

Modelación matemática y estadística de las notificaciones de infectados por viruela símica en Colombia

Francisco Antonio García Jaramillo

PERTINENTE CREATIVA INTEGRADORA

 @uniquindio  unquindioconectada  unquindioconectada

www.uniquindio.edu.co



UNIVERSIDAD
DEL QUINDÍO



Universidad del Quindío

Facultad de Ciencias de la Educación

Licenciatura en Matemáticas

Modelación matemática y estadística de las notificaciones de infectados por viruela símica en Colombia

Informe escrito de trabajo de grado

Francisco Antonio García Jaramillo

Director: PhD Gladys Elena Salcedo Echeverry



UNIVERSIDAD
DEL QUINDÍO



Modalidad: Trabajo de investigación

Línea de investigación: Modelos para la información autocorrelacionada en el tiempo y en el espacio

Área de profundización: Estadística

Junio 2023



Índice

1. Introducción.....	8
2. Estado del arte.....	11
3. Marco Conceptual	17
3.1 El Modelo de crecimiento logístico	17
3.2 El Modelo SIR.....	17
3.3 Modelos de Series de tiempo.....	19
3.3.1 Estacionariedad	19
3.3.2 Diferenciación de una serie temporal	20
3.3.3 Transformaciones de Box-Cox.....	20
3.3.4 Autocovarianza y Autocorrelación	23
3.3.5 Función de Autocorrelación Parcial (FAP)	24
3.3.6 Modelos Autorregresivos (AR).....	26
3.3.7 Modelos de Medias Móviles (MA)	27
3.3.8 Modelo ARMA	28
3.3.9 Modelo ARIMA	29
3.3.10 Procesos estacionales SARIMA.....	30
3.4 Modelo de redes neuronales autorregresivas (ARNN)	31
3.5 Enfoque Híbrido de Modelación.....	33
3.6 Selección y comparación de modelos	35
4. Aplicación y ajuste de los modelos.....	36
4.1 Los Datos.....	36
4.2 Análisis exploratorio de la serie temporal de las notificaciones diarias	36
4.3 Modelación de la serie temporal de las notificaciones acumuladas	39
4.3.1 Ajuste del modelo de crecimiento logístico	41
4.3.2 Ajuste del modelo SIR.....	42
4.3.3 Ajuste del modelo ARNN	44
4.3.4 Ajuste del modelo de series temporales.....	44
4.3.5 Ajuste de modelos híbridos.....	47
4.4 Comparación de modelos.....	51
5. Discusión, conclusiones y trabajos futuros	54
Bibliografía.....	56

Agradecimientos

Este trabajo de investigación es dedicado a todas las personas que han influido en mi proceso académico de una u otra manera, a mis seres queridos y en especial a mi madre por su apoyo incondicional, siendo ellos pilar fundamental en mi crecimiento personal y académico. Un agradecimiento especial a mi directora Gladys Elena Salcedo Echeverry por su apoyo y sus enseñanzas con las cuales desarrollé un gusto especial por la estadística y la modelación. Además, reconocer el invaluable apoyo de la profesora Valentina Zuluaga Zuluaga en mi proceso de realización del trabajo de grado.



Resumen

Debido al rebrote de viruela símica en el año 2022, cuando aún no se superaban totalmente las consecuencias dejadas por la pandemia del Covid-19, surge un gran interés por investigar y monitorear el comportamiento de esta enfermedad resurgente. En este trabajo de investigación, se modelan los datos acumulados del número de infectados diarios por viruela símica en Colombia mediante modelos matemáticos, estadísticos y computacionales. Se ajustan a los datos cuatro tipos de modelos: un modelo SARIMA de la teoría de series de tiempo, un modelo autorregresivo neuronal de la teoría de aprendizaje de máquinas y dos modelos matemáticos, el modelo de crecimiento logístico y el modelo SIR; también se explorarán algunas mezclas de modelos desde un enfoque híbrido con el fin de evaluar si mejoran las predicciones y si disminuyen las métricas de precisión. Basado en dichas métricas se selecciona el modelo que mejor se ajusta a las notificaciones de casos. Los modelos obtenidos permiten entender el comportamiento de la viruela símica y provee información relevante para la toma de decisiones con respecto a la enfermedad.

Palabras clave: Datos de viruela del mono, crecimiento logístico, series de tiempo, redes neuronales, modelo SIR.

Abstract

Due to the resurgence of monkeypox in 2022, when the consequences left by the Covid-19 pandemic had not been fully overcome, it is of great interest to research and monitor the behavior of this resurgent disease. In this research work, the cumulative notifications of the daily infected cases by monkeypox in Colombia are modeled through some mathematical, statistical and computational models. Four types of models are fitted to the data: a SARIMA



UNIVERSIDAD
DEL QUINDÍO



model from time series theory, an autoregressive neural network model from machine learning theory, and two mathematical models, the logistic growth model and the SIR model; Some mixtures of models from a hybrid approach will be explored in order to evaluate if they improve predictions and decrease accuracy metrics. Based on these metrics, the model that best fits the cumulative notifications is selected. The models allow us to understand the behavior of monkeypox data and provide relevant information for decision-making regarding the disease.

Keywords: Monkeypox data, logistic growth, time series, neural networks, SIR model.



1. Introducción

Desde el año 2022 el rebrote de la viruela símica generó preocupación global debido a las altas tasas de contagio alrededor del mundo. En Colombia, el INS (Instituto Nacional de salud, 2022) registra al día 5 de noviembre de 2022 más de 3500 casos, considerándose una tasa muy alta teniendo en cuenta que los primeros tres casos confirmados fueron el 23 de junio del 2022, además, a escala global, según datos del último reporte de la OMS ya existen más de 79.000 casos confirmados a nivel mundial. Por esta razón, se ha generado un interés particular en estudiar esta problemática que ya empieza a afectar en gran medida a países de América y Europa.

Según la OMS (Organización mundial de la salud, 2022), la viruela símica es un virus miembro del Orthopoxvirus que se esparce debido a la exposición con un animal infectado o por un contacto estrecho entre humanos, generando así preocupación a nivel mundial ya que el contagio puede llegar a ser similar al Covid-19 del que se guarda un cierto temor debido a los problemas que esto conllevó.

Teniendo en cuenta lo anterior, el poco interés de generar una preocupación alta en la población ha hecho que no se exponga de manera eficiente esta nueva problemática resurgente, lo que es peligroso, ya que informar y estudiar sobre una nueva posible pandemia puede ayudar a controlar y prevenir posibles consecuencias a futuro. Contribuyendo a esto, la OPS (Organización Panamericana de la salud, 2021) expone que una correcta difusión de la información es “clave para disponer de evidencia para la acción, tomar decisiones lo más



informadas posibles y adecuar políticas que permitan una mejor inteligencia en acciones de salud” (p. 2).

Esta investigación surge de la necesidad de estudiar el comportamiento de la viruela símica en Colombia, ya que el 14 de noviembre de 2022 fue categorizado por la OMS como el séptimo país con más infectados por dicha enfermedad con un total de 3630 casos; se plantea este estudio con la intención de desarrollar unos pilares que faciliten comprender la enfermedad y su expansión, además, ajustar modelos que nos permitan pronosticar las notificaciones futuras, facilitando así la toma de decisiones que regulen la expansión del virus.

La importancia de modelar la dinámica de las enfermedades infecciosas es expuesta en diferentes trabajos de investigación, como en el caso de (Montesinos & Hernández, 2007) los cuales afirman que modelar enfermedades estadística y matemáticamente “contribuye a prevenir futuras situaciones patológicas, además, determina la prevalencia e incidencia para tomar decisiones objetivas en el control de enfermedades infecciosas” (p. 8). A partir de esto, se plantea esta investigación con el fin de modelar el número de infectados diarios por la viruela símica en Colombia mediante algunos modelos matemáticos y estadísticos a partir de información extraída del Instituto Nacional de Salud de Colombia con el fin de describir el comportamiento de los datos y proveer información relevante acerca de la enfermedad.

En este trabajo se presentan algunos modelos útiles en la actualidad para pronosticar series de tiempo: Un modelo SARIMA de la teoría de series de tiempo, un modelo autorregresivo



neuronal de la teoría de aprendizaje de máquinas y dos modelos matemáticos, el modelo de crecimiento logístico y el modelo compartimental SIR; además se exploran algunas mezclas de modelos desde un enfoque híbrido con el fin de evaluar si mejoran las predicciones y si disminuyen ciertas métricas de precisión utilizadas; más específicamente, la calidad del ajuste de cada modelo se evalúa a través de las métricas MAE (error medio absoluto), MAPE (error porcentual absoluto medio) y el RMSE (error cuadrático medio). Por tratarse de una serie temporal, previamente se evalúan la presencia de componentes como la estacionariedad, los componentes estacionales y la autocorrelación de las observaciones. El mejor modelo se selecciona con base en las métricas de precisión las cuales se calculan tanto para el conjunto de entrenamiento (131 primeros registros) como para el conjunto de prueba (7 registros adelante) de las notificaciones de casos.

El resto del documento se organiza de la siguiente manera, en el Capítulo dos se presentan varios trabajos relacionados, en el Capítulo tres aparece el marco conceptual en el cual se detalla la teoría de cada uno de los modelos ajustados, en el Capítulo cuatro se exploran gráficamente y analíticamente los datos y se ajustan los modelos discutidos, en el Capítulo cinco se presentan las discusiones, conclusiones y posibles trabajos futuros.



2. Estado del arte

Como antecedentes que den validez a la investigación se encontraron los siguientes trabajos:

- Olumuyiwa et al.(2022) en su artículo titulado “Fractional order mathematical model of monkeypox transmission dynamics” construyeron un modelo matemático determinístico usando el modelo clásico y fraccional de ecuaciones diferenciales ajustado por mínimos cuadrados, donde se esperaba identificar las interacciones que hacen que el virus se descontrole. Además, de generar pautas a la población para darle un buen manejo a una enfermedad de la cual se tiene desconocimiento y que puede ser un riesgo a la salud mundial; esta investigación propone una forma de controlar la propagación de la viruela símica. Se concluyó con un modelo fraccionario que se ajustaba bien a los datos y que daba la oportunidad de predecir el comportamiento de la enfermedad y su progresión en Nigeria. Se concluye, presentando diferentes simulaciones según algunos parámetros para investigar las dinámicas por las cuales se esparce el virus en la población Nigeriana.
- Fuller et al.(2011) en una investigación llamada “Using Remote Sensing to Map the Risk of Human Monkeypox Virus in the Congo Basin” hacen uso de los ENMs, el algoritmo Maxent y de la regresión logística para modelar datos entre 2005 y 2007 en una población de la República Democrática del Congo, el distrito de Sankuru y la provincia de Kasai-Oriental. Este estudio tuvo como



objetivo identificar las poblaciones más expuestas al virus, teniendo en cuenta unas variables específicas como la presencia de ardillas arborícolas y las ratas gambianas que son los animales que esparcen el virus Orthopoxvirus, además de tener en cuenta factores del entorno, como por ejemplo los bosques o la selva tropical, donde es más posible contraer este tipo de viruela si se está próximo a bosques densos debido a que son el hábitat predilecto de las ardillas. Además, se comprobó que la ardilla arborícola es una de las causas más comunes para contraer la viruela símica; el estudio sugiere un monitoreo del hábitat para el control de una nueva posible pandemia.

- Olumuyiwa et al.(2021) publicaron el artículo “Transmission dynamics of Monkeypox virus: a mathematical modelling approach” donde se propone un modelo matemático determinístico, más específicamente un modelo compartimental no lineal cuyo objetivo es entender cómo se transmite y se esparce la viruela símica. En el estudio se tuvieron en cuenta las personas expuestas (E_h), aisladas (Q_h), infectados (I_h) y recuperados (R_h); todas estas variables fueron utilizadas en simulaciones numéricas a través de ecuaciones diferenciales determinando un modelo que establece las condiciones detonantes de la enfermedad y concluyendo que el aislamiento de las personas infectadas reduce la probabilidad de esparcimiento del Orthopoxvirus.
- Eid et al.(2022) en un artículo titulado “Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox



Cases” utilizan una serie de datos con infectados y no infectados alrededor del mundo la cual fue analizada en una red profunda conocida como *Long Short-Term Memory* (LSTM), dentro de ésta usaron algunos algoritmos de optimización como el Al-Biruni Earth Radius (BER) cuyo objetivo era encontrar un nuevo enfoque o modelo que permitiera predecir los casos de viruela símica. Posterior a esto, para probar la efectividad del enfoque, realizaron seis pruebas con modelos de Machine Learning donde comprobaron que el modelo es el que se adapta mejor a los datos.

En adición, para comprobar la significancia estadística fueron utilizados ANOVA y pruebas de Wilcoxon, esto para darle solidez a los resultados obtenidos.

- Khan et al.(2022) realizaron una investigación llamada “Stochastic modeling of the Monkeypox 2022 epidemic with cross-infection hypothesis in a highly disturbed environment” en la cual se estudió la dinámica por la cual se propaga el virus a través de un sistema compartimental en forma de Itô-Lévy, siendo este un proceso estocástico de ecuaciones diferenciales. Para estos modelos se tomaron en cuenta dos modos de transmisión, la infección entre humanos-animales y la transmisión humana horizontal; a partir de esto, se realizaron unas técnicas analíticas para probar algunas propiedades asintóticas como la erradicación y la continuación de la media de infectados por la viruela símica, esto para entender por qué en algunos países está resurgiendo el virus y analizar la dinámica de esparcimiento del Orthopoxvirus. Se concluye que



este tipo de estudios innovan la manera de estudiar pandemias y que son un pilar para comprender enfermedades resurgentes y ayudar a su posterior erradicación.

