

Estimación de casos de COVID-19 en países de Suramérica empleando modelos ARIMA (Autorregresivo Integrado de Promedio Móvil)

Esther D. Gutiérrez

Escuela Superior Politécnica del Litoral
Centro de Física, Instituto Venezolano de Investigaciones Científicas
orcid: 0000-0001-7579-0711
sterguti@gmail.com
Venezuela –Ecuador

Rafael Puche

Unidad de Estudios Genéticos y Forenses
Instituto Venezolano de Investigaciones Científicas
orcid: 0000-0001-5392-2429
rpucheq@gmail.com
Venezuela

Fernando Hernández

Centro de Medicina Experimental
Instituto Venezolano de Investigaciones Científicas
orcid: 0000-0003-4988-1454
fernandoh76@gmail.com
Venezuela

Fecha de recepción: 10 - 06 - 2020 Fecha de aceptación: 18- 07- 2020

Resumen

El objetivo principal de este trabajo es emplear modelos ARIMA para la estimación de nuevos contagios usando datos públicos disponibles para Venezuela y la región suramericana, actualmente foco principal de un segundo brote de la COVID-19. Se realiza la predicción a 30 días del número de casos de Covid-19 en países suramericanos usando los datos públicos disponibles. Se emplearon modelos ARIMA para estimar el impacto de nuevos contagios en las dinámicas de infección para Suraméri-

ca. Desde la aparición del primer caso de la nueva neumonía Covid-19 en China, esta enfermedad se ha convertido en un problema de salud pública global y representa un gran reto el control de la infección para los países de Suramérica. Al 24 de junio de 2020 un total de 1.866.090 casos han sido detectados en la región y en el caso particular de Venezuela un total de 4.365 casos. El rápido incremento en el número de casos y la alta tasa de contagios asociado con el virus han llevado al desarrollo de distintas aproximaciones matemáticas, tales como: modelos SIR, SEIR, redes

neuronales y regresiones lineales que permitan predecir la probable evolución de la epidemia. Los modelos ARIMA han sido empleados con éxito en otras infecciones como influenza, malaria, SARS, entre otras. Los resultados de las estimaciones realizadas empleando estos modelos muestran que aún en la región hacen falta mayores esfuerzos que conlleven al control de la epidemia.

Palabras clave: Covid-19; pandemia; predicción; modelos ARIMA

Estimation of COVID-19 cases in South American countries using ARIMA models

Abstract

The main objective of this work is to use ARIMA models for the estimation of new contagions using public data available for Venezuela and the South American region, currently the main focus of a second COVID19 outbreak. A 30-day prediction is made for the number of Covid-19 cases in South American countries using available public data. ARIMA models were used to estimate the impact of new contagions on infection dynamics for South America. Since the appearance of the first case of

the new Covid-19 pneumonia in China, which has become a global public health problem and the great challenge that the infection has represented for the countries of South America to June 24, 2020, a total of 1,866,090 cases have been detected and in the particular case of Venezuela a total of 4,365 cases have been detected for the same date. The rapid increase in the number of cases and the high rate of contagion associated with the virus have led to the development of different mathematical approaches, such as: SIR, SEIR models, neural networks and linear regressions

that allow predicting the probable evolution of the epidemic. The ARIMA model has been successfully used in other infections such as influenza, malaria, SARS, among others. In the following work, the 30-day prediction of the number of Covid-19 cases in South American countries is made using public data available. The results of the estimates made using these models show that even in the region, greater efforts are needed to control the epidemic.

Key words: Covid-19; pandemic; prediction; ARIMA model

Introducción

A finales del 2019, la Organización Mundial de la Salud (OMS) informó sobre la existencia de un grupo de 27 casos de neumonía, con etiología desconocida, en la ciudad de Wuhan, Provincia de Hubei, en China. Posteriormente, se determinó que la causa fue provocada por un nuevo tipo de corona virus denominado SARS-Cov2. Luego de esto, la propagación del virus se ha incrementado de manera exponencial en todo el planeta, convirtiéndose en una pandemia que, hasta la fecha, ha infectado a más de 8 millones de personas en el mundo. Un estudio preliminar proporcionó una estimación temprana del 3% para la tasa global de letalidad de los casos infectados (Wang, Horby, Hayden, & Gao, 2020).

Diversas aproximaciones estadísticas han sido empleadas para la predicción del número de infectados en una epidemia entre los que destacan regresión lineal multivariada (Du et al., 2020; Ghosal, Sengupta, Majumder, & Sinha, 2020), redes neuronales (Al-Najjar & Al-Rousan, 2020), modelos de simulación basados en el SIR (Anastassopoulou, Russo, Tsakris, & Siettos, 2020; Giordano et al., 2020) S (población susceptible), I (población infectada) y R (población recuperada) e incluso ya se han creado paquetes de software de redes neuronales profundas que pretenden detectar la enfermedad por medio del uso de imágenes de rayos X (Khan, Shah, & Bhat, 2020). Sin embargo, a pesar de la apabullante cantidad de información sobre la predicción de la enfermedad, es difícil obtener resultados confiables debido a que solo muestran

tendencias y una alta aleatoriedad.

En este sentido, se ha implementado el modelo Autorregresivo Integrado de Promedio Móvil (ARIMA) (Box, Jenkins, & Reinsel, 2013) con éxito en diferentes áreas de la ciencia y por primera vez en el campo de salud pública para la evaluación histórica de muertes causadas por influenza (CHOI & THACKER, 1981), varios estudios emplean dicho modelo en la predicción de la prevalencia e incidencia en enfermedades infecciosas tales como Malaria, Síndrome Agudo Respiratorio Severo (SARS), influenza, tuberculosis y hepatitis. Y más recientemente se ha usado en Europa para la estimación de nuevos casos en países como España, Italia y Francia (Bekiros, Kouloumpou, & Ceylan, 2020).

