmetric matrix to characterize the behavior of the time series for the states of the United States (parameter $\alpha = 0.5$). The greater the similarity, the smaller the distance between the series (being the diagonal of this matrix of zero value).

It is important to highlight the existence of various clusters with only one state (corresponding with District of Columbia, New Jersey and Rhode Island). Cluster 7 (New Jersey, listed as 48 in Table 1) links directly to cluster 8 ((49) Connecticut, (50) Massachusetts, (51) New York), which denote similar behavior between these states. For cluster 6 (District of Columbia, number 47), the linkage is performed for both cluster 3 ((32) Michigan, (33) Pennsylvania) and cluster 4 ((34) Delaware, (35) Illinois, (36) Louisiana, (37) Maryland). There are two large clusters (cluster 2 and cluster 5) that contain 29 and 9 states, respectively. Its linkage is performed through cluster 1, which contains only two states (Nebraska and South Dakota).

The similarities and distances between the different states and clusters obtained can be analyzed using the results presented in the hierarchical clustering (Figure 3) and distance matrix (Figure 2). Figure 4 presents a geographical representation of all the states according to the cluster grouping obtained in Table 1, using the similarity between the clusters obtained using the dendrogram in Figure 3.

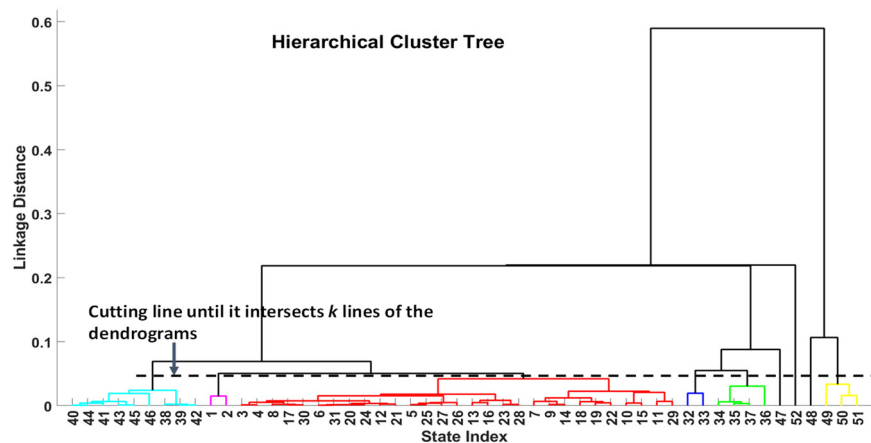


Figure 3. Hierarchical cluster tree obtained using as the distance metric the $DTW_\alpha(S_A, S_B)$ and $\alpha = 0.5$.

Informed Consent Statement: Not applicable.

Acknowledgments: This contribution has been partially supported by the projects with reference RTI2018-101674-B-I00 and CV20-64934.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Roques, L.; Klein, E.; Papaix, J.; Sar, A.; Soubeyrand, S. Using Early Data to Estimate the Actual Infection Fatality Ratio from COVID-19 in France. *Biology* **2020**, *9*, 97. [[CrossRef](#)] [[PubMed](#)]
2. Wu, K.; Darcet, D.; Wang, Q.; Sornette, D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. *Nonlinear Dyn.* **2020**. [[CrossRef](#)] [[PubMed](#)]
3. Acuña-Zegarra, M.; Santana-Cibrian, M.; Velasco-Hernandez, J. Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. *Math. Biosci.* **2020**, *325*, 108370. [[CrossRef](#)] [[PubMed](#)]
4. Karako, K.; Song, P.; Chen, Y.; Tang, W. Analysis of COVID-19 infection spread in Japan based on stochastic transition model. *BioSci. Trends* **2020**, *14*, 134–138. [[CrossRef](#)] [[PubMed](#)]
5. Zhao, Z.; Li, X.; Liu, F.; Zhu, G.; Ma, C.; Wang, L. Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Sci. Total Environ.* **2020**, *729*, 138959. [[CrossRef](#)] [[PubMed](#)]
6. Ribeiro, M.H.D.M.; da Silva, R.G.; Mariani, V.C.; Coelho, L.d.S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos Solitons Fractals* **2020**, *135*, 109853. [[CrossRef](#)] [[PubMed](#)]
7. Yousaf, M.; Zahir, S.; Riaz, M.; Hussain, S.M.; Shah, K. Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos Solitons Fractals* **2020**, *138*, 109926. [[CrossRef](#)] [[PubMed](#)]
8. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ.* **2020**, *729*, 138817. [[CrossRef](#)] [[PubMed](#)]
9. Perkins, A.; Cavany, S.M.; Moore, S.M.; Oidtman, R.J.; Lerch, A.; Poterek, M. Estimating unobserved SARS-CoV-2 infections in the United States. *Proc. Natl. Acad. Sci. USA* **2020**. [[CrossRef](#)] [[PubMed](#)]
10. Fauver, J.R.; Petrone, M.E.; Hodcroft, E.B.; Shioda, K.; Ehrlich, H.Y.; Watts, A.G.; Vogels, C.B.F.; Brito, A.F.; Alpert, T.; Grubaugh, N.D.; et al. Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell* **2020**, *181*, 990–996. [[CrossRef](#)] [[PubMed](#)]
11. Aghabozorgi, S.; Shirkhorshidi, A.; Teh Ying, W. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
12. Johnpaul, C.I.; Prasad, M.V.N.K.; Nickolas, S.; Gangadharan, G.R. Trendlets: A novel probabilistic representational structures for clustering the time series data. *Expert Syst. Appl.* **2020**, *145*, 113119.
13. Taoying, L.; Xu, W.; Zhang, J. Time Series Clustering Model based on DTW for Classifying Car Parks. *Algorithms* **2020**, *13*, 57.
14. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *3099*, 19–20. [[CrossRef](#)]
15. Keogh, E.; Ratanamahatana, C.A. Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **2005**, *7*, 358–386. [[CrossRef](#)]
16. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Sign. Process.* **1978**, *26*, 43–49. [[CrossRef](#)]
17. Bandara, K.; Bergmeir, C.; Smyl, S. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Syst. Appl.* **2020**, *140*, 112896. [[CrossRef](#)]
18. Kaufman, L.; Rousseeuw, P. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 344.