- Suthep et al.(2023) en “Scaled Conjugate Gradient for the Numerical Simulations of the Mathematical Model-Based Monkeypox Transmission” se aplicaron derivadas de orden fraccional para obtener mejores resultados construyendo un modelo no lineal que describe las dinámicas de la viruela símica; para llevar a cabo esta investigación se tuvieron en cuenta ocho variables, donde cinco de ellas estaban relacionadas con la población humana y otras tres con roedores, como herramienta de modelación se utilizó inteligencia artificial a través del uso de redes neuronales basadas en el gradiente conjugado, concluyendo con unas simulaciones obtenidas a partir de diferentes procesos que describen la transmisión del virus.
- Usman & Adamu (2017) realizaron una investigación titulada “Modeling the Transmission Dynamics of the Monkeypox Virus Infection with Treatment and Vaccination Interventions”, la cual responde a la problemática de viruela símica resurgente en Nigeria. Los investigadores emplearon un modelo al que llamaron “disease-free and endemic”, en su trabajo modelan el comportamiento del virus antes y después de aplicar mecanismos de control como la vacunación y otros tratamientos en humanos y primates. La



modelación fue realizada a partir de parámetros ya existentes en otras investigaciones. Se obtuvo que el modelo era estable si el número básico de reproducción era de $R_0 < 1$, durante el tiempo de estudio la tasa de infectados va decreciendo tanto que tiende a desaparecer en el tiempo gracias a los mecanismos de control. Sin embargo, si R_0 no cumplía la condición anterior, el modelo era inestable y poco significativo. Concluyeron que con un seguimiento y control del virus es posible que se erradique la viruela en un tiempo considerablemente más pequeño que sin estos tratamientos y vacunación.

- Arotolu et al.(2022) investigan la viruela símica en un artículo titulado “Spatial modeling and ecological suitability of monkeypox disease in Southern Nigeria” cuyo objetivo era entender el comportamiento de esta enfermedad con el uso de un modelo de máxima entropía (MAXENT). Por otra parte, este trabajo abarcó todo el sur de Nigeria para identificar las poblaciones más expuestas a contraer el virus. Este modelo inicialmente constaba de muchas variables ambientales, más exactamente 69, sin embargo, sólo se utilizaron 32 en el modelo; las variables representaban características del ambiente en el que se encontraba cada población. El algoritmo MAXENT predijo las poblaciones que necesitan alta vigilancia y control del virus, esto gracias a que se identificaron las variables ambientales que generan una mayor susceptibilidad a contraer la enfermedad, las cuales fueron: la temperatura mínima en diciembre, la temperatura máxima en marzo, la



elevación, la precipitación del trimestre más seco y la densidad de población humana.

- Nakazawa et al.(2013) en su artículo llamado “Mapping Monkeypox Transmission Risk through Time and Space in the Congo Basin” examina las correlaciones potenciales entre la viruela símica y el entorno; para esto llevó a cabo un tipo de modelación conocido como *Ecological Niche Models* (ENMs) con unas bases de datos de los años 1980 y 2000. En la investigación se construyeron dos algoritmos que corresponderían al ENMs, uno de ellos es un algoritmo genético (GARP) y el otro de máxima entropía (Maxent). Gracias a estos modelos fue posible conocer las poblaciones más vulnerables a contraer el virus, además fue visible que entre 20 a 30 años de diferencia, el ambiente se ha hecho más susceptible a que se desarrolle más la enfermedad.

3. Marco Conceptual

En esta sección se presenta la teoría que se utiliza en la realización del trabajo de investigación, se amplía la información acerca de los modelos matemáticos y estadísticos, como también los modelos computacionales y el enfoque híbrido.

3.1 El Modelo de crecimiento logístico

La función logística es una herramienta muy utilizada para modelar el crecimiento de plagas y enfermedades que se producen en una población en un tiempo determinado, y describe la manera en que una problemática crece o se expande.

Este modelo matemático es de tipo no lineal y está definido según (Bovas & Ledolter, 2006) como

$$f(t) = \frac{\alpha}{1 + \beta \exp(-\gamma t)} \quad (3.1.1)$$

donde $f(t)$ representa la tendencia media esperada del crecimiento, α es un valor que delimita el crecimiento, γ representa un coeficiente de proporcionalidad y β es un valor determinado por unas condiciones iniciales.

3.2 El Modelo SIR

El modelo SIR es un modelo matemático clásico mediante el cual se modela la dinámica del desarrollo de un virus o enfermedad durante un tiempo t , como lo explica (Zill, 2008).

Suponga que los habitantes de un lugar cualquiera son susceptibles a contraer una cierta enfermedad; en el inicio de la propagación de la enfermedad aparecerían los primeros infectados y no habrían recuperados. Mientras transcurre el tiempo, los infectados se van

incrementando como también las personas recuperadas, además, podría ser que existan muertes o no debido a la enfermedad. Basado en estos supuestos, en cualquier instante t , la población estará dividida en individuos susceptibles (S), infectados (I) y removidos (R). Si no se registran muertes, los removidos corresponden a los recuperados, y en la presencia de muertes los removidos representan los recuperados y los muertos. El total de la población, n , es la suma de los susceptibles, infectados y recuperados.

El modelo SIR consiste en un sistema de ecuaciones diferenciales de primer orden como se muestra a continuación:

$$\begin{aligned}\frac{dS}{dt} &= -k_1SI \\ \frac{dI}{dt} &= -k_2I + k_1SI \\ \frac{dR}{dt} &= k_2I\end{aligned}\tag{3.2.1}$$

donde k_1 representa la tasa de infectados y k_2 la tasa de removidos, ambas constantes siendo positivas. En este tipo de modelos dinámicos, se puede obtener un parámetro de especial relevancia denominado el R_0 , el cual representa el número de infectados por un mismo individuo en una población susceptible a ser infectada, y es dado por $R_0 = \frac{k_1}{k_2}$, con esto se puede reescribir las ecuaciones en (3.4.1) como:

$$\begin{aligned}\frac{dS}{dt} &= -k_2R_0 \frac{SI}{n} \\ \frac{dI}{dt} &= k_2R_0 \frac{SI}{n} - k_2I \\ \frac{dR}{dt} &= k_2I\end{aligned}\tag{3.2.2}$$



3.3 Modelos de Series de tiempo

Una serie de tiempo es una secuencia de observaciones ordenadas en el tiempo t ; el estudio de las series de tiempo forma una rama de la estadística la cual permite analizar y modelar problemáticas en distintas áreas del conocimiento con el fin de predecir y estudiar fenómenos que son registrados, generalmente igualmente espaciados, en el tiempo.

La formalización del concepto de serie temporal es descrita en la teoría de procesos estocásticos como dado en (Wei, 1990) donde un proceso estocástico es una familia de variables aleatorias indexadas en el tiempo, esto es $X(\omega, t)$ donde ω representa un espacio muestral y t el tiempo. Cuando tenemos un t fijo, $X(\omega, t)$ representa una variable aleatoria, sin embargo, si se tiene un ω fijo, $X(\omega, t)$ representa una función de la muestra o la realización de una serie de tiempo, concluyendo así, que una serie temporal es una función o una realización de un proceso estocástico.

Algunas características importantes de una serie de tiempo son las siguientes.

3.3.1 Estacionariedad

La Estacionariedad es una propiedad de las series de tiempo mediante la cual ciertas componentes se mantienen constantes y en equilibrio durante el tiempo. Como lo describe (Hurtado T. & Salcedo E., 1996) para que una serie sea estacionaria debe cumplir con las siguientes tres condiciones.

1. Su media permanecer constante.
2. Su varianza permanecer constante.
3. La covarianza no depender del tiempo.

Expresado matemáticamente, se dice que una serie $\{X_t, t \in T \subset \mathbb{R}\}$ es estacionaria si



$$\mu_t = E(X_t) = \mu, \quad \forall t \in T$$

$$\sigma_t = Var(X_t) = \sigma, \quad \forall t \in T$$

$$Cov(X_{t_1}, X_{t_1+k}) = Cov(X_{t_2}, X_{t_2+k}) = \gamma_k, \quad \forall t_1, t_2 \in T \wedge k \in \mathbb{Z}$$

Cabe recalcar que, en la práctica esto no es lo más común, es muy probable que en la modelación de situaciones reales las series sean no estacionarias y se deban realizar algunas transformaciones a los datos para volverla estacionaria. Las dos transformaciones más comunes son la transformación en primeras o segundas diferencias de la serie para estabilizar la media o la transformación de Box y Cox para estabilizar la varianza (Hurtado T. & Salcedo E., 1996).

3.3.2 Diferenciación de una serie temporal

Sea X_t una serie temporal no estacionaria en media y sea B el operador de rezagos tal que $B^d X_t = X_{t-d}$ retrasa la serie d etapas atrás.

$(1 - B)^d X_t = X_t - X_{t-d}$ es la transformación de las primeras d diferencias de la serie. Si $d = 0$, la serie no requiere ser diferenciada. Si $d = 1$, se tiene que, $(1 - B)X_t = X_t - X_{t-1}$ es la transformación de las primeras diferencias en tiempo discreto de la serie. Generalmente, $d = 1$ y muy rara vez $d = 2$.

3.3.3 Transformaciones de Box-Cox

Es probable encontrarse con situaciones en las cuales los datos tengan un comportamiento demasiado *variable* y sea difícil generar modelos viables sin una transformación de la escala de los datos. Una forma de corregir algunos comportamientos “caóticos” de los datos como asimetrías, varianzas desiguales y presencia de no linealidades, bien sea de una función de regresión o de las series temporales, es utilizando o llevando a cabo una transformación de

Box-Cox la cual consiste en aplicar una función potencia a los datos que permita de cierta forma homogenizar su comportamiento.

La familia de las transformaciones de potencia como lo expresa (Kutner, Nachtstein, Neter, & Li, 2004) viene dada por

$$Y' = Y^\lambda \quad (3.3.3.1)$$

donde lambda (λ) representa el parámetro y debe ser estimado a partir de los datos.

Para ilustrar la transformación considere el modelo de regresión simple clásico de error normal, y suponga que la variable respuesta viene dada con una transformación de potencia, esto es

$$Y_i^\lambda = \beta_0 + \beta_1 X_i + \varepsilon_i. \quad (3.3.3.2)$$

En el modelo anterior, λ es un nuevo parámetro a ser estimado o identificado. Un método mediante el cual se puede estimar este parámetro es el método de máxima verosimilitud, siendo posible de manera alterna hallar los parámetros restantes del modelo como lo son β_0 , β_1 y la varianza de los errores, σ^2 .

A continuación se describe la transformación de Box-Cox en el contexto de una serie temporal no estacionaria como aparece en (Hurtado T. & Salcedo E., 1996).

Sea X_t una serie temporal no estacionaria, inicialmente se considera que la varianza de la serie es una función de su media, es decir, $\sigma_t^2 = f(\mu_t)$. Considere luego que existe una función (o transformación) T de la serie y cuya primera derivada existe; entonces, mediante una aproximación de Taylor de primer orden se tiene que

$$T(x_t) \approx T(\mu_t) + \left(\frac{dT}{dx_t} \Big|_{X_t = \mu_t} \right) (X_t - \mu_t).$$

Aplicando el operador varianza a ambos lados de la ecuación anterior se obtiene

$$\text{Var}(T(x_t)) \approx \left(\frac{dT}{dx_t} \Big|_{X_t = \mu_t} \right)^2 f(\mu_t).$$

Si se desea que esta varianza sea constante, el primer término se denota por C^2 , esto es

$$C^2 = \left(\frac{dT}{dx_t} \Big|_{X_t = \mu_t} \right)^2 f(\mu_t)$$

$$\left(\frac{dT}{dx_t} \Big|_{X_t = \mu_t} \right)^2 = \frac{C^2}{f(\mu_t)}$$

$$\left(\frac{dT}{dx_t} \Big|_{X_t = \mu_t} \right) = \frac{C}{\sqrt{f(\mu_t)}}$$

Ahora, integrando adecuadamente a ambos lados de la ecuación anterior se obtiene

$$T(\mu_t) = \int \frac{C}{\sqrt{f(\mu_t)}} d\mu_t$$

Asumiendo en particular que $f(\mu_t) = \mu_t^{2(1-\alpha)}$, se tiene

$$T(\mu_t) = \int \frac{C}{\sqrt{\mu_t^{2(1-\alpha)}}} d\mu_t$$

$$T(\mu_t) = \int \frac{C}{\mu_t^{(1-\alpha)}} d\mu_t$$

de donde

$$T(\mu_t) = \begin{cases} \frac{C\mu_t^\alpha}{\alpha} & \alpha \neq 0 \\ C \ln(\mu_t) & \alpha = 0 \end{cases}$$

Este resultado indica que una manera de estabilizar la varianza es mediante la siguiente transformación:

$$T(X_t) = \begin{cases} \frac{X_t^\alpha}{\alpha} & \alpha \neq 0 \\ \ln(X_t) & \alpha = 0 \end{cases} \quad (3.3.3.3)$$

conocida como la transformación de Box-Cox.