El trabajo se organiza así: a) Introducción, b) Metodología; c) Implementación el modelo ARIMA; d) Resultados; e) Conclusiones. Finalmente, se presenta la bibliografía que fundamenta el desarrollo del trabajo.

Metodología

Obtención de los datos

Para obtener datos epidemiológicos de países de Suramérica se usó la librería COVID-19 de Emanuele Guidotti, disponible en: <https://covid19datahub.io/>, que provee una base de datos unificada de los casos de la COVID-19 a nivel mundial (Guidotti & Ardia, 2020).

Los datos específicos para Venezuela se obtuvieron de una página oficial del

Gobierno Venezolano, específicamente de la API de acceso libre, disponible en: <https://covid19.patria.org.ve>

Implementación del Modelo ARIMA

El modelo Autorregresivo Integrado de Promedio Móvil, ARIMA (por sus siglas en inglés) es una metodología econométrica basada en modelos dinámicos que utiliza datos de series temporales para describir la evolución temporal como una función lineal de datos previos y errores debidos al azar. El método utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Los análisis fueron realizados empleando el lenguaje de programación estadístico R (R Core Team, 2020) (versión 3.6.3), en conjunto con el entorno interactivo de desarrollo R Studio (RStudio Team, 2020) (versión 1.2.533).

Se define un modelo ARIMA (p, d, q), donde los parámetros p , d y q son números enteros positivos que indican el orden de las distintas componentes del modelo: p denota el número de términos autorregresivos, d denota el número de veces que la serie debe ser diferenciada para hacerla estacionaria, q es el orden de las Medias móviles o el número de términos de la media móvil invertible.

En general, se dice que una serie temporal Y_t admite una representación autorregresiva integrada y de medias móviles de órdenes p , d y q respectivamente, y se expresa de la forma:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Donde y_t , es la serie temporal diferenciada (uno puede diferenciar más de una vez), los parámetros θ 's son las medias móviles, los parámetros ϕ 's son los terminos autoregresivos y ε_t es ruido blanco.

La construcción de los modelos se realiza de manera iterativa mediante la metodología de Box y Jenking (Asteriou & Stephen G., 2011) en el que se distinguen cuatro fases:

- **Identificación:** Utilizando los datos ordenados cronológicamente se intenta sugerir un modelo posible que sigue la serie. El objetivo es determinar los valores que sean apropiados para reproducir la serie temporal.

- **Análisis y diferenciación de la serie temporal:** Consiste en examinar la estacionariedad, los diagramas de auto-

correlación, también conocidos como ACF y PACF, y elección del orden del modelo.

- **Ajuste del modelo ARIMA:** Obtención de los coeficientes de determinación. Es el diagnóstico, en donde se comprueba que los residuos no poseen estructura de dependencia y siguen un ruido gaussiano.

- **Predicción:** Una vez seleccionado el mejor modelo candidato se pueden hacer predicciones con el mismo.

Resultados

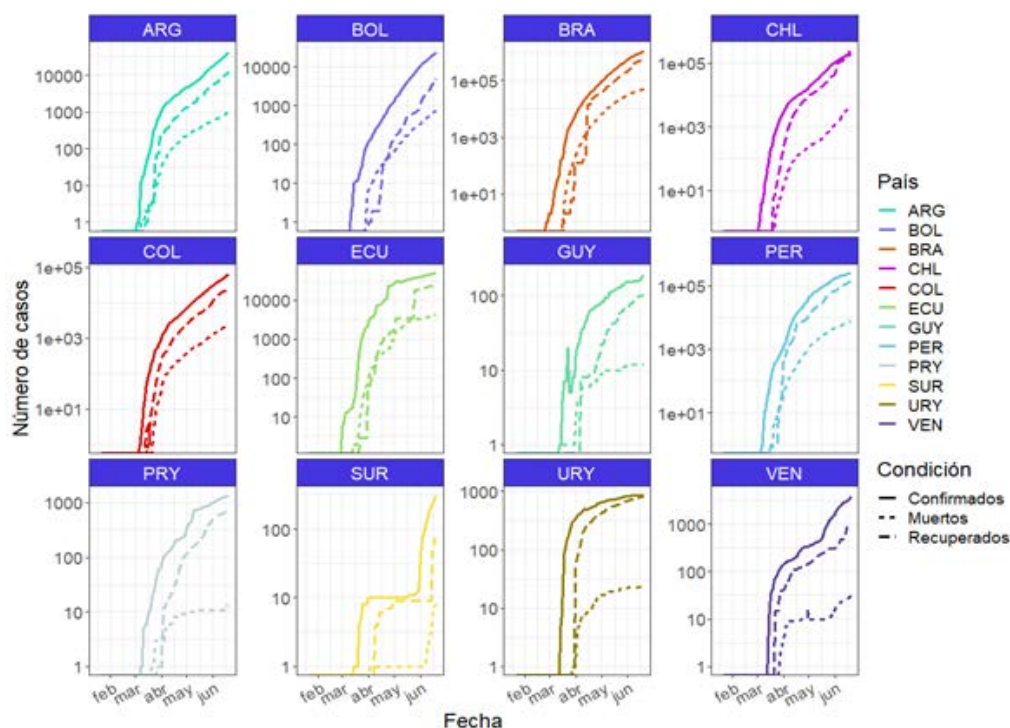
Casos acumulados para países de Suramérica

Tal y como se puede observar en la Figura 1, el número de casos confirma-

dos posee una evolución temporal de tipo exponencial en todos los países salvo en Uruguay. En este último se podría inferir que a pesar de tener un incremento rápido en la población infectada pareciera haber tomado las medidas necesarias para detener el crecimiento del número de casos.

Otra observación importante es que el crecimiento de los casos recuperados es proporcional al número de casos confirmados por país a excepción de Venezuela que evidencia un retraso en la curva de los casos recuperados en comparación con el número de casos confirmados diarios, lo que da cuenta de un crecimiento más rápido en la difusión de la enfermedad, lo que indica que se hace necesaria la toma de medidas para el frenado de la pendiente.

