Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Implementación del Filtro Kalman a un Modelo Epidemiológico SIRD

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

Licenciado en Física Aplicada

por

Enrique Conde Parodi

Asesorado por

Dr. Jorge Velázquez Castro & Dr. Uvencio José Giménez Mujica

Puebla Pue.
Mayo 2024/ 10 Junio 2024

Título: Implementación del Filtro Kalman a un Modelo Epidemiológico SIRD
Estudiante: Enrique Conde Parodi

COMITÉ

Dra. Beatriz Bonilla Capilla
Presidente

Dr. Enrique Varela Carlos
Secretario

Dr. Oliveros Oliveros José Jacobo
Vocal

Dr. Jorge Velázquez Castro
Asesor

Dr. Uvencio José Giménez Mujica
Asesor

Índice general

Capítulo 1

Introducción

En el campo de la modelización epidemiológica, se ha vuelto necesaria una mejora continua de los modelos, como el SI, SIR, SEAIRD, SIRD [19], así como de los enfoques propuestos por figuras destacadas como John Snow y Kermack y McKendrick. Abordando la teoría de propagación de enfermedades, explicando cómo se transmiten a través de una población y cómo diversos factores afectan esta propagación, incluyendo la tasa de contacto entre individuos susceptibles e infectados, la virulencia del agente infeccioso y la inmunidad adquirida [20]. Esta necesidad de mejora refleja una amplia variedad de estrategias utilizadas para mejorar los resultados y su fiabilidad, como las técnicas de procesamiento de datos, las cuales se explorarán con más detalle más adelante. En consonancia con esta premisa, la presente tesis se dedica a la evaluación de la eficacia relativa derivada de la aplicación del filtro de Kalman en la determinación de parámetros asociados a un modelo epidemiológico.

El filtro de Kalman es reconocido por sus diversas aplicaciones, que incluyen la estimación mensual del producto interno bruto de un país, la modelización de un camino aleatorio y la predicción de posicionamiento, entre otras [7, 9, 21, 22, 23]. Siendo sometido a un análisis específico en esta investigación. Se enfocará en la aplicación del filtro de Kalman para evaluar su eficiencia al procesar datos antes de incorporarlos en el modelo epidemiológico, en contraste con la utilización exclusiva del modelo sin pretratamiento de datos mediante el filtro. A pesar de la eficacia intrínseca del modelo epidemiológico, se destaca la importancia de explorar vías para potenciar dicha eficiencia. En el contexto epidemiológico, donde la calidad de los datos influye de manera crucial en la precisión de los modelos predictivos, la inclusión del filtro de Kalman como herramienta de procesamiento podría conferir una ventaja significativa al mitigar el impacto de datos ruidosos o incompletos.

A lo largo de esta tesis, se llevará a cabo una comparativa exhaustiva de ambas metodologías, es decir, el desarrollo del análisis de los datos sin tener en cuenta el filtro, posteriormente el análisis de los datos filtrados y en consecuencia, ver cómo se comportan con el modelo mediante el análisis de conjuntos de datos reales de la pandemia de COVID-19 proporcionados por el CONAHCYT [2].

Aunque existen múltiples modelos epidemiológicos eficientes [24, 25, 26], este trabajo se basa en una variante del modelo SIR, denominada modelo SIRD, debido a la disponibilidad de datos específicos sobre defunciones. Esta elección se consideró óptima al obtener todos los datos de una misma fuente, minimizando potenciales sesgos derivados de la obtención de datos de fuentes diversas. Cabe mencionar que, aunque se haya adoptado un modelo SIRD como paso intermedio, este puede evolucionar en futuras actualizaciones hacia variantes más complejas, como un modelo SEAIRD. Los resultados obtenidos a lo largo de este estudio proporcionan una visión integral sobre la eficacia relativa de ambas metodologías.

En última instancia, esta investigación aspira a contribuir al desarrollo y optimización de herramientas de modelización epidemiológica. Se destaca la importancia de considerar enfoques diversos, como el filtro de Kalman, para mejorar la calidad de los datos y, por ende, la eficiencia de los mo-

delos en la predicción y gestión de situaciones pandémicas.