3.3.4 Autocovarianza y Autocorrelación

En el análisis de las series de tiempo estacionarias es muy importante profundizar en conceptos como la función de autocovarianza (FAC) y la función de autocorrelación (FAS). Para ello, se define inicialmente la covarianza y la correlación entre dos variables aleatorias X y Y , como lo expresa (Bisgaard & Kulahci, 2011)

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{\sigma_x^2} \sqrt{\sigma_y^2}} \quad (3.3.4.3)$$

Así que la autocovarianza entre una serie de tiempo estacionaria y la misma rezagada k lugares se define por $Cov(X_{t+k}, X_t) = E[(X_{t+k} - \mu)(X_t - \mu)]$. Esto es,

$$\gamma_k = E[(X_{t+k} - \mu)(X_t - \mu)].$$

Si se tiene en cuenta que la varianza de la serie es $\gamma(0)$ y reemplazando en (3.3.4.3) se obtiene la forma de la función de autocorrelación para una serie estacionaria, esto es

$$\rho_k = \frac{\gamma_k}{\sqrt{\gamma_0} \sqrt{\gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

Para estimar la función de autocorrelación en situaciones reales, es decir a partir de los datos se utilizan las siguientes expresiones matemáticas

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (X_{t+k} - \bar{X})(X_t - \bar{X}) \quad (3.3.4.4)$$

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (3.3.4.5)$$

Note que, mientras k aumenta la separación entre las observaciones se vuelve más distante, y por lo tanto en procesos estacionarios, se espera que la autocorrelación se desvanezca.

3.3.5 Función de Autocorrelación Parcial (FAP)

Otra función relevante en el análisis de las series temporales es la función de autocorrelación parcial, en este caso se trata de la autocorrelación entre X_t y X_{t+k} , después de eliminar la dependencia lineal que X_t y X_{t+k} tengan de las variables intermedias $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$.

A partir de esto, se plantea la siguiente correlación condicional

$$\text{Corr}(X_t, X_{t+k} | X_{t+1}, \dots, X_{t+k-1})$$

En forma más explícita, como dado en (Wei, 1990), sea X_t un proceso estacionario con $E(X_t) = 0$, y defina un modelo de dependencia lineal entre X_{t+k} y $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$.

La estimación lineal más aproximada, en el sentido del menor error cuadrático medio, denotado \hat{X}_{t+k} es expresada como:

$$\hat{X}_{t+k} = \alpha_1 X_{t+k-1} + \alpha_2 X_{t+k-2} + \dots + \alpha_{k-1} X_{t+1}$$

Siendo α_i los coeficientes de la regresión lineal estimados al minimizar

$$E(X_{t+k} - \hat{X}_{t+k})^2 = E(X_{t+k} - \alpha_1 X_{t+k-1} - \dots - \alpha_{k-1} X_{t+1})^2 \quad (3.3.5.1)$$

A partir de un método de minimización se obtiene el siguiente sistema de ecuaciones lineales:

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \dots + \alpha_{k-1} \gamma_{i-k+1} \quad (1 \leq i \leq k-1) \quad (3.3.5.2)$$

y también,

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + \dots + \alpha_{k-1} \rho_{i-k+1} \quad (1 \leq i \leq k-1)$$

expresado matricialmente de la siguiente manera

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix}$$

Análogamente, para $\hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \dots + \beta_{k-1} X_{t+k-1}$ donde β_i es obtenido al minimizar

$$E(X_t - \hat{X}_t)^2 = E(X_t - \beta_1 X_{t+1} - \dots + \beta_{k-1} X_{t+k-1})^2$$

y siguiendo los procesos realizados a partir de (3.3.5.1), se obtiene que

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \dots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix}$$

Descubriendo así que $\alpha_i = \beta_i$.

Si calculamos la varianza y la covarianza como en la ACF, tendríamos que

$$\begin{aligned} \text{Var}(X_{t+k} - \hat{X}_{t+k}) &= E[(X_{t+k} - \alpha_1 X_{t+k-1} - \dots - \alpha_{k-1} X_{t+1})^2] \\ &= E[X_{t+k}(X_{t+k} - \alpha_1 X_{t+k-1} - \dots - \alpha_{k-1} X_{t+1})] \end{aligned}$$

Una vez resuelto el cuadrado, los demás términos restantes se reducen a cero gracias a

(3.3.5.2). Considerando lo anterior, podemos igualar las varianzas y obtener:

$$\begin{aligned} \text{Var}(X_{t+k} - \hat{X}_{t+k}) &= \text{Var}(X_t - \hat{X}_t) \\ \text{Var}(X_t - \hat{X}_t) &= \gamma_0 - \alpha_1 \gamma_1 - \dots - \alpha_{k-1} \gamma_{k-1} \end{aligned} \quad (3.3.5.3)$$

Además, podemos plantear lo siguiente

$$\begin{aligned} &\text{Cov}[(X_t - \hat{X}_t), (X_{t+k} - \hat{X}_{t+k})] \\ &= E[(X_t - \alpha_1 X_{t+1} - \dots - \alpha_{k-1} X_{t+k-1})(X_{t+k} - \alpha_1 X_{t+k-1} - \dots - \alpha_{k-1} X_{t+1})] \\ &= E[(X_t - \alpha_1 X_{t+1} - \dots - \alpha_{k-1} X_{t+k-1})X_{t+k}] \\ &= \gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1 \end{aligned} \quad (3.3.5.4)$$

Con la información obtenida podríamos utilizar el proceso (3.3.4.3), esto nos dice que

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\sigma_x^2} \sqrt{\sigma_y^2}}$$

Observe que ya no se tendrá en cuenta sólo la autocorrelación $(X_t - \hat{X}_t)$ sino que se estudiará

en simultáneo con $(X_{t+k} - \hat{X}_{t+k})$, así obtenemos que la PACF de p_k está dada por:

$$P_k = \frac{Cov[(X_t - \hat{X}_t), (X_{t+k} - \hat{X}_{t+k})]}{\sqrt{Var(X_t - \hat{X}_t)}\sqrt{Var(X_{t+k} - \hat{X}_{t+k})}} \quad (3.3.5.5)$$

Y como ya obtuvimos los parámetros de P_k reemplazando la covarianza y la varianza en (3.3.5.5) obtenemos que:

$$P_k = \frac{\gamma_k - \alpha_1\gamma_{k-1} - \dots - \alpha_{k-1}\gamma_1}{\gamma_0 - \alpha_1\gamma_1 - \dots - \alpha_{k-1}\gamma_{k-1}} \quad (3.3.5.6)$$

siendo P_k la función de autocorrelación parcial (PACF).

3.3.6 Modelos Autorregresivos (AR)

Cuando se trata de modelar una serie de tiempo, existen múltiples herramientas matemáticas, estadísticas y computacionales. Una familia muy amplia de modelos para series de tiempo es la familia de modelos autorregresivos (AR); este tipo de modelos son muy usados en la práctica porque son fáciles de estimar e interpretar. Sea \tilde{X}_t una serie estacionaria corregida por su media, esto es $\tilde{X}_t = X_t - \mu$, la serie sigue un modelo autorregresivo de orden $p \in \mathbb{Z}^+$ ($AR(p)$) si

$$\tilde{X}_t = \phi_1\tilde{X}_{t-1} + \phi_2\tilde{X}_{t-2} + \dots + \phi_p\tilde{X}_{t-p} + a_t \quad (3.3.6.1)$$

que en términos del polinomio característico se expresa como

$$\phi(B)\tilde{X}_t = a_t \quad (3.3.6.2)$$

donde $\phi(B) = 1 - \phi_1B - \phi_2B^2 - \dots - \phi_pB^p$ es un polinomio de grado p en términos del operador de retrasos B , los parámetros del modelo son $\phi_1, \phi_2, \dots, \phi_p$, y σ_a^2 los cuales deben ser estimados a partir de los datos; σ_a^2 , es la varianza de un proceso ruido blanco. Si el polinomio $\phi(B) = 0$, contiene todas sus raíces por fuera del círculo unitario se dice que \tilde{X}_t es un proceso estacionario.

En la práctica, es común que se usen estos modelos para un p pequeño. Considere por ejemplo el modelo $AR(1)$, un modelo autorregresivo de orden $p = 1$, se trata de la regresión de \tilde{X}_t sobre la serie de sus primeros rezagos como se observa a continuación

$$\tilde{X}_t = \phi \tilde{X}_{t-1} + a_t \quad (3.3.6.3)$$

Un resultado muy interesante viene de sustituir n veces, y en forma sucesiva, los términos \tilde{X}_{t-j} , de donde se obtiene la ecuación

$$\tilde{X}_t = \phi^{n+1} \tilde{X}_{t-n-1} + a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \dots + \phi^n a_{t-n} \quad (3.3.6.4)$$

La anterior ecuación a su vez puede ser escrita como una serie en términos de los rezagos del ruido blanco $\{a_t\}$, de la siguiente manera

$$\tilde{X}_t = \sum_{j=0}^{\infty} \phi^j a_{t-j} \quad (3.3.6.5)$$

Definiendo $\psi_j = \phi^j$, $j \geq 1$ con $|\phi| < 1$, y teniendo en cuenta que n tiende al infinito, se puede concluir que \tilde{X}_t describe una serie infinita convergente, siendo un caso particular de un modelo $MA(\infty)$.

Con base en lo anterior, un modelo $AR(1)$ descrito en (3.3.6.3) es equivalente al $MA(\infty)$

$$\tilde{X}_t = \phi^{-1}(B)a_t = \psi(B)a_t \quad (3.3.6.6)$$

donde $\psi(B) = \phi^{-1}(B) = \sum_{j=0}^{\infty} \psi^j B^j$.

3.3.7 Modelos de Medias Móviles (MA)

Note que el modelo autorregresivo (3.3.6.1), es un modelo de regresión lineal en el cual la serie \tilde{X}_t depende de sus propios rezagos más un error aleatorio $\{a_t\}$ del tipo ruido blanco. Este modelo no tiene en cuenta otras perturbaciones representadas por los rezagos del propio ruido blanco. Por su parte, un modelo que sólo incluye estas características es el modelo de

medias móviles o de promedio móvil (MA), en el cual \tilde{X}_t es una combinación lineal de los rezagos de $\{a_t\}$, esto es

$$\tilde{X}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (3.3.7.1)$$

Esta ecuación puede ser expresada también en términos del operador de rezagos en la forma

$$\tilde{X}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (3.3.7.2)$$

o de manera abreviada

$$\tilde{X}_t = \theta(B) a_t \quad (3.3.7.3)$$

y es denominado modelo de medias móviles de orden q , $MA(q)$. Este modelo tiene unas características importantes como por ejemplo, su esperanza es cero $E[\tilde{X}_t] = 0$. Además, su varianza también es constante y se puede describir como $Var[\tilde{X}_t] = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_a^2$ y los parámetros $\theta_1, \theta_2, \dots, \theta_q$ y σ_a^2 serán estimados a partir de los datos.

Este modelo $MA(q)$ se puede combinar con el modelo $AR(p)$ generando así un modelo mixto de rezagos de la serie y rezagos del ruido blanco, el cual es muy conocido en la teoría de los procesos estocásticos y del análisis de series de tiempo denominado el modelo $ARMA(p, q)$ o modelo autorregresivo de medias móviles de orden (p, q) .

3.3.8 Modelo ARMA

Un modelo ARMA más conocido como modelo autorregresivo de medias móviles es una combinación de los dos modelos descritos anteriormente, esto es

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \dots + \phi_p \tilde{X}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

donde los parámetros ϕ_i 's, θ_j 's y la varianza del ruido blanco serán estimados a partir de los datos. La manera simplificada de expresar este modelo es dada por

$$\phi(B) \tilde{X}_t = \theta(B) a_t \quad (3.3.8.1)$$

o

$$\tilde{X}_t = \phi^{-1}(B)\theta(B)a_t = \psi(B)a_t \quad (3.3.8.2)$$

con $\psi(B) = \phi^{-1}(B)\theta(B)$, siempre que el proceso sea estacionario, es decir que las raíces del polinomio $\phi(B)$ estén por fuera del círculo unitario.

3.3.9 Modelo ARIMA

Cuando una serie de tiempo no tiene un comportamiento estacionario en el sentido que el polinomio $\phi(B)$ presenta raíces unitarias, el modelo ARMA incluye otro componente que da información sobre estas raíces, esto es un nuevo componente d , conocido como el parámetro de diferenciación, (Box, Jenkins, & Reinsel, 2008). Se define así un modelo no estacionario homogéneo en términos del operador autorregresivo $\varphi(B)$ escrito como

$$\varphi(B)\tilde{X}_t = \phi(B)(1 - B)^d X_t = \theta(B)a_t \quad (3.3.9.1)$$

o de otra forma

$$\phi(B)w_t = \theta(B)a_t \quad (3.3.9.2)$$

con

$$w_t = (1 - B)^d \tilde{X}_t = \nabla^d \tilde{X}_t \quad (3.3.9.3)$$

Donde la serie w_t es estacionaria. Estas dos ecuaciones permiten crear un modelo más completo que permita estudiar el comportamiento de series de tiempo estacionarias y no estacionarias, este se conoce como modelo autorregresivo integrado de medias móviles más conocido como ARIMA y denotado $ARIMA(p, d, q)$ con 3 parámetros, donde p hace referencia al orden del polinomio autorregresivo estacionario, d al orden de diferenciación de la serie y q al orden del polinomio de medias móviles invertible (González Casimiro,

2009); observe que si $d = 0$, obtendríamos el modelo $ARMA(p, q)$; en general, p, d, q no suelen ser más grandes que 2.

