



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **Análisis de estrategias de aprendizaje del modelo SIR (Susceptible – Infectado - Recuperado) para el pronóstico del Covid-19 en Colombia**

**Jasmín Alejandra Marín Pérez**

Universidad Nacional de Colombia  
Facultad de Minas  
Área Curricular de Sistemas e Informática  
Medellín, Colombia  
2022

# **Análisis de estrategias de aprendizaje del modelo SIR (Susceptible – Infectado - Recuperado) para el pronóstico del Covid-19 en Colombia**

**Jasmin Alejandra Marín Pérez**

Tesis de Maestría presentado como requisito parcial para optar al título de:  
**Magister en Ingeniería - Analítica**

Director:

Juan David Velásquez Henao, MSc, PhD

Línea de Investigación:

Analítica predictiva

Grupo de Investigación:

Big Data & Data Analytics

Universidad Nacional de Colombia

Facultad de Minas, Área Curricular de Sistemas e Informática

Medellín, Colombia

2022

*(Dedicatoria o lema)*

*A Dios por permitirme cumplir esta meta y por regalarme sabiduría, capacidad y entendimiento.*

*A mi madre por todo el apoyo brindado y su incondicionalidad en cada momento de mi vida.*

*A mi asesor Juan David Velásquez, por su enseñanza, compromiso y acompañamiento en este proceso.*

## Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



Jasmín Marín P

Nombre: Jasmín Alejandra Marín Pérez

Fecha 10/01/2023

# Resumen

En esta tesis de maestría se presenta un modelo para pronosticar la enfermedad infecciosa Covid-19 a nivel nacional y departamental por medio del modelo SIR (Susceptible – Infectado - Recuperado) incorporando la información del desarrollo de la enfermedad para el caso colombiano; para ello, se analizaron diversos escenarios de la contingencia del Covid-19 y se contrasta o compara el modelo con los resultados de otros trabajos de investigación frente al mismo universo y al mismo periodo de evaluación. El modelo propuesto es el SSE-SIR que relaciona el modelo SIR con el Suavizado Exponencial Simple, en el que presenta las características de la historia natural de la enfermedad y permite auto-adaptarse a nuevos cambios en la dinámica del sistema como la apertura y cierre de fronteras, cuarentenas generales o parciales o programas de vacunación escalados. Así mismo, se analiza un segundo modelo que parte de la Regresión Lineal y de la aplicación de un modelo de combinación de pronósticos entre la Regresión Lineal y el modelo adaptado SSE-SIR; como resultado, el modelo adaptado presenta resultados más favorables que aquellos publicados en los casos de comparación. Además, los datos de contagios predichos presentan mayor acierto con los valores reales de la pandemia. Se identificó, también desaciertos en el modelo, cuando se evalúan periodos críticos o picos de contagios y cuando el periodo de evaluación para la predicción es amplio. Finalmente, se concluye que, el modelo adaptado SSE-SIR se adapta al comportamiento y a la dinámica de Colombia, arrojando mejores resultados en periodos de alcance corto y estimaciones optimas frente a la desagregación territorial.

**Palabras claves:** Enfermedad Infecciosa, Covid-19, Suavizado Exponencial Simple, Modelo SIR, Regresión Lineal, predicción

# Analysis of learning strategies of the SIR model (Susceptible – Infected – Recovered) for the prognosis of Covid-19 in Colombia

## ABSTRACT

This master's thesis presents a model to predict the infectious disease Covid-19 at the national and departmental level through the SIR model (susceptible - infected - recovered) incorporating the information of the development of the disease for the Colombian case; To do this, different scenarios of the Covid-19 contingency were analyzed and the model is compared with the results of other research works against the same universe and the same evaluation period. The proposed model is the SSE-SIR that relates the SIR model with the Simple Exponential Smoothing, in which it presents the characteristics of the natural history of the disease and allows self-adaptation to new changes in the dynamics of the system such as the opening and closing of borders, general or partial quarantines or scaled vaccination programs. Likewise, a second model is analyzed that starts from the linear regression and the application of a model of combination of forecasts between the linear regression and the adapted SSE-SIR model; as a result, the adapted model presents more favorable results than those published in the comparison cases. In addition, the predicted contagion data are more accurate with the actual values of the pandemic. We also identified failures in the model when critical periods or peaks of contagion are evaluated and when the evaluation period for prediction is large. Finally, it is concluded that the adapted SSE-SIR model adapts to the behavior and dynamics of Colombia, yielding better results in short-range periods and optimal estimates against territorial disaggregation.

**Keywords:** Infectious disease, Covid-19, Simple Exponential Anti-Aliasing, SIR, Prediction, Linear regression

## Contenido

1.1	INTRODUCCIÓN	15
1.2	FORMULACIÓN DEL PROBLEMA	15
1.2.1	<i>Definición del problema real</i>	16
1.2.2	<i>Definición del problema desde el punto de vista analítica</i>	17
1.2.3	<i>Definición del problema de investigación</i>	18
1.3	HIPÓTESIS	18
1.4	OBJETIVOS	19
1.4.1	<i>Objetivo General</i>	19
1.4.2	<i>Objetivos Específicos</i>	19
1.5	MAPA DEL DOCUMENTO	19
<b>2.</b>	<b>REVISIÓN DE LITERATURA</b>	<b>20</b>
2.1	INTRODUCCIÓN	20
2.2	METODOLOGÍA	22
2.3	RESULTADOS OBTENIDOS	23
2.4	RESULTADOS GENERALES	23
2.4.1	<i>Análisis temático por palabra claves del autor</i>	29
2.5	DISCUSIÓN	33
<b>3.</b>	<b>INCORPORACIÓN DE MECANISMO ADAPTATIVO EN EL MODELO SIR</b>	<b>36</b>
3.1	INTRODUCCIÓN	36
3.2	MODELO PROPUESTO	37
3.2.1	<i>Modelo SIR</i>	37
3.2.2	<i>Suavizado exponencial</i>	38
3.2.3	<i>Método propuesto SSE-SIR</i>	40
3.3	EXPERIMENTACIÓN, ANÁLISIS Y RESULTADOS	41
3.3.1	<i>Caso 1</i>	41
3.3.2	<i>Caso 2</i>	43
3.3.3	<i>Caso 3</i>	47
3.3.4	<i>Caso 4</i>	48
3.4	CONCLUSIONES	49
<b>4.</b>	<b>PREDICCIÓN INCORPORANDO INFORMACIÓN DE OTROS PAÍSES</b>	<b>50</b>
4.1	INTRODUCCIÓN	50
4.2	MODELO PROPUESTO	50
4.2.1	<i>Regresión lineal</i>	50
4.3	SELECCIÓN DE VARIABLES RELEVANTES	51
4.4	INFORMACIÓN UTILIZADA	52
4.5	PREPROCESAMIENTO DE LOS DATOS	52
4.6	EXPERIMENTACIÓN, ANÁLISIS Y RESULTADOS	53
4.6.1	<i>Caso 1</i>	56
4.6.2	<i>Caso 2</i>	57
4.6.3	<i>Caso 3</i>	59
4.6.4	<i>Caso 4</i>	60

4.7	CONCLUSIONES	62
<b>5.</b>	<b>PRONÓSTICO DE LA EVOLUCIÓN DEL COVID-19 A NIVEL NACIONAL Y DEPARTAMENTAL</b>	<b>63</b>
5.1	INTRODUCCIÓN	63
5.2	MODELO APLICADO	63
5.3	ANÁLISIS Y RESULTADOS	63
5.3.1	<i>Colombia</i>	64
5.3.2	<i>Departamento</i>	68
5.4	CONCLUSIONES	69
<b>6.</b>	<b>CONCLUSIONES</b>	<b>71</b>
<b>6.1</b>	<b>CUMPLIMIENTO DE OBJETIVOS</b>	<b>72</b>
6.1.1	<i>Objetivo Específico 1</i>	72
6.1.2	<i>Objetivo Específico 2</i>	72
6.1.3	<i>Objetivo Específico 3</i>	73
6.1.4	<i>Objetivo Específico 4</i>	73
<b>7.</b>	<b>REFERENCIAS</b>	<b>74</b>



## Lista de tablas

<b>Tabla 1.</b> Representación de las fuentes de revistas, numero de documentos publicados y total de citasiones por año. ....	25
<b>Tabla 2.</b> Representación del total de documentos publicados por autor y año. ....	27
<b>Tabla 3.</b> Representación de autores por documentos publicados y total de citasiones...	28
<b>Tabla 4.</b> Impacto de las publicaciones por cantidad de autores.....	29
<b>Tabla 5.</b> Representación de las palabras claves por Clúster. ....	33
<b>Tabla 6.</b> Relación modelo SIR adaptado Suavizado exponencial.....	42
<b>Tabla 7.</b> Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 1.....	43
<b>Tabla 8.</b> Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 2, Modelo para el periodo entre Marzo y abril del 2020. ....	44
<b>Tabla 9.</b> Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 2, Modelo para el periodo entre Marzo y Julio del 2020. ....	45
<b>Tabla 10.</b> Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 2, Modelo para el periodo entre Marzo y Agosto del 2020.....	46
<b>Tabla 11.</b> Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 3.....	47
<b>Tabla 12.</b> Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 4.....	48
<b>Tabla 13.</b> Relación modelo SSE-SIR y Regresión Lineal .....	56
<b>Tabla 14.</b> Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 1. ....	57
<b>Tabla 15.</b> Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 2. Modelo para el periodo entre Marzo y abril del 2020. ....	58
<b>Tabla 16.</b> Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 2. Modelo para el periodo entre Marzo y julio del 2020.....	58
<b>Tabla 17.</b> Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 2. Modelo para el periodo entre Marzo y Agosto del 2020. ....	59
<b>Tabla 18.</b> Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 3. ....	59
<b>Tabla 19.</b> Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 4. ....	61
<b>Tabla 20.</b> Resultados por estimador $R^2$ o coeficiente de determinación para cinco departamentos.....	69

## Lista de figuras

<b>Figura 1.</b> Total de documentos publicados por año. ....	24
<b>Figura 2.</b> Representación de total de publicaciones por país. ....	26
Figura 3. Visualización de las palabras claves por clusters. ....	29
Figura 4. Representación del modelo SIR adaptado Suavizado exponencial caso 1. ....	42
<b>Figura 5.</b> Representación del modelo SIR adaptado Suavizado exponencial caso 2, Modelo para el periodo entre Marzo y abril del 2020. ....	44
<b>Figura 6.</b> Representación del modelo SIR adaptado Suavizado exponencial caso 2, Modelo para el periodo entre Marzo y Julio del 2020. ....	45
<b>Figura 7.</b> Representación del modelo SIR adaptado Suavizado exponencial caso 2, Modelo para el periodo entre Marzo y Agosto del 2020. ....	46
<b>Figura 8.</b> Representación del modelo SIR adaptado Suavizado exponencial caso 3. ....	47
<b>Figura 9.</b> Representación del modelo SIR adaptado Suavizado exponencial caso 4. ....	48
<b>Figura 10.</b> Representación trimestre II del año 2020 de la predicción Colombia a nivel global. ....	53
<b>Figura 11.</b> Representación trimestre III del año 2020 de la predicción Colombia a nivel global. ....	54
<b>Figura 12.</b> Representación trimestre I del año 2021 de la predicción Colombia a nivel global. ....	54
<b>Figura 13.</b> Representación trimestre II del año 2021 de la predicción Colombia a nivel global. ....	55
<b>Figura 14.</b> Representación trimestre IV del año 2021 de la predicción Colombia a nivel global. ....	55
<b>Figura 15.</b> Representación de la clase Infectados del modelo SSE-SIR y Regresión Lineal, Caso 1. ....	57
<b>Figura 16.</b> Representación de la clase Infectados del modelo SSE-SIR y Regresión Lineal, Caso 3. ....	60
<b>Figura 17.</b> Representación de la clase Infectados del modelo SSE-SIR y Regresión Lineal, Caso 4. ....	61
<b>Figura 18.</b> Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio inicial. ....	64
<b>Figura 19.</b> Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio II. ....	65
<b>Figura 20.</b> Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio III. ....	66
<b>Figura 21.</b> Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio IV. ....	67
<b>Figura 22.</b> Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio V. ....	67

# Lista de Símbolos y abreviaturas

## Abreviaturas

<b>Abreviatura</b>	<b>Término</b>
CRISP-DM	Cross Industry Standard Process for Data Mining
SIR	Susceptible - Infectado - Recuperado
NRMSE	Error Cuadrático Medio Normalizado
R2	Coefficiente de Determinación
RMSE	Error Cuadrático Medio
CV	Validación Cruzada o Cross Validation

## 1.1 Introducción

La pandemia del COVID-19 ha sacudido al mundo entero, provocando impactos significativos en la dinámica política, social y económica de más de 190 países. En respuesta a esta crisis sin precedentes, se han implementado diversas medidas drásticas a nivel global para mitigar los efectos del virus. En este contexto, la capacidad de predecir y pronosticar se ha vuelto fundamental para anticipar riesgos y tomar decisiones informadas.

La predicción basada en datos actuales e históricos ha demostrado ser una herramienta invaluable en la lucha contra el COVID-19, especialmente en el ámbito de la salud pública. Al aprovechar técnicas y modelos de pronóstico, se han logrado tomar medidas de mitigación, promoción y prevención en diferentes territorios, lo que ha ayudado a reducir el impacto del virus y proporcionar una atención más eficiente a los afectados.

Uno de los aspectos claves que se ha destacado es la necesidad de analizar los datos desagregados a nivel territorial. Cada país y departamento ha experimentado la llegada del virus en momentos distintos, lo que ha generado la importancia de realizar análisis específicos y adaptados a las características locales. Este enfoque ha permitido obtener una comprensión más precisa de la evolución de la enfermedad, identificar patrones y tendencias específicas, y diseñar estrategias efectivas para hacer frente a la pandemia.

En este contexto, el análisis y pronóstico del COVID-19 a nivel global y territorial se ha convertido en herramientas esenciales para mitigar riesgos y tomar medidas oportunas. La combinación de datos actuales e históricos, así como el uso de técnicas cuantitativas y cualitativas, incluyendo modelos de aprendizaje de máquina, ha permitido anticipar la propagación del virus, identificar posibles escenarios futuros y diseñar estrategias adecuadas para abordar la situación.

Con el objetivo de contribuir a este campo de estudio, este trabajo de tesis de maestría se enfoca en el desarrollo de un modelo de aprendizaje de máquina para pronosticar la tasa de infección del COVID-19 en Colombia. Para lograrlo, se utiliza un modelo basado en el concepto SIR (Susceptible - Infectado - Recuperado), adaptado y ajustado utilizando datos históricos. Este enfoque permitirá obtener información valiosa sobre la evolución de la enfermedad y validar la efectividad del modelo en la predicción de contagios en el contexto colombiano.

## 1.2 Formulación del Problema

En este Capítulo se presenta la formulación del problema de investigación de la Tesis de Maestría titulada “Análisis de estrategias de aprendizaje del modelo SIR (Susceptible – Infectado - Recuperado) para el pronóstico del Covid-19 en Colombia”. La formulación presentada en este Capítulo se basa en la metodología Cross Industry Standard Process for Data Mining (CRISP-DM), adicionando los elementos necesarios que son requeridos para presentar una propuesta de investigación.

## 1.2.1 Definición del problema real

A finales del año 2019 en la capital Wuhan de la provincia de Hubei en China Central, comenzó el brote del Coronavirus 2019 (COVID -19), enfermedad declarada por la OMS (Organización Mundial de la Salud) como una emergencia de Salud Pública de preocupación internacional, extendiéndose a un total de 188 países, que corresponde al 97% de los países para el año 2020. Así mismo y de acuerdo a las cifras de la Universidad de John Hopkins, hasta el 08 de diciembre del año 2022, se han confirmado un total de 647.697.949 casos y más de 6.648.906 muertes. A lo largo de la historia, muchas han sido las pandemias que han dejado millones de contagios y muertes como la Viruela, el Sarampión, el VIH, la gripe A, el Síndrome Respiratorio Agudo Grave (SARS), entre otras (Christophersen, 1989).

El COVID 19 pertenece a la categoría taxonómica de los Coronaviridae, que es una familia de virus de más de doce patógenos. El cuadro clínico implica fiebre, escalofríos, tos, fatiga y una insuficiencia respiratoria como disnea, hipoxia y neumonía, y la gravedad o severidad de la enfermedad depende del tipo de paciente, lo que puede provocar Sepsis y shock séptico. Es considerada una enfermedad aguda y a la vez presenta una tasa de letalidad del 4,4% (Xu, et al., 2020).

Los primeros casos en América Latina y el Caribe aparece en Brasil el 25 de febrero de 2020 y durante ese año, dicho país ocupó el segundo lugar con mayor número de casos a nivel mundial; para el año 2022, la clasificación la encabeza Estados Unidos con mayor número de casos positivos confirmados, siguiéndole, India, Francia, Alemania y Brasil.

El primer caso en Colombia se presenta en la primera semana de marzo; a partir de los primeros casos, se abre camino a medidas urgentes cada vez más severas en el territorio – como en todos los países – considerando inicialmente precauciones en todos los puntos de entrada, campañas de concientización sobre esta enfermedad, medidas en materia agropecuaria y en materia de ingresos y campañas para identificar síntomas. Además, desde las primeras semanas de marzo se llevaron a cabo acciones a nivel nacional como la adopción de medidas preventivas y sanitarias de aislamiento y cuarentena, la cancelación de todos los eventos públicos, la suspensión del tránsito y desembarco de cruceros, el aislamiento preventivo de ciudadanos extranjeros en hoteles, habilitación de centros para adultos mayores y para grupos étnicos; igualmente, se realizó la suspensión temporal de visita a los sitios de reclusión, la suspensión de clases presenciales, el acceso a la canasta nutricional para algunas poblaciones, el cierre de pasos marítimos, terrestres y fluviales de fronteras con Panamá, Ecuador, Perú y Brasil, el cierre temporal y el ingreso de visitantes a los parques arqueológicos. En fin, estas han sido algunas de las medidas preventivas nacionales estipuladas para la disminución del contagio de la enfermedad. En cuanto a las medidas de aislamiento, Colombia entró desde el 17 de marzo de 2020 en estado de emergencia con directrices para toda la población, en especial para los adultos mayores; también se dieron medidas de protección al empleo, aumento del presupuesto al sector salud y modificaciones transitorias al comportamiento político y económico del país (Estado Colombiano–GOV.CO, 2023).

A junio del 2022, Colombia se posiciona entre los primeros diez (10) países del mundo con más decesos a causa de la enfermedad; las zonas de Colombia que más infecciones acumulan son Bogotá, Antioquia y Valle del Cauca, según lo revelaron datos de la Universidad de Johns Hopkins (infobae, 2021).

En este contexto, se han realizado durante la pandemia diversos ejercicios enfocados en pronosticar valores futuros de una serie de tiempo, al igual que el uso de modelos matemáticos epidemiológicos, aun así, no se han determinado ejercicios de pronóstico a partir del uso de métodos clásicos epidemiológicos, generando niveles de incertidumbre bajos, que permitan reacciones inmediatas y decisiones en Salud Pública, con el fin de implementar medidas de mitigación y generar acciones de promoción y prevención en el territorio. Así mismo, adelantarse en este fenómeno permite tomar decisiones en cuanto al Sistema de Salud Colombiano como acciones para reducir el impacto y brindar de forma eficiente atención a los contagiados, identificar requerimientos en camas hospitalarias y unidades de cuidado intensivo (UCI), definir rutas de remisión y traslados, identificar inexistencia o posible alerta en zonas del territorio sin capacidad instalada, entre otras.

### **1.2.2 Definición del problema desde el punto de vista analítica**

El problema que se aborda en esta tesis de maestría, desde el punto de vista de analítica se clasifica como un problema de analítica predictiva, en el que se desea aplicar técnicas de predicción al comportamiento del Covid-19, teniendo como base métodos clásicos epidemiológicos y otras técnicas integradas.

En la actualidad, a partir de las revisiones identificadas y de la exposición de los datos de esta enfermedad, se hallan diversas aplicaciones desde modelos matemáticos de epidemias como lo son los modelos SIR (Susceptibles- Infectados- Recuperados), SIRS (Susceptibles- Infectados- Recuperados-Susceptibles), SIS (Susceptibles- Infectados- Susceptibles), SEIS (Susceptibles- Expuestos- Infectados- Susceptibles), SEIR (Susceptibles- Expuestos- Infectados- Recuperados), MSIR (Infantes con inmunidad- Susceptibles- Infectados- Recuperados) y MSEIR (Infantes con inmunidad- Susceptibles- Expuestos- Infectados- Recuperados), como la aplicación de modelos de series de tiempo e Inteligencia Artificial para el pronóstico de nuevos casos (Christophersen, 1989).