Capítulo 2

Preliminares

A continuación, presentaremos las herramientas esenciales para llevar a cabo este trabajo. En primer lugar, abordaremos el modelo epidemiológico SIR para posteriormente poder entender de manera adecuada y comprensible el modelo epidemiológico SIRD, una vez mencionado el modelo, serán proporcionados los conceptos fundamentales necesarios que respalden el filtro Kalman, un método ampliamente utilizado en la estimación y predicción de estados en sistemas dinámicos. Será descrita su estructura, destacando su relevancia en la resolución de problemas complejos. Mostrando cómo el filtro de Kalman puede ser aplicado de manera avanzada para mejorar la estimación de los parámetros del modelo utilizando datos del mundo real. Este enfoque contribuirá significativamente al entendimiento y la capacidad para abordar desafíos epidemiológicos cruciales.
La COVID-19 es la enfermedad infecciosa causada por una nueva variante del coronavirus. Tanto este nuevo virus como la enfermedad que provoca eran desconocidos antes de que estallara el brote en Wuhan (China) en diciembre de 2019 [1]. Por tanto, fue una pandemia que afectó a muchos países de todo el mundo [16]. Pese a que ya han pasado más de tres años, aún sigue evolucionando el virus. Al ser la causa de una pandemia a nivel global, se han hecho numerosos estudios de distinta índole para explicar su comportamiento y tratar de predecir los distintos tipos de impactos que dicha enfermedad tendrá. En México, siendo más específicos en Puebla, esta enfermedad afectó la ciudad de una forma tan grande que registro picos de 522 casos de ingresos a hospitales por día [2], tomando en cuenta solo la primera ola, que tuvo una duración estimada de 210 días. Las personas al momento de contraer la enfermedad y presentar los síntomas de la misma [5], tardaban aproximadamente 14 días en curarse, aunque claro, existía la probabilidad de que la persona que contraía la enfermedad no se llegase a recuperar y terminara por fallecer. Dado que era una enfermedad nueva, los casos no hacían nada más que aumentar y la población no solo dudaba que la enfermedad fuera real [3], sino que además ignoraban los protocolos de cuidados que proporcionaban las autoridades. En consecuencia a esto, el factor infeccioso jugo un papel fundamental para el desarrollo y la propagación de la enfermedad. Una forma de poder analizar el comportamiento de este tipo de enfermedades, es a través de modelos matemáticos epidemiológicos [6] basados en ecuaciones en diferencias, autómatas celulares, ecuaciones estocásticas o ecuaciones diferenciales, siendo este último de gran interés para este trabajo, debido a la importancia que este otorga para poder describir el comportamiento que presenta la epidemia en tiempo real, analizando los datos con los que se parametricen las ecuaciones.
En referencia a la relevancia de los datos, su importancia radica en corroborar si el modelo planteado es el correcto y presenta coherencia con los datos, a su vez, es destacable que los datos filtrados presentan mayor fiabilidad. Por otro lado, el filtrado de Kalman [4] posee un método clásico de estimación de estado empleado en diversas aplicaciones, como el procesamiento de señales y los sistemas autónomos, también es utilizado para resolver problemas en sistemas informáticos, como el voltaje y la temperatura [7]. Aunque en la literatura existe una abundancia de descripciones del filtrado de Kalman, estas suelen referirse principalmente a sistemas específicos, como los autónomos

de sistemas lineales con ruido gaussiano [8], este caso no será muy diferente, el filtrado tendrá un objetivo específico, eliminar en lo mayor de lo posible el ruido presentado en los datos y poder presentar una visión más óptima de la presentación de los mismos.

2.1. Modelo Epidemiológico SIR

En esta sección, se abordará el modelo epidemiológico SIR [6] y se hablará un poco de los autores responsables de su formulación. Sin embargo, el enfoque principal de nuestra investigación recae en el modelo SIRD, una versión que incorpora ajustes específicos. Por lo tanto, es esencial contextualizar nuestro estudio mencionando el modelo original, lo que nos permitirá comprender y apreciar mejor las modificaciones y mejoras introducidas en nuestra versión del modelo SIRD.