Figura 1. Número acumulado de casos al 20 de junio del 2020 para diversos países de Suramérica



Fuente: Elaboración propia del autor, (2020)

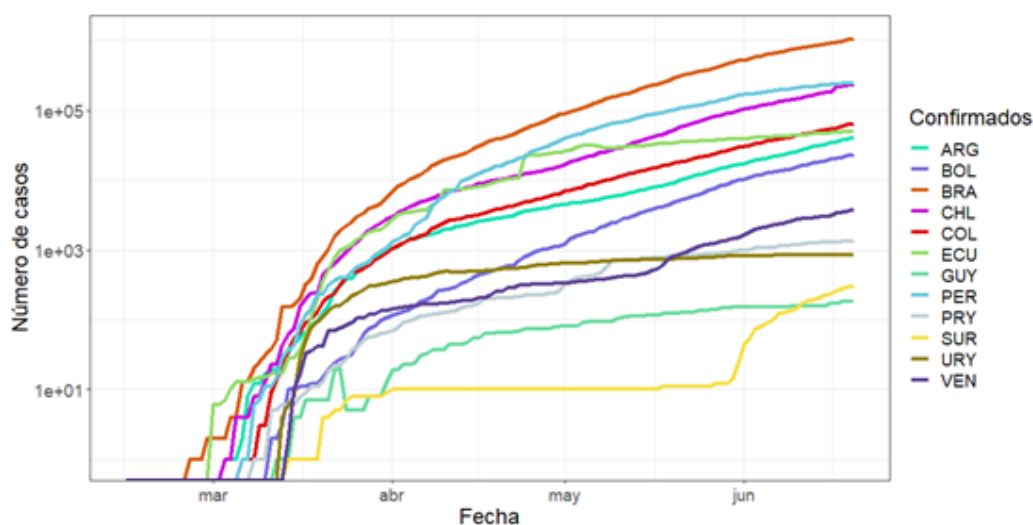
En la Figura 1 se presentan el número de acumulado de casos en función del tiempo.

Curvas de crecimiento

En la Figura 2 (2a, 2b y 2c) se puede observar que el comportamiento exponencial es independiente del número de casos que existan en el país y la fecha del inicio de la pandemia en el mismo.

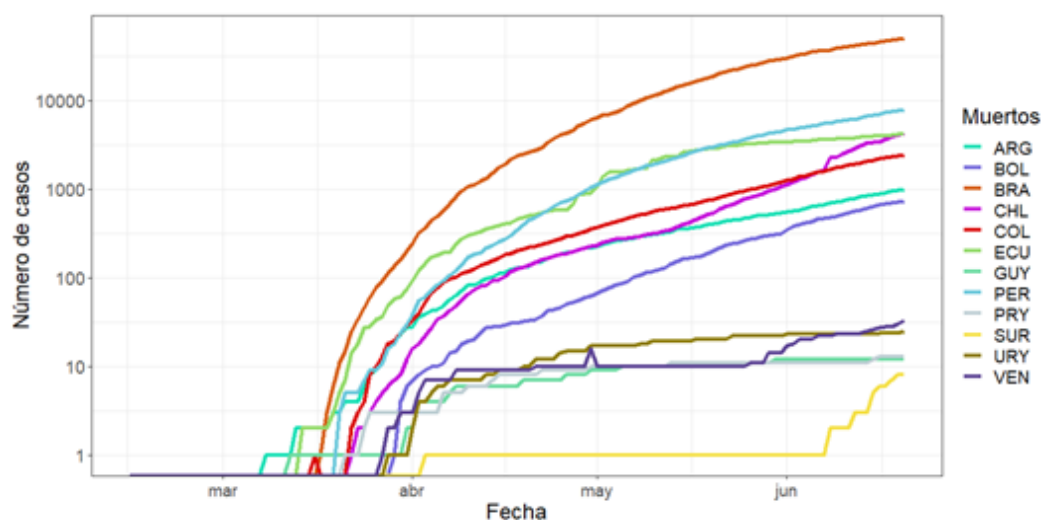
Figura 2. Número total de casos, en escala logarítmica, al 20 de junio de 2020 en diversos países de Suramérica

Figura 2a. Casos confirmados en países de Suramérica



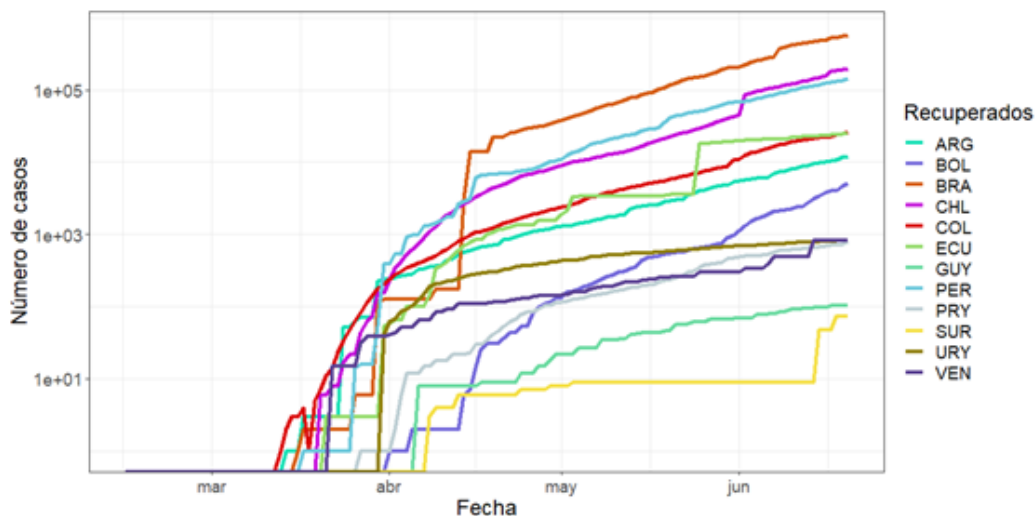
Fuente: Elaboración propia del autor, (2020)

Figura 2b. Muertes confirmadas en países de Suramérica



Fuente: Elaboración propia del autor, (2020)

Figura 2c. Casos recuperados en países de Suramérica



Fuente: Elaboración propia del autor, (2020)

Otra forma de representar los datos es a través de la inclusión del número total de casos acumulados para las categorías de confirmados, muertos y recuperados añadiendo adicionalmente el número de pruebas realizadas por país tal como aparece en la librería COVID-19 (Figura 3).