A partir de lo anterior, este trabajo identificará dos enfoques adaptados desde un modelo clásico epidemiológico como modelos predictivos, para pronosticar la evolución de la enfermedad y verificar la validez a partir de los datos reales de la pandemia y contrastando los resultados a partir de estudios investigativos con objetivos similares para el territorio Colombiano. Los modelos que se relacionan en el trabajo incorporan, a diferencia de otras predicciones, la evolución de la enfermedad, como el cambio de las tasas de infección y de recuperación.

### **1.2.3 Definición del problema de investigación**

Consecuentemente con lo anterior, esta tesis de maestría responde la siguiente las siguientes preguntas de investigación:

- ¿Un modelo adaptado de pronóstico basado en un modelo matemático epidemiológico puede adaptarse a la dinámica del desarrollo de la enfermedad infecto-contagiosa (COVID -19) para estimar la tasa de infección y la tasa de recuperación en Colombia y por departamentos?
- ¿Un modelo adaptado de pronóstico basado en la tasa de contagio de los casos detectados en otros países, puede adaptarse a la dinámica del desarrollo de la enfermedad COVID -19 para estimar el comportamiento en Colombia?

Estas preguntas de investigación permitirán detectar las áreas de desarrollo a futuro, y corroborar el papel de la ciencia de datos y la analítica frente a otras áreas, sobre todo con el área de la salud. Fortaleciendo las técnicas epidemiológicas existentes como es el modelo SIR y aprovechando las bondades, asociándolas a la aplicación de modelos de previsión, de esta forma se adaptan dinámicamente las tasas de infección y recuperación de la enfermedad con el fin de reflejar los cambios relacionados con las intervenciones del sistema, con las medidas de mitigación y de contención.

## **1.3 Hipótesis**

La hipótesis de esta tesis de maestría es la siguiente:

Es posible desarrollar un modelo matemático epidemiológico que se adapte a la dinámica del desarrollo de la enfermedad infecto-contagiosa (COVID -19) para estimar la tasa de infección y de recuperación en Colombia y que tenga en cuenta las tasas de contagio de los casos detectados en otros países pronostica el comportamiento en Colombia.

## **1.4 Objetivos**

### **1.4.1 Objetivo General**

Pronosticar la enfermedad infecciosa Covid-19 a nivel nacional y departamental por medio del modelo SIR (Susceptible – Infectado - Recuperado) incorporando la información del desarrollo de la enfermedad para el caso colombiano.

### **1.4.2 Objetivos Específicos**

Esta tesis de maestría tiene los siguientes objetivos específicos, con el fin de poder cumplir con el objetivo general propuesto:

1. Ajustar el modelo estadístico epidemiológico SIR (Susceptible – Infectado - Recuperado) para que se adapte al comportamiento de la Historia Natural de Enfermedad.
2. Proponer un mecanismo que permita ajustar la cantidad total de casos detectados, en donde se incorpore la información del desarrollo de la enfermedad infecto-contagiosa (COVID -19) de otros países por medio de las tasas de contagio.
3. Implementar el modelo al caso de Colombia con el fin de pronosticar la tasa semanal e identificar el nivel de acierto y desacierto de los resultados.
4. Adaptar el modelo a nivel departamental con el fin de pronosticar la tasa semanal y encontrar persistencias, disminución y aumento del número de contagios.

## **1.5 Mapa del documento**

El resto de este documento está organizado de la siguiente manera: En el Capítulo 2 se realiza una revisión de la literatura relacionada con los conceptos abordados en este trabajo. En el Capítulo 3 se presenta una propuesta de adaptación del modelo SIR, que tiene en cuenta las tasas de infección y recuperación específicas de Colombia, y se compara con técnicas de predicción existentes. El Capítulo 4 evalúa una propuesta para pronosticar la cantidad de infectados en Colombia, basándose en el desarrollo de la enfermedad en otros países. En el Capítulo 5 se analizan los resultados obtenidos con el modelo presentado en el Capítulo 3 a nivel de departamentos, enfocándose en los cinco principales departamentos de Colombia, y se evalúa la calidad del ajuste del modelo utilizando medidas estadísticas. Por último, en el Capítulo 6 se presentan las conclusiones obtenidas a partir de la aplicación del modelo y los resultados obtenidos.



## 2. Revisión de literatura

En este Capítulo se presenta la caracterización bibliográfica de la literatura seleccionada para realizar el análisis bibliométrico desarrollado en esta tesis de maestría.

### 2.1 Introducción

Las enfermedades que han surgido en los últimos años han dado lugar a un papel importante para las enfermedades infecciosas o contagiosas; es por eso que los modelos matemáticos se han popularizado para analizar la propagación y el control de las enfermedades.

En general, los modelos proporcionan variables y parámetros, resultados conceptuales como umbrales, números básicos de reproducción, número de contacto y números de reemplazo (Christophersen, 1989). Hethcote (2000) indica que estos modelos presentan las siguientes ventajas:

- Son instrumentos experimentales útiles para probar teorías.
- Evalúan conjeturas cuantitativas.
- Responden preguntas.
- Estiman parámetros claves.
- Permiten la comprensión de características de transmisión de las enfermedades en las comunidades, regiones y países.
- Posibilitan la disminución de la transmisión de las enfermedades.
- Contribuyen al diseño y análisis de las encuestas epidemiológicas.
- Identifican tendencias.
- Estiman la incertidumbre de los pronósticos.

Históricamente, el uso de modelo matemáticos epidemiológicos inician en 1760, con el matemático Daniel Bernoulli que presenta por primera vez el uso de un modelo para estudiar la difusión de una enfermedad infecciosa, esto se realiza para la técnica de inoculación contra la viruela que azotaba a la población. Bernoulli, identifica la ecuación diferencial y evalúa la efectividad de la técnica de inoculación, dando lugar al modelado de epidemiología determinista que inicia en el siglo XX.

Luego, en 1906, Hamer formuló y analizó el primer modelo de tiempo discreto para comprender el comportamiento de las epidemias de Sarampión teniendo en cuenta la incidencia como el producto de las densidades de los susceptibles y los infectados, indicando que, la tasa en que una enfermedad se propaga es proporcional al número de individuos susceptibles por el número de individuos infecciosos (Christophersen, 1989).

Luego, se reconocen los trabajos de Kermack y McKendrick realizados durante el periodo de 1927 a 1939, en el que desarrollaron el famoso teorema del Umbral. En este teorema, la introducción de individuos infecciosos dentro de una población de susceptibles puede originar una epidemia, sólo si la densidad de los susceptibles rebasa el umbral. Aunque

este trabajo sólo se toma veinte (20) años más tarde ha tenido influencia para el desarrollo posterior de una gran variedad de modelos matemáticos determinantes que identifican la densidad de los susceptibles (Xu, et al., 2020).

Los modelos matemáticos, actualmente son usados para el estudio de problemas en medicina, biología, fisiología, epidemiología, entre otras áreas, para describir y explicar fenómenos, sin embargo, su uso es limitado. Un modelo matemático epidemiológico permite revelar relaciones que no son obvias y que se encuentran ocultas, como la inmunidad pasiva, la pérdida gradual de la vacuna, la inmunidad adquirida por la enfermedad, las etapas de infección, los vectores de la enfermedad, la estructura de la edad, la vacunación y la cuarentena, entre otros.

Los modelos etiquetados M, S, E, I y R se han utilizado a menudo para las clases epidemiológicas. La ruta de los modelos parte de la etiqueta M, que hace alusión a los bebés con inmunidad pasiva temporal a una infección. Cuando los anticuerpos maternos desaparecen del cuerpo del bebé, este pasa a la etiqueta S de susceptibles, es decir, los que pueden infectarse. Luego, se encuentran los E (Expuestos) que son aquellos susceptibles que entran en contacto con un I (Infeccioso) para que se produzca una transmisión. Esta etiqueta (I) dura el periodo de latencia de la enfermedad, es decir, están infectados, pero aún no son infecciosos. Después de que el periodo de latencia finalice, el individuo entra a la clase I de Infecciosos, que son capaces de transmitir una infección. Finalmente, cuando el periodo de Infección finaliza a partir del tratamiento, cuidados y controles, el individuo entra en la clase R de Recuperados que son aquellos con inmunidad permanente adquirida por la infección (Christophersen, 1989).

Todo lo anterior depende directamente de la enfermedad concreta que se modela y del propósito del modelo. Los acrónimos de los modelos de epidemias se basan a menudo en los patrones como MSEIR, MSEIRS, SEIR, SEIRS, SIR, SIRS, SEI, SEIS, SI y SIS.

Ahora bien, en el siglo XVII, se inicia la toma de decisiones basadas en pronósticos, dando lugar a técnicas como los análisis de regresiones, descomposición, suavizamiento y promedio móviles. Las diversas técnicas de elaboración de pronósticos parten de la identificación y comprensión de patrones históricos en los datos y su uso se resalta inicialmente para los pronósticos financieros, de energía, industrial, mientras el sector salud, es uno de los últimos sectores en el uso y la incursión de estas técnicas, sobre todo, se evidencia durante el tiempo de contingencia de la pandemia (Montgomery et al., 2008).

A partir de lo anterior, se identifica la necesidad de realizar una caracterización bibliográfica de la literatura identificando principales modelos matemáticos y epidemiológicos usados, los tipos de enfermedades a las que se le aplican estos modelos, adaptación de los modelos a la evolución de las enfermedades infecciosas y medidas e impacto en la propagación y transmisión de las enfermedades.

## 2.2 Metodología

En esta Capítulo se utiliza la metodología de Revisión Sistemática de Literatura (RSL). La especificación de la metodología empleada es la siguiente:

- Se utilizó la Base de Datos Bibliográfica Scopus.
- La búsqueda fue realizada con la siguiente cadena:  
( TITLE-ABS-KEY ( model OR "epidemiological model" ) AND TITLE-ABS-KEY ( mseir OR mseirs OR seir OR sir OR sirs OR sei OR seis OR si OR sis ) AND TITLE-ABS-KEY ( epidemic OR pandemic OR endemic ) AND TITLE-ABS-KEY ( behavior OR behaviour OR performance OR demeanor OR bearing ) AND TITLE-ABS-KEY ( forecast OR predicting OR forecasting OR projection OR prevision OR adapt OR adjust OR adjusted OR tailor OR customize OR gear OR adjust OR set OR tune OR trim OR revise OR redress ) ) AND ( EXCLUDE ( SUBJAREA , "PHYS" ) OR EXCLUDE ( SUBJAREA , "SOCI" ) OR EXCLUDE ( SUBJAREA , "BUSI" ) OR EXCLUDE ( SUBJAREA , "EART" ) )
- Se restringió la búsqueda a las siguientes áreas de conocimiento: Salud pública, la epidemiología y las ciencias de la computación.
- Se incluyen predicción o pronóstico de los contactos de la enfermedad infecto contagiosa.
- No se establece límite del periodo de tiempo de publicación.
- Se incluye el reconocimiento de la epidemiología a partir del área geográfica como pandemia, epidemia y endemia.
- Se incluyen los modelos matemáticos ajustados con parámetros poblacionales y a partir de la taxonomía de la enfermedad.
- Se aplicaron los siguientes criterios de exclusión: 1) Áreas de conocimiento que no eran relacionadas con las técnicas computacionales, de salud o epidemiología, como las áreas de negocios, Ciencias sociales, Energía, Ingeniería Química, Astronomía, entre otros; 2) Publicaciones con alcances de prevención o de medidas preventivas, debido a que no corresponden a la finalidad de la siguiente tesis de maestría y 3) Publicaciones alrededor de enfermedades crónicas.
- Se recolectó la siguiente información bibliográfica: Autores, ID por autor, Título de la publicación, año, volumen, número del artículo, tiempo de citación, link de publicación, afiliación, autor con afiliación, resumen, palabras claves por autor, referencias, tipo de documento, número de autores, país de publicación, país de procedencia del primer autor, institución o instituciones, institución del primer autor, palabras claves, resumen, entre otros.

## 2.3 Resultados obtenidos

Al aplicar la cadena de búsqueda, se recuperaron 244 artículos de forma automática. Los artículos recuperados cumplen con los criterios.

El archivo descargado de Scopus fue procesado, generando lo siguiente:

- Remoción de acentos.
- Formateo de nombre de autores.
- Conteo de la cantidad de autores por documento.
- Conteo de la cantidad de documentos por autor.
- Desambiguación de nombres de los autores.
- Extracción de los nombres de los países.
- Extracción de países de afiliación.
- Extracción del país del primer autor nombrado.
- Extracción del primer autor.
- Extracción de las instituciones del primer autor.
- Unificación de los países.
- Transformación de las palabras claves del índice a minúsculas.
- Extracción de las palabras del resumen o abstract.
- Extracción de las palabras del título.

Las palabras claves fueron sometidas a un proceso de limpieza que consiste en agrupar términos que tienen el mismo significado para los campos "Author Keywords" e "Index Keywords", con el fin de unificar términos para poder realizar el análisis de los conceptos asociados.

## 2.4 Resultados Generales

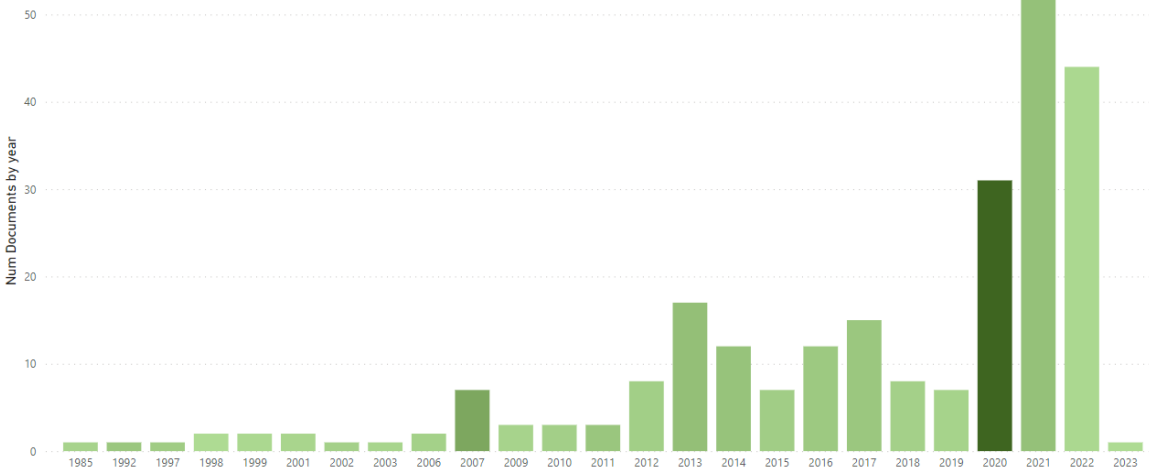
La búsqueda de artículos en el motor de búsqueda arrojó un total de 244 documentos publicados entre 1985 y 2023 por 916 autores; así mismo, el tipo de publicación en un 80.33% corresponde a los Artículos, el 14.75% a las conferencias y el 4.92% restante a los Capítulos de libros y revisiones.

El crecimiento del número de publicaciones por año se presenta sobre todo a partir del año 2012, lo que indica que progresivamente se ha prestado mayor atención o se tiene mayor conocimiento para el abordaje de estos temas de investigación. Antes del año 2012, el porcentaje de las publicaciones era del 11.89% de los artículos de revista aquí nombrados, entre estos se destacan las publicaciones de los años 1985 y 1992 que señalan un tiempo crucial en las enfermedades infecciosas como la malaria o también denominada el paludismo. Igualmente, se identificaron en los años 1992-1997 publicaciones que hacen alusión a las epidemias sostenidas del dengue en algunas zonas y a nivel mundial, las enfermedades infantiles como el sarampión, las paperas y la varicela (Newton & Reiter, 1992; Schwartz, 1992).

En la Figura 1 se muestra la evolución del número total de documentos publicados por año. Se observa un incremento constante en la cantidad de artículos de revista publicados desde 2007 hasta 2022, lo que refleja la creciente relevancia de la investigación en el campo de las enfermedades contagiosas o infecciosas.

En los años 2020, 2021 y 2022 se reporta el mayor número de publicaciones (128 publicaciones), siendo estas las más citadas en la actualidad. Además, el 74.21% de las publicaciones en estos años se relacionan con la pandemia Covid-19.

Ahora bien, se destacan las publicaciones de los años 1992, 2007, 2011, 2013 y 2014 con un mayor número de citas, así como las publicaciones del 2020. Es necesario resaltar el año 1992, que a pesar de contar con una publicación elaborada por Reiter & Newton fue citada por 229 publicaciones.



**Figura 1.** Total de documentos publicados por año.

Igualmente es relevante identificar las revistas de publicación que presentan el mayor número de artículos publicados, resaltando:

- “PLoS ONE” y “Epidemics” con nueve y ocho publicaciones, respectivamente.
- “Journal of Mathematical Biology” con siete publicaciones.
- “Mathematical Biosciences”, “Nonlinear Dynamics” y “Lecture Notes in Computer Science” con cinco (5) publicaciones cada una.
- “IEEE Access” y “Scientific Reports” con cuatro (4) publicaciones cada una.

En la Tabla 1 se muestran las diferentes fuentes de revistas, destacando aquellas con mayor número de documentos publicados. Se incluyen un total de 172 revistas fuentes, y se observa que la revista más citada es "PLoS ONE" con nueve publicaciones y 330 citas. Sin embargo, es importante destacar que el número de publicaciones en estas revistas no es constante a lo largo del tiempo. Esto también se aplica a la revista "Epidemics".

Indistintamente, se destacan revistas que solo publicaron en un año como por ejemplo “American Journal of Tropical Medicine and Hygiene”, “Nonlinear Analysis: Real World Applications”, “Journal of Differential Equations”, “PLoS Medicine”, “Advances in Applied Probability”, entre otros.

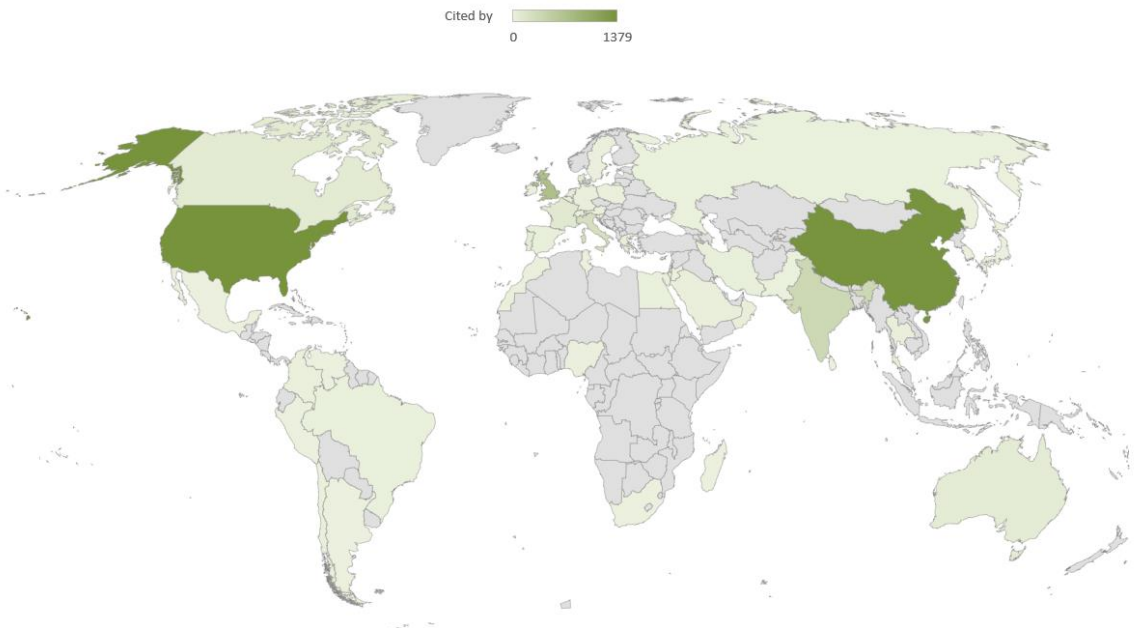
Las revistas más recientes en publicaciones de este tipo de temas son: “PLoS ONE” que ha empezado a publicar artículos desde el 2011 y su última publicación es del año 2022, además es uno de los más citados, lo que demuestra un nivel de conocimiento y de interés específico sobre este tipo de temas. Así mismo, se observa la fuente “Infectious Disease Modeling” que presentó dos publicaciones en los años 2018 y 2020 y aunque el enfoque de la revista está relacionado con la temática desarrollada sobre todo con enfermedades contagiosas e infecciosas, el reconocimiento hasta el momento, es bajo. Lo mismo ocurre con la revista “Computational Intelligence” y “Computational Methods in Applied Sciences”.