2.1.1. Clasificación de enfermedades infecciosas

Una enfermedad infecciosa es una enfermedad clínicamente evidente resultante de la presencia de un agente microbiano patógeno. El agente microbiano que causa la enfermedad puede ser bacteriano, viral, fúngico, parasitario, o puede ser proteínas tóxicas [6], llamadas priones, A su vez, hay enfermedades transmisibles, que son enfermedades infecciosas que pueden transmitirse de una persona infecciosa a otra, directa o indirectamente. A menudo, no distinguimos entre enfermedades infecciosas y enfermedades transmisibles, ya que muchas de las enfermedades infecciosas son de hecho enfermedades transmisibles. Sin embargo, hay enfermedades que son infecciosas pero no transmisibles. El tétanos es un ejemplo de este tipo de enfermedades. Las enfermedades transmisibles son enfermedades infecciosas que pueden transmitirse de una persona a otra por vías no naturales. No obstante, la distinción entre enfermedades infecciosas, enfermedades transmisibles y enfermedades contagiosas es sutil. Por tanto, la clasificación de enfermedades infecciosas se clasifica por:

- **Las enfermedades de transmisión de persona a persona**: Son enfermedades que requieren un contacto directo o contacto directo o indirecto. El contacto directo incluye el contacto físico o sexual (VIH o gonorrea). El contacto indirecto incluye el intercambio de un objeto infectado, sangre u otros fluidos corporales (como la gripe).
- **La transmisión por vía aérea**: Se produce al inhalar aire infectado (gripe, varicela o tuberculosis).
- **Las enfermedades transmitidas por los alimentos y el agua**: Se transmiten por la ingestión de alimentos o agua contaminados (cólera o la salmonela).
- **Las enfermedades transmitidas por vectores**: Son transmitidas por un vector, generalmente un artrópodo como un mosquito o una garrapata (malaria o el dengue).
- **La transmisión vertical**: Se produce cuando una enfermedad se transmite a través de la placenta de la madre al niño antes o en el momento del nacimiento (Hepatitis B o sífilis).

A efectos de modelización, distinguimos cuatro tipos de transmisión:

- Directa (persona a persona).
- Transmitida por vectores (vector a ser humano).
- Transmisión ambiental (Infección por un patógeno presente en el medio ambiente).
- Vertical.

Hay una serie de conceptos en epidemiología estrictamente relacionados con las enfermedades infecciosas:

- **Individuos susceptibles**. Cuando un individuo sano que es vulnerable a contraer una enfermedad entra en contacto con un posible transmisor de la misma, ese individuo se expone. Los individuos pueden o no desarrollar la enfermedad. Estos individuos no suelen ser infecciosos. En los modelos matemáticos, solemos suponer que todos los individuos susceptibles acaban desarrollando la enfermedad.
- **Individuos infectados e infecciosos**. Si el patógeno se establece en un individuo expuesto, este se infecta. Los individuos infectados que pueden transmitir la enfermedad se denominan infecciosos. Los individuos infectados pueden no ser infecciosos durante todo el tiempo que estén infectados.
- **Individuos latentes**. Son individuos que están infectados, pero aún no son infecciosos. El periodo de latencia se define como el tiempo que transcurre desde la infección hasta que el huésped es capaz de transmitir el agente infeccioso a otro individuo.
- **Periodo de incubación**. El periodo de incubación es el que transcurre entre la exposición a un agente infeccioso y la aparición de los síntomas de la enfermedad. En las enfermedades infecciosas, el periodo de incubación es el tiempo necesario para que el agente infeccioso se multiplique hasta el umbral necesario para producir síntomas o pruebas de laboratorio de la infección. El periodo de incubación no coincide necesariamente con el periodo de latencia. Por ejemplo, en el caso de la gripe, los individuos se vuelven infecciosos aproximadamente un día antes de presentar síntomas visibles de gripe.
- **Incidencia**. La incidencia se define como el número de personas que enferman durante un intervalo de tiempo determinado (por ejemplo, un año). A veces, la incidencia es el número de individuos que se enferman durante un intervalo de tiempo específico dividido por la población total. En la mayoría de los casos, la incidencia se determina a partir del número de casos clínicos, lo que subestima la verdadera incidencia, ya que ignora los casos sub clínicos.
- **Prevalencia**. La prevalencia de una enfermedad es el número de personas que tienen la enfermedad en un momento determinado. A veces, la prevalencia se define como el número de personas que tienen la enfermedad en un momento determinado dividido por el tamaño total de la población. tamaño.
- **Proporción de letalidad (PPC)**. La proporción de letalidad se da como la relación entre las personas que mueren de una enfermedad y las que la contraen. Por ejemplo, a fecha de 27 de junio de 2014, se han diagnosticado 667 personas con gripe aviar H5N1, y 393 de ellas han muerto. El PPC es de 0.59.
- **Mortalidad inducida por la enfermedad**. La mortalidad inducida por la enfermedad es el número de personas que han muerto por la enfermedad en una unidad de tiempo (por ejemplo, un año) dividida por la población total.