Figura 3. Número total acumulado en escala logarítmica de individuos confirmados, recuperados, muertos por la Covid-19, al igual que el número total de pruebas realizadas para el 20 de junio de 2020



Fuente: Elaboración propia del autor, (2020)

Podemos observar que el país con mayor número de casos confirmados en Suramérica es Brasil, seguido por Perú y Chile. Al poseer el mayor número de casos confirmados y por la dinámica propia de la enfermedad, estos países poseerán el mayor número de recuperados, tal como se evidencia en la Figura 3.

Los países que han tomado mayor número de pruebas son Chile, Brasil y Colombia, de mayor a menor, respectivamente. Para Guyana, Surinam y Venezuela la librería consultada no muestra el número de pruebas realizadas.

En orden de magnitud, Venezuela, Paraguay, Uruguay y Guyana son los países con menor número de casos.

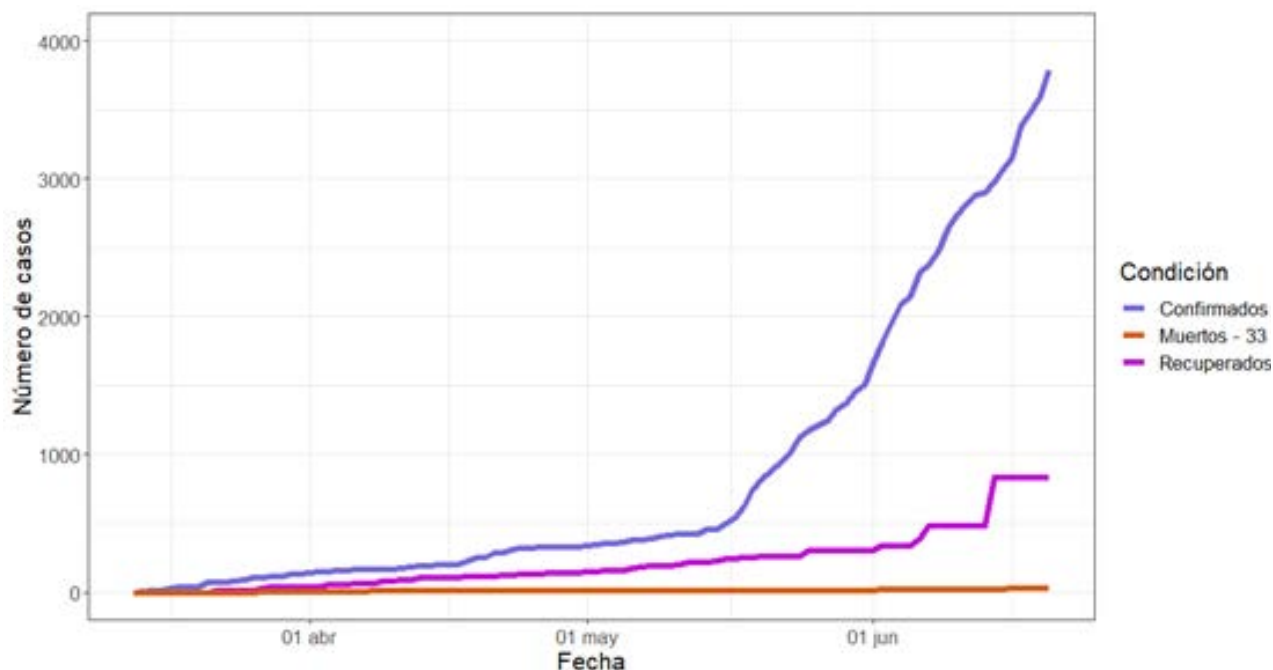
A pesar de Uruguay poseer un número de casos suficientemente bajo ha realizado un número de pruebas significativas, lo que puede ser un indicio claro de la efectiva gestión de las autoridades en cuanto a la evolución de la enfermedad. Hasta la fecha se evidencia una convergencia en el número de casos confirmados diarios.

Es de nuestro interés detallar el comportamiento de la enfermedad en Venezuela, por lo que descargamos la información del país con la librería COVID-19 y observamos que solo nos muestra información del país como región, pero no incluye un análisis más detallado sobre los casos de contagios por estado ni ciudades.

Dicha información está disponible en el sitio web del gobierno alojado en la plataforma PATRIA.

En la Figura 4 se muestra la evolución temporal de los casos confirmados, muertos y recuperados en Venezuela al 20 de junio, evidenciándose una evolución exponencial. Aunque el número de muertos solo representa el 0.87% de los contagiados, el incremento del número de casos día a día es alarmante.

Figura 4. Número de casos acumulados para Venezuela hasta el 20 de junio de 2020



Fuente: Elaboración propia del autor, (2020)

Estimación de la evolución temporal en Venezuela

Haciendo uso del modelo ARIMA se estudió la evolución temporal de la enfermedad en Venezuela analizando las series temporales desde el punto de vista estocástico. Algunos puntos importantes a definir considerando la literatura disponible es que el periodo de incubación del coronavirus es de entre 5 a 10 días (Lauer et al., 2020), es decir, desde que se contrae el virus hasta que aparecen los primeros síntomas característicos de la infección; posteriormente desde esta fase hasta la aparición de la sintomatología severa en el paciente pasa aproximadamente una semana adicional, y finalmente desde esta última fase hasta la recuperación o fallecimiento puede transcurrir otra semana más. Posteriormente, establecimos la media móvil cada cinco días, lo cual suaviza la serie temporal en algo más estable y, por lo tanto, más predecible.