**Tabla 1.** Representación de las fuentes de revistas, numero de documentos publicados y total de citaciones por año.

	1985	1992	1997	1998	1999	2001	2002	2003	2006	2007	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
PLoS ONE9:330													1		1		1		2			1	2	1	
Epidemics8:90																		1			2		3	2	
Journal of Theoretical Biology7:135											1	1	1					2					1	1	
Mathematical Biosciences5:303					1		1		1									1		1					
Nonlinear Dynamics5:39																						1	2	2	
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)5:17										1					1	1			1			1			
IEEE Access4:35																						3	1		
Scientific Reports4:13																						1	2	1	
Journal of the Royal Society Interface3:420											1											1	2		
Proceedings of the Royal Society B: Biological Sciences3:221			1		1																		2		
PLoS Computational Biology3:194																1		1		1			1		
Stochastic Analysis and Applications3:93														1		1									1
Bulletin of Mathematical Biology3:59						1									1										1
Advances in Difference Equations3:25																1		1				1			1
International Journal of Environmental Research and Public Health3:21																						1	1	1	
Journal of Thoracic Disease2:786																						2			
Applied Mathematics and Computation2:181										1									1						
Algorithms2:177																							1	1	
Applied Mathematics Letters2:120																						1			
Journal of Mathematical Biology2:108	1																	1							

Un total de 50 países ha contribuido en la publicación de artículos, destacándose Estados Unidos con el 19.26%, siguiéndole China (18.85%), India (7.79%), Reino Unido (6.15%) e Italia (5.33%), estos países representan un acumulado del 57.38% del total de documentos, considerándose como los países más productivos en el campo de esta investigación. En la Figura 2 se muestra un mapa cloroplético que representa el total de publicaciones por país. Se destaca la comparación del número de publicaciones entre diferentes países. Se observa que los artículos más citados desde 1985 hasta 2023 corresponden a países como Estados Unidos, China, Reino Unido, India e Italia. Estos países se relacionan directamente con un mayor número de publicaciones en el campo de estudio.

Se identifica que algunos países como Puerto Rico, Dinamarca, Israel y Países Bajos que no presentan un gran número de publicaciones, se caracterizaron por tener documentos de interés, tanto, que superan en el número de citas; un ejemplo de esto es Puerto Rico que presenta una publicación en el tema, pero ha sido citado hasta 229 veces, este documento hace alusión a un modelo determinista de transmisión del dengue a partir del modelo SEIR, en donde exploran el comportamiento de la epidemia (Newton & Reiter, 1992).



**Figura 2.** Representación de total de publicaciones por país.

En cuanto a las instituciones que son aquellas que acreditan oficialmente la publicación, como lo puede ser un departamento, una institución o una universidad, se presenta en la consulta 184 instituciones, siendo 85 las instituciones del primer autor. Las instituciones que presentan un mayor número de publicaciones son el Departamento de matemáticas relacionados en varios países como Japón, China, Arabia Saudita, Egipto, Estado Unidos, entre otros que presentan un total de 23 publicaciones y 316 citas, resaltando los años 2012 al 2019.

Otra institución destacada es la Escuela de Matemáticas y Estadística en Estados Unidos que presenta cinco publicaciones y 88 citas. Este tipo de instituciones se considera reciente en el tema del modelado matemático epidemiológico y su adaptación y reconocimiento ha sido positivo. Así mismo, se observa el Politécnico de Milán de Italia que presenta cuatro publicaciones y 119 citas, pero la diferencia radica en que en los últimos años no ha generado publicaciones alrededor de este tema.

La revisión de las publicaciones presenta un total de 916 autores. En general, el promedio de autores por documento es de tres, aunque 16 publicaciones se caracterizaron por tener un único autor.

En la tabla 2 se presenta la lista de autores y el año de publicación, así como el total de publicaciones por año. Los autores más productivos en términos de número de publicaciones son Wang Y, Liu X y Wang K. Destaca especialmente Wang Y por tener un alto número de citaciones. Además, los autores más citados en general son Wang K, Liu X y Li Y.

Se destaca un grupo de autores que realizan publicaciones recientes como Liu H, Zhang S y Zhou Y, perfilándose como nuevos investigadores en el área de modelados epidemiológicos.

**Tabla 2.** Representación del total de documentos publicados por autor y año.

	1985	1992	1997	1998	1999	2001	2002	2003	2006	2007	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	
Zhou Y.																		1	3			1				
Zhang C.															2	1										
Hu Z.																2	2	1								
Teng Z.																2		1								
Liu H.															1									2		
Zhang X.-S.															2		1									
Wang H.																	1							2		
Wang K.															1								1			
Wang Z.																							2			
Zhang S.																					1	1				
Mishra B.K.										1		1														
Ball F.							1				1															
Prakash B.A.														1				1								
Keeling M.J.			1		1																					
Moura J.M.F.																1	1									
Moneim I.A.										2																
Zhang L.																2										
Capone F.															1	1										
Hossain M.S.																								2		
Lazzaris S.				2																						
Piccardi C.				2																						
Yang W.																	1							1		
Scoglio C.													1										1			
Schwartz I.B.	1														1											
Shaman J.																1								1		



**Tabla 3.** Representación de autores por documentos publicados y total de citaciones.

Autor	Total de documentos publicados	Número de citaciones	Autor	Total de documentos publicados	Número de citaciones
Wang K.	2	857	Zhong N.	1	785
Wang Y.	5	805	Bansal S.	1	413
Li Y.	2	788	Grenfell B.T.	1	413
Liu X.	3	788	Meyers L.A.	1	413
Zhang S.	2	788	Newton E.A.C.	1	229
Cao X.	1	785	Reiter P.	1	229
Gao Z.	1	785	Keeling M.J.	2	215
Guan W.	1	785	Mishra B.K.	2	208
He J.	1	785	Ball F.	2	188
Li F.	1	785	Morris A.J.	1	180
Li S.	1	785	Rand D.A.	1	180
Liang J.	1	785	Ardabili S.F.	1	169
Liang W.	1	785	Atkinson P.M.	1	169
Liu B.	1	785	Ferdinand F.	1	169
Liu P.	1	785	Ghamisi P.	1	169
Luo S.	1	785	Mosavi A.	1	169
Mai Z.	1	785	Rabczuk T.	1	169
Wang Z.	2	785	Reuter U.	1	169
Wong S.-S.	1	785	Varkonyi-Koczy A.R.	1	169
Xie Y.	1	785	Balcan D.	1	148
Yang Y.	2	785	Gonçalves B.	1	148
Yang Z.	1	785	Perra N.	1	148
Ye F.	1	785	Vespignani A.	1	148
Zanin M.	1	785	Neal P.	2	132
Zeng Z.	1	785	Jha N.	1	125

A partir de la Tabla 3, se presentan aquellas publicaciones realizadas con uno o más investigadores y las citaciones reportadas de la publicación. En general, el 85.9% de los artículos se realizan en grupos superiores a tres investigadores, aunque estos autores no cuentan con un mayor número de publicaciones. La tabla 4 muestra un resumen del total de publicaciones en función del número de autores. Se observa que a medida que aumenta el número de colaboradores involucrados en una publicación, la trayectoria individual de los autores en temas relacionados con la aplicación de modelos en enfermedades infecciosas tiende a ser menor.

**Tabla 4.** Impacto de las publicaciones por cantidad de autores.

Número de autores	Total de publicaciones	%	Acum. Número de autores	% acum
1	16	1,68%	16	1,68%
2	118	12,38%	134	14,06%
3	195	20,46%	329	34,52%
4	164	17,21%	493	51,73%
5	135	14,17%	628	65,90%
6	72	7,56%	700	73,45%
7	21	2,20%	721	75,66%
8	72	7,56%	793	83,21%
9	9	0,94%	802	84,16%
10	30	3,15%	832	87,30%
11	11	1,15%	843	88,46%
12	24	2,52%	867	90,98%
13	26	2,73%	893	93,70%
14	14	1,47%	907	95,17%
20	20	2,10%	927	97,27%
26	26	2,73%	953	100,00%

### 2.4.1 Análisis temático por palabra claves del autor

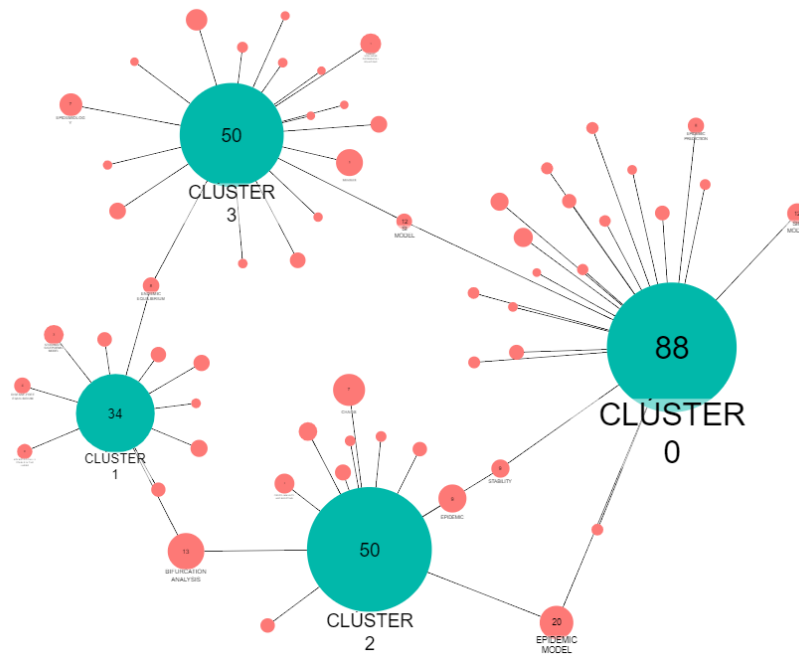


Figura 3. Visualización de las palabras claves por clusters.

Las diversas publicaciones que aquí se enuncian son caracterizadas principalmente por palabras claves como modelos epidemiológicos en especial los modelos SI (Susceptibles - Infectados), SIR (Susceptibles – Infectados - Recuperados), SIS (Susceptibles – Infectados - Susceptibles) y SEIR (Susceptibles – Expuestos – Infectados - Recuperados), análisis de bifurcación, epidemia, epidemiología y predicción epidemiológica.

El objetivo de esta sección es analizar en detalle el mapa temático de la investigación de modelos matemáticos epidemiológicos mediante el uso de una técnica de análisis de conglomerados o análisis de clúster. Para lograrlo, se aplicó la técnica K-Means para agrupar y clasificar las palabras claves por autor. El resultado de este análisis se puede observar en la Figura 3, donde se representa cada clúster con las palabras claves. En total se identificaron 332 palabras claves y luego de generar la respectiva limpieza, 274 son las palabras claves que se presentan para el análisis. A partir de la Tabla 5 y de la revisión generada, se identifica cierto grado de homogeneidad en base a las palabras claves, generando así 6 conglomerados, los centros son:

#### **2.4.1.1 Clúster 0: Modelos matemáticos epidemiológicos**

A partir de las palabras claves, el clúster cero hace alusión como tal a los modelos clásicos más relevantes y marcados, como por ejemplo el modelo SIS, SIR, SEIR, SIRS. Además, en el clúster se identifican los diversos tratamientos de los parámetros en el proceso de salud-enfermedad, como los infectados, periodo de latencia, el número básico de reproducción y susceptibles de contagio. En general, es uno de los clústers más fuertes porque las palabras claves mencionadas hacen alusión a los modelos matemáticos epidemiológicos que tiene el fin de predecir y explicar cómo puede la tasa de contacto afectar el comportamiento de la dinámica de la enfermedad.

Autores como Santillana et al. (2018) y Abuhasel et al. (2022), señalan que el modelo SIR y sus adaptaciones permiten controlar y prevenir el brote de infecciones, a través de características de las diferentes enfermedades infecciosas, como el comportamiento de la epidemia, los picos de brote periódicos, la reproducción de la enfermedad y el impacto de las técnicas de contención utilizadas. Estos modelos responden a las medidas de control y no sólo al agotamiento de los individuos susceptibles, teniendo en cuenta la historia Natural de la enfermedad, como la dinámica de la transmisión de la enfermedad.

Los artículos relacionados a este clúster resaltan la importancia de los modelos matemáticos en epidemiología y su aplicación para comprender y controlar la propagación de enfermedades. Proporcionan una revisión exhaustiva de los diferentes tipos de modelos utilizados, como los modelos compartimentales, basados en ecuaciones diferenciales, estocásticos y de redes complejas. Los modelos epidemiológicos basados en ecuaciones diferenciales permiten representar la dinámica de la propagación de enfermedades, mientras que los modelos estocásticos tienen en cuenta la incertidumbre y la aleatoriedad en la transmisión. Los modelos de redes complejas brindan una comprensión más detallada de cómo las enfermedades se propagan a través de las interacciones sociales. En conjunto, estos artículos contribuyen a mejorar la comprensión

y la capacidad de evaluación de intervenciones en epidemiología, lo que resulta fundamental para abordar y controlar eficazmente las enfermedades.

#### **2.4.1.2 Clúster 1: Métodos estocásticos**

Este clúster a diferencia del anterior, homogeneiza el tema de los modelos estocásticos de enfermedades infecciosas que muestra en general la forma que afecta la estructura de la población al tamaño de la epidemia de la enfermedad infecciosa. En el caso de los modelos epidemiológicos, los métodos estocásticos son una alternativa a los modelos tradicionales, utilizando sistemas de ecuaciones diferenciales para determinar las variaciones de la enfermedad.

Algunos autores como Alahakoon et al. (2022) y Malloy & Brandeau (2022), afirman que la simulación de una enfermedad puede lograr una mayor precisión frente al desvanecimiento u oleadas de infecciones recurrentes o a cambios en los parámetros que dirige la dinámica de la enfermedad, pero con mayores requisitos computacionales, recursos, datos y menor interpretabilidad.

En general, los artículos destacan que los métodos estocásticos permiten simular escenarios en los cuales se consideran diferentes factores aleatorios que pueden influir en la propagación de enfermedades, como la interacción social, los patrones de transmisión y la efectividad de las intervenciones. Estos modelos pueden ayudar a comprender la dinámica de una enfermedad, evaluar la eficacia de diferentes estrategias de control y predecir la propagación futura.

Al utilizar métodos estocásticos en epidemiología, es posible tener en cuenta la variabilidad inherente en la transmisión de enfermedades, lo que puede llevar a estimaciones más realistas de la propagación y el impacto de una enfermedad. Además, estos modelos pueden proporcionar intervalos de confianza y medidas de incertidumbre, lo que es crucial para tomar decisiones informadas y planificar estrategias de control.

#### **2.4.1.3 Clúster 2: Parámetro de tiempo**

Es el conjunto de palabras que hacen alusión al tiempo de latencia e infección de la enfermedad, aquellas utilizadas popularmente para detectar la patogénesis, la propagación y el control de las enfermedades. Es decir, las publicaciones relacionadas con este tipo de palabras claves eligen el parámetro del tamaño del paso de tiempo con la propagación, la tasa de incidencia estándar, la incubación, entre otros.

En conclusión, los artículos destacan la importancia del parámetro de tiempo en los modelos matemáticos y su papel fundamental en la comprensión y predicción de fenómenos en diversas disciplinas. El tiempo se reconoce como una variable crucial en la dinámica de los sistemas y su inclusión en los modelos permite capturar la evolución temporal de los mismos.

En el contexto de los modelos de predicción, se resalta que el parámetro de tiempo es un elemento clave para predecir eventos futuros y tendencias, utilizando la información histórica y la evolución temporal de las variables relevantes.

En el ámbito de los modelos epidemiológicos, se destaca la sensibilidad del parámetro de tiempo y cómo pequeños cambios en su valor pueden tener un impacto significativo en los resultados y las predicciones. Esto enfatiza la importancia de calibrar y ajustar adecuadamente este parámetro para obtener estimaciones precisas.

En los modelos de simulación de sistemas, el parámetro de tiempo desempeña un papel esencial al permitir representar la evolución temporal de los eventos y las interacciones dentro de un sistema. Esto es fundamental para comprender el comportamiento del sistema y realizar análisis de escenarios.

En conjunto, estos artículos resaltan la importancia de considerar y comprender adecuadamente el parámetro de tiempo en los modelos y las investigaciones científicas. Su inclusión adecuada permite capturar la dinámica temporal de los sistemas, mejorar la precisión de las predicciones y obtener una comprensión más completa de los fenómenos estudiados en diversos campos.

#### **2.4.1.4 Clúster 3: Medidas de prevención**

Las palabras claves del tercer clúster indican las diversas medidas correctivas para evitar el alto número de contagios con respecto a enfermedades infecto-contagiosas, haciendo alusión a las dinámicas sociales, a las medidas de vacunación, a la adaptación de estrategias para el cumplimiento de las medidas de prevención.

En conclusión, los artículos resaltan la importancia de implementar medidas de prevención en diversos ámbitos para proteger la salud y prevenir la propagación de enfermedades. Algunos aspectos destacados incluyen:

**Salud pública:** Las medidas preventivas, como la vacunación, el lavado de manos, el uso de mascarillas y el distanciamiento social, son herramientas claves para prevenir la propagación de enfermedades infecciosas en la población.

**Reducción de enfermedades:** La implementación adecuada de medidas preventivas, como la promoción de estilos de vida saludables y la detección temprana de enfermedades, puede contribuir a disminuir la incidencia y la gravedad de diversas enfermedades.

**Lugar de trabajo:** La adopción de políticas y prácticas de seguridad ocupacional, junto con la provisión de equipos de protección personal, la promoción de la salud y el bienestar de los empleados, y las evaluaciones de riesgos, son fundamentales para prevenir accidentes y enfermedades laborales.

**Ámbito escolar:** La promoción de entornos seguros y saludables en las escuelas, la implementación de protocolos de higiene, el fomento de la vacunación y la promoción de prácticas de prevención entre estudiantes y personal docente son esenciales para reducir la propagación de enfermedades y mejorar el bienestar general.

**Comunidad:** La participación comunitaria en la prevención de enfermedades, la promoción de estilos de vida saludables, la vigilancia epidemiológica, la educación en salud y la colaboración con organismos de salud pública son importantes para garantizar la efectividad de las medidas preventivas a nivel comunitario.

**Tabla 5.** Representación de las palabras claves por Clúster.

ID	CLÚSTER 0: MODELOS MATEMÁTICOS EPIDEMIOLÓGICOS	CLÚSTER 1: MÉTODOS ESTOCÁSTICOS	CLÚSTER 2: PARÁMETRO DE TIEMPO	CLÚSTER 3: MEDIDAS DE PREVENCIÓN.		
0	SIR MODEL 12:083	EPIDEMIC MODEL 17:178	ENDEMIC EQUILIBRIUM 05:056	BIFURCATION ANALYSIS 10:199	SI MODEL 09:050	EPIDEMIOLOG Y 07:102
1	EPIDEMIC PREDICTIO N 06:055	STABILITY 06:069	DISEASE-FREE EQUILIBRIUM 04:056	EPIDEMIC 09:139	MEASLES 03:131	DYNAMICAL SYSTEMS 02:058
2	SEIR MODEL 05:016	PARAMETER ESTIMATION 03:021	EXTINCTION 04:055	CHAOS 07:167	BISTABILITY 02:092	STATIONARY DISTRIBUTION 02:051
3	DYNAMIC MODEL 04:068	EBOLA 02:076	STOCHASTIC SIS EPIDEMIC MODEL 03:080	SIS MODEL 06:073	GAME THEORY 02:008	SIRS MODEL 02:017
4	INFECTIOUS DISEASE 04:027	DISCRETE EPIDEMIC MODEL 02:046	STOCHASTIC SIR MODEL 03:061	PERIOD DOUBLING 03:053	EPIDEMIC SPREADING 02:006	SOCIAL DYNAMICS 02:008
5	DENGUE FEVER 03:023	FLIP BIFURCATIO N 02:044	STOCHASTIC SIRS MODEL 03:043	CRITICAL TRANSITION 02:042	PERCOLATIO N THRESHOLD 02:000	NONLINEAR DYNAMICS 02:006
6	INFLUENZA 03:008	NUMERICAL SIMULATION 02:041	PERSISTENCE 03:041	TRANSCRITICA L BIFURCATION 02:042	PREVENTION 02:000	POPULATION DYNAMICS 02:004
7	FUZZY SYSTEMS 03:008	ABSORBING SETS 02:024	ASYMPTOTICALL Y STABLE IN THE LARGE 02:049	BAYESIAN 02:016	QUARANTINE 02:000	NONLINEAR COUPLED OSCILLATORS 01:058
8	COVID-19 03:001	DIRECT LYAPUNOV METHOD 02:024	STOCHASTIC LYAPUNOV FUNCTION 02:049	LIMIT CYCLE 02:015	SCALING LAW 02:000	
9	PERIODICIT Y 02:026	REACTION-DIFFUSION SYSTEM 02:024	ASYMPTOTIC PROFILE 02:009	CROSS-IMMUNITY AND BOOSTING 01:081	FORCED NONLINEAR DIFFERENTIAL EQUATIONS 01:088	

## 2.5 Discusión

Los modelos matemáticos epidemiológicos han evolucionado desde 1985 hasta la fecha, siendo inicialmente aceptados como modelos simples que predicen epidemias de larga duración. Actualmente, la epidemiología matemática ha evolucionado para combinar datos complejos de diversas fuentes con el fin de estudiar y generar resultados más adaptados a la evolución general de la enfermedad. Aun así, entre las publicaciones revisadas que no relacionan técnicas o modelos matemáticos epidemiológicos, se presentan parámetros adaptados, es decir, básicamente se está estudiando la dinámica endógena del modelo, pero no se ha encontrado evidencias de usar la evolución de la enfermedad o la Historia Natural de la misma a partir del territorio en el que se presenta el brote.