2.1.2. Derivación del modelo epidémico de Kermack-McKendrick

Cuando una enfermedad se propaga en una población, la divide en clases diferentes. En uno de los escenarios más sencillos, hay tres clases de este tipo [6, 17, 18]:

- La clase de individuos que están sanos, pero pueden contraer la enfermedad. Se denominan individuos susceptibles o susceptibles. El tamaño de esta clase se suele se denota por S.

- La clase de individuos que han contraído la enfermedad y ahora están enfermos con ella, llamados individuos infectados. En este modelo, se supone que los individuos infectados son también infecciosos. El tamaño de la clase de individuos infecciosos/infectados se denota por I.

- La clase de individuos que se han recuperado y no pueden contraer la enfermedad se denominan individuos eliminados/recuperados. La clase de individuos recuperados se suele denominar R.

Este modelo supone varias situaciones antes de comenzar con un análisis matemático:

- La población analizada en el modelo es constante y su tamaño se denota por la letra "N", es decir, que las tasas de nacimiento y muerte durante el proceso en que la epidemia es la misma, además de que el tiempo de la epidemia es corto.

- Los fenómenos demográficos no son tomados en cuenta, además de asumir que el tiempo del brote infeccioso, es lo suficientemente rápido para que las hipótesis mencionadas sean válidas.

- La población es cerrada, es decir, que no hay migraciones.

- La población está homogéneamente mezclada, ya que nuestro conjunto de susceptibles es mayor a seis millones y medio. La manera en que esta enfermedad se transmite está regida por la ley de acción de masas de la epidemiología, la cual dice que la tasa por la que una enfermedad se propaga es proporcional al número de individuos susceptibles por el número de individuos infecciosos.

- El tiempo que tarda una persona en convertirse en infeccioso desde que estuvo expuesto a la enfermedad es tan pequeño como para que no sea considerado.

- Los individuos infecciosos se convierten en recuperados con una tasa γ, donde γ es el inverso del tiempo de recuperación estimado de la enfermedad, para ser más preciso:

$$\gamma = \frac{1}{\text{tiempo de recuperación en días}}. \tag{2.1}$$

Con lo anterior en cuenta y que el número de individuos de cada una de estas clases cambia con el tiempo, es decir, $S(t), I(t)$ y $R(t)$ son funciones del tiempo t. El tamaño total de la población N es la suma de los tamaños de estas tres clases:

$$N = S(t) + I(t) + R(t). \tag{2.2}$$

Uno de los modelos más sencillos incluye la dinámica de individuos susceptibles, infecciosos y recuperados. Cuando un individuo susceptible entra en contacto con un individuo infeccioso, ese individuo susceptible se infecta con una cierta probabilidad y pasa de la clase susceptible a la clase infectada. La población susceptible disminuye en una unidad de tiempo por todos los individuos que se infectan en ese tiempo. En ese mismo instante, la clase de infectados aumenta en el mismo número de individuos recién infectados. El número de individuos que se infectan por unidad de tiempo en epidemiología se denomina incidencia, y la tasa de cambio de la clase susceptible viene dada por:

$$S'(t) = -\text{Incidencia}.$$

También debe considerarse que no necesariamente el contacto de un individuo susceptible y uno infeccioso hará que la epidemia se propague. Si se define $\lambda(t) = \beta I$, el número de individuos que se infectan por unidad de tiempo es igual a $\lambda(t)S$. La función $\beta I(t)$ se denomina fuerza de infección. El coeficiente β es la constante de proporcionalidad denominada constante de la tasa de

transmisión. El número de individuos infectados en la población $I(t)$ se denomina prevalencia de la enfermedad.