3.3.10 Procesos estacionales SARIMA

Cuando una componente “estacional” está presente en una serie de tiempo, dicha componente se puede observar a través del gráfico de la función de autocorrelación la cual muestra un cierto patrón repetitivo cada s períodos de tiempo. En los modelos de series temporales anteriores sólo se tenía en cuenta la estacionariedad, pero cuando entra en juego la estacionalidad de una serie, es necesario realizar una extensión del modelo. El nuevo modelo se conoce como el autorregresivo estacional integrado de medias móviles (SARIMA).

El modelo $ARIMA(p, d, q)$ se complementará con nuevos parámetros que permitan modelar el comportamiento estacional de los datos, este nuevo modelo se denota como $SARIMA(p, d, q) \times (P, D, Q)_s$ donde P, D, Q son los parámetros estacionales y s describe el período en que se repite el patrón. Un $SARIMA(p, d, q) \times (P, D, Q)_s$ está expresado según (Brockwell & Davis, 2002) como

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D X_t = \theta(B)\Theta(B^s)a_t \quad (3.3.10.1)$$

donde los polinomios $\phi(B)$, $\Phi(B^s)$, $\theta(B)$ y $\Theta(B^s)$ en términos de X corresponden a

$$\phi(X) = 1 - \phi_1 X - \dots - \phi_p X^p$$

$$\Phi(X) = 1 - \Phi_1 X - \dots - \Phi_P X^P$$

$$\theta(X) = 1 + \theta_1 X + \dots + \theta_q X^q$$

$$\Theta(X) = 1 + \Theta_1 X + \dots + \Theta_Q X^Q$$

Además, es posible expresar un proceso $SARIMA(p, d, q) \times (P, D, Q)_s$ estacionario solamente en términos de rezagos del ruido blanco $\{a_t\}$, teniendo en cuenta que si Y_t es dado por

$$\begin{aligned} (1 - B)^d(1 - B^s)^D X_t &= Y_t \\ \phi(B)\Phi(B^s)Y_t &= \theta(B)\Theta(B^s)a_t \\ Y_t = \psi(B)a_t &= \sum_{j=0}^{\infty} (\psi_j B^j) a_t = \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)} a_t. \end{aligned} \quad (3.3.10.3)$$

Note que los ARIMA y SARIMA conforman una familia de modelos lineales para las series de tiempo. En el caso de series no lineales existe una amplia literatura de modelación; una familia particular de modelos no lineales provenientes de la teoría de inteligencia artificial o aprendizaje computacional son los modelos de redes neuronales artificiales, denotados por la sigla ANN, de su nombre en inglés *Artificial Neural Network*. Una adaptación de estos modelos a las series de tiempo es dada cuando las entradas de la red son distintos rezagos de la serie observada y es denominado el modelo ARNN del inglés *Autoregressive Neural Network*.

3.4 Modelo de redes neuronales autorregresivas (ARNN)

Las redes neuronales artificiales provienen de los sistemas de inteligencia artificial; según las características de los datos existen ciertas variaciones que incluyen el uso de series temporales, para el caso particular de esta investigación se hará referencia al modelo de redes neuronales autorregresivas (ARNN).

En el caso puntual de la modelación, una neurona es una función alimentada por datos que se transforman en una respuesta. Por lo general, una red neuronal multicapa consta de una o

varias capas ocultas que permiten generar aproximaciones y pronósticos de comportamientos no lineales.

Una red neuronal autorregresiva consta de un vector entrada $X = (X_{t-1}, \dots, X_{t-p})^T$ que corresponde a los rezagos, cada entrada genera cierta cantidad de pesos w_{ij} que permiten transportar de una primer capa donde pertenece x , a una capa oculta donde se genera un vector $Y = (Y_1, \dots, Y_M)^T$ que corresponde a las observaciones no visibles pertenecientes a la capa oculta.

Una vez reconocido el vector y por el software, pasan por la función de activación de las neuronas, por lo general, estas son

- a) Función logística sigmoidea

$$f(x) = \frac{1}{1 + e^{-x}}$$

- b) Función tangente hiperbólica

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

La función de activación, aporta la cantidad de neuronas que seguirán activas y esto permite asignar unos nuevos pesos que generan unas respuestas, este conjunto se conoce como la capa de salida denotada por el vector $Z = (Z_1, \dots, Z_k)^T$.

La estructura de una red neuronal se puede ver de manera clara en el siguiente gráfico:

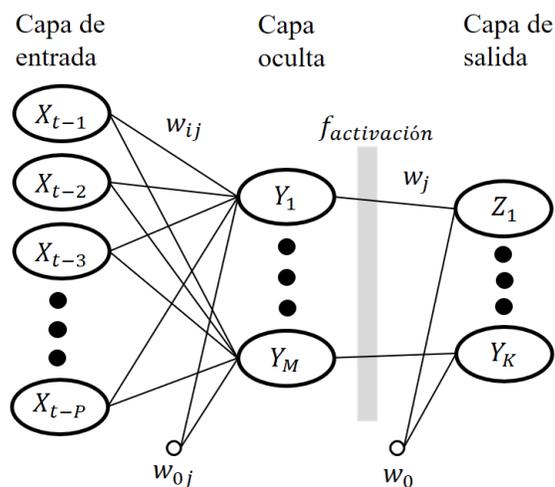


Figura 3.4.1 Estructura del modelo ARNN

donde w_{0j} y w_0 corresponden a los sesgos aplicados inicialmente.

Generalizando, el modelo $ARNN(p, q)$ donde p hace alusión a las entradas retrasadas y q a los nodos de la capa oculta, se describe matemáticamente como:

$$Z_t = w_0 + \sum_{j=1}^q w_j \cdot f \left(w_{0j} + \sum_{i=1}^p w_{ij} \cdot X_{t-i} \right) + \varepsilon_t$$

donde $f(\cdot)$ corresponde a la función de activación y ε_t a los residuales.

Otra manera de expresar abreviadamente un modelo ARNN, está dado de la siguiente manera

$$z_t = f(w_0 + w_1 X_{t-1} + w_2 X_{t-2} + \dots + w_n X_{t-n}) + \varepsilon_t$$

3.5 Enfoque Híbrido de Modelación

Bajo el supuesto que *todo modelo puede ser mejorado*, una vez se ajusta un modelo a un conjunto de datos siguen otras etapas importantes como, medir la adecuación del modelo, compararlo con otros y si es el caso, mejorarlo en cuanto a su calidad predictiva. La comparación de los modelos puede hacerse mediante las métricas discutidas en la Sección

3.6. Por su parte, una metodología moderna para tratar de mejorar las predicciones en una

serie de tiempo es aplicando un enfoque híbrido, el cual combina varios modelos en búsqueda de un mejor ajuste. La mezcla puede ser de distintos tipos, por ejemplo, entre un modelo lineal y uno no lineal, entre un modelo en tiempo continuo y uno en tiempo discreto, o entre un modelo determinístico o computacional y uno estocástico.

Los pasos para ajustar un modelo híbrido son los siguientes: primero se ajusta un modelo (por ejemplo, determinístico no lineal) que sea adecuado para la serie temporal observada $\{X_t, t = 01, 2, \dots, n\}$ y a partir de dicho modelo se obtiene la serie predicha $\{\hat{X}_t, t = 01, 2, \dots, n\}$. Luego se obtiene la serie

$$e_t = X_t - \hat{X}_t \quad (3.5.1)$$

la cual corresponde a la serie residual, o de los errores estimados, y viene dada por la diferencia entre los datos reales (X_t) y los predichos (\hat{X}_t).

Una vez obtenida la serie residual se debe evaluar si contiene otros componentes por ejemplo de autocorrelación, estacionales, cíclicos, etc. Ante la presencia de alguna de estas componentes la propuesta consiste en ajustar un nuevo modelo (por ejemplo un modelo lineal estocástico) para la serie de los errores y cuyas predicciones generadas por dicho modelo se denotan \hat{e}_t . Los valores predichos de e_t se combinan con las predicciones del primer modelo y se obtiene la predicción híbrida para la serie original dada por

$$w_t = X_t + \hat{e}_t \quad (3.5.2)$$

Las predicciones del modelo híbrido serán mejores si las métricas de calidad del modelo son significativamente menores a las del modelo inicial.

3.6 Selección y comparación de modelos

Con el fin de seleccionar el modelo que mejor se ajuste a una serie temporal se debe disponer de algunas herramientas que permitan cuantificar la precisión de cada modelo ajustado. Los modelos se comparan mediante métricas que permiten evaluar de cierta forma, la variabilidad de los errores obtenidos en cada modelo. Los errores se estiman mediante la diferencia entre los datos reales obtenidos y los pronósticos de los datos a partir de cada modelo.

Para esto, se pueden utilizar las siguientes tres métricas cuyas siglas vienen de su nombre en inglés: RMSE (root mean square error), MAE (mean absolute error) y MAPE (mean absolute error porcentual) donde

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{X}_t - X_t)^2} \quad (3.6.1)$$

$$MAE = \frac{\sum_{t=1}^n |\hat{X}_t - X_t|}{n} \quad (3.6.2)$$

$$MAPE = \frac{\sum_{t=1}^n |(\hat{X}_t - X_t)/X_t|}{n} \quad (3.6.3)$$

donde \hat{X}_t representa el pronóstico dado por el modelo en cuestión y X_t representa el dato real para cada observación $t = 1, 2, \dots, n$.

Posteriormente se comparan los resultados obtenidos con las métricas y se define el “mejor modelo” como aquel cuyas métricas se aproximan más a cero, minimizando el error.

4. Aplicación y ajuste de los modelos

4.1 Los Datos

Los datos usados en esta investigación consisten en las notificaciones diarias por viruela símica en Colombia entre el 23 de junio y el 31 de octubre del 2022, los cuales están disponibles en la página web <https://www.ins.gov.co/Noticias/Paginas/Enfermedades-emergentes.aspx>.

4.2 Análisis exploratorio de la serie temporal de las notificaciones diarias

La gráfica de la serie temporal que representa las notificaciones diarias de viruela símica en Colombia desde el 23 de junio al 31 de octubre de 2022 es la siguiente:

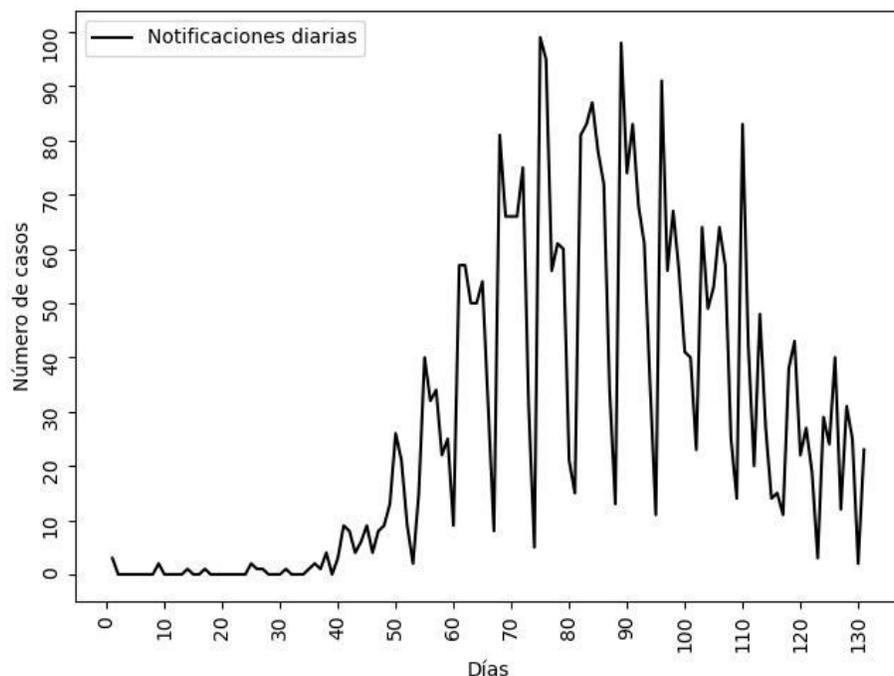


Figura 4.1 Notificaciones diarias de viruela símica en Colombia desde el 23 de junio al 31 de octubre de 2022

Note que en los primeros 40 días, tanto las notificaciones como la variabilidad en los registros son muy bajas. Sin embargo, a partir del registro 40 la serie temporal no sólo presenta un aumento de los casos, sino una mayor variabilidad en los registros y aparentemente, un comportamiento estacional el cual podría ser espurio. Para explorar un poco mejor esta serie temporal se presenta inicialmente en la Tabla 4.1 y en la Figura 4.2, el comportamiento de las medias y varianzas locales de la serie, calculadas por tramos de diez observaciones; como puede observarse tanto las medias como las varianzas no permanecen constantes en el tiempo lo cual se asocia con un comportamiento no estacionario de la serie; adicionalmente, la varianza local aumenta con la media lo cual sugiere la necesidad de una posible transformación para estabilizar la varianza. Por otra parte, en las Figuras 4.3 y 4.4 aparecen las gráficas de autocorrelación simple y parcial de la serie temporal, notándose en particular unos valores altos en los rezagos 7, 14 y 21 de la FAS y en el rezago 7 de la FAP, esto podría estar sugiriendo la presencia de una componente estacional en las notificaciones.