Se observa, además, que las publicaciones revisadas precisan modelos endémicos básicos sin estructura de edad, pero bien se sabe que el tamaño poblacional cambia exponencialmente y si el alcance final del modelo es estudiar las enfermedades infecciosas, implica sobre todo las estructuras de edad, porque las tasas de contagio, en general, dependen de las edades de los individuos. Igualmente, debe ser un dato razonable indagar por la demografía poblacional, como los datos actualizados de fecundidad como de mortalidad.

Por otra parte, se han encontrado publicaciones que hacen alusión a aquellos datos sobre pandemias, que generalmente, son ruidosos, contaminados, sesgados y procedentes de múltiples fuentes. En los últimos 10 años, se identifica una característica de las publicaciones y es la importancia de mecanismos de generación de datos con el fin de sintetizar múltiples datos imperfectos (Birrell et al., 2020).

En cuanto a los tipos de enfermedades relacionadas por cada modelo en los artículos revisados, se identifica lo siguiente:

Para el caso de la aplicación de modelos matemáticos epidemiológicos, se encontró que las enfermedades infecciosas como el COVID-19, la gripe, el Ébola, la tuberculosis y el dengue fueron más frecuentes. Estos modelos permiten comprender la dinámica de la transmisión, estimar las tasas de infección y recuperación, y evaluar estrategias de intervención. Además, se encontró que las enfermedades transmitidas por vectores, como la malaria, la fiebre del Nilo Occidental y la enfermedad de Lyme, también se abordaron en estos modelos. Estos ayudan a estudiar la interacción entre los vectores (como mosquitos y garrapatas) y los humanos, y a evaluar estrategias de control y prevención. Por último, se mencionaron las enfermedades crónicas, como la diabetes, la obesidad, las enfermedades cardiovasculares y el cáncer. Estos modelos ayudan a comprender los factores de riesgo, la progresión de la enfermedad y el impacto de las intervenciones en la salud de la población.

Por otra parte, los modelos matemáticos de predicción se relacionaron principalmente con las enfermedades emergentes como el COVID-19, el Ébola y el Zika. Estos modelos permitieron utilizar los datos existentes y tendencias para proyectar la propagación futura, estimar el número de casos y ayudar en la planificación de respuestas de salud pública. También se abordaron las enfermedades crónicas y no transmisibles, como la diabetes, la enfermedad cardiovascular y el cáncer. Por último, se mencionaron las enfermedades ambientales relacionadas con la contaminación del aire, la calidad del agua y la exposición a productos químicos.

Para el caso de las medidas de impacto en la propagación, los artículos revisados reconocen varias medidas. Entre ellas se encuentra el número básico de reproducción ( $R_0$ ), utilizado para estimar el número promedio de nuevas infecciones causadas por un individuo infectado. Además, se utiliza el número efectivo de reproducción ( $R_t$ ), que tiene en cuenta cambios en la susceptibilidad de la población debido a inmunidad adquirida o intervenciones de control. También se considera la tasa de ataque, que proporciona la proporción de individuos que se infectan durante un brote o epidemia. Por último, se

emplea el Índice de reproducción efectivo (RE) para representar el número promedio de casos secundarios producidos por un caso primario en una población que contiene individuos susceptibles y no susceptibles. Estas medidas permiten comprender y evaluar la dinámica de la propagación de enfermedades.

Finalmente, es preciso destacar la importancia de incorporar la imprecisión de los parámetros en los modelos epidémicos y el papel de la intervención de la información en el control de las enfermedades infecciosas (Bao et al., 2019).



## 3. Incorporación de mecanismo adaptativo en el modelo SIR

### 3.1 Introducción

El pronóstico anticipa y reduce el riesgo basado en datos actuales e históricos para determinar un hecho o fenómeno en un tiempo. El pronóstico se define conceptualmente como:

*Un proceso de conocimiento que investiga en forma metódica el pasado, el presente y el futuro, haciendo uso de técnicas cuantitativas y cualitativas, así como de la experiencia y habilidad de la persona que lo realiza, con el propósito de generar acciones en el tiempo para diseñar el futuro. (Montgomery et al., 2008, p.20).*

En el área de la salud, la predicción ayuda a prevenir y anticipar problemas de salud pública. Actualmente, se presenta un auge en diversos modelos o técnicas que permiten apoyar la toma de decisiones clínicas, fundamentar intervenciones o incluso incidir en desafíos operativos y administrativos. La pandemia del Covid-19 presenta una gran variedad de aplicaciones de diversas técnicas y modelos que permiten pronosticar y mitigar los efectos.

Predecir el comportamiento de una enfermedad permite generar reacciones inmediatas y decisiones en Salud Pública, con el fin de implementar medidas de mitigación y generar acciones de promoción y prevención en el territorio. Siendo importante al permitir reducir el impacto y brindar de forma eficiente atención a los contagiados, identificar requerimientos en camas hospitalarias y unidades de cuidado intensivo (UCI), número de vacunas, entre otras.

A partir de lo anterior, las técnicas y modelos indicados para el pronóstico deben tener en cuenta la distribución, frecuencia y factores determinantes de la enfermedad. Es por eso que el propósito de este Capítulo es presentar un proceso de modelización adaptando el modelo básico SIR (Susceptible – Infectado – Recuperado) de Kermack y McKendrick con el modelo Suavizado Exponencial, con el fin de presentar un modelo ajustado que permita predecir nuevos casos de la enfermedad infecciosa Covid-19 a nivel nacional incorporando la información del desarrollo de la enfermedad para el caso colombiano.

A pesar de la amplia gama de modelos epidemiológicos disponibles, este trabajo de tesis de Maestría se enfoca en el uso del modelo SIR debido a sus ventajas y simplificaciones. El modelo SIR asume que la población se puede dividir en tres grandes categorías, lo cual simplifica la realidad y permite un análisis más manejable de la propagación de la enfermedad. Además, el modelo SIR ofrece la capacidad de estimar parámetros claves, como la tasa de contagio y la tasa de recuperación, lo cual es fundamental para comprender la dinámica de la enfermedad. Otra ventaja importante del modelo SIR es su

reconocimiento y aceptación en la comunidad científica, lo que facilita la validación y comparación de los resultados obtenidos en este estudio con otros trabajos relacionados.

El modelo se centró en aplicar la metodología propuesta en casos infectados, susceptibles y recuperados arrojados desde el modelo SIR, utilizando las librerías y el ambiente de Python como una herramienta importante para el desarrollo de la tesis de maestría.

## 3.2 Modelo propuesto

### 3.2.1 Modelo SIR

En el campo de las matemáticas y la epidemiología existen modelos que buscan emular el comportamiento de una epidemia u enfermedad de propagación rápida para tener resultados simulados de cómo podría afectar a la población en estudio. Los modelos epidemiológicos presentan variantes que contemplan la posibilidad de que las personas tomen el mismo estado después de un tiempo, debido al contacto con la enfermedad.

Para este trabajo se tendrá en cuenta el modelo matemático aplicando ecuaciones diferenciales (SIR) propuesto por Roland Ross en 1902 y ajustado en 1927 por Kermack y McKendrick, que divide a la población en tres grandes grupos: Susceptibles, Infectados y Recuperados. En el primer grupo se presentan las personas del territorio que serán objeto de estudio. En el segundo grupo denominado "Infectados", se consideran todas las personas portadoras de la enfermedad y que pueden infectar a los susceptibles y, finalmente, los Recuperados, que son las personas que tienen inmunidad a la enfermedad o que fallecieron, por lo cual no tienen la capacidad de contagiar a más individuos (Montesinos & Hernández, 2007).

El modelo SIR presenta dos tasas, un  $\beta$  que describe qué tan rápido se transmite la infección de un individuo a otro. Mientras que  $\gamma$  o tasa de recuperación describe qué tan rápido un individuo infectado se recupera. Es posible calcular los valores de estos parámetros analizando el comportamiento de la enfermedad, lo cual depende de varios factores: la tasa de transmisión que depende de la interacción entre individuos y de la probabilidad de que una persona se enferme; mientras que la tasa de recuperación puede depender del tratamiento médico y la resistencia de cada individuo (Bill & Melinda Gates Foundation, 2022).

Cada etapa del modelo presenta una ecuación diferencial, que indica su comportamiento en el tiempo. Las ecuaciones del modelo SIR son las siguientes:

$$dS/dt = (-\beta SI)/N \quad (1)$$

$$dI/dt = \beta SI/N - \gamma I \quad (2)$$

$$dR/dt = \gamma I \quad (3)$$

La ecuación (1) representa el cambio instantáneo en un tiempo  $t$ , es decir, la variación con respecto al tiempo. La variación del número de personas susceptibles con respecto al

tiempo debe ser negativo, porque la cantidad de personas susceptibles tiende a disminuir. Esto depende del número inicial de personas susceptibles y su interacción con las personas infectadas en proporción con la población total. No todas las personas susceptibles resultaran afectadas, esto está indicado por la tasa de transmisión que depende de la interacción entre personas y la probabilidad de que una persona se enferme.

La ecuación (2) presenta la variación con respecto al tiempo del grupo de infectados que depende de la cantidad de casos susceptibles que ingresan a este grupo diariamente y de la cantidad de infectados que diariamente han pasado al grupo de recuperados.

La ecuación (3) presenta la cantidad de infectados que se han recuperado.

Estas ecuaciones forman un sistema de ecuaciones diferenciales que permiten modelar enfermedades infecciosas.

Así mismo, existe una cuarta ecuación, donde  $N$  significa el número total de personas en la población estudiada que debe ser igual a los Susceptibles, Infectados y Recuperados.

$$N = S + I + R \quad (4)$$

Otro factor importante en los modelos epidemiológicos es  $R_0$  o el número básico de reproducción de la enfermedad; este es un valor que cuantifica el número de personas que es capaz de infectar una persona contagiada por el virus. Si el  $R_0$  es igual a 1, quiere decir que una persona infectada es capaz de infectar a otra.

Para estimar  $R_0$  se debe calcular como la relación de  $\beta$  sobre  $\gamma$ , o también, la tasa de transmisión sobre la tasa de recuperación, esto se observa en la ecuación (5).

$$R_0 = \beta/\gamma \quad (5)$$

El modelo SIR es básico y útil para modelar y predecir el comportamiento de enfermedades infecciosas. Aunque, en su estructura omite factores como los nacimientos o muertes en la población y no se considera la edad de las personas o la estructura social.

### 3.2.2 Suavizado exponencial

Este modelo se basa en el término medio imparcial de la observación nueva y la anterior, dando lugar al valor suavizado para una serie. Los conjuntos de datos en general, constan de dos componentes: la señal y el ruido. Se entiende señal como el patrón causado por la dinámica del proceso, en el que se pueden ver patrones muy simples como complejos (Montgomery et al., 2008).

El proceso constante, por ejemplo, se representa como:

$$y_t = \alpha + \varepsilon_t \quad (6)$$

Donde  $\alpha$  representa el nivel constante y  $\varepsilon_t$  es el ruido en el tiempo  $t$ .

El suavizado se puede relacionar como la técnica para separar la señal del ruido y para obtener una estimación de la señal.

Un proceso constante puede suavizarse sustituyendo la observación actual por la mejor estimación de  $\mu$ . En el caso de que el proceso sea constante, el suavizador es eficaz. Mientras que, si el proceso no es constante y presenta un patrón más complicado, la hipótesis del proceso constante no es válida (Montgomery et al., 2008).

El Suavizamiento Exponencial Simple pondera los datos históricos exponencialmente para que los datos considerados más recientes tengan más peso en el promedio móvil. La predicción valida el último periodo más una porción  $\alpha$  que consiste en la diferencia entre el valor de la demanda real del periodo anterior y el pronóstico del periodo anterior. Sus características son: 1) Maneja un comportamiento histórico horizontal; 2) Calcula el promedio a través de un sistema de autocorrección que pretende ajustar el pronóstico en orientación contraria a las desviaciones del pasado; 3) Es un método muy utilizado para pronosticar ya que necesita pocos datos; 4) Tiene la superioridad sobre el promedio móvil ponderado y 5) No requiere de mucho aumento de períodos y aprobaciones para conseguir un óptimo resultado (Montgomery et al., 2008).

La estimación de los valores futuros parte de los pesos de los valores pasados, el valor mayor del peso  $\alpha$  donde ( $0 < \alpha < 1$ ) es otorgado a la observación más reciente y cada vez disminuye el peso para las observaciones pasadas  $\alpha(1 - \alpha)$ ,  $\alpha(1 - \alpha)^2$ ,  $\alpha(1 - \alpha)^3$  y así sucesivamente.

El valor para predecir es la sumatoria de la multiplicación de los pesos por los valores anteriores, así:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (7)$$

donde,

$\hat{Y}_{t+1}$  = Valor del pronóstico suavizado para el siguiente periodo.

$\alpha$  = constante de suavización que determina la ponderación relativa puesta en la observación de la demanda actual ( $0 < \alpha < 1$ ). Estrictamente el valor de  $\alpha$ , no puede tomar el valor de cero o el valor de uno.

$Y_t$  = nueva observación o el valor real de la serie en el periodo  $t$ .

$\hat{Y}_t$  = último valor suavizado o el pronóstico del periodo  $t$ .

$(1 - \alpha)$  = Peso asignado a las observaciones pasadas de la demanda.

Si el valor de  $\alpha$  se acerca a 1 significa que existe mayor influencia en el pronóstico actual por parte de la observación anterior, esto significa que los pronósticos anteriores tendrán menor influencia en el resultado (Montgomery et al., 2008).

### **3.2.3 Método propuesto SSE-SIR**

El modelo epidemiológico SIR se utiliza para describir la propagación de enfermedades infecciosas en una población. En este modelo, los individuos susceptibles pueden infectarse, luego se vuelven infectados y, finalmente, se recuperan o fallecen. Sin embargo, el modelo SIR considera las tasas de infección y recuperación como constantes, lo que limita su capacidad para tener en cuenta el impacto de medidas como el aislamiento o la vacunación.

En esta tesis de maestría se propone el método SSE-SIR para mejorar este modelo. El suavizado exponencial simple es una técnica que permite estimar valores futuros a partir de datos pasados y presentes. En el método SSE-SIR, las tasas de infección y recuperación, que eran consideradas constantes en el modelo SIR, son reemplazadas por valores estimados dinámicamente utilizando el suavizado exponencial simple.

En el método SSE-SIR, las tasas  $\beta$  y  $\gamma$  son estimadas de forma dinámica utilizando el suavizado exponencial simple, en función de los datos disponibles en cada punto en el tiempo. Esto permite que las tasas de infección y recuperación se adapten a los cambios en la dinámica del sistema, como las medidas de control implementadas (aislamiento, vacunación, etc.). Los nuevos valores estimados de  $\beta$  y  $\gamma$  se utilizan en el modelo SIR para realizar pronósticos más precisos y ajustados a la realidad.

En general, el método SSE-SIR combina el modelo epidemiológico SIR con el suavizado exponencial simple. Reemplaza las tasas de infección y recuperación constantes por valores estimados dinámicamente utilizando el suavizado exponencial simple. Esto permite que el modelo se adapte a los cambios en la dinámica del sistema, mejorando la precisión de los pronósticos y teniendo en cuenta el impacto de las medidas de control implementadas.

#### **3.2.3.1 Procesamiento de datos**

Inicialmente, la carga del dataset se realiza a partir de la publicación en datos abiertos con la disponibilidad diaria de la información actualizada con las cifras del Covid-19 para Colombia. El conjunto de datos presenta 21 atributos dentro de los cuales se destacan: Nombre departamento, Estado, Recuperado, Fecha de notificación, Fecha de muerte, Fecha de diagnóstico, Fecha de recuperación y fecha reporte web.

A partir de lo anterior, se crean los casos confirmados, muertes, recuperados y activos. Dando lugar a la identificación de los susceptibles, Infectados y Recuperados, a partir de las siguientes ecuaciones:

$$S = N - C_t \quad (8)$$

$$I = C_t - R_t - F_t \quad (9)$$

$$R = R_t + F_t \quad (10)$$

La ecuación (8) presenta los Susceptibles que corresponde a la diferencia entre la población reportada al inicio de la pandemia y los casos confirmados reportados diarios.

La ecuación (9) presenta los Infectados que corresponde a la diferencia entre los casos confirmados, los casos recuperados y los casos fallecidos.

La ecuación (10) presenta los Recuperados que es la adición entre los casos recuperados y los casos fallecidos.

Luego, se define la ecuación diferencial del modelo SIR (Susceptible, Infectado, Recuperado), indicados en las ecuaciones (1), (2) y (3) y se estima un  $\beta$  para la tasa de infección y  $\gamma$  o tasa de recuperación.

Seguidamente, se aplica la ecuación (7) del modelo Suavizado exponencial simple en el que el valor de amortización toma los valores de beta y gamma.

### 3.3 Experimentación, análisis y resultados

El método propuesto SSE-SIR se ejecuta a partir de diversos escenarios de la contingencia del Covid-19. A continuación, se contrasta o compara el método con los resultados de otros trabajos de investigación frente al mismo universo o territorio objetivo y al mismo periodo de evaluación. Así mismo, se validan en los términos enunciados en cada trabajo investigativo.

#### 3.3.1 Caso 1

En Díaz (2020a) se presenta un pronóstico de los infectados con el modelo ARIMA (0,1,0), estimando que, al finalizar el mes de diciembre del año 2021 se tendría un número de 4.973.547 personas contagiadas, 4.784.987 personas recuperadas, y 110.159 personas fallecidas. La predicción realizada presenta un corte al 5 de noviembre de 2020 con pronóstico de un año y un periodo de evaluación de 421 días.

Para este primer caso de contraste, debido a que el documento no identifica medidas o criterios de evaluación, se revisan los datos en valores absolutos. En la Tabla 6, se presentan los resultados de predicción, inicialmente se encuentran los datos reales, luego se observan los datos arrojados por Díaz (2020a) y finalmente, se presentan los resultados del método propuesto SSE-SIR.

La Figura 4 muestra los valores originales, valores ajustados y los valores predichos de los Susceptibles, Infectados y Recuperados para el método propuesto; y finalmente, en la Tabla 7 se presentan las medidas estadísticas de la bondad de ajuste o fiabilidad del método SSE-SIR.

**Tabla 6.** Relación modelo SIR adaptado Suavizado exponencial.

Mes	Número de infectados	Arima	SIR – SE
Enero	243.178	1.914.750	192.459
Febrero	393.494	2.171.176	388.234
Marzo	577.932	2.455.076	585.978
Abril	837.681	2.729.819	783.871
Mayo	1.127.415	3.013.719	980.108
Junio	1.490.170	3.288.461	1.173.058
Julio	1.660.813	3.572.362	1.361.000
Agosto	1.752.064	3.856.262	1.542.477
Septiembre	1.821.137	4.131.004	1.716.520
Octubre	1.882.765	4.414.905	1.881.899
Noviembre	1.941.278	4.689.647	2.038.179
Diciembre	1.997.934	4.973.547	2.187.472

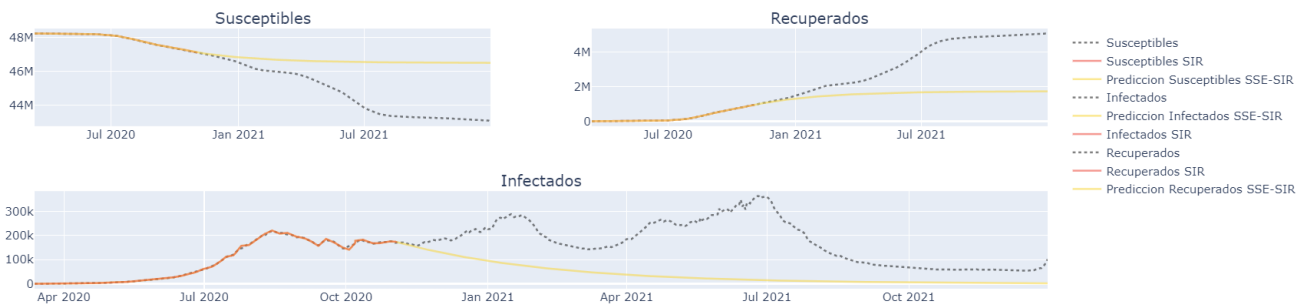


Figura 4. Representación del modelo SIR adaptado Suavizado exponencial caso 1.