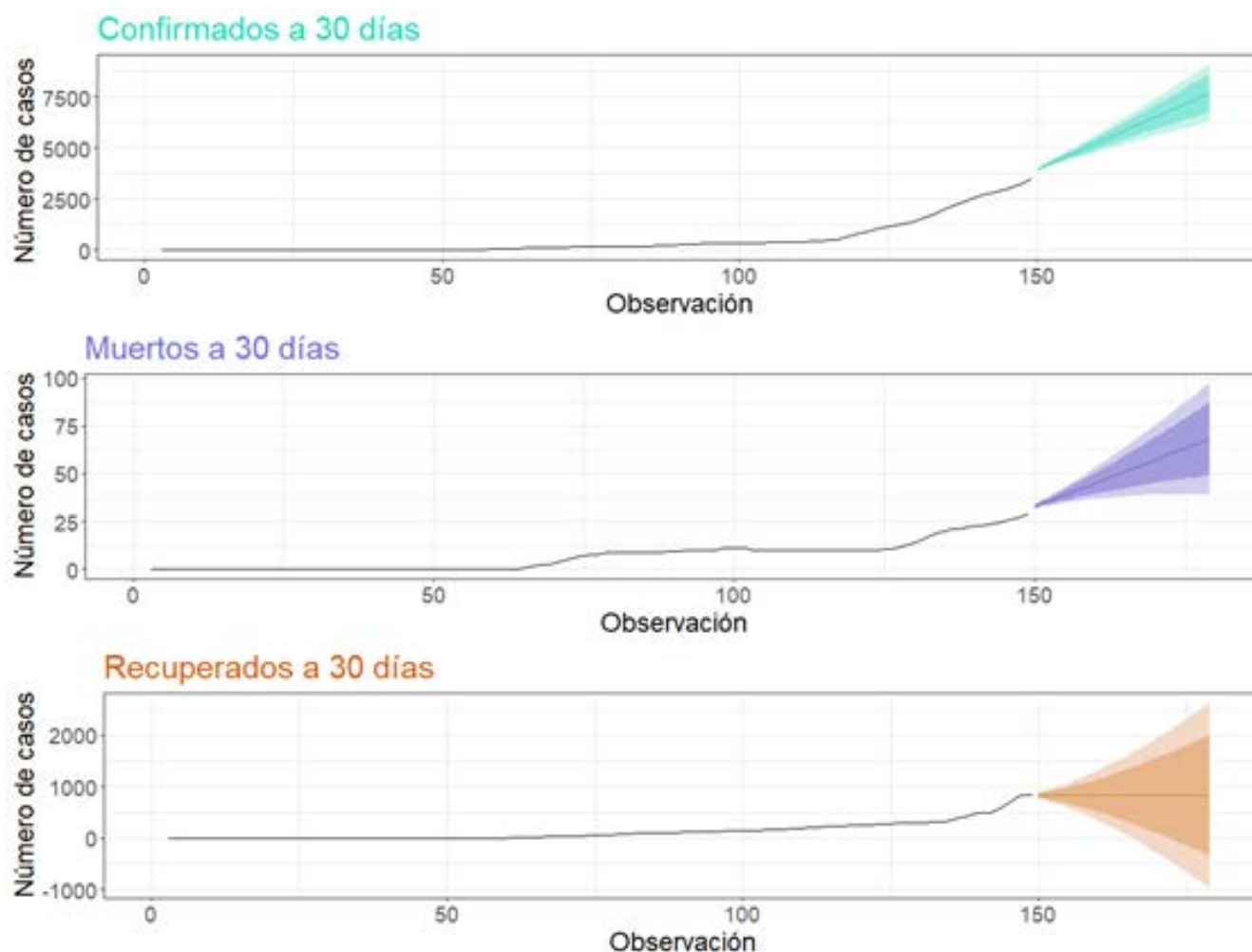
Un punto importante para notar es que a simple inspección las series temporales que se están estudiando muestran comportamiento no periódico (Figura 4). La instalación de un modelo ARIMA, sin embargo, requiere que la serie sea estacionaria (media, varianza y auto covarianza invariante en el tiempo). Para probar la estacionariedad de las series temporales se realiza la prueba Augmented Dickey- Fuller (ADF) (Said & Dickey, 1984) con ayuda de la librería *tseries* (Trapletti & Hornik, 2019) y se obtiene que para las tres series temporales (casos confirmados, muertos y recuperados) los valores obtenidos concluyen que estas no son estacionarias.

Tal y como se menciona, para realizar un modelo ARIMA, la serie temporal debe ser estacionaria, por lo que se hace necesario diferenciar las series temporales con el fin de obtener series con las características apropiadas. Los valores obtenidos de las pruebas ADF muestran que es necesario diferenciar dos veces para obtener estacionariedad con patrones erráticos alrededor de 0, sin una tendencia fuerte visible. Esto sugiere que la diferenciación de los términos del orden 2 es suficiente y debe incluirse en el modelo.

El modelo ARIMA necesita identificar los coeficientes y número de regresiones que se utilizarán, tal como hemos probado previamente. La función **auto.arima()** es una variación del algoritmo (Hyndman & Khandakar, 2008) de la librería *forecast* para R, que realiza una búsqueda sobre el modelo posible dentro de las restricciones de orden proporcionadas.

Cuando se toma la media móvil cada cinco días, las series se suavizan y vuelven más estables y, por lo tanto, más predecibles. En la Figura 5 se muestra la media móvil de cada 5 días junto con la predicción de los casos confirmados, muertos y recuperados a 30 días. Se obtiene que a menos que se tomen decisiones que controlen la difusión de la enfermedad esta seguirá con su incremento exponencial.