MEDIA	VARIANZA
0,35	0,66
0,80	1,33
4,15	26,69
4,50	12,05
9,80	62,17
22,40	305,62
17,70	142,221
36,75	418,72
58,25	866,09
54,45	580,58

Tabla 4.1 Medias versus varianzas locales de la serie de notificaciones diarias

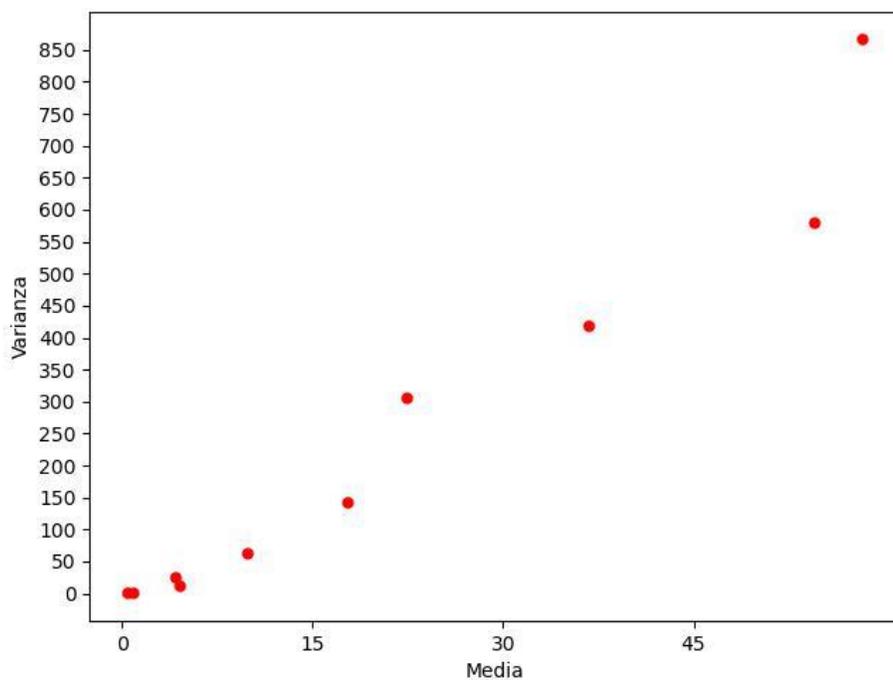


Figura 4.2 Medias versus varianzas locales de la serie de notificaciones diarias

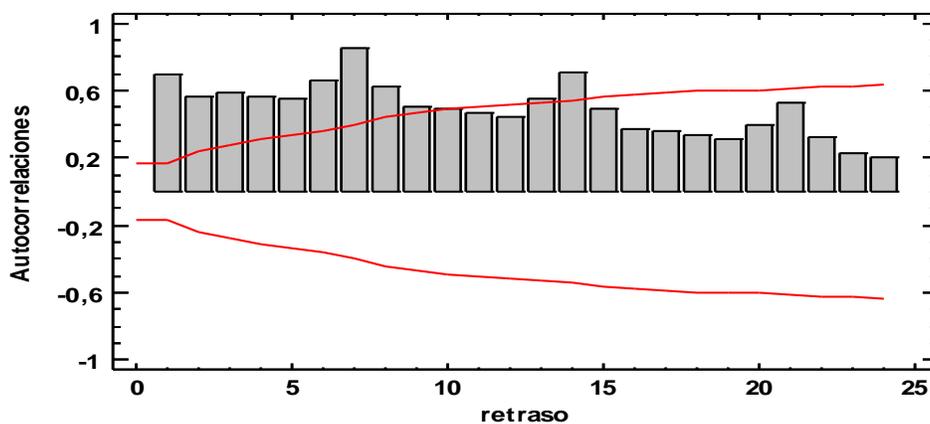


Figura 4.3 Función de autocorrelación de la serie de notificaciones diarias

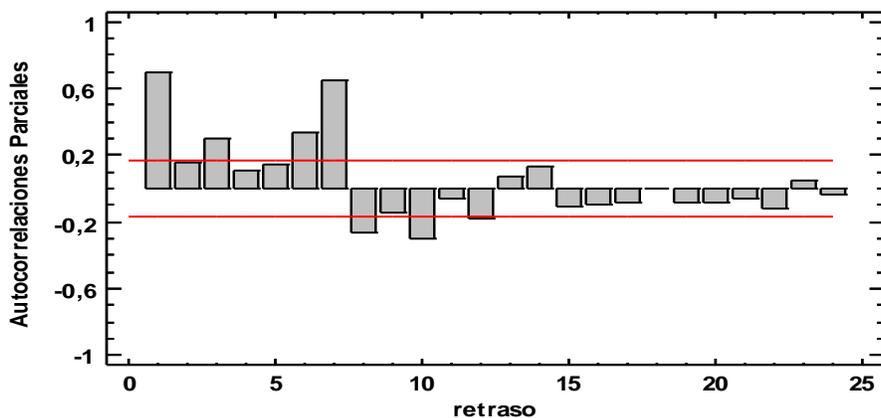


Figura 4.4 Función de autocorrelación parcial de la serie de notificaciones diarias

4.3 Modelación de la serie temporal de las notificaciones acumuladas

Para tener una idea global de cómo ha sido el comportamiento del aumento de los casos de viruela símica en Colombia desde la primera notificación hasta el 31 de octubre de 2022, aparece en la Figura 4.5 el comportamiento de la serie temporal de las notificaciones acumuladas diariamente.

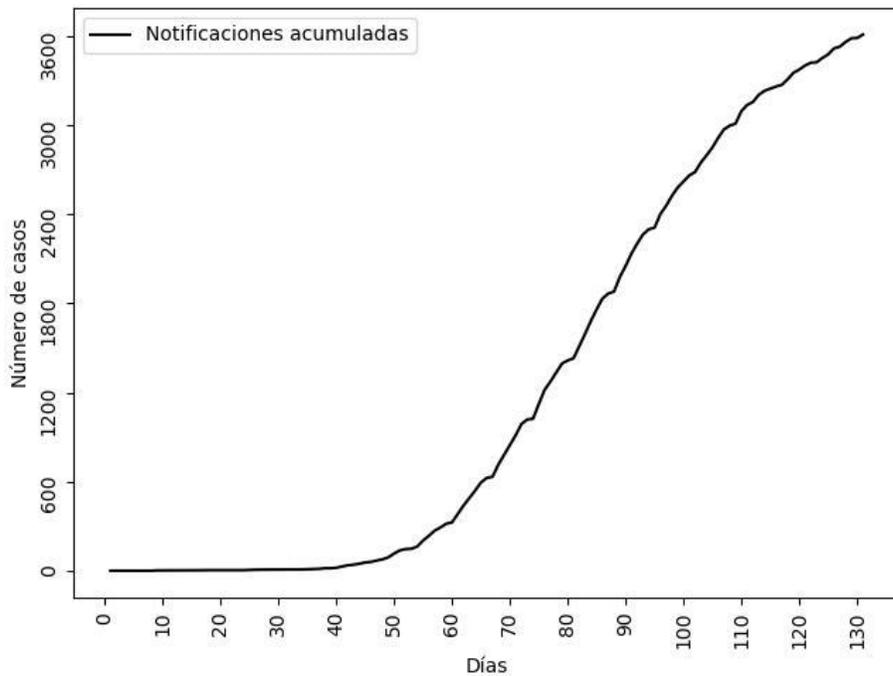


Figura 4.5 Notificaciones diarias acumuladas de la viruela símica en Colombia desde el 23 de junio al 31 de octubre de 2022

Una forma de explorar el aumento de la infección es comparando la curva de la Figura 4.5 con las curvas de los casos acumulados de otros países importantes de América como Estados Unidos, Brasil, Canadá, Chile y Perú, entre el 23 de junio hasta el 31 de octubre del 2022. Note que, Estados Unidos, Brasil y Perú son los países con mayor número de notificaciones en América hasta la fecha de estudio, mientras que Canadá y Chile presentaron en general bajas notificaciones. Colombia alcanza los niveles de Perú al final del período de estudio.

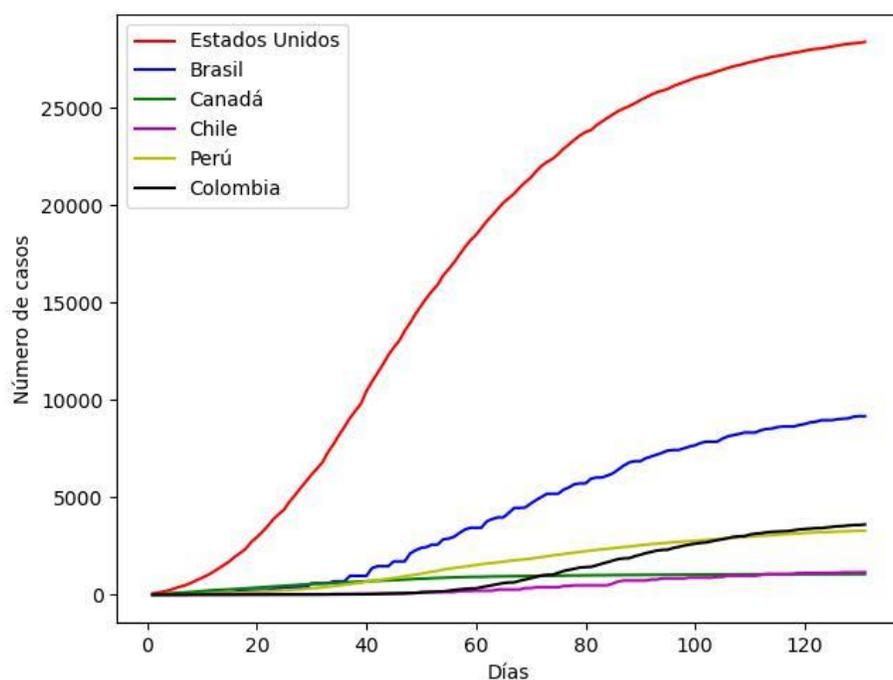


Figura 4.6 Notificaciones acumuladas de algunos países de América desde el 23 de junio al 31 de octubre de 2022

Por practicidad, en este trabajo se modelan los datos acumulados con el fin de analizar el crecimiento del número de infectados; la modelación se hizo con la ayuda de las herramientas computacionales StatGraphics Centurion XVI, Python y R. La idea de usar un software u otro fue solamente por conveniencia.

4.3.1 Ajuste del modelo de crecimiento logístico

El ajuste del modelo básico de crecimiento poblacional logístico a la serie de las notificaciones acumuladas diarias se realizó mediante un código adaptado en Python, esto con el fin de estimar los parámetros α , γ y β del modelo (3.1.1). El modelo ajustado dio como resultado los siguientes parámetros:

$$\alpha = 3624.95, \quad \gamma = 0.077, \quad \beta = 740.497$$

Su respectivo gráfico junto con los datos reales acumulados aparece en la Figura 4.7.

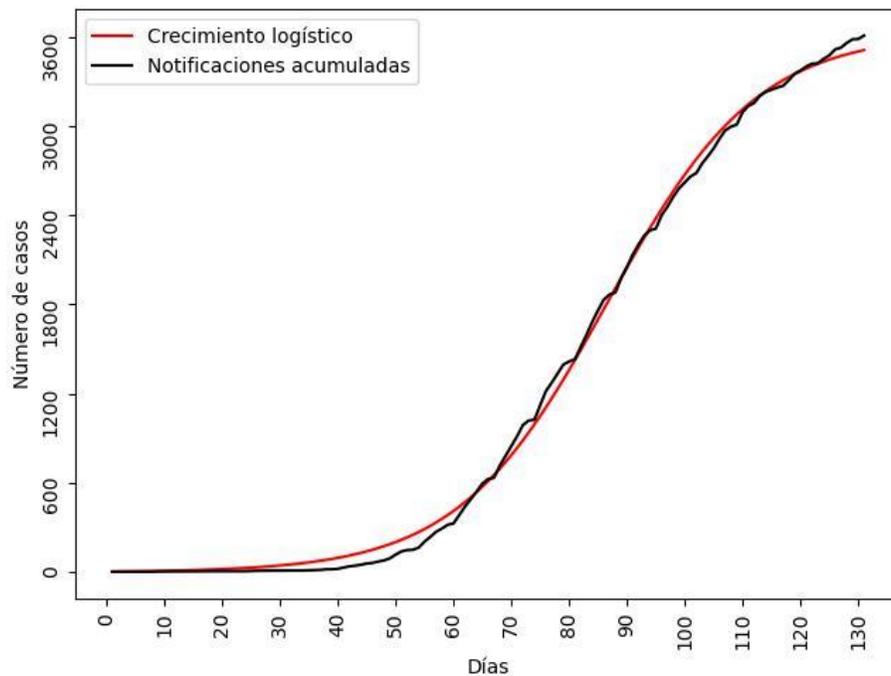


Figura 4.7 Modelo de crecimiento logístico comparado a los datos reales

Observe que el modelo ajusta razonablemente bien, sin embargo, su adecuación no depende sólo del gráfico, en la Tabla 4.2 adelante aparecen las respectivas métricas de precisión para ser comparada con la de los otros modelos ajustados en este trabajo.