**Tabla 7.** Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 1.

Clase epidemiológica	R2 Díaz (2020a)	R2 SSE-SIR	Error Cuadrático Medio (RMSE) SSE-SIR
Susceptibles	No aplica	0,077	5,27
Infectados	0,452	-0,853	2,55
Recuperados	0,926	0,148	4,8

Se observa que, el estado de contagios de COVID-19 para el mes de enero del 2021 era de 243.178 infectados y al finalizar el año de 1.997.934 infectados. En contraste a los resultados, el modelo aplicado por Díaz (2020a) inicia con 1.914.750 infectados y finaliza con 4.973.547 infectados. Mientras que el método SSE-SIR inicia con 192.459 infectados y finaliza con 2.187.472 Infectados. A partir de lo anterior se concluye que, los resultados propuestos del modelo ARIMA planteado por el autor presenta mayor diferencia al valor real del total de infectados. Mientras que el modelo SSE-SIR se aproxima con una diferencia menor a los resultados reales. Aún así, es necesario resaltar que, al ser un modelo adaptado de suavización, la precisión está directamente relacionada a periodos de evaluación cortos, condición que no aplica en el caso 1 debido a que presenta un periodo de evaluación de 421 días. En cuanto a los resultados del coeficiente de determinación  $R^2$ , la medida para la clase de infectados no se ajusta del todo a la variable real.

### 3.3.2 Caso 2

En Ramírez et al. (2020) se presentan tres (3) modelamientos aplicados en diferentes periodos de evaluación:

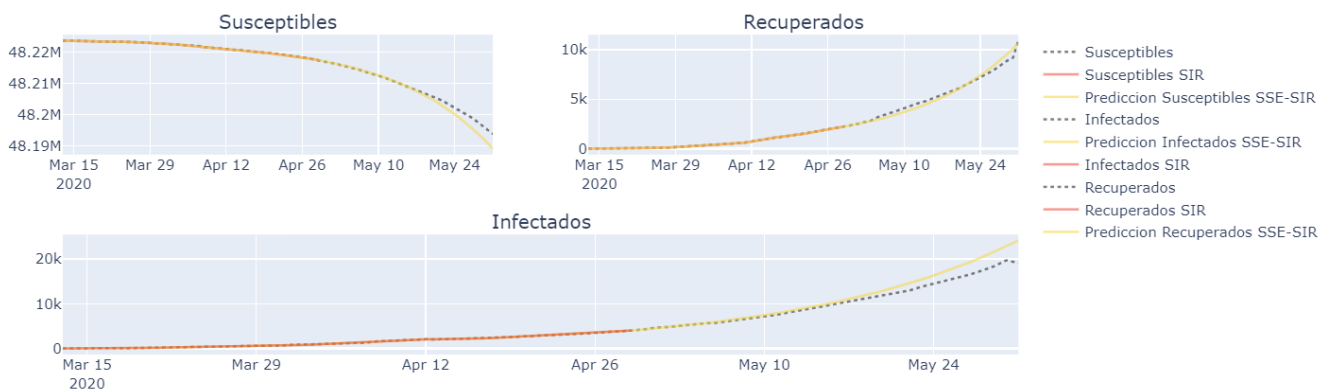
- Modelos entre marzo y abril: en este periodo se realizaron cuatro modelos para contagios, recuperados y fallecidos y se ajustaron las curvas de pronóstico para predecir el comportamiento del Covid-19 para el mes de mayo, los modelos aplicados hacen referencia al modelo exponencial, polinómico, Potencial y Gaussiano.
- Modelos entre marzo y julio: para este segundo se aplican los cuatro modelos para contagios, recuperados y fallecidos y se ajustaron curvas de pronóstico para predecir el comportamiento del Covid-19 en el mes de agosto.
- Modelos entre marzo y agosto: se aplican dos modelos gaussianos con los datos y se ajustaron curvas de pronóstico para la predicción del mes de septiembre.

A continuación, se presentan los resultados de la comparación con el método propuesto SSE-SIR, la medida de fiabilidad del modelo aplicado será el coeficiente de determinación:

**Modelo entre Marzo y abril:**



Los resultados para este periodo de tiempo, con mejor ajuste según el Coeficiente de Determinación, fue el gaussiano. En la Figura 5 se muestran los valores originales, ajustados y predichos de los grupos de Susceptibles, Infectados y Recuperados utilizando el método propuesto. Se observa un buen ajuste en el comportamiento durante este periodo de evaluación. Además, en la Tabla 8 se presentan los resultados obtenidos para este modelo, con un valor de 0.868 para los infectados y 0.897 para los recuperados. Por otro lado, para el método SSE-SIR se obtuvieron resultados de 0.991 para los recuperados y 0.979 para los infectados. A partir de estos resultados, se puede inferir que el método SSE-SIR realiza predicciones precisas para ambas clases epidemiológicas, mostrando mejores resultados en términos de bondad de ajuste, tanto en el coeficiente de determinación como en el Error Cuadrático Medio (RMSE).



**Figura 5.** Representación del modelo SIR adaptado Suavizado exponencial caso 2, Modelo para el periodo entre Marzo y abril del 2020.

**Tabla 8.** Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 2, Modelo para el periodo entre Marzo y abril del 2020.

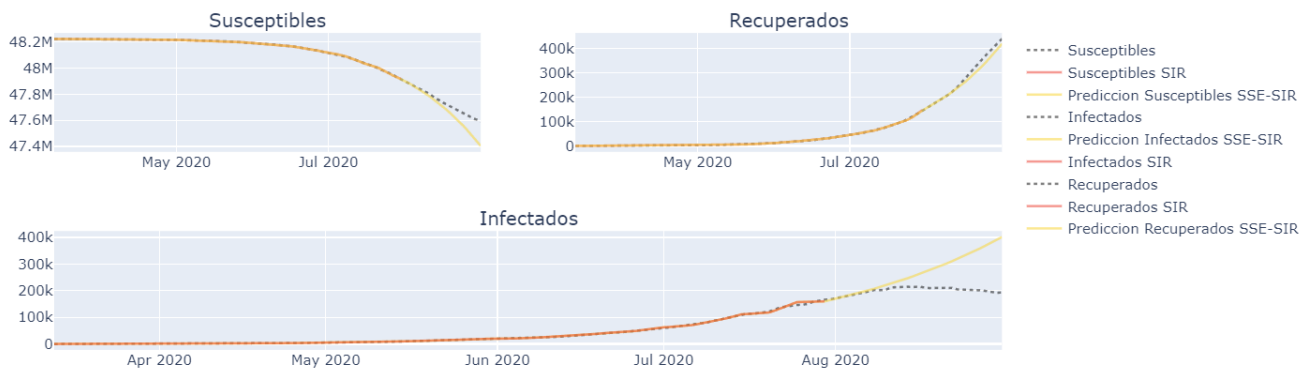
Clase epidemiológica	R <sup>2</sup> Ramírez et al., 2020	R <sup>2</sup> SSE-SIR	Error Cuadrático Medio (RMSE) SSE-SIR
Susceptibles	No aplica	0,991	1,57
Infectados	0,868	0,979	0.000164
Recuperados	0,897	0,991	1,819

### Modelo entre Marzo y Julio

En este periodo de evaluación, se observa que el modelo Gaussiano muestra el mejor ajuste según el Coeficiente de Determinación, con valores de 0.981 para los contagiados y 0.987 para los recuperados. La Figura 6 muestra los valores originales, ajustados y predichos de los grupos de Susceptibles, Infectados y Recuperados utilizando el método

propuesto, evidenciando un buen ajuste en el comportamiento durante este periodo de evaluación. En la Tabla 9, se presentan los resultados del modelo SSE-SIR, donde se destaca la clase epidemiológica de Recuperados con resultados significativos, con un valor de 0.994 en comparación con el modelo Gaussiano. Sin embargo, los Contagiados presentan un coeficiente de determinación de 0.718, inferior al modelo comparativo.

Los resultados indican que la predicción del modelo SSE-SIR es precisa hasta mediados de agosto, correspondiente al primer pico de la pandemia que ocurrió entre el 16 de julio y el 17 de agosto de 2020 (Semana, 2021). A partir de ese punto, se observa una sobreestimación en las predicciones, probablemente debido a los planes de contingencia adoptados durante ese período de tiempo.



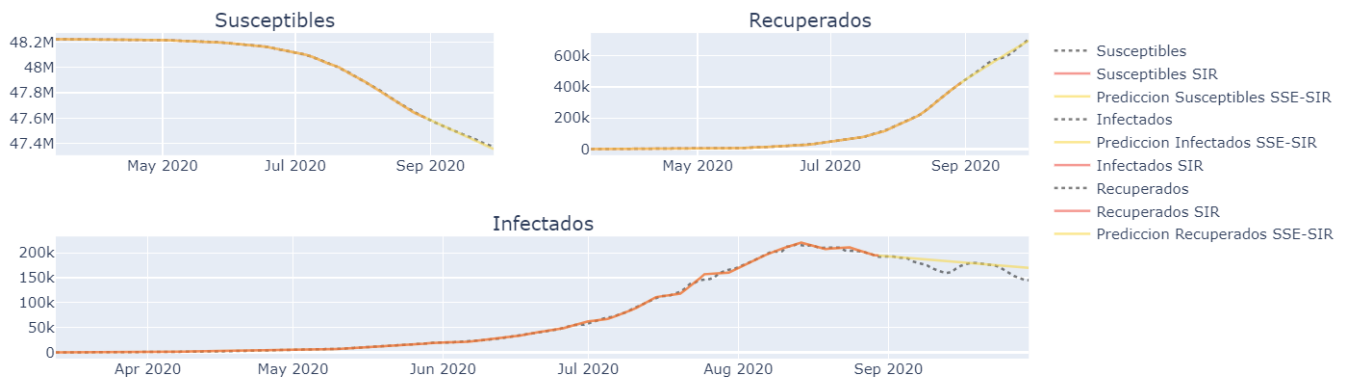
**Figura 6.** Representación del modelo SIR adaptado Suavizado exponencial caso 2, Modelo para el periodo entre Marzo y Julio del 2020.

**Tabla 9.** Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 2, Modelo para el periodo entre Marzo y Julio del 2020.

Clase epidemiológica	R2 Ramírez et al., 2020	R2 SSE-SIR	Error Cuadrático Medio (RMSE) SSE- SIR
Susceptibles	No aplica	0,973	0,005
Infectados	0,9814	0,717	0,900
Recuperados	0,9873	0,989	0.0007

## Modelo entre Marzo y Agosto

Durante este periodo de evaluación, se destaca el modelo Gaussiano como el mejor en términos de ajuste, con un Coeficiente de Determinación de 0.971 para la clase de Contagiados y 0.995 para la clase de Recuperados. La Figura 7 muestra los valores originales, ajustados y predichos de los grupos de Susceptibles, Infectados y Recuperados utilizando el método propuesto, evidenciando un buen ajuste en el comportamiento durante este periodo de evaluación. En la Tabla 10, se presentan los resultados del modelo SSE-SIR, que muestran una mayor significancia en comparación con los resultados del modelo Gaussiano para todas las clases epidemiológicas. Además, se observa que los valores del Error Cuadrático Medio (RMSE) son bajos, lo cual indica un mejor ajuste del modelo propuesto.



**Figura 7.** Representación del modelo SIR adaptado Suavizado exponencial caso 2, Modelo para el periodo entre Marzo y Agosto del 2020.

**Tabla 10.** Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 2, Modelo para el periodo entre Marzo y Agosto del 2020.

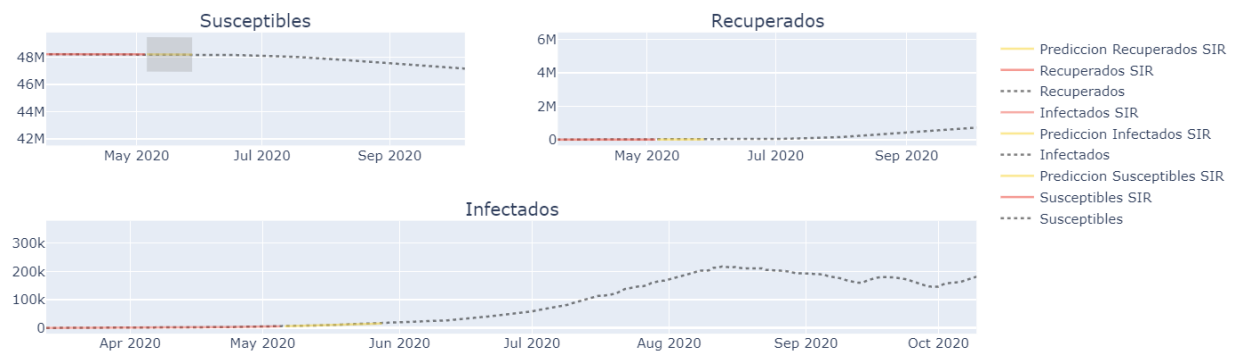
Clase epidemiológica	$R^2$ Ramírez et al., 2020	$R^2$ SSE-SIR	Error Cuadrático Medio (RMSE) SSE- SIR
Susceptibles	No aplica	0,999	3,064
Infectados	0,9715	0,995	0,019
Recuperados	0,9958	0,999	0,00016

En conclusión, la aplicación de los modelos no lineales Gaussianos permitió elaborar un análisis acertado para Ramírez et al., 2020. Estos modelos son una práctica que permite describir el comportamiento de las curvas y cuenta con diferentes formas de evaluación del comportamiento de la pandemia. En el caso del modelo SSE-SIR adaptado para este trabajo, los coeficientes indican un mayor acierto de este modelo para los periodos analizados en comparación con el modelo Gaussiano. Se destaca la predicción de infectados para el periodo de evaluación entre Marzo y Julio, debido a que presenta una predicción mayor a la real explicado por el pico de contagio y el aislamiento preventivo implementado en el país.

### 3.3.3 Caso 3

En el estudio de Bravo et al. (2020), se desarrollaron dos modelos para proyectar los casos confirmados de infección en Colombia. El modelo Sigmodial se utilizó para predecir el comportamiento de la infección por Covid-19 en el país entre el 7 de mayo y el 28 de mayo de 2020, basándose en los datos reportados del 6 de marzo al 6 de mayo de 2020. Para evaluar el desempeño del modelo Sigmodial en términos de ajuste y estimación, se empleó el Error Cuadrático Medio Normalizado (NRMSE).

La Figura 8 muestra en la zona sombreada el NRMSE, así como los valores originales, ajustados y predichos. El periodo de evaluación para la predicción de los casos confirmados abarcó 21 días, desde el 7 de mayo de 2020 hasta el 28 de mayo de 2020. El error de predicción obtenido fue del 3.5%. En la Tabla 11 se presentan los resultados de la predicción, y se observa que el NRMSE para el modelo SSE-SIR es más bajo en comparación, lo que demuestra que este modelo se ajusta mejor a la predicción de los casos confirmados relacionados con la pandemia.



**Figura 8.** Representación del modelo SIR adaptado Suavizado exponencial caso 3.

**Tabla 11.** Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 3.

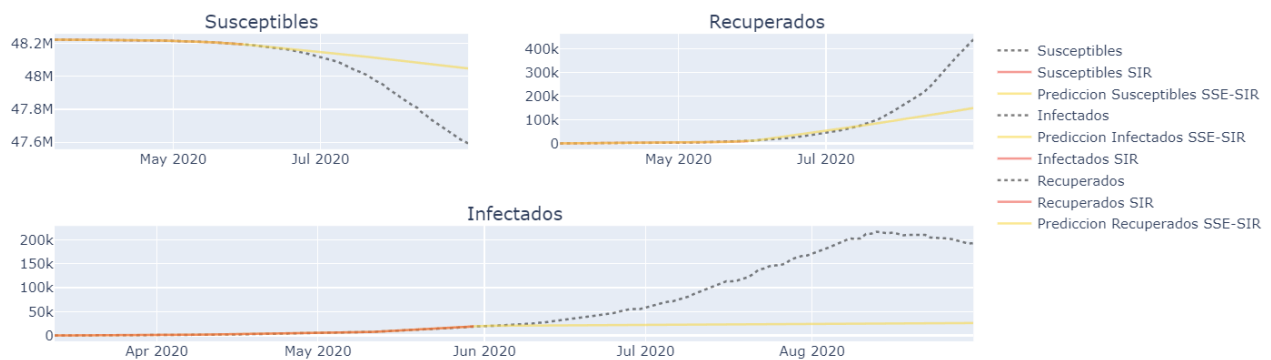
clase epidemiológica	Error cuadrático medio normalizado (NRMSE) Bravo et al.,2020	Error cuadrático medio normalizado (NRMSE) SSE-SIR	R <sup>2</sup> SSE-SIR	Error Cuadrático Medio (RMSE) SSE-SIR
Susceptibles	No aplica	No aplica	0,995	7,447
Infectados	3,5	1,286	0,977	0,0001
Recuperados	No aplica	No aplica	0,985	3,095

En conclusión, el modelo SSE-SIR según el Error cuadrático medio normalizado (NRMSE) presenta un nivel de aproximación suficiente para evaluar el comportamiento del contagio del Covid-19. Igualmente, las medidas de Coeficiente de Determinación ( $R^2$ ) y Error Cuadrático Medio (RMSE) son significativos frente al periodo de evaluación.

### 3.3.4 Caso 4

En el estudio de Díaz (2020a) se presentan modelos predictivos para la dinámica de transmisión del COVID-19 en Colombia, específicamente para los casos confirmados, activos, recuperados y fallecidos. El método utilizado fue el modelo lineal de Brown, considerando el registro de la primera persona infectada el 6 de marzo de 2020 y la última fecha considerada el 5 de mayo de 2020. Además, se realizaron predicciones para los meses de mayo, junio y julio.

Al analizar los resultados mostrados en la Figura 9, que representa los valores originales, ajustados y predichos de los grupos de Susceptibles, Infectados y Recuperados utilizando el método propuesto, y en la Tabla 12 que muestra la relación por estimador, se observa una mayor diferencia entre los resultados del modelo lineal de Brown y el modelo adaptado SSE-SIR. El modelo SSE-SIR proporciona un alto nivel de precisión en la predicción hasta mediados de junio. Sin embargo, a partir del coeficiente de determinación  $R^2$ , se evidencia que la clase de infectados no se ajusta completamente a la variable real. Esto puede estar asociado al inicio del pico de contagio del COVID-19 que ocurrió entre mayo y agosto de 2020. (Semana, 2021).



**Figura 9.** Representación del modelo SIR adaptado Suavizado exponencial caso 4.

**Tabla 12.** Resultados por estimador par el modelo SIR adaptado Suavizado Exponencial caso 4.

Clase epidemiológica	$R^2$ Díaz (2020a)	$R^2$ SSE-SIR	Error Cuadrático Medio (RMSE) SSE-SIR
Susceptibles	No aplica	0,291	0,04452
Infectados	0,452	-0,248	1,36338
Recuperados	0,926	0,534	0,0107

### 3.4 Conclusiones

En la pandemia del Covid-19, los modelos matemáticos a partir de patrones epidemiológicos permiten describir las interacciones y explicar la dinámica de la infección y recuperación, así como predecir posibles escenarios que pueden presentarse con la introducción de medidas de mitigación de la pandemia como el distanciamiento Social. En este Capítulo se relaciona un método adaptado para predecir, denominado SSE-SIR, que se aplicó para la estimación de las clases epidemiológicas de Susceptibles, Infectados y Recuperados de la enfermedad del Covid-19 a partir de cuatro escenarios.

El método se basa en el modelo epidemiológico SIR (Susceptibles, Infectados y Recuperados) y el suavizamiento exponencial simple que genera pronósticos con base en el promedio de valores pasados. Los resultados de este modelo, valorados a partir de medidas como el coeficiente de determinación  $R^2$  y el Error Cuadrático Medio (RMSE), son adecuados, con mayor ajuste a los datos reales de la pandemia y sobre todo, resultados más óptimos en comparación a los casos revisados por los autores.

En la revisión se identifica que el modelo adaptado presentó dos características en las predicciones: 1) Las cantidades inherentemente más alta de inexplicable variabilidad, debido al periodo de imprecisión, es decir, los resultados de predicción para los periodos de evaluación que relacionan los picos o puntos críticos de la pandemia a nivel nacional, no fueron del todo acertados y 2) en el caso de que el periodo de evaluación para la predicción sea amplio la medida no se ajusta del todo a la variable real.

## 4. Predicción incorporando información de otros países

### 4.1 Introducción

El Covid-19 a nivel global produjo grandes efectos en la dinámica política, social y económica de más de 190 países. Cada país víctima de la pandemia ha tomado acciones drásticas con el fin de mitigar los efectos del Covid-19. Algunos cierran las fronteras, confinan a la población, implementan medidas de precaución, entre otras acciones.