Figura 5. Estimaciones a 30 días para Venezuela empleando el modelo ARIMA con media móvil de cada 5 días para casos confirmados, muertos y recuperados con sus intervalos de confianza al 95%



Fuente: Elaboración propia del autor, (2020)

Este tipo de método, por su complejidad, requiere el uso de series temporales largas, aunque se observa que se ajusta correctamente al comportamiento observado. Box y Jenkins recomiendan como mínimo 50 observaciones de la serie temporal diarios y para la fecha, la extensión de los datos en las series de Venezuela es de 141 datos.

En la Tabla 1 se pueden observar los modelos predictivos seleccionados para cada serie temporal de casos confirmados, muertos y recuperados en Venezuela. Se observa que estos se ajustan a modelos de suavizado exponencial lineal o procesos integrados de segundo orden y de media móvil de primer y segundo orden. Estos estiman tanto el

nivel local como la tendencia local en la serie. Las predicciones de largo alcance de este modelo convergen a una línea recta cuya pendiente depende de la tendencia media observada hacia el final de la serie.

Tabla 1. Modelo ARIMA para la estimación de casos en Venezuela

Serie	Método	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Confirmados	ARIMA (0,2,1)	2.818	24.990	11.878	2.522	5.573	0.470	-0.036
Muertos	ARIMA (0,2,2)	0.059	0.860	0.376	2.197	7.529	1.254	-0.003
Recuperados	ARIMA (0,2,2)	4.844	48.078	10.938	2.639	6.198	1.269	0.006

Fuente: Elaboración propia del autor, (2020)

ME: Error medio**RMSE:** Error cuadrático medio**MAE:** Error absoluto medio**MPE:** Error porcentual medio**MAPE:** Error porcentual absoluto medio**MASE:** Error medio absoluto escalado**ACF1:** Coeficiente de autocorrelación de primer orden

Se observa que el comportamiento de las series temporales en Venezuela es similar a las de la región.

Estimación de la evolución temporal en países de Suramérica

Ya habiendo establecido la estimación de casos para Venezuela empleando el modelo ARIMA, ahora trataremos de hacer una estimación para los casos confirmados en todos los países de Suramérica que estamos analizando a 30 días y se utilizará una media móvil de 5 días para las series temporales.

En la Tabla 2 se pueden observar los modelos seleccionados para las series temporales, empleando una media móvil de 5 días, de casos confirmados en los Países Suramericanos. Se obtiene que para:

- Bolivia y Colombia (modelo IMA), la predicción se ajusta a un modelo de suavizado exponencial lineal o proceso integrado de segundo orden y media móvil de primer y segundo orden, respectivamente.

- Argentina, Surinam, Uruguay y Venezuela (modelo ARI), la predicción se ajusta a procesos autorregresivos de diferentes órdenes, pero integrados de segundo orden.

- Brasil y Chile (modelo I), la predicción se ajusta a un modelo integrado de segundo orden, lo que implica que el valor medio no se mantiene a lo largo del tiempo, la varianza depende del tiempo y tiende a ser infinito cuando este tiende a infinito. En general, se puede decir que estas series poseen una velocidad variable en el tiempo, pero con aceleración constante.

- Ecuador, Guyana, Perú y Paraguay siguen procesos más complejos y del tipo ARIMA.

Tabla 2. ARIMA para la estimación de casos en Suramérica

País	Método	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Argentina	ARIMA (1,2,0)	4.372	24.303	13.952	1.426	3.133	0.054	-0.020
Bolivia	ARIMA (0,2,1)	3.967	28.110	12.564	1.716	2.694	0.074	-0.014
Brasil	ARIMA (0,2,0)	149.659	1458.012	592.086	2.540	3.129	0.080	0.033
Chile	ARIMA (0,2,0)	25.660	720.770	150.471	2.628	2.774	0.091	-0.028
Colombia	ARIMA (0,2,2)	3.908	54.788	26.082	1.498	3.112	0.056	-0.012
Ecuador	ARIMA (2,2,3)	17.353	267.608	106.989	2.767	3.521	0.296	-0.047
Guyana	ARIMA (3,2,3)	0.085	0.771	0.454	2.278	4.055	0.325	0.024
Perú	ARIMA (3,2,4)	9.629	215.460	110.196	2.308	2.873	0.064	0.029
Paraguay	ARIMA (2,2,2)	0.088	3.664	1.919	1.412	3.008	0.206	0.085
Surinam	ARIMA (2,2,0)	-0.001	0.873	0.331	1.246	2.817	0.158	0.021
Uruguay	ARIMA (1,2,0)	0.0277	1.679	0.951	1.784	2.127	0.162	-0.009
Venezuela	ARIMA (5,2,0)	0.680	5.893	2.972	2.074	2.826	0.110	-0.016

Fuente: Elaboración propia del autor, (2020)

ME: Error medio**RMSE:** Error cuadrático medio**MAE:** Error absoluto medio**MPE:** Error porcentual medio**MAPE:** Error porcentual absoluto medio**MASE:** Error medio absoluto escalado**ACF1:** Coeficiente de autocorrelación de primer orden

Adicionalmente, en la Tabla 2 se puede observar que las predicciones que poseen un mayor error corresponden a los países Guyana, Ecuador y Brasil, en orden decreciente. La mejor predicción es la realizada para Uruguay.

Para Guyana, la tasa de infectados representa un 0.02% de la población, para Ecuador representa el 0.3%, un orden de magnitud mayor que para Guyana. Mientras que para Brasil representa el

0.5% de la población. Como se puede ver en la Tabla 3, la enfermedad apareció en Brasil antes que en Ecuador y Guyana.

En Uruguay la enfermedad apareció el 13 de marzo del presente año (Tabla 3), y la tasa de contagios representa el 0.02% de la población. Se evidenció un incremento muy rápido inicialmente, con una posterior disminución de la pendiente en la tasa de contagios que

aún se mantiene (Figura 2) y que se refleja en una mejor predicción a futuro del número de casos.