4.3.2 Ajuste del modelo SIR

Notando la ecuación (3.4.1), el modelo SIR consta de tres ecuaciones diferenciales y dos parámetros, así que uno de los problemas a los que nos enfrentamos debido a las pocas investigaciones relacionadas con los datos de la Viruela del mono, es el desconocimiento de las tasas de infectados y de recuperación, k_1 y k_2 , respectivamente. Por lo tanto, la estimación de estos parámetros es un paso inicial en el ajuste del modelo a los datos reales.

La estimación se realiza minimizando el error cuadrático medio entre los datos observados y los datos ajustados según el modelo (3.2.2) para distintos valores de los parámetros. El algoritmo de optimización Runge Kutta obtuvo los siguientes estimadores

$$R_0 = 1.04$$

$$k_2 = 0.81.$$

Reemplazado R_0 y k_2 en $R_0 = \frac{k_1}{k_2}$ se obtiene que

$$k_1 = 0.8424$$

Estos parámetros estimados se reemplazan en el sistema de ecuaciones diferenciales (3.2.2).

$$\frac{dS}{dt} = -0.81 \cdot 1.04 \left(\frac{SI}{n} \right)$$

$$\frac{dI}{dt} = 0.81 \cdot 1.04 \left(\frac{SI}{n} \right) - 0.81 \cdot I$$

$$\frac{dR}{dt} = 0.81 \cdot I$$

Es decir,

$$\frac{dS}{dt} = -0.8424 \left(\frac{SI}{n} \right)$$

$$\frac{dI}{dt} = 0.8424 \left(\frac{SI}{n} \right) - 0.81 \cdot I$$

$$\frac{dR}{dt} = 0.81 \cdot I$$

Interpretando ligeramente los parámetros estimados, se sabe que si $R_0 > 1$, la enfermedad podría aumentar rápidamente en el tiempo; sin embargo, como su estimación es ligeramente mayor a 1 podría decirse que, en el período de análisis, la enfermedad se encuentra en crecimiento pero aún bajo control. Por su parte, con respecto a la tasa de infección de 0.84 es relativamente alta, pero dado que la tasa de recuperación también es relativamente alta (0.81), la enfermedad en Colombia está bajo control y no alcanza registros alarmantes.

Con el modelo SIR es posible obtener los gráficos de infectados y recuperados, sin embargo, nosotros sólo estamos interesados en los infectados, por lo tanto, resolviendo este sistema con ayuda de un software se obtienen los pronósticos y se comparan con los infectados notificados acumulados, obteniendo el siguiente gráfico:

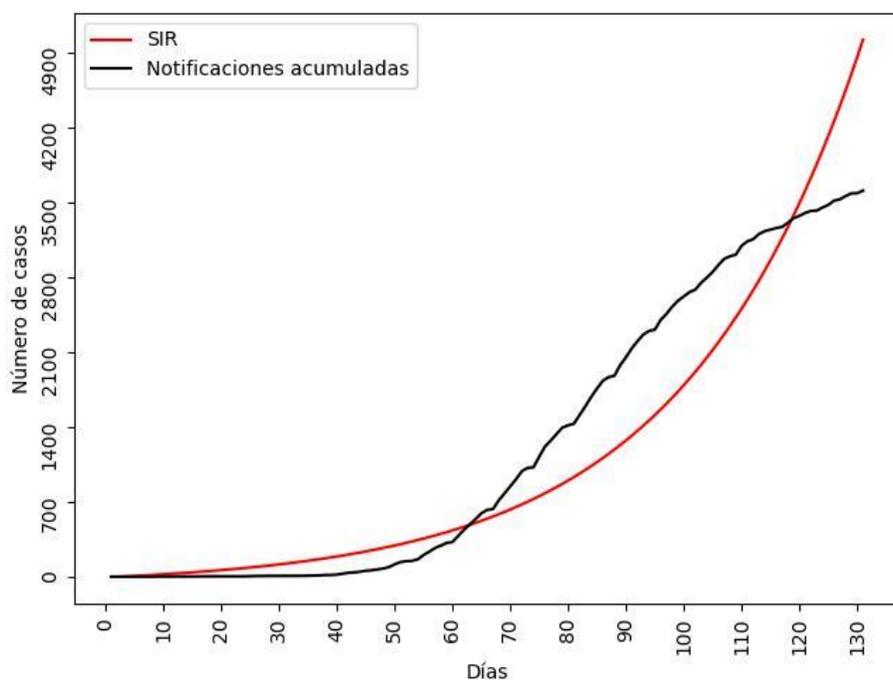


Figura 4.8 Modelo SIR comparado a los datos reales

4.3.3 Ajuste del modelo ARNN

El ajuste del modelo autorregresivo de redes neuronales se realizó con ayuda de R. El modelo estimado fue el ARNN(12,1), donde 12 corresponde al número de entradas en la primera capa las cuales corresponden a los primeros 12 retrasos de la serie observada y 1 es la única variable de salida. La expresión matemática que describe el modelo está dada por

$$X_t = f(w_0 + w_1X_{t-1} + w_2X_{t-2} + \dots + w_{12}X_{t-12} + e)$$

donde f es la función de activación. La serie original y la pronosticada aparecen en la Figura 4.9.

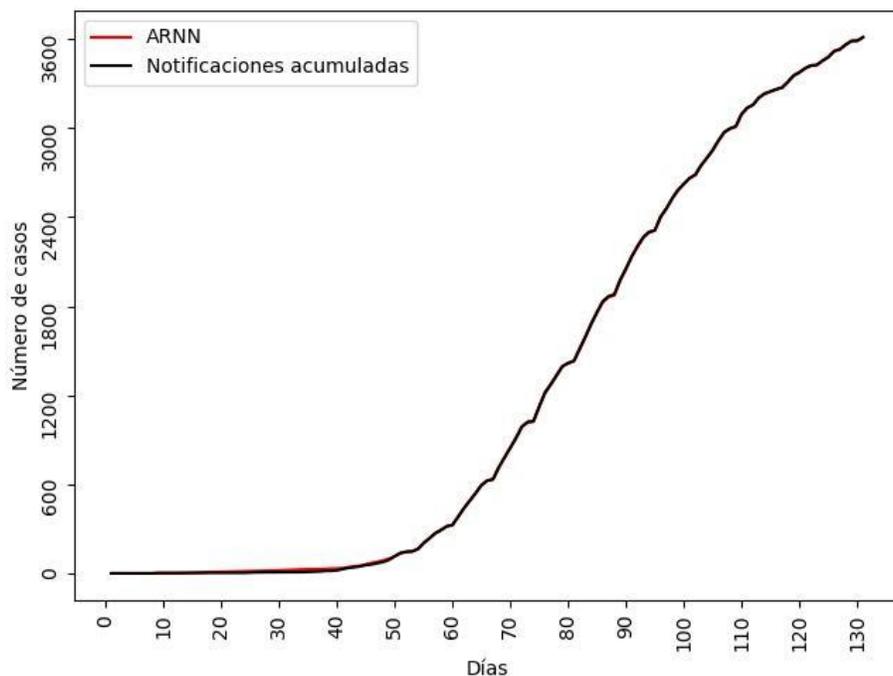


Figura 4.9 Modelo ARNN comparado a los datos reales

4.3.4 Ajuste del modelo de series temporales

Los ajustes de los modelos de series de tiempo fueron realizados con ayuda del StatGraphics Centurion XVI. Los modelos AR, MA Y ARMA no proporcionaron un buen resultado, lo

cual era de esperarse ya que son modelos que no tienen en cuenta ciertas características presentes en los datos. Con los modelos ARIMA observamos que el ruido blanco y las métricas ya iban mejorando, pero observando la gráfica de la función de autocorrelación observamos que existía un patrón repetitivo cada 7 períodos de tiempo (días), sugiriendo una componente estacional de 7 días. Se ajusta entonces un modelo SARIMA y se evalúan los supuestos estadísticos, más específicamente la no autocorrelación y la normalidad residual. Varios modelos SARIMA ajustados incumplían el supuesto de Normalidad mediante la prueba de Kolmogorov-Smirnov. Además, la varianza de los residuales no era estacionaria, por lo tanto, se realiza una transformación de Box y Cox a la serie original. Para identificar la potencia de la transformación se realiza un gráfico de dispersión de la varianza muestral versus la media muestral para distintas ventanas de observación, como aparece en la Figura 4.10.

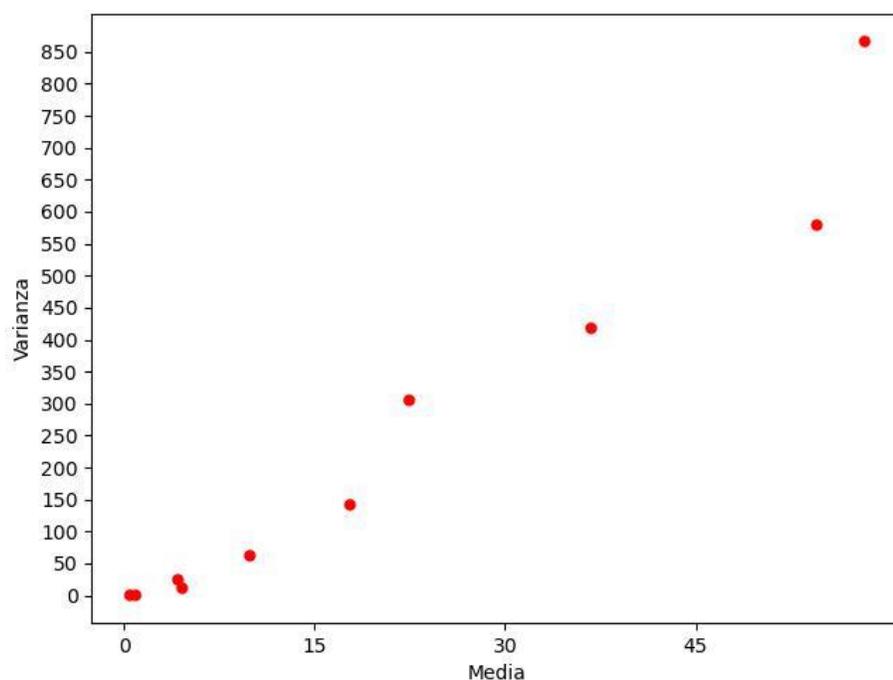


Figura 4.10 Medias versus varianzas locales de la serie de notificaciones diarias

Del comportamiento de la Figura anterior se dedujo que la serie temporal requiere una transformación de Box-Cox, con un valor de $\lambda = \frac{2}{3}$. El modelo SARIMA ajustado a la serie transformada es el $SARIMA(1,1,0)X(1,0,0)_7$, con muy buena calidad de ajuste. En adición, podemos decir que los residuales son practicamente no autocorrelacionados como se muestra en la Figura 4.11 y 4.12; además, se satisface el supuesto de normalidad de los residuales como era de esperarse gracias a la transformación de potencia, solucionando así uno de los problemas con los modelos no estacionales ajustados previamente.

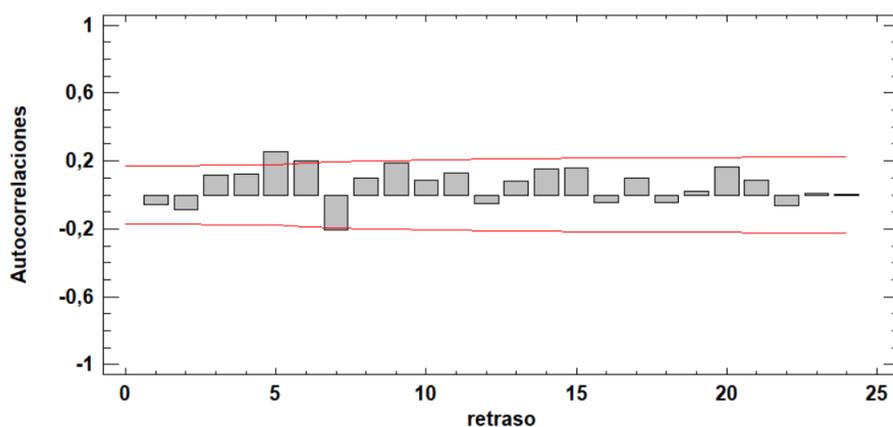


Figura 4.11 Función de autocorrelación residual del $SARIMA(1, 1, 0) \times (1, 0, 0)_7$

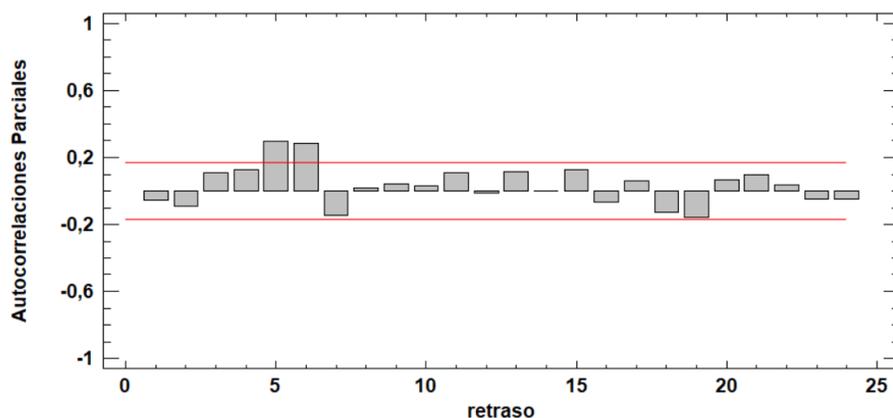


Figura 4.12 Función de autocorrelación parcial residual del $SARIMA(1, 1, 0) \times (1, 0, 0)_7$

Finalmente, el modelo para la serie de casos acumulados de Viruela del mono, X_t , es

$$(1 - \varphi_1 B)(1 - \Phi_1 B^7)(1 - B) \left(X_t^{\frac{2}{3}} \right) = a_t$$

donde $X_t^{\frac{2}{3}}$ representa la transformación de Box-Cox y posteriormente aplicando unas diferencias de orden 1 a la serie transformada.