Es por eso que, sin llegar a evaluar el total poblacional, densidad territorial, características, económicas, sociales, culturales o políticas, se pueden identificar frente a los datos históricos de la dinámica del Covid-19, los casos nuevos, contagios, activos y muertes, lo que posibilita determinar estimaciones frente de la incidencia de la enfermedad, como en el caso de la Tasa de Infección.

El presente capítulo busca determinar si es posible pronosticar la tasa de infección para Colombia, usando como datos de entrada las tasas de infección en otros países y las tasas pasadas de infección en Colombia. Esta hipótesis surge de la observación de que algunos países iniciaron de forma anticipada el proceso de contagio, y por tanto, podría esperarse que esa información sirva para pronosticar la evolución de la enfermedad para el caso colombiano. Es decir, la fecha de llegada al primer pico de la pandemia ocurrió primero en algunos países, por lo que esta información fue usada por expertos para pronosticar la evolución de otros países que estaban más rezagados.

En este capítulo se pretende construir un modelo de aprendizaje de maquina que pronostique la tasa de infección para Colombia a partir de las tasas de infección históricas de otros países. Los parámetros de este modelo fueron verificados usando validación cruzada.

### 4.2 Modelo propuesto

#### 4.2.1 Regresión lineal

Es una técnica paramétrica utilizada para predecir una variable dependiente a partir de un vector de variables independientes. Se denomina paramétrica porque permite hacer suposiciones basadas en conjuntos de datos.

Matemáticamente la regresión usa una función lineal para predecir las variables dependientes como:

$$y = \alpha x + b \quad (1)$$

donde,

$y$  es la variable dependiente o la variable para predecir.

$x$  es la variable independiente.

$a$  = Pendiente o el valor determinado. También conocido como coeficiente o magnitud de cambio que pasa por  $y$  cuando  $x$  cambia.

$b$  = Es la constante. Se le conoce como intercepción porque cuando  $x$  es igual a 0, entonces  $y = b$ .

Si únicamente hay un valor en el vector  $x$  se considera unidimensional y el modelo se expresa como la ecuación 1, donde el modelo de regresión lineal simple asume una relación lineal entre el valor medio de la respuesta y el valor de la variable de entrada. Si por el contrario la entrada es multidimensional el método es denominado regresión lineal múltiple, la cual se puede expresar como se indica en la ecuación 2. y si la salida es a su vez también multidimensional este método es denominado regresión lineal multivariable (M.Ross, 2007).

$$y = w_0 + w_1x_1 + w_2x_2 + w_dx_d \quad (2)$$

Adicionalmente, se realiza un modelo de combinación de pronósticos, dado que se puede obtener un mejor resultado al combinar la predicción desde diferentes modelos.

La combinación de pronósticos permite minimizar la varianza del error del predictor, ya sean errores independientes o correlacionados. En general los modelos de combinación permitirán mejores y óptimos resultados.

### 4.3 Selección de variables relevantes

La selección de características es un proceso relevante que permite mejorar la precisión y el rendimiento en los problemas de datasets con muchas dimensiones. Inicialmente se seleccionaron aquellas características que son relevantes, que permiten la identificación de las tasas de contagio y que contribuyen a la predicción.

Se emplea una técnica muy utilizada en Machine Learning denominada validación cruzada o cross validation (CV) que consiste en calcular la media aritmética de las medidas de precisión de validación sobre diferentes particiones del dataset. De esta manera se asegura que los resultados sean independientes de la partición entre datos de entrenamiento y validación. La validación simple es el método más sencillo y usado para estimar la predicción del error. Este método utiliza una parte de los datos para entrenar el modelo y otra parte para probarlo.

En la validación de los datos, las tasas por países se contrastan bajo las tasas de infección de la fecha de corte de evaluación  $(t - 1)$ ,  $(t - 2)$ , ...,  $(t - n)$ . Este proceso se realiza hasta 40 días de retraso para evitar tendencias. A partir de la correlación se otorgan los resultados y se evalúan desde 0,5 (correlación intermedia) hasta el máximo de correlación obtenida para encontrar los países con mayor correlación al caso colombiano.



## 4.4 Información utilizada

Inicialmente, la carga del DataSet se realiza a partir de una recopilación de los datos diarios del Covid-19 mantenidos por Our World in Data.

A continuación, se describe el proceso para convertir las tasas de infección de los países que reportan casos diarios que pueden ser utilizados por los modelos y así, encontrar el modelo ideal. Se explica, primero, lo relativo a técnicas exploratorias y al preprocesamiento requerido para este tipo de información. Luego, se describe el modelo usado para el aprendizaje supervisado. Por último, se explican las técnicas para entrenar el modelo, la selección y validación.

## 4.5 Preprocesamiento de los datos

En el presente trabajo, antes de realizar cualquier análisis, se proponen las siguientes etapas de preprocesamiento:

- Eliminar atributos que no aportan al ejercicio como Continente, país, fecha, casos totales, casos nuevos, muertes totales y muertes nuevas.
- Generar diferentes tablas para cada país.
- Obtiene la tasa de infección para cada día, siendo considerada la siguiente ecuación:

$$Tasa\ de\ Infección_t = \frac{CN_t}{S_t}$$

- En el caso de los valores nulos o faltantes, los correspondientes al inicio de la pandemia tomaran un valor de cero, considerada como tasa de infección nula. Mientras que, para los registros más recientes, se utilizaba el promedio de los últimos siete (7) días.

## 4.6 Experimentación, análisis y resultados

En este apartado se relacionan los resultados a partir del método adaptado SSE-SIR revisado en el Capítulo 2 y contrastando con otros trabajos investigativos frente al mismo universo o territorio objetivo y al mismo periodo de evaluación.

Inicialmente, se presenta el comportamiento de los modelos propuestos en el presente trabajo y se pone a prueba ciclos de predicción de 45 días desde el inicio de la pandemia, con el fin de verificar las predicciones obtenidas y el ajuste de los datos con respecto a los reales.

La Figura 10 presenta los infectados acumulados para tres escenarios, el modelo propuesto SSE-SIR, el modelo de Regresión Lineal y un promedio entre los dos, el R2 o coeficiente de determinación para estos modelos son los siguientes: 1) SSE-SIR: 0,827 ,2) Regresión Lineal: 0,835 y 3) Promedio: 0,832.

Se identificó que, desde el año 2020 hasta la fecha el modelo adaptado SSE-SIR y el modelo de Regresión Lineal, según el coeficiente de determinación, presenta mayor bondad de ajuste. Esto se observa en todos los periodos de evaluación, con excepción de los siguientes trimestres:

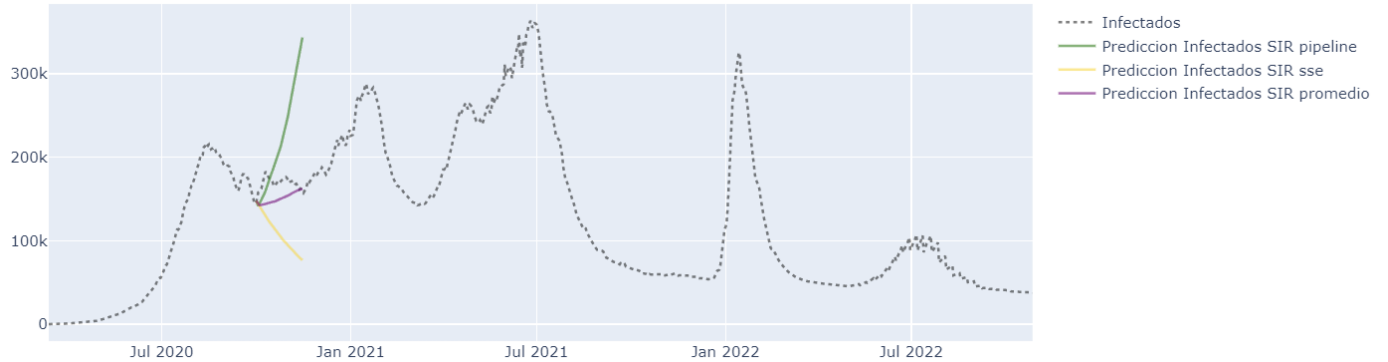
En la Figura 10 se muestra la representación del trimestre II del 2020 en el que los modelos aplicados presentan resultados notables, según el coeficiente de determinación. Pese a eso, los resultados de la predicción reportan menos contagios a los casos reales y aunque, se incrementan el numero de casos, no genera en el periodo necesario la alerta a las autoridades para incidir en el impacto. Esto se relaciona directamente con el primer pico de contagio en Colombia.



**Figura 10.** Representación trimestre II del año 2020 de la predicción Colombia a nivel global.

La Figura 11 presenta los infectados acumulados para el trimestre III del 2020 para tres escenarios. El R2 o coeficiente de determinación de infectados para estos modelos son los siguientes: 1) SSE-SIR: 0,879, 2) Regresión Lineal: 0,812 y 3) Promedio: 0,987.

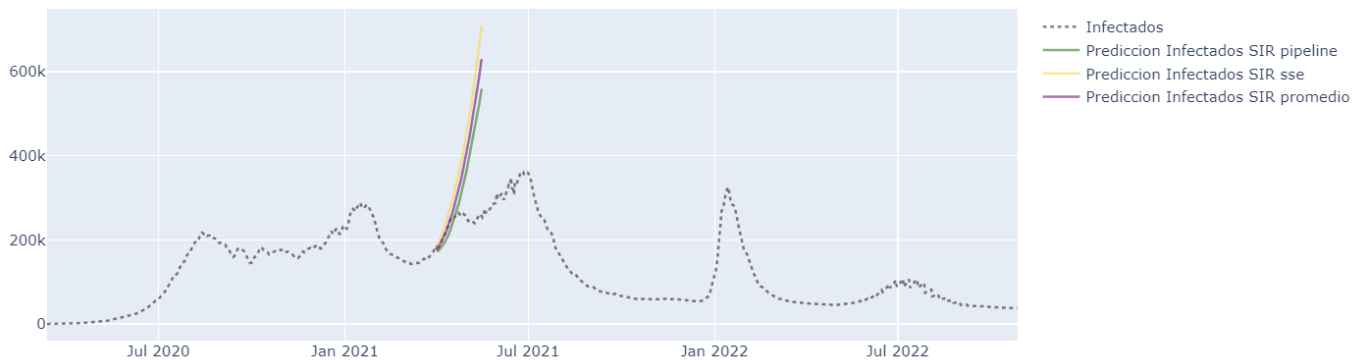
Se identifica que, para el trimestre, aumenta el número de contagios dando lugar al segundo pico de Colombia. En este caso, el modelo que mejor explica el comportamiento en el periodo de evaluación es el Promedio. Sin embargo, no se predice el aumento del número de casos de los primeros 10 días, se ajusta al final del periodo de predicción.



**Figura 11.** Representación trimestre III del año 2020 de la predicción Colombia a nivel global.

La Figura 12 presenta los infectados acumulados para el trimestre I del 2021 para tres escenarios. El R2 o coeficiente de determinación para estos modelos son los siguientes: 1) SSE-SIR: 0,405, 2) Regresión Lineal: 0,761 y 3) Promedio: 0,617.

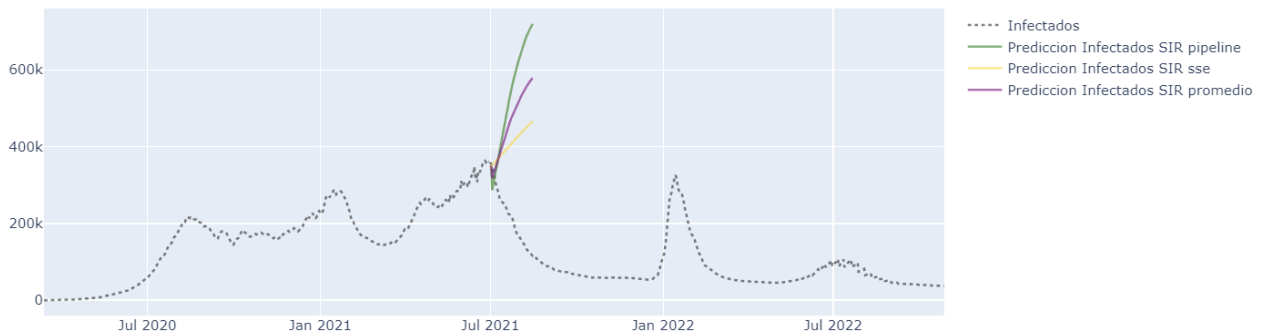
Se identifica que, para este trimestre, aumenta el número de contagios. En este caso, el modelo que mejor explica el comportamiento en el periodo de evaluación es el de Regresión Lineal, si bien su predicción es acertada hasta el 20 de abril de 2021, luego se mantiene la representación de forma continua en un crecimiento exponencial.



**Figura 12.** Representación trimestre I del año 2021 de la predicción Colombia a nivel global.

La Figura 13 relaciona los infectados acumulados para el trimestre III del 2021 para tres escenarios. El R2 o coeficiente de determinación para estos modelos son los siguientes: 1) SSE-SIR: 0,537, 2) Regresión Lineal: -0,356 y 3) Promedio: 0,194.

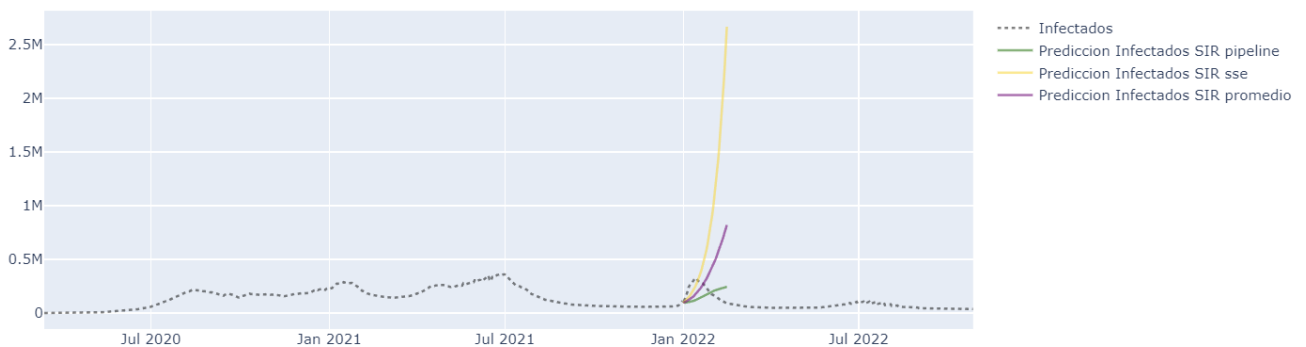
Una vez finaliza el aumento de casos a nivel nacional del mes de junio del año 2021, el coeficiente de determinación para el SSE-SIR refleja mayor bondad de ajuste. No obstante, el resultado no representa la variabilidad en el comportamiento a nivel Nacional del Covid-19, esto debido a las acciones de mitigación por parte del gobierno Nacional.



**Figura 13.** Representación trimestre II del año 2021 de la predicción Colombia a nivel global.

La Figura 14 presenta los infectados acumulados para el trimestre IV del 2021 para tres escenarios. El R2 o coeficiente de determinación para estos modelos son los siguientes: 1) SSE-SIR: -6,144, 2) Regresión Lineal: 0,893 y 3) Promedio 0,286.

En este periodo de tiempo, para la evaluación del cuarto trimestre, el modelo de Regresión Lineal fue el más acertado, el ajuste del modelo a la variable es significativo y, por lo tanto, se considera en este periodo el modelo más fiable.



**Figura 14.** Representación trimestre IV del año 2021 de la predicción Colombia a nivel global.

### 4.6.1 Caso 1

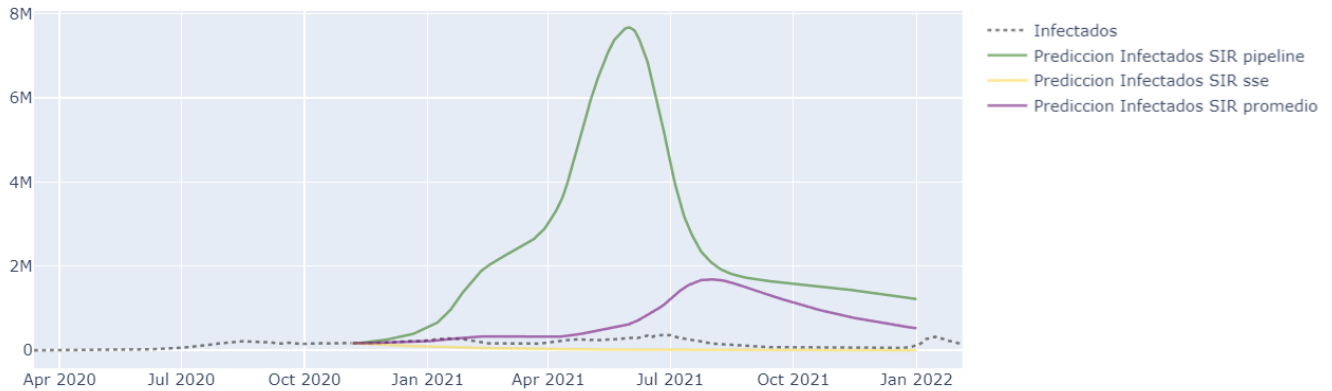
En el estudio realizado por Díaz (2020b), se utilizó el modelo ARIMA (0,1,0) para realizar pronósticos sobre la situación de contagios de COVID-19. Según el pronóstico obtenido, se estimó que para finales de diciembre de 2021 habría un total de 4.973.547 personas contagiadas, 4.784.987 personas recuperadas y 110.159 personas fallecidas.

En la Tabla 13 se presentan los resultados de los pronósticos de contagios. Se muestran los datos reales, los datos proporcionados por el autor, los resultados del método propuesto SSE-SIR y los resultados obtenidos mediante el modelo de Regresión Lineal.

En comparación con los resultados presentados en el Capítulo II, se observa que el modelo adaptado SSE-SIR se acerca más a los datos reales, aunque su precisión se ve afectada debido al periodo de evaluación que abarca más de 421 días. Las predicciones realizadas por este modelo para el periodo de tiempo tienen un promedio de aproximación del 93% por mes. Por otro lado, la Figura 15 muestra las predicciones realizadas por los tres modelos, se observa que el modelo de Regresión Lineal no proporciona resultados precisos, generando una alerta de aumento de contagios entre los meses de abril y julio de 2021 que es seis veces mayor al número real. En la Tabla 14 se presentan los estimadores para los tres modelos mencionados.

**Tabla 13.** Relación modelo SSE-SIR y Regresión Lineal

Mes	Datos Infectados	ARIMA Díaz (2020)	SIR - SE	SIR-Pipeline
Enero	243.178	1.914.750	192.459	1.577.294
Febrero	393.494	2.171.176	388.234	2.280.892
Marzo	577.932	2.455.076	585.978	3.012.365
Abril	837.681	2.729.819	783.871	5.761.782
Mayo	1.127.415	3.013.719	980.108	7.667.797
Junio	1.490.170	3.288.461	1.173.058	4.442.483
Julio	1.660.813	3.572.362	1.361.000	2.082.956
Agosto	1.752.064	3.856.262	1.542.477	1.696.103
Septiembre	1.821.137	4.131.004	1.716.520	1.580.436
Octubre	1.882.765	4.414.905	1.881.899	1.481.734
Noviembre	1.941.278	4.689.647	2.038.179	1.346.795
Diciembre	1.997.934	4.973.547	2.187.472	1.217.068



**Figura 15.** Representación de la clase Infectados del modelo SSE-SIR y Regresión Lineal, Caso 1.

**Tabla 14.** Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 1.

Modelo	Grupo epidemiológico	MSE	R <sup>2</sup>
Pipeline	Susceptibles	554,666	-96,616
	Infectados	929,527	-670,250
	Recuperados	469,611	-82,752
SSE-SIR	Susceptibles	5,27	0,077
	Infectados	2,55	-0,853
	Recuperados	4,8	0,148
Promedio	Susceptibles	13,396	-1,357
	Infectados	54,165	-38,114
	Recuperados	9,332	-0,664

#### 4.6.2 Caso 2

Ramírez et al. (2020) se presenta tres (3) modelamientos aplicados en diferentes periodos de evaluación. Para abordar el comparativo con el modelo de Regresión Lineal, la medida de fiabilidad del modelo aplicado será el coeficiente de determinación. A continuación, se presentan los resultados:

##### Modelo entre marzo y abril:

Los resultados para este periodo de tiempo, con mejor ajuste según el R<sup>2</sup> o Coeficiente de Determinación, fue el modelo Gaussiano. En la Tabla 16, se observan los resultados para las medidas propuesta en esta tesis de maestría. Para este modelo los resultados fueron infectados 0,868 y Recuperados 0,897. A partir de la revisión anterior en el Capítulo II, el método SSE-SIR indica mejores resultados de bondad de ajuste para las clases epidemiológicas tanto en el R<sup>2</sup> como en el Error Cuadrático Medio (RMSE). Las medidas de Regresión Lineal no se ajustan como en el caso del modelo SSE-SIR, en esta revisión el primer modelo adaptado tiene mayor precisión en comparación a los resultados del trabajo de Ramírez et al., 2020.