Chile es el país con la mayor tasa de infectados, alcanzando el 1.26%, seguido por Perú con el 0.79% y Ecuador con el 0.29%. Venezuela al tener sólo el 0.01% de tasa de infectados refleja una tasa aún manejable comparativamente, pero al mostrar un evidente crecimiento exponencial si hace necesario la toma

de medidas de control más estrictas, tomando en cuenta la estimación a 30 días alcanzando un aproximado de 7.500 casos para el 20 de Julio del presente año.

La predicción para todos los países posee una alta exactitud en la predicción, tal como se observa en la Tabla 2, los valores del error absoluto medio (MAE: mean average error) son menores al 10%.

Para encontrar cual fue el mejor modelo de pronóstico seleccionamos

a aquellas medidas de errores absolutos en lugar de los cuadráticos ya que éstos penalizan en mayor medida los errores grandes. Se seleccionan las medidas de Error porcentual absoluto medio (MAPE: mean absolute percentage error) y Error porcentual medio (MPE: mean percentage error) ya que poseen la ventaja de que son libre de unidades.

En cuanto al MPE (error porcentual medio), la media del error porcentual es una métrica simple, que sirve para ver si el error de la predicción tiene un sesgo positivo o negativo. También nos

dice si el pronóstico está subestimado o sobrestimado. Un pronóstico sobre estimado puede generar un uso exagerado en las políticas para el control de la difusión de la enfermedad, afectando la calidad de vida, creando pérdidas económicas, medidas de aislamiento que pueden ser perjudiciales. Un pronóstico subestimado en cambio puede producir una reducción en las medidas de control, que puede llegar a originar rebrotes de la enfermedad.

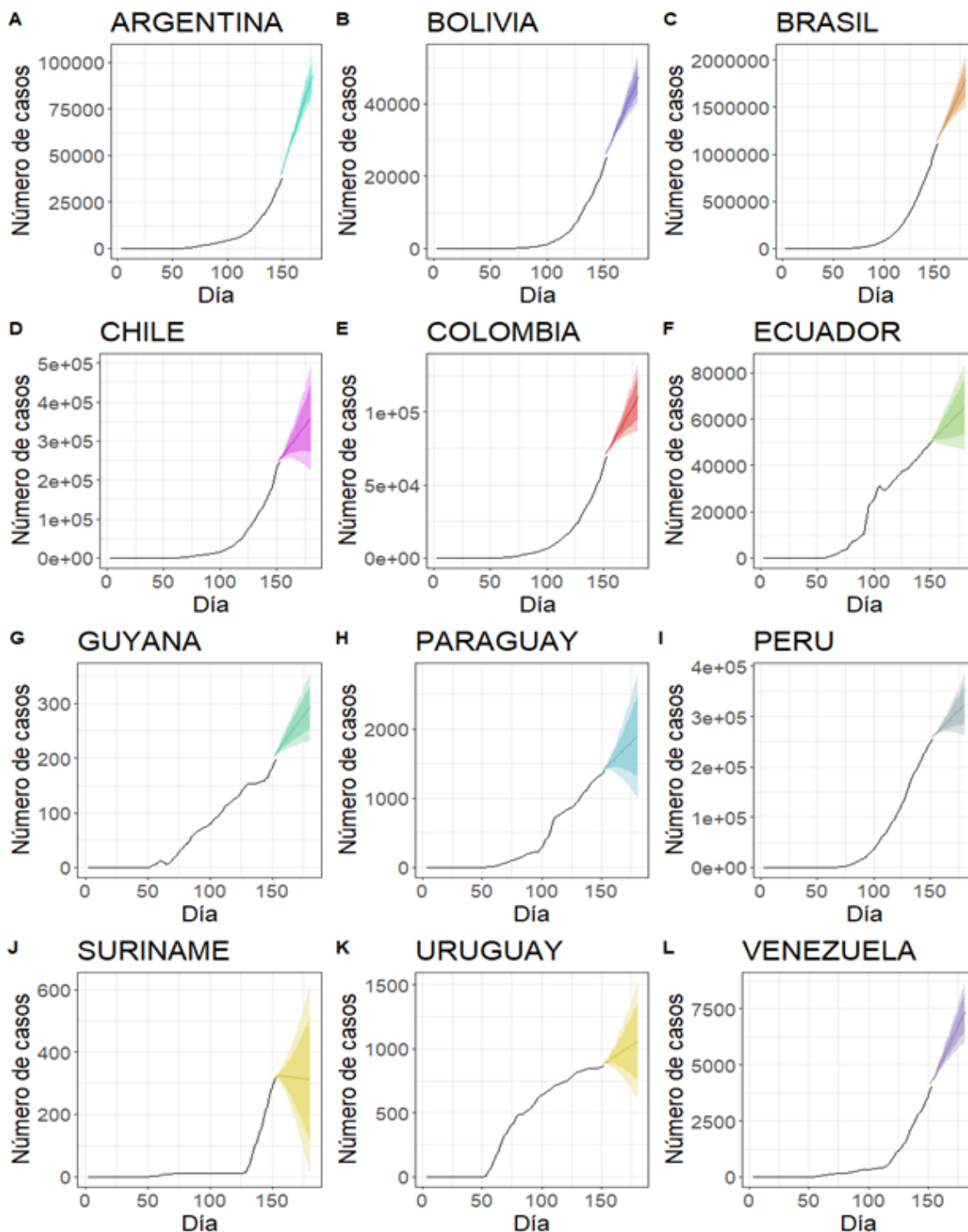
Tabla 3. Fecha de aparición de casos índice en Suramérica

País	Fecha Inicial	Población	Confirmados	Muertos	Recuperados
Argentina	2020-03-03	44.494.502	41.204	992	12.206
Bolivia	2020-03-11	11.353.142	23.512	740	5.086
Brasil	2020-02-26	209.469.333	1.067.579	49.976	576.779
Chile	2020-03-03	18.729.160	23.6748	4295	196.609
Colombia	2020-03-06	49.648.685	64.634	2397	26.156
Ecuador	2020-03-01	17.084.357	49.731	4156	24.446
Guyana	2020-03-12	779.004	183	12	102
Perú	2020-03-06	31.989.256	251.338	7861	143.017
Paraguay	2020-03-08	6.956.071	1.362	13	791
Suriname	2020-03-14	575.991	303	8	74
Uruguay	2020-03-13	3.449.299	859	25	815
Venezuela	2020-03-14	28.870.195	3.789	33	835

Fuente: Elaboración propia del autor, (2020)

En la Figura 6, se muestra la evolución temporal junto con la predicción a 30 días de los casos confirmados, empleando una media móvil de 5 días, en los países suramericanos.