El gráfico 4.13 representa la serie observada en rojo y la serie de los casos acumulados en negro, notándose un muy buen ajuste.

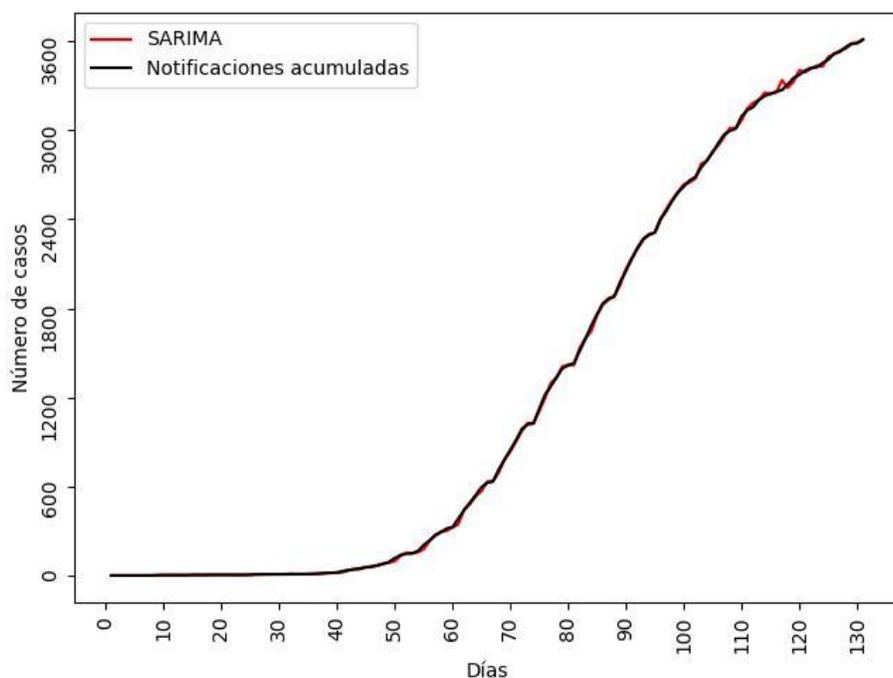


Figura 4.13 Modelo SARIMA comparado a los datos reales

4.3.5 Ajuste de modelos híbridos

Los modelos híbridos se obtuvieron combinando un modelo matemático no lineal para la serie de casos acumulados y un modelo lineal de la familia SARIMA para la serie del error.

Los procesos se hacen con ayuda del Excel y el StatGraphics Centurion XVI. Los modelos no lineales son el modelo logístico y el SIR.

Modelo de crecimiento logístico + SARIMA

Una vez ajustado el modelo logístico, la serie residual sigue un modelo $SARIMA(0,1,0) \times (1,0,0)_7$, que cumple con la normalidad de los residuales y su no autocorrelación como se observa en las Figuras 4.14 y 4.15. Al adicionar los pronósticos de este SARIMA al ajuste del modelo logístico inicial se obtienen los pronósticos del modelo híbrido. Estos pronósticos aparecen en la Figura 4.16 en color rojo, la curva de color negro corresponde a los casos observados acumulados.

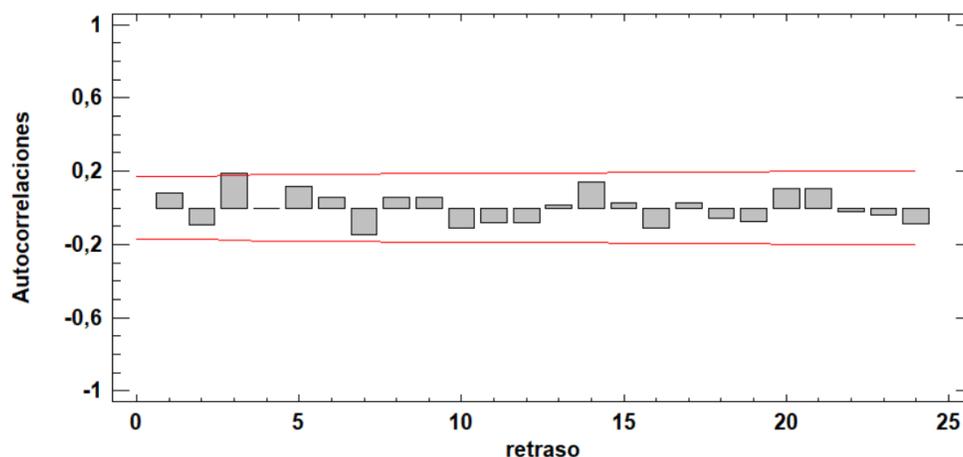


Figura 4.14 Función de autocorrelación residual del $SARIMA(0, 1, 0) \times (1, 0, 0)_7$

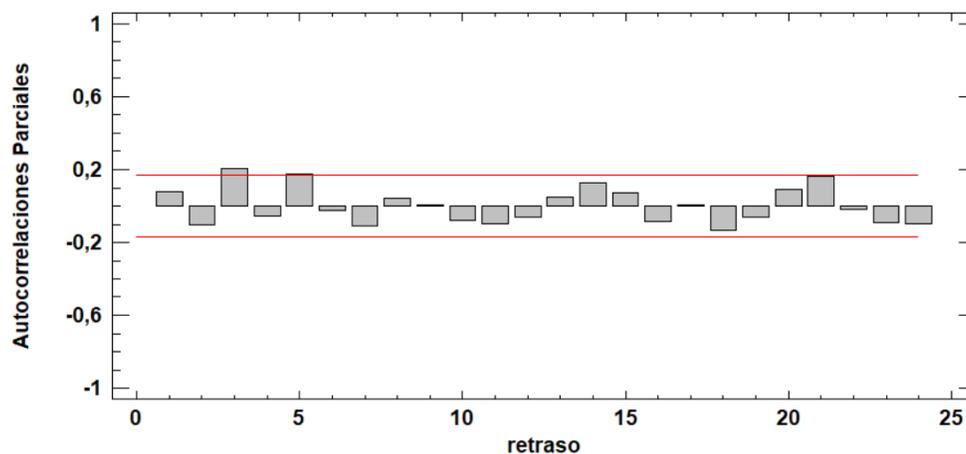
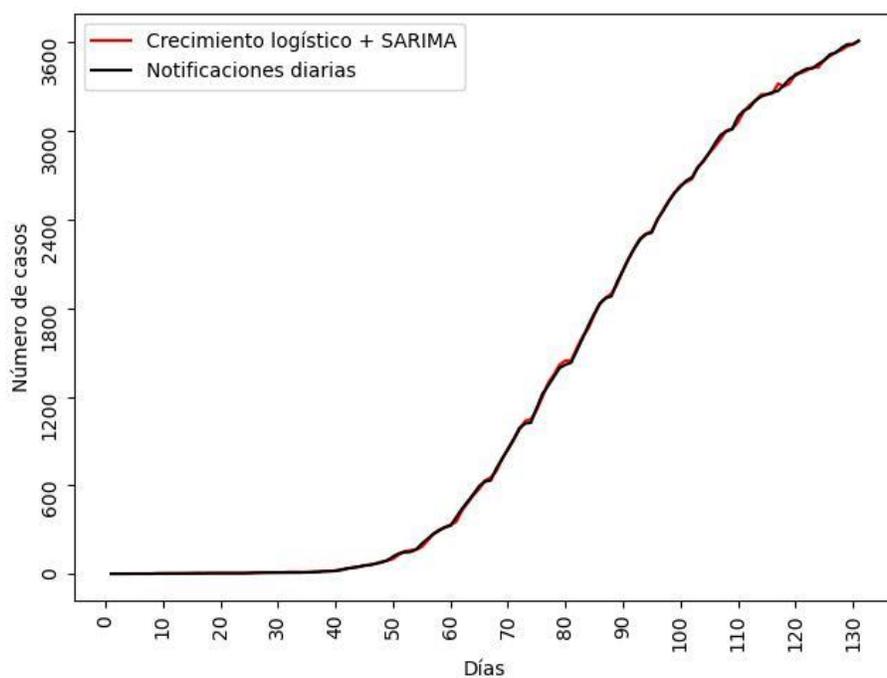


Figura 4.15 Función de autocorrelación parcial residual del $SARIMA(0, 1, 0) \times (1, 0, 0)_7$



4.16 Modelo de crecimiento logístico + SARIMA comparado a los datos reales

Modelo SIR + SARIMA

Análogamente al caso anterior, los residuales del modelo SIR ajustado en la sección 4.3.2 siguen un $SARIMA(0,2,1) \times (1,1,0)_7$. Observe que este modelo tuvo que ser diferenciado dos veces con el objetivo de conseguir la estacionariedad, además, cumple con los supuestos estadísticos, más específicamente con la no autocorrelación de los residuos como se muestra en las Figuras 4.17 y 4.18. Al adicionar estas nuevas predicciones con las del modelo SIR ajustado se obtiene la curva roja de 4.19. La curva negra corresponde a los casos acumulados.

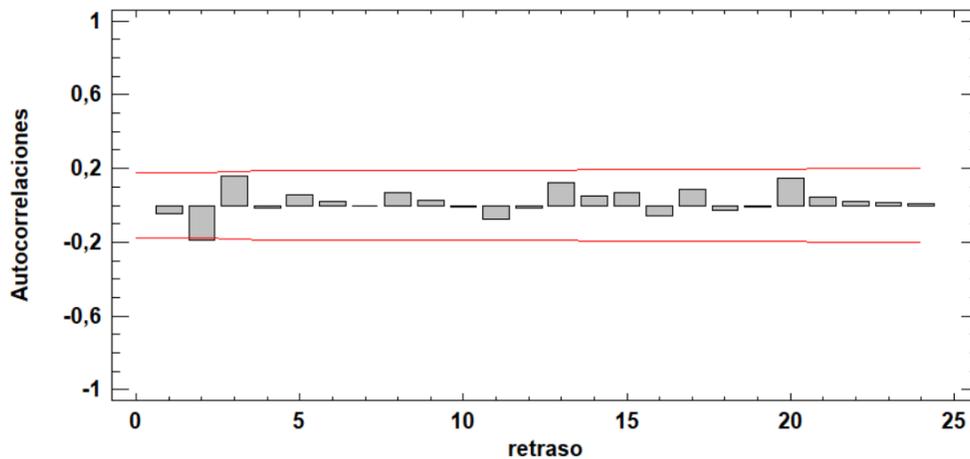


Figura 4.17 Función de autocorrelación residual del $SARIMA(0, 2, 1) \times (1, 1, 0)_7$