**Tabla 15.** Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 2. Modelo para el periodo entre Marzo y abril del 2020.

<b>Modelo</b>	<b>Grupo epidemiológico</b>	<b>MSE</b>	<b>R<sup>2</sup></b>
Pipeline	Susceptibles	0,0001	0,237
	Infectados	0,0061	0,216
	Recuperados	1,562	0,274
SSE-SIR	Susceptibles	1,57	0,991
	Infectados	0,00016	0,979
	Recuperados	1,819	0,991
Promedio	Susceptibles	4,912	0,730
	Infectados	0,001	0,750
	Recuperados	6,734	0,687

### Modelos entre marzo y julio

Durante este periodo de tiempo, se observa que el modelo Gaussiano muestra un mejor ajuste, según el R<sup>2</sup> o Coeficiente de Determinación. En la Tabla 17 se presentan los resultados por estimador para cada modelo identificado, donde se obtienen valores de 0,868 para los infectados y 0,897 para los recuperados. Al realizar la revisión del método adaptativo SSE-SIR, se encontraron resultados significativos y un mejor ajuste a los datos, en comparación con los resultados del trabajo de investigación y del modelo de Regresión Lineal.

**Tabla 16.** Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 2. Modelo para el periodo entre Marzo y julio del 2020.

<b>Modelo</b>	<b>Grupo epidemiológico</b>	<b>MSE</b>	<b>R<sup>2</sup></b>
Pipeline	Susceptibles	0,001	0,990
	Infectados	1,170	0,630
	Recuperados	0,006	0,90
SSE-SIR	Susceptibles	0,005	0,970
	Infectados	0,900	0,717
	Recuperados	0,0007	0,989
Promedio	Susceptibles	0,001	0,989
	Infectados	0,890	0,717
	Recuperados	0,003	0,952

### Modelo entre marzo y agosto

Los resultados para este periodo de evaluación indica que el modelo Gaussiano es el más acertado con resultados para la clase de Contagiados de 0,971 y de Recuperados de 0,995. En la Tabla 18 se identifica que el modelo SSE-SIR presenta resultados más significativos en comparación con los resultados del modelo Gaussiano y del modelo de Regresión Lineal.

**Tabla 17.** Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 2. Modelo para el periodo entre Marzo y Agosto del 2020.

Modelo	Grupo epidemiológico	MSE	R <sup>2</sup>
Pipeline	Susceptibles	0,0001	0,990
	Infectados	1,010	0,760
	Recuperados	0,008	0,970
SSE-SIR	Susceptibles	3,060	0,990
	Infectados	0,010	0,990
	Recuperados	0,00016	0,990
Promedio	Susceptibles	6,014	0,990
	Infectados	0,250	0,930
	Recuperados	0,002	0,990

### 4.6.3 Caso 3

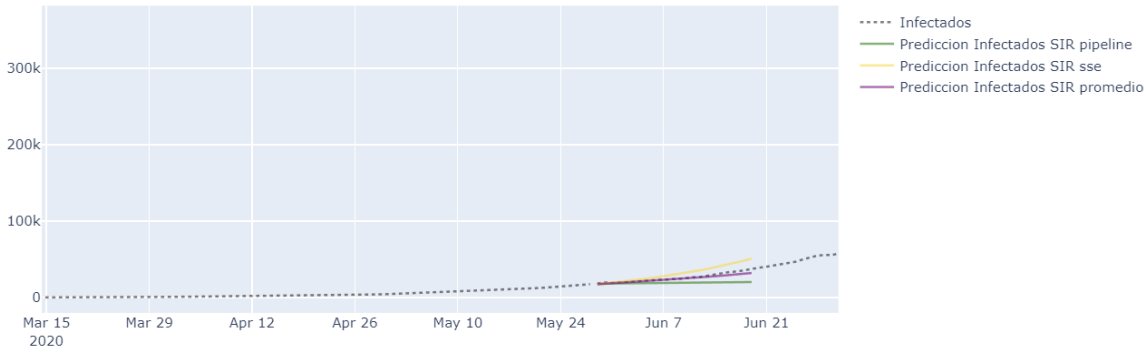
En el estudio realizado por Bravo et al. (2020) se presenta la predicción de casos confirmados de la enfermedad Covid-19 en Colombia. En este estudio, se utiliza el parámetro del Error Cuadrático Medio Normalizado (NRMSE), el cual alcanza un valor de 3,5%. A partir de los resultados del Capítulo II Caso 3, se observa que el modelo SSE-SIR muestra un valor más bajo en comparación, lo que demuestra su utilidad en la predicción de casos confirmados relacionados con la pandemia. Este resultado se corrobora adicionalmente con el Coeficiente de Determinación (R<sup>2</sup>) y el Error Cuadrático Medio (RMSE), en comparación con el modelo aplicado por Bravo et al. (2020) y el modelo de Regresión Lineal. Estos resultados se presentan en la Tabla 19, que muestra los resultados por estimador para cada modelo y grupo epidemiológico. Además, en la Figura 16 se muestran los resultados de la clase de infectados y las predicciones asociadas a cada modelo.

**Tabla 18.** Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 3.

Modelo	Grupo epidemiológico	MSE	R <sup>2</sup>
Pipeline	Susceptibles	0,0003	0,758
	Infectados	0,006	0,847
	Recuperados	9,433	0,591
SSE-SIR	Susceptibles	7,447	0,995



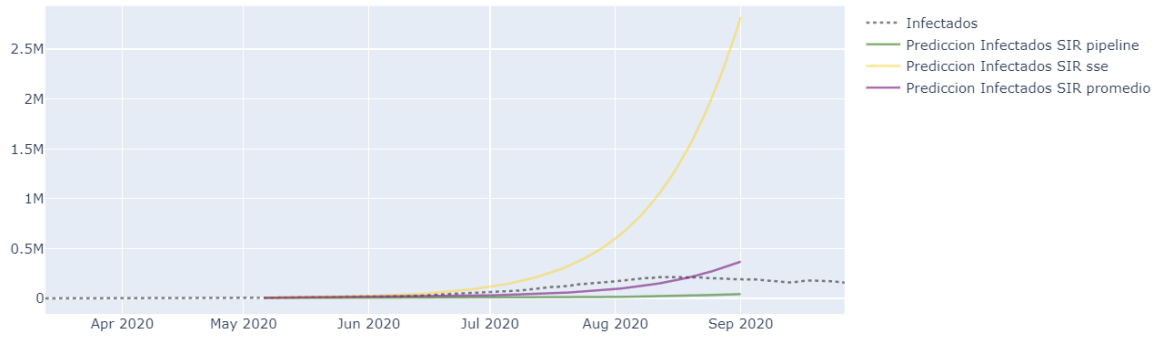
	Infectados	0,0001	0,977
	Recuperados	3,095	0,985
Promedio	Susceptibles	7,812	0,941
	Infectados	0,0004	0,989
	Recuperados	4,794	0,792



**Figura 16.** Representación de la clase Infectados del modelo SSE-SIR y Regresión Lineal, Caso 3.

#### 4.6.4 Caso 4

Díaz (2020a) presenta la predicción utilizando el modelo lineal de Brown, teniendo en cuenta el registro de la primera persona infectada el 6 de marzo de 2020 y la última fecha considerada el 5 de mayo de 2020, y su respectiva predicción para los meses de mayo, junio y julio. Los resultados preliminares del Capítulo II indican que, el método SSE-SIR, proporciona un alto nivel de precisión para pronósticos. En la tabla 21 se identifican los estimadores para cada modelo y para cada grupo epidemiológico, se resalta el Promedio con los mejores resultados para este periodo de tiempo con un coeficiente de determinación de: 1) Susceptible: 0,738, 2) Infectado: 0,677 y 3) Recuperado: 0,436. Estos valores se pueden identificar en la figura 17, en donde la predicción del modelo SIR-SSE aumenta de forma exponencial siendo la que mayor diferencia presenta frente a los resultados reales.



**Figura 17.** Representación de la clase Infectados del modelo SSE-SIR y Regresión Lineal, Caso 4.

**Tabla 19.** Resultados por estimador para el modelo SIR- SE y Regresión Lineal, Caso 4.

Modelo	Grupo epidemiológico	MSE	R <sup>2</sup>
Pipeline	Susceptibles	0,066	-0,318
	Infectados	1,146	-0,336
	Recuperados	0,024	-0,299
SSE-SIR	Susceptibles	0,044	0,291
	Infectados	1,363	-0,248
	Recuperados	0,010	0,534
Promedio	Susceptibles	0,013	0,738
	Infectados	0,276	0,677
	Recuperados	0,010	0,436

## 4.7 Conclusiones

En este capítulo, se propuso un modelo adaptado bajo las relaciones entre los países utilizando solamente las tasas de contagio durante la emergencia sanitaria y omitiendo otras variables que explicarían la dinámica política, social y económica que atraviesan los países. Cada país víctima de la pandemia generó acciones drásticas con el fin de mitigar los efectos del Covid-19; los resultados de estas acciones se pueden visualizar también en las tasas de contagio y, aunque, las medidas de prevención fueron determinadas en cada país a partir del impacto de la enfermedad, se presentan relaciones sustanciales entre la dinámica de Colombia y otros países.

El modelo adaptado aplicado en la presente tesis de maestría explora por medio de la tasa de infección hasta 40 días de retraso para evitar tendencias. Los resultados arrojan que este modelo adaptado con base en dinámicas externas a la de Colombia presenta aciertos en ciertos periodos de tiempo, como en la predicción del trimestre III del 2020, que predice un aumento sustancial del número de casos en los primeros 10 días, lo que permitiría en un tiempo de contingencia alertar un aumento significativo de contagios para tomar las medidas de prevención. Así mismo, ocurre con el periodo de evaluación del trimestre I del 2021 en el que el modelo adaptado asocia un aumento significativo de contagios para este periodo al igual que el modelo SSE-SIR.

El modelo adaptado se caracterizó por tener una aproximación en los aumentos significativos de la pandemia, aunque la interpretación predice un mayor impacto a la real, lo que es pertinente en tiempos de contingencia para tomar medidas de acción, pero los resultados provocarían una mayor crisis a nivel nacional que impactaría la economía, política, desempleo y salud.

A partir de lo anterior, se concluye que, el modelo adaptado SSE-SIR revisado en el Capítulo II, se adapta mejor al comportamiento y a la dinámica de Colombia, arrojando mejores resultados en periodos de alcance corto.

# **5. Pronóstico de la evolución del Covid-19 a nivel nacional y departamental**

## **5.1 Introducción**

El análisis de predicción del Covid-19 a nivel Nacional y departamental motiva para mitigar riesgos y actuar en el momento preciso. Durante la pandemia se evidenció que las medidas no pueden enfrentarse de igual forma en todos los lugares y es por eso que, los análisis con desagregación a nivel territorial son útiles.

El Covid-19 llega en diferente orden a cada departamento, inicia con el departamento del Huila el 12 de marzo de 2020, mientras que para el departamento de Amazonas el reporte de casos inicia el 22 de abril de 2020, siendo uno de los últimos departamentos con inicios de reporte de contagios.

En este capítulo se hace la revisión de casos aplicando el modelo adaptado SSE-SIR validado y contratado como el más preciso frente a los datos reales. Este modelo se aplica a nivel general, para los picos de contagio reportados en el transcurso de la pandemia y para algunos departamentos seleccionados. Los departamentos seleccionados presentaron durante la pandemia un mayor reporte de contagios, como lo son Antioquia, Cundinamarca, Boyacá, Meta y Santander que, para mediados de junio del año 2020, acogían alrededor del 80.1% de los contagiados. Los resultados arrojados en este capítulo, permitirán verificar si el modelo puede llegar a predecir y encontrar persistencias, disminución y aumento del número de contagios frente a los datos reales.

## **5.2 Modelo Aplicado**

Para este Capítulo, se aplica el modelo presentado en el Capítulo II.

## **5.3 Análisis y resultados**

A partir de la revisión de casos del Capítulo II, se retoma el modelo adaptado SSE-SIR para realizar la estimación para cada clase epidemiológica a partir de los picos de contagios anunciados por las entidades competentes desde el año 2020.

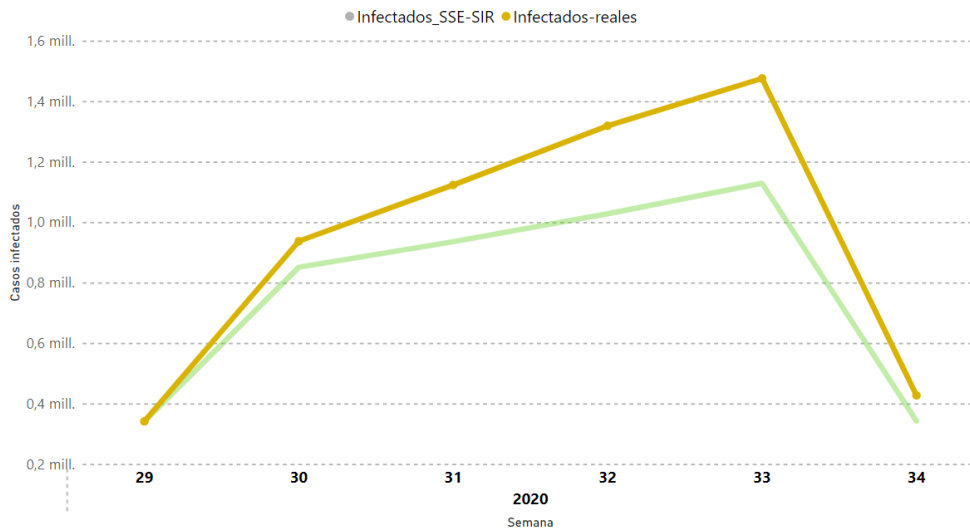
El análisis del modelo por departamentos permite identificar si el método expuesto se adapta al nivel departamental con el fin de pronosticar la tasa semanal y encontrar persistencias, disminución y aumento del número de contagios.

### 5.3.1 Colombia

Inicialmente, se representa los datos a nivel de Colombia para los puntos críticos denominados picos de contagios del Covid-19. La Tabla 23 expone el coeficiente de determinación para identificar el ajuste a los datos y la bondad de ajuste de las cinco (5) fases de alarma por contagio, dando así los siguientes resultados:

#### 5.3.1.1 Pico de contagio fase inicial

La representación del primer pico reportado se muestra en la Figura donde se presenta la estimación obtenida para cada uno de los días de avance de la enfermedad, desde el 16 de julio de 2020 hasta el 17 de agosto de 2020, lo que corresponde a un período de 32 días. Durante este tiempo, el primer pico de contagio se extendió durante 24 días, abarcando desde el 21 de julio hasta el 12 de agosto (el tiempo, 2021). El modelo SSE-SIR presenta mejores resultados y logra entender el comportamiento volátil de la variable a predecir para las clases epidemiológicas de Susceptibles, Infectados y Recuperados. En la Figura 18 se presenta la relación entre el grupo de infectados reales con respecto a la predicción del modelo adaptado, esto se puede visualizar por semana, se concluye que la predicción para los 32 días es favorable con un promedio de certeza del 83,7% frente a los datos reales.

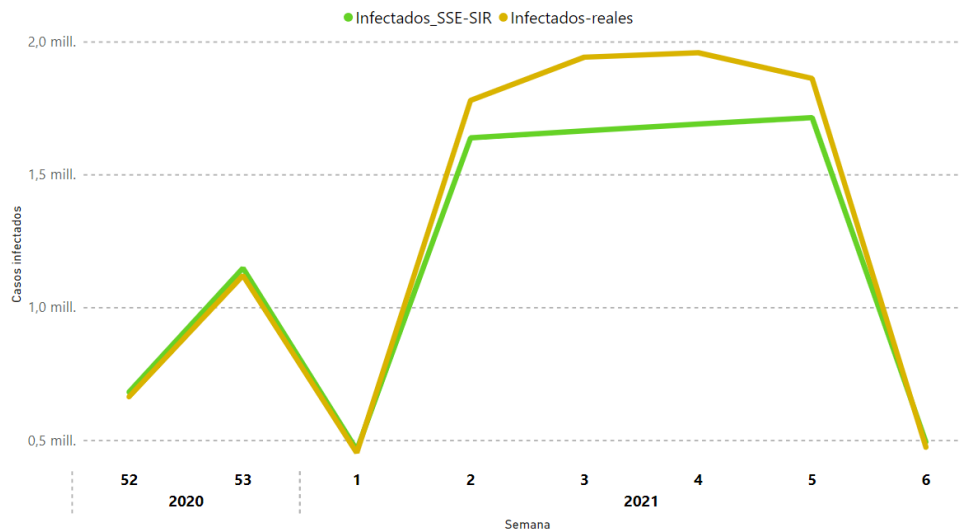


**Figura 18.** Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio inicial.

#### 5.3.1.2 Pico de contagio 2

Para representar el segundo pico reportado se muestra la figura donde se presenta la estimación obtenida para cada uno de los días de avance de la enfermedad para el pico 2 a partir del 24 de diciembre del 2020 hasta el 01 de febrero de 2021 (39 días). Lo anterior debido a que la notificación de este segundo pico indica un aumento definitivo el 29 de diciembre de 2020 y comienza a descender a partir del 27 de enero de 2021 (el tiempo, 2021). El método adaptado SSE-SIR se ajusta y presenta estabilidad para las tres clases epidemiológicas (Susceptibles, Infectados y Recuperados). En la Figura 19 se presenta la relación entre el grupo de infectados reales con respecto a la predicción del modelo

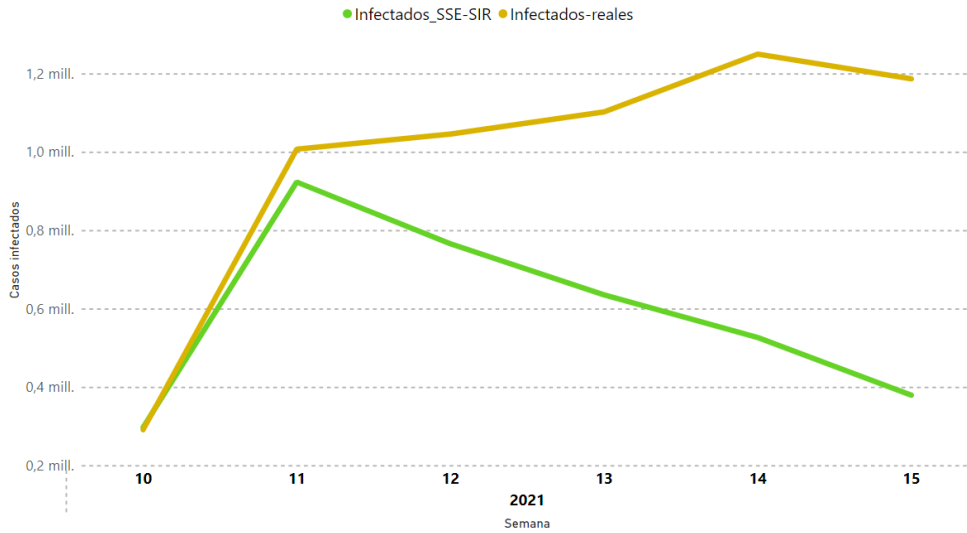
adaptado, esto se puede visualizar por semana, se concluye que la predicción para los 39 días es precisa con un promedio de certeza del 93.3% frente a los datos reales.



**Figura 19.** Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio II.

### 5.3.1.3 Pico de contagio 3

Para representar el tercer pico reportado, se muestra la Figura, que presenta la estimación obtenida para cada uno de los días de avance de la enfermedad desde el 05 de marzo de 2021 hasta el 09 de abril de 2021, lo que corresponde a un período de 35 días. Estas estimaciones son una representación del comportamiento de la enfermedad durante ese período. Estas condiciones se tienen en cuenta debido a que la tercera ola inicia el 10 de marzo, desde esta fecha hasta el 30 de marzo de 2021. El método SSE-SIR es significativo y la predicción se relaciona con el dato real en los primeros 22 días de evaluación, para los últimos 13 días, el modelo SSE-SIR no es preciso en la predicción con relación a los datos reales. En la Figura 20 se presenta la relación entre el grupo de infectados reales con respecto a la predicción del modelo adaptado, esto se puede visualizar por semana, se concluye que la predicción para los 35 días tiene un promedio de certeza del 62.58% frente a los datos reales.