Figura 6. Estimaciones a 30 días para Suramérica empleando el modelo ARIMA para el número total de casos confirmados empleando una media móvil de 5 días.



Fuente: Elaboración propia del autor, (2020)

Finalmente, es importante destacar que el Api, del cual se extraen los datos específicos de Venezuela, no cuenta con datos de la evolución temporal de los casos por estado, sino que solo entrega el resumen de los casos diarios, razón por la cual es difícil identificar la evolución de la enfermedad en cada estado.

La utilización de las medias móviles es útil, si se quiere calcular la tendencia de la serie temporal, ya que ofrece una visión suavizada de la serie puesto que, al promediar los valores, se elimina la irregularidad de la serie. Sin embargo, a mayor sea la longitud de las medias móviles mejor se eliminarán las irregularidades, pero el coste informativo será mayor. Por lo que a futuro quedaría por indagar sobre la dependencia de la predicción de estas series con la elección del suavizado de la serie temporal.

Conclusión

Los resultados muestran que la estimación para el total acumulado de casos en la región suramericana, tiene en su mayoría un claro crecimiento exponencial lo cual podría sugerir que es aún temprano para adoptar medidas de desconfinamiento, a la par que hace necesario reestructurar las medidas de distanciamiento físico.

La implementación de los modelos ARIMA es una aproximación matemática que permite hacer estimaciones de la dinámica de las enfermedades infecciosas en función del tiempo, debido a su fácil estructura y rápida aplicabilidad. En este caso particular, nuestro trabajo aporta por primera vez una estima-

ción a 30 días del número de infectados con Covid-19 de la región suramericana, con particular énfasis en Venezuela, contribuyendo con datos confiables a la adopción de políticas públicas que disminuyan la acelerada propagación de esta pandemia.

Adicionalmente, con los datos obtenidos a través del API disponible en patria se desarrolló una aplicación que presenta de manera concisa y precisa un conjunto de datos y gráficos acerca del reporte diario y acumulado de casos para Venezuela. La aplicación se encuentra alojada en: https://sterguti.shinyapps.io/AppCovidVzla_102/

Referencias

- Al-Najjar, H., & Al-Rousan, N. (2020). A classifier prediction model to predict the status of Coronavirus COVID-19 patients in South Korea. *European Review for Medical and Pharmacological Sciences*, 24(6), 3400–3403. Recuperado en: https://doi.org/10.26355/eur-rev_202003_20709
- Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE*, 15(3), 1–21. Recuperado en: <https://doi.org/10.1371/journal.pone.0230405>
- Asteriou, D., & Stephen G., H. (2011). *Applied Econometrics*. Palgrave Macmillan (Second Edi). Hampshire: Palgrave Macmillan.
- Bekiros, S., Kouloumpou, D., & Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *Chaos, Solitons and Fractals*, 729, 138817. Recuperado en: <https://doi.org/10.1016/j.chaos.2020.109828>
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2013). *Time series analysis: Forecasting and control: Fourth edition. Wiley Series in Probability and Statistics* (Vol. 37). Wiley. Recuperado en: <https://doi.org/10.1002/9781118619193>
- CHOI, K., & THACKER, S. B. (1981). An Evaluation of Influenza Mortality Surveillance, 1962–1979. *American Journal of Epidemiology*, 113(3), 215–226. Recuperado en: <https://doi.org/10.1093/oxford-journals.aje.a113090>
- Du, R. H., Liang, L. R., Yang, C. Q., Wang, W., Cao, T. Z., Li, M., Shi, H. Z. (2020). Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: A prospective cohort study. *European Respiratory Journal*, 55(5), 2000524. Recuperado en: <https://doi.org/10.1183/13993003.00524-2020>
- Ghosal, S., Sengupta, S., Majumder, M., & Sinha, B. (2020). Prediction of the number of deaths in India due to SARS-CoV-2 at 5–6 weeks. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 14(4), 311–315. Recuperado en: <https://doi.org/10.1016/j.dsx.2020.03.017>

- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 26(June). Recuperado en: <https://doi.org/10.1038/s41591-020-0883-7>
- Guidotti, E., & Ardia, D. (2020). *COVID-19 Data Hub*. Recuperado en: <https://doi.org/10.13140/RG.2.2.11649.81763>
- Hyndman, R., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software, Articles*, 27(3), 1–22. Recuperado en: <https://doi.org/10.18637/jss.v027.i03>
- Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, 196, 105581. Recuperado en: <https://doi.org/10.1016/j.cmpb.2020.105581>
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9), 577–582. Recuperado en: <https://doi.org/10.7326/M20-0504>
- R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria. Recuperado en: <https://www.r-project.org/>
- RStudio Team. (2020). RStudio: Integrated Development Environment for R. Boston, MA. Recuperado en: <http://www.rstudio.com/>
- Said, S. E., & Dickey, D. A. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, 71(3), 599–607. Recuperado en: <https://doi.org/10.2307/2336570>
- Trapletti, A., & Hornik, K. (2019). tseries: Time Series Analysis and Computational Finance. Recuperado en: <https://cran.r-project.org/package=tseries>
- Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*. Recuperado en: [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)