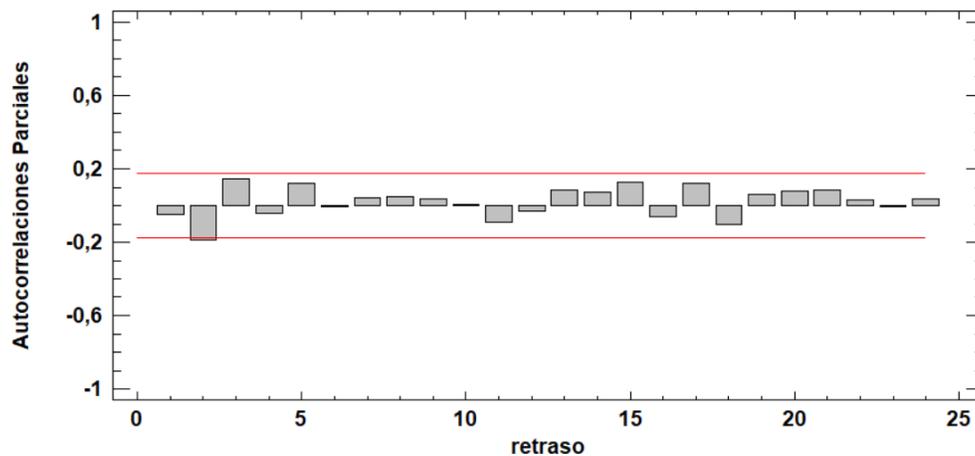
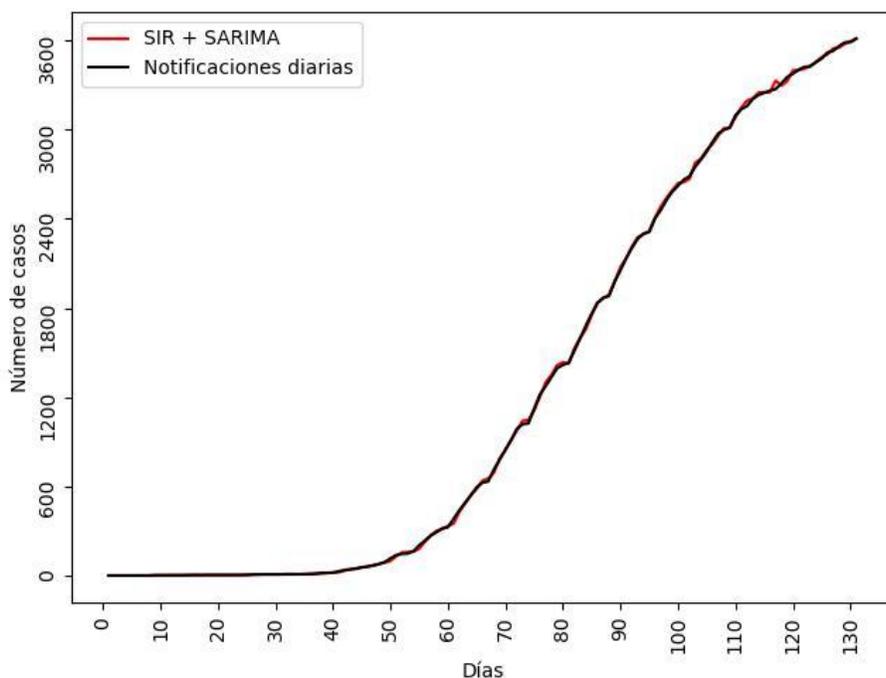


Figura 4.18 Función de autocorrelación residual del $SARIMA(0, 2, 1) \times (1, 1, 0)_7$



4.19 Modelo SIR + SARIMA comparado a los datos reales

4.4 Comparación de modelos

Como se mencionó anteriormente, la comparación de los modelos para la selección del mejor ajuste se hace a través de las métricas RMSE, MAE y MAPE. Estas medidas para cada modelo considerado aparecen en la siguiente tabla:

MODELO	RMSE	MAE	MAPE
Crecimiento Logístico	53,52	42,84	0,83%
SIR	464,43	335,94	2,56 %
ARNN	5,65	3,50	0,19 %
SARIMA(1,1,0)x(1,0,0)₇	12,70	7,99	0,04 %
Híbrido Log + SARIMA	11,64	7,73	0,05 %
Híbrido SIR + SARIMA	12,18	7,93	0,03 %

Tabla 4.2 Métricas de calidad de ajuste o de precisión de los modelos

Dado que los modelos de mejor precisión fueron el ARNN, el SARIMA, el de crecimiento logístico y los híbridos, se pronostican el número de infectados acumulados para los siete (7) días siguientes utilizando los cinco modelos, dichos pronósticos aparecen en la Tabla 4.3.

REALES	ARNN	SARIMA	CRECIMIENTO LOGÍSTICO	Híbrido Log + SARIMA	Híbrido SIR + SARIMA
3628	3598	3628	3524	3629	3637
3638	3590	3664	3532	3656	3670
3642	3582	3675	3538	3645	3649
3680	3576	3703	3545	3663	3667
3696	3570	3725	3550	3697	3703
3698	3565	3727	3556	3696	3698
3700	3561	3748	3561	3714	3724

Tabla 4.3 Predicción del número de infectados los 7 días siguientes

Una vez obtenidas las proyecciones para el número acumulado de infectados en los próximos (7) días, el objetivo reside en identificar los modelos que mostraron mayor concordancia con los valores reales. Este proceso permitirá identificar que modelos presentan un desempeño sobresaliente al anticipar la evolución temporal de la viruela símica. Para esto, se hará uso de las métricas de precisión como se muestra en la Tabla 4.4.

MODELO	RMSE	MAE	MAPE
Crecimiento Logístico	126,372	125,077	0,034 %
ARNN	100,437	91,507	0,025 %
SARIMA(1,1,0)x(1,0,0)₇	29,919	26,919	0,007 %
Híbrido Log + SARIMA	10,659	7,923	0,002 %
Híbrido SIR + SARIMA	16,925	13,322	0,003 %

Tabla 4.4 Métricas de calidad de ajuste o de precisión de las predicciones



5. Discusión, conclusiones y trabajos futuros

Según los resultados presentados en la Tabla 4.2 se observa que los modelos que mejor se ajustan a los datos acumulados por viruela símica en Colombia son los híbridos SIR + SARIMA y crecimiento logístico + SARIMA; esto era de esperarse ya que, tratándose del crecimiento de infectados, la serie temporal puede presentar componentes lineales y no lineales a la vez. Por su parte, el modelo autorregresivo de redes neuronales (ARNN) y el autorregresivo estacional integrado de medias móviles (SARIMA) presentaron un buen ajuste en el conjunto de entrenamiento; sin embargo, las predicciones generadas por los modelos para el conjunto de prueba no concordaron con las expectativas acerca de la evolución futura de la serie temporal como se muestra en la Tabla 4.4. En particular, aunque el modelo SIR y el de crecimiento logístico presentaron un desempeño razonable, los modelos matemáticos no fueron la mejor opción pues al final del período de estudio el ajuste es muy poco preciso.

Con todo, los modelos discutidos en este trabajo permiten describir razonablemente bien el comportamiento o tendencia de las notificaciones de la viruela símica en Colombia entre el 21 de junio hasta el 31 de octubre de 2022. Se espera entonces que esta investigación sea un punto de partida y sirva de referencia en los estudios de modelación de la dinámica de esta enfermedad en Colombia. Con el objetivo general de modelar la expansión de este virus, este trabajo ofrece herramientas importantes para visualizar el comportamiento de la enfermedad; y en un momento dado, este trabajo puede ser útil a científicos, gobernantes y administradores de salud pública como una base para la toma de buenas decisiones y/o propuestas de control.



Finalmente, basados en el comportamiento gráfico de los distintos modelos ajustados y mediante el uso de las métricas RMSE, MAE y MAPE, se concluye lo siguiente:

- (i) Los modelos ARNN, SIR + SARIMA y crecimiento logístico + SARIMA son los que mejor describen el comportamiento de los datos de la viruela símica en Colombia. Sin embargo, entre ellos son preferibles los modelos híbridos debido al hecho que a partir de este modelo se pueden obtener pronósticos hacia adelante con mejor precisión.
- (ii) A pesar de que gráficamente el modelo de crecimiento logístico presenta un muy buen desempeño, las métricas de precisión no lo favorecen como el mejor modelo ajustado.
- (iii) Aunque el modelo epidemiológico SIR no presenta el mejor desempeño, los estimadores de los parámetros no contradicen la realidad; es decir, la tasa de infección en el período de estudio es alta pero la tasa de recuperación también lo es, y al no registrarse muertes hasta la fecha de estudio, la enfermedad no llega a generar una alarma en el sector de la salud pública.

Como trabajos futuros proponemos la consideración de otros modelos determinísticos como por ejemplo el SIR con parámetros funcionales combinados con modelos ARMA con parámetros variando en el tiempo.



Bibliografía

- Arotolu, T., Afe, A., Wang, H., Lv, J., Shi, K., Huang, L., & Wang, X. (2022). Spatial modeling and ecological suitability of monkeypox disease in Southern Nigeria. *PLOS ONE*, 17(9). Obtenido de <https://doi.org/10.1371/journal.pone.0274325>
- Bisgaard, S., & Kulahci, M. (2011). *Time Series Analysis and Forecasting by Example*. New Jersey: John Wiley & Sons, Inc. ISBN: 978-1-118-05694-3
- Bovas, A., & Ledolter, J. (2006). *Introduction to regression modeling*. Brooks/Cole, Cengage Learning. ISBN-10: 0-534-42075-3
- Box, G., Jenkins, G., & Reinsel, G. (2008). *Time Series Analysis: Forecasting and control (Fourth Edition)*. New Jersey: John Wiley & Sons, Inc. ISBN-10: 0470272848
- Brockwell, P., & Davis, R. (2002). *Introduction to time series and forecasting (Second edition)*. Springer-Verlag NewYork, Inc. ISBN: 0-387-95351-5
- Eid, M., El-Kenawy, E.-S., Khodadadi, N., Mirjalili, S., Khodadadi, E., Abotaleb, M., . . . Khafaga, D. (17 de Octubre de 2022). Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases. *Mathematics* 2022, 10. Obtenido de <https://doi.org/10.3390/math10203845>
- Fuller, T., Thomassen, H., Mulembakani, P., Johnston, S., Lloyd-Smith, J., Lutete, T., . . . Rimoin, A. (2011). Using Remote Sensing to Map the Risk of Human Monkeypox Virus in the Congo Basin . *EcoHealth*, 14-25. Obtenido de <https://doi.org/10.1007/s10393-010-0355-5>
- González Casimiro, M. (2009). *Análisis de series temporales: Modelos ARIMA*. País Vasco: Universidad del País Vasco (UPV-EHU). ISBN 978-84-692-3814-1



UNIVERSIDAD
DEL QUINDÍO



Hurtado T., L., & Salcedo E., G. (Octubre de 1996). *Series temporales con aplicaciones a la epidemiología y la ecología*. Armenia, Colombia: AAS Publicidad Editorial.

Instituto Nacional de salud. (2022). *Casos positivos de Viruela símica en Colombia*. Bogota DC. Recuperado el 5 de Noviembre de 2022, de <https://www.datos.gov.co/es/Salud-y-Proteccion-Social/Casos-positivos-de-Viruela-s-mica-en-Colombia/tmet-yeek>

Khan, A., Sabbar, Y., & Din, A. (Septiembre de 2022). Stochastic modeling of the Monkeypox 2022 epidemic with cross-infection hypothesis in a highly disturbed environment. *Mathematical biosciences and engineering: MBE*, 19(12). doi:10.3934/mbe.2022633

Kutner, M., Nachtsteim, C., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models (Fifth Edition)*. New York, US: McGraw-Hill/Irwin. ISBN-10: 007310874X

Montesinos, O., & Hernández, C. (Julio de 2007). Modelos matemáticos para enfermedades infecciosas. *Salud Pública de México*, 49(3), 218-226. ISSN 0036-3634

Nakazawa, Y., Lash, R., Carroll, D., Damon, I., Kareem, K., Reynolds, M., . . . Peterson, A. (2013). Mapping Monkeypox Transmission Risk through Time and Space in the Congo Basin. *PLOS ONE*, 8(9). Obtenido de <https://doi.org/10.1371/journal.pone.0074816>

Olumuyiwa, J. P., Festus, A. O., Mayowa, M., Abdulmumin, O. O., Rashid, J., & Ilyas, K. (2022). Fractional order mathematical model of monkeypox transmission



UNIVERSIDAD
DEL QUINDÍO



dynamics. *Physica Scripta*, 97(8). Obtenido de <https://doi.org/10.1088/1402-4896/ac7ebc>

Olumuyiwa, J., Kumar, S., Kumari, N., Festus, A., Kayode, O., & Musa, R. (2021).

Transmission dynamics of Monkeypox virus: a mathematical modelling approach.

Modeling Earth Systems and Environment, 8, 3423–3434. Obtenido de

<https://doi.org/10.1007/s40808-021-01313-2>

Organización mundial de la salud. (2022). *Viruela símica [Internet] 2022*. Recuperado el 5

de Noviembre de 2022, de <https://www.who.int/es/news-room/factsheets/detail/monkeypox>

Organización Panamericana de la salud. (2021). *Covid-19 y la importancia de fortalecer los*

sistemas de información. Hoja informativa 13. Obtenido de

<https://iris.paho.org/handle/10665.2/52128>

Suthep, S., Zulqurnain, S., Muhammad, U., & Watcharaporn, C. (2023). Scaled Conjugate

Gradient for the Numerical Simulations of the Mathematical Model-Based

Monkeypox Transmission. *Fractal Fract*, 7(63). Obtenido de

<https://doi.org/10.3390/fractalfract7010063>

Usman, S., & Adamu, I. (2017). Modeling the Transmission Dynamics of the. *Journal of*

Applied Mathematics and Physics, 5, 2335-2353. ISSN: 2327-4379

Wei, W. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. Addison-

Wesley Publishing Company, Inc. ISBN 0-201-15911-2

Zill, D. (2008). *A first course in differential equations with modeling applications* (Ninth

Edition ed.). US: Brooks/Cole, Cengage Learning. ISBN-13 : 978-0495108245



UNIVERSIDAD
DEL QUINDÍO



LICENCIATURA EN MATEMÁTICAS

Tel: (57) 6 735 9300 Ext
Carrera 15 Calle 12 Norte
Armenia, Quindío – Colombia
licenciaturaenmatematicas@uniquindio.edu.co

PERTINENTE CREATIVA INTEGRADORA



@uniquindio



uniquindioconectada



uniquindioconectada