**Figura 20.** Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio III.

#### 5.3.1.4 Pico de contagio 4

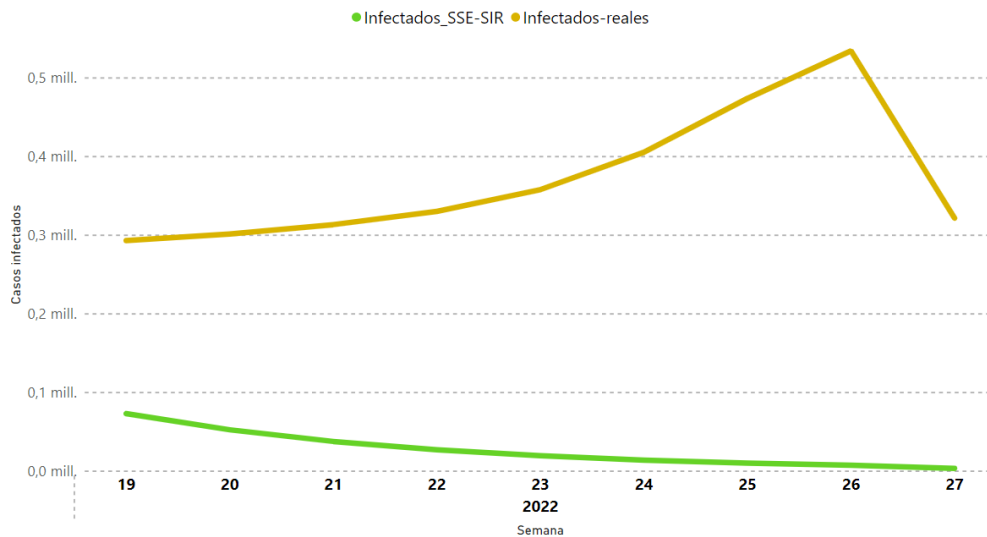
Para representar el cuarto pico reportado se presenta la figura, que muestra la estimación obtenida para cada uno de los días de avance de la enfermedad entre el 26 de noviembre de 2021 hasta el 31 de diciembre (35 días). Aunque no se reportan muchas notificaciones alrededor de este pico, se identifica un aumento significativo a partir del mes de diciembre del año 2021. En este periodo de evaluación, el método SSE-SIR no se ajustó a las clases susceptibles y recuperados. Mientras que, para la clase de infectados la relación con el dato real es más acertada para los primeros 24 días de evaluación, para los últimos 11 días, el modelo SSE-SIR no es preciso en la predicción con relación a los datos reales. En la Figura 21 se presenta la relación entre el grupo de infectados reales con respecto a la predicción del modelo adaptado, se concluye que la predicción para los 35 días tiene un promedio de certeza del 54,14% frente a los datos reales.



**Figura 21.** Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio IV.

### 5.3.1.5 Pico de contagio 5

A continuación, se representa el quinto pico reportado con la estimación obtenida para cada uno de los días de avance de la enfermedad a partir del 01 de mayo de 2022 hasta el 29 de junio (59 días). Estas condiciones se tienen en cuenta debido a que esta última ola inicia el 6 de mayo, se estabiliza, forma una meseta desde el 24 de junio hasta el 15 de julio, y de ahí, un descenso hasta el 11 de agosto del 2022. Para este caso, el modelo adaptado SSE-SIR representa la variable real y su estimación es óptima para el caso de los infectados, sobre todo para los 23 días de evaluación. En la Figura 22 se presenta la relación entre el grupo de infectados reales con respecto a la predicción del modelo adaptado, se concluye que la predicción para los 59 días tiene un promedio de certeza del 69,6% frente a los datos reales.



**Figura 22.** Representación semanal de los datos reales y los datos de infectados del modelo adaptado SSE-SIR. Pico de contagio V



### 5.3.2 Departamento

En la representación de los datos a nivel departamental se consideran, al igual que en el punto anterior, los puntos críticos de contagios del Covid-19. En la Tabla 23 se muestran los coeficientes de determinación para identificar la bondad de ajuste a los datos para cinco departamentos, estos son: 1) Cundinamarca, 2) Boyacá, 3) Santander, 4) Meta y 5) Antioquia.

La medida de bondad de ajuste o fiabilidad del modelo SSE-SIR para los departamentos en evaluación es significativa. En la Tabla 23 se presentan los coeficientes de determinación para cada clase epidemiológica y departamento a partir del pico de contagio del Covid-19 a nivel Nacional. Se identificaron en algunos periodos de contagio que la clase de infectados no presenta resultados favorables en la predicción con el método SSE-SIR. A continuación, se relacionan algunas características relacionadas con las predicciones para cada departamento:

- Cundinamarca incrementa el número de contagios del Covid-19 a partir del 23 de marzo de 2020, siendo uno de los departamentos que presentó el mayor reporte de casos de contagio. El modelo SSE-SIR es pertinente en la obtención de datos para las clases de susceptibles y Recuperados. En el caso de los infectados, el parámetro no explica el comportamiento en todos los periodos de alerta, esto se observa sobre todo en el Pico 3 que inicia el 05 de marzo de 2021. Igualmente, se puede destacar, que el modelo adaptado permitió para la fase 4 (26 de noviembre de 2021) un mayor acierto entre los datos predichos y los reales, con un promedio del 77,4%.
- El departamento de Meta comienza a acelerar el número de casos reportados a partir del 13 de abril de 2020 y a partir de esta fecha se consideró para el mes de abril como una de las regiones con mayor número de contagios. El modelo SSE-SIR, aplicado para este departamento es preciso, sobre todo para las clases Susceptibles y los recuperados. La clase de infectados, según el coeficiente de determinación, no logra ajustarse a las variables sobre todo para la fase 2 que inicia el 24 de diciembre de 2020. Igualmente, se puede destacar que la fase 4 presenta un mayor acierto entre los datos predichos y los reales, con un promedio del 92%.
- El departamento de Boyacá presenta mayor reporte de contagios a partir del 25 de marzo de 2020. El modelo adaptado indica que, a nivel de departamento, se pueden generar predicciones con mayor fiabilidad para los casos de susceptibles y Recuperados. En el caso del reporte de infectados, se puede observar una mayor relación entre los datos reales y los predichos, esto se presenta, al igual que en los otros departamentos, en el pico 4 con un promedio de acierto de 83,5%.
- El departamento de Santander presenta un mayor reporte de casos a partir del 16 de marzo de 2020. Para este caso el modelo SSE-SIR predice para las clases de

susceptibles y Recuperados con una mayor fiabilidad, con excepción de los infectados. Este grupo, para el pico 3 presenta un porcentaje promedio de acierto del 34% entre los datos predichos y los reales en el periodo de evaluación.

Finalmente, el departamento de Antioquia presenta un mayor número de contagios a partir del 19 de marzo de 2020, este al igual que otros departamentos, fueron considerados como las regiones de mayores casos de contagio. Para este caso el modelo adaptado SSE-SIR tiene resultados óptimos en los picos de la pandemia, con excepción de la clase de infectados que presenta mayor variabilidad, sobre todo para el pico 2. El pico 4, presenta un mayor porcentaje de acierto para los infectados con el 29,5% entre los datos predichos y los reales en el periodo de evaluación.

**Tabla 20.** Resultados por estimador  $R^2$  o coeficiente de determinación para cinco departamentos.

Pico de contagio	Clase epidemiológica	Antioquia	Boyacá	Cundinamarca	Meta	Santander
Fase 1	Susceptibles	0,997	0,962	0,976	0,932	0,960
	Infectados	0,769	0,938	0,741	0,924	0,930
	Recuperados	0,911	0,986	0,964	0,928	0,993
Fase 2	Susceptibles	0,988	0,995	0,994	0,992	0,997
	Infectados	0,018	0,501	0,520	0,004	0,215
	Recuperados	0,991	0,987	0,984	0,999	0,996
Fase 3	Susceptibles	0,915	0,985	0,976	0,988	0,975
	Infectados	0,904	0,021	0,098	0,442	0,529
	Recuperados	0,923	0,991	0,996	0,991	0,997
Fase 4	Susceptibles	0,855	0,784	0,743	0,940	0,985
	Infectados	0,366	0,222	0,344	0,380	0,014
	Recuperados	0,998	0,990	0,976	0,996	0,988
Fase 5	Susceptibles	0,913	0,846	0,874	0,819	0,882
	Infectados	0,924	0,729	0,807	0,542	0,796
	Recuperados	0,847	0,961	0,938	0,963	0,909

## 5.4 Conclusiones

Las predicciones a un nivel de desagregación territorial permiten generar acciones más precisas como las acciones para reducir el impacto, brindar de forma eficiente atención a los contagiados, identificar requerimientos en camas hospitalarias y unidades de cuidado intensivo (UCI), definir rutas de remisión y traslados, entre otros. Aunque los departamentos presentan una variabilidad y su dinámica no está necesariamente influenciada a la dinámica de país es necesario desagregar este tipo de ejercicios.

Inicialmente, se realiza un análisis a nivel de Colombia para cada pico de contagio, se identifica que el modelo adaptado SSE-SIR presenta mejores resultados y logra entender

el comportamiento volátil de la variable a predecir para las clases epidemiológicas de Susceptibles, Infectados y Recuperados. Igualmente, se logró identificar que los datos reales de infectados y los predichos están relacionados en todos los picos de la pandemia, con excepción del tercer pico que inicia el 05 de marzo de 2021 hasta el 09 de abril de 2021.

En el caso de los departamentos, se seleccionan cinco que se caracterizaron por tener el mayor número de registro de contagios durante el año 2020. Los resultados indican que el modelo SSE-SIR se adapta a una desagregación territorial permitiendo generar resultados acordes, donde la variable real y su estimación es óptima para los grupos de Susceptibles y Recuperados. En el caso de los infectados se presentan variaciones en todos los departamentos y para todos los picos de la pandemia; sin embargo, el cuarto pico de contagio que inicia el 26 de noviembre de 2021 es el que presenta mejores resultados frente a los valores reales y los predichos para todos los departamentos.

## 6. Conclusiones

Durante la pandemia se realizan ejercicios enfocados en pronosticar valores futuros utilizando series de tiempo, a través de herramientas como lo son los modelos matemáticos epidemiológicos. En la caracterización bibliográfica revisada en la presente tesis de maestría se encontraron publicaciones que estudian la dinámica endógena del modelo, pero no se han encontrado evidencias de usar la evolución de la enfermedad o la Historia Natural de la misma a partir del territorio en el que se presenta el brote.

Los modelos matemáticos epidemiológicos han evolucionado hasta la fecha dando lugar a diversas adaptaciones que consideran otros grupos y variables como los nacimientos, muertes y la edad o estructura social de la población. Para este trabajo el modelo epidemiológico usado como base es el modelo SIR básico que permite modelar y predecir el comportamiento de las enfermedades infecciosas. Así mismo, se adapta el modelo SIR con el Suavizado Exponencial Simple, lo que permitió adaptar un modelo que presente las características de la historia natural de la enfermedad y que posibilite la autocorrección de los betas y gamas para ajustar el pronóstico en orientación contraria a las desviaciones del pasado; posibilitando que el modelo sea capaz de auto-adaptarse a nuevos cambios en la dinámica del sistema como la apertura y cierre de fronteras, cuarentenas generales o parciales o programas de vacunación escalados.

Así mismo, se analiza un segundo modelo que parte de la regresión lineal y de la aplicación de un modelo de combinación de pronósticos entre la regresión lineal y el modelo adaptado SSE-SIR. Esto se realiza con el fin de identificar el modelo con los resultados óptimos, con mayor ajuste y con el mínimo error de predicción.

Inicialmente, el modelo adaptado SSE-SIR se analiza bajo los cinco (5) picos de contagio y bajo comparaciones con otros casos que planteaban predicciones en el mismo territorio y en diferentes periodos de tiempo. Se concluye que, el modelo adaptado presenta resultados más favorables que aquellos publicados en los casos de comparación. Además, los datos de contagios predichos presentan mayor acierto con los valores reales de la pandemia. Aun así, se halló lo siguiente:

Con relación a los periodos críticos o picos de contagios, el modelo adaptado SSE-SIR realiza una predicción de casos de susceptible, infectados y recuperados mayor a los casos real, como en el caso del primer pico de la pandemia que inició el 16 de julio de 2020 hasta el 17 de agosto del mismo año. Esto debido al periodo de imprecisión

En el caso de que el periodo de evaluación para la predicción sea amplio la medida no se ajusta del todo a la variable real.

Igualmente, en la predicción asociada al modelo de regresión Lineal se presentan aproximaciones en los aumentos significativos de la pandemia, aunque la interpretación predice un mayor impacto a la real, lo que es pertinente en tiempos de contingencia para

tomar medidas de acción, pero los resultados provocarían una mayor crisis a nivel nacional que impactaría la economía, la política, el desempleo y la salud.

Finalmente, se concluye que, el modelo adaptado SSE-SIR revisado en el Capítulo II, se adapta mejor al comportamiento y a la dinámica de Colombia, arrojando mejores resultados en periodos de alcance corto. En cuanto a la predicción a nivel departamental, los resultados indican que el modelo SSE-SIR se adapta a una desagregación territorial permitiendo generar resultados acordes, donde la variable real y su estimación es óptima para los grupos de Susceptibles y Recuperados. En el caso de los infectados se presentan variaciones en todos los departamentos y para todos los picos de la pandemia; sin embargo, el cuarto pico de contagio que inicia el 26 de noviembre de 2021 es el que presenta mejores resultados frente a los valores reales y los predichos.

## **6.1 Cumplimiento de objetivos**

### **6.1.1 Objetivo Específico 1**

Sí, se logró el objetivo específico 1 de ajustar el modelo estadístico epidemiológico SIR (Susceptible - Infectado - Recuperado) para que se adapte al comportamiento de la Historia Natural de Enfermedad. En el Capítulo 3, se desarrolló y presentó el método adaptado denominado SSE-SIR, el cual permite predecir y estimar las clases epidemiológicas de Susceptibles, Infectados y Recuperados de la enfermedad del Covid-19.

Los resultados obtenidos con este modelo fueron evaluados utilizando medidas como el coeficiente de determinación  $R^2$  y el Error Cuadrático Medio (RMSE), y se encontró que los resultados son adecuados, con un mejor ajuste a los datos reales de la pandemia en comparación con otros casos revisados.

### **6.1.2 Objetivo Específico 2**

Sí, se logró el objetivo específico 2 de proponer un mecanismo que permita ajustar la cantidad total de casos detectados, incorporando la información del desarrollo de la enfermedad infecto-contagiosa (COVID-19) de otros países a través de las tasas de contagio. Este objetivo fue desarrollado en el Capítulo 4, donde se propuso un modelo adaptado que utiliza las tasas de contagio durante la emergencia sanitaria para establecer relaciones entre los países.

Sin embargo, se observó que el modelo adaptado tendió a sobreestimar el impacto de la pandemia en comparación con la realidad. Aunque esta aproximación exagerada puede ser pertinente en tiempos de contingencia para tomar medidas de acción, también puede generar una mayor crisis a nivel nacional y afectar la economía, la política, el desempleo y la salud.

### **6.1.3 Objetivo Específico 3**

Sí, se logró el objetivo específico 3 de implementar el modelo al caso de Colombia con el fin de pronosticar la tasa semanal e identificar el nivel de acierto y desacierto de los resultados. Este objetivo fue desarrollado en el Capítulo 5, donde se realizaron predicciones a nivel de desagregación territorial para generar acciones más precisas en la reducción del impacto del COVID-19.

Se logró el objetivo de implementar el modelo adaptado al caso de Colombia, pronosticando la tasa semanal y evaluando el nivel de acierto y desacierto de los resultados.

### **6.1.4 Objetivo Específico 4**

Sí, se logró el objetivo específico 4 de adaptar el modelo a nivel departamental con el fin de pronosticar la tasa semanal y encontrar persistencias, disminución y aumento del número de contagios. Este objetivo fue desarrollado en el Capítulo 5, donde se seleccionaron cinco departamentos que se caracterizaron por tener el mayor número de registros de contagios durante el año 2020.

Los resultados obtenidos indican que el modelo SSE-SIR se adapta adecuadamente a la desagregación territorial, generando resultados acordes a la realidad. La estimación de las clases epidemiológicas de Susceptibles y Recuperados es óptima, mostrando una buena concordancia entre los valores reales y los predichos.

## 7. Referencias

- Abuhasel, K., Khadr, M., & Alquraish, M. (2022). Analyzing and forecasting COVID-19 pandemic in the Kingdom of Saudi Arabia using ARIMA and SIR models. *Comput Intell*, 38(3), 770–783. doi: 10.1111/coin.12407.
- Alahakoon, P., McCaw, J., & Taylor, P. (2022). Estimation of the probability of epidemic fade-out from multiple outbreak data. *Epidemics*, 28, 1-10. doi: 10.1016/j.epidem.2022.100539.
- Bao, K., Rong, L., & Zhang, Q. (2019). Analysis of a stochastic SIRS model with interval parameters. *American institute of mathematical sciences*, 24(9), 814-838. Doi: 10.3934/dcdsb.2019033.
- Bill & Melinda Gates Foundation. (2022). SIR and SIRS models. [https://docs.idmod.org/projects/emod-generic/en/2.20\\_a/model-sir.html](https://docs.idmod.org/projects/emod-generic/en/2.20_a/model-sir.html).
- Birrell, P., Wernisch, L., Tom, B., Held, L., Roberts, G., Pebody, R., & De Angelis, A. (2020). Efficient real-time monitoring of an emerging influenza pandemic: How feasible? *Annals of Applied Statistics*, 14(1), 74-93. DOI: 10.1214/19-AOAS1278.
- Bravo, A., Vera, M., & Huérfano, Y. (2020). Modelos matemáticos estimadores de la infección por COVID-19: Consideraciones esenciales y proyecciones en Colombia. *Revista de Salud Pública*, 22(3), 316–322. <https://doi.org/10.15446/rsap.v22n3.87813>.
- Christophersen, E. (1989). Injury control. *Injury control. American Psychologist*, 44(2), 237–241. <https://doi.org/10.1037/0003-066X.44.2.237>.
- Díaz, J. (2020). Perspectiva del COVID-19 en Colombia para el año 2021. *REPERT MED CIR*, 29 (Núm. Supl.1), 128-133. <https://revistas.fucsalud.edu.co/index.php/repertorio/article/view/1136/1411>.
- Díaz, E. (2020). Proyección de la propagación del COVID-19 en Colombia. *Revista Med*, 28(1), 11–20. <https://doi.org/10.18359/rmed.4702>.
- Hethcote, H. (2000). The Mathematics of Infectious Diseases. *SIAM Review*, 42(4), 599–653. <https://epubs.siam.org/doi/10.1137/S0036144500371907>.
- infobae. (10 de Junio de 2021). *Colombia está entre los 10 países del mundo con más fallecimientos reportados por covid-19*. Obtenido de infobae: <https://www.infobae.com/america/colombia/2021/06/17/colombia-entre-los-10-paises-del-mundo-con-mas-fallecimientos-reportados-por-covid-19/>

- Malloy, G., & Brandeau, M. (2022). When Is Mass Prophylaxis Cost-Effective for Epidemic Control? A Comparison of Decision Approaches. *Medical Decision Making*, 42(8), 1052-1063. doi: 10.1177/0272989X221098409.
- Montesinos, O., & Hernández, C. (2007). Modelos matemáticos para enfermedades infecciosas. *Salud pública Méx*, 49(3), 218-226.  
[https://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0036-36342007000300007](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342007000300007).
- Montgomery, D., Jennings, C., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. Wiley.
- M.Ross, S. (2007). *Introducción a la estadística* (Elsevier Inc.).
- Newton, E., & Reiter, P. (1992). A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ULV) insecticide applications on dengue epidemics. *American Journal of Tropical Medicine and Hygiene*, 47(6), 709–720. <https://doi.org/10.4269/ajtmh.1992.47.709>.
- Ramírez, L., Puerto, K., & López, G. (2020). Análisis de curvas de covid-19 en Colombia utilizando ajuste por mínimos cuadrados. *Ingeniare*, 29, 41-55.  
<https://dialnet.unirioja.es/descarga/articulo/8051529.pdf>.
- Santillana, M., Tuite, A., Nasserie, T., Fine, P., Champredon, D., Chindelevitch, L., . . . Fisman, D. (2018). Relatedness of the incidence decay with exponential adjustment (IDEA) model, "Farr's law" and SIR compartmental difference equation models. *Infectious Disease Modelling*, 9(3), 1-12. doi: 10.1016/j.idm.2018.03.001.
- Schwartz, I. (1992). Small amplitude, long period outbreaks in seasonally driven epidemics. *Journal of Mathematical Biology*, 30(5), 473-491. doi: 10.1007/BF00160532.
- Semana. (4 de Febrero de 2021). *El segundo pico de la pandemia de covid-19 ha sido más mortal*. Obtenido de Semana: <https://www.semana.com/vida-moderna/salud/articulo/el-segundo-pico-de-la-pandemia-de-covid-19-ha-sido-mas-mortal/202114/>
- Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., . . . Wang, F. (2020). Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med*, 8(4), 420-422. doi: 10.1016/S2213-2600(20)30076-X.