O dicho de otra forma:

$$S'(t) = -\beta S(t)I(t). \tag{2.3}$$

A su vez, el término γI indica la salida de la clase infecciosa, con γ como la tasa de ganancia de individuos recuperados:

$$I'(t) = \beta S(t)I(t) - \gamma I(t), \tag{2.4}$$

y la ecuación de los individuos recuperados:

$$R'(t) = \gamma I(t). \tag{2.5}$$

Formando un sistema de ecuaciones que será analizado en la siguiente sección, por lo que solo será presentado:

$$\begin{aligned} S' &= -\beta S(t)I(t)\,, & (2.6)\\ I' &= \beta S(t)I(t) - \gamma I(t)\,, & (2.7)\\ R' &= \gamma I(t). & (2.8) \end{aligned}$$

2.1.3. Análisis del modelo SIR

Debido a que $\forall t > 0$ se cumple que el total de la población permanece constante, la ecuación (2.2) se puede reescribir como:

$$R(t) = N - S(t) - I(t), \tag{2.9}$$

por lo que el modelo se puede resumir de la siguiente manera:

$$\begin{aligned} S'(t) &= -\beta SI, & (2.10)\\ I'(t) &= \beta SI - \gamma I. & (2.11) \end{aligned}$$

A continuación, será realizado un análisis cualitativo del sistema (2.10)-(2.11). Para una condición inicial dada por $S(0)$, $I(0)$ y $R(0)$ se tiene lo siguiente:

1. Si $I' > 0$ en el tiempo inicial t_0, entonces, $I'(0) = \beta S(0)I(0) - \gamma I(0) > 0$, de aquí es posible deducir que, $S(0) > \dfrac{\gamma}{\beta}$, lo que nos dice que el número de individuos infecciosos aumentará y habrá epidemia. Por tanto, para algún tiempo $t > 0$, existirá un brote epidémico, si $I'(t) > 0$. Dado que $I'(t) > 0$ en el tiempo t_0, se puede reescribir la ecuación (2.11) como:

   $$I'(0) = I(0)(\beta S(0) - \gamma), \tag{2.12}$$

   y además, $I'(0) = I(0)(\beta S(0) - \gamma) > 0$, lo que nos dice que $\beta S(0) > \gamma$, es decir, si $\dfrac{\beta S(0)}{\gamma} > 1$, habrá epidemia.

2. Ahora bien, si $I'(t) < 0$ en el tiempo inicial t_0, entonces, $I'(0) < 0$, de manera similar al caso 1, podemos ver que $S(0) < \dfrac{\gamma}{\beta}$, lo que nos dice que no conllevará una pandemia futura, también $S'(t) \leq S(0)$ en cualquier instante $t \geq 0$. Además de llegar a la conclusión de que $\dfrac{\beta S(0)}{\gamma} < 1$.

Estos dos cocientes encontrados ($\dfrac{\beta S(0)}{\gamma} > 1$ ó $\dfrac{\beta S(0)}{\gamma} < 1$) son conocidos como número reproductivo básico, del cuál será abordado a detalle más adelante.

2.1.4. Modelo SIRD

Cuando formulamos un modelo, tenemos que preocuparnos por las unidades de las cantidades implicadas. Las unidades de ambos lados de las ecuaciones anteriores deben ser las mismas. Todas las derivadas tienen unidades de número de personas por unidad de tiempo. Sin embargo, en este trabajo el modelo utilizado es una adaptación, aquí, además de tomar en cuenta a las personas susceptibles, infectados y recuperados, se contarán con aquellos que fallecieron en esos instantes de tiempo, denotados por la letra D. Por lo que es necesaria una adaptación a la hipótesis de (2.2):

$$N = S(t) + I(t) + R(t) + D(t), \tag{2.13}$$

en consideración con lo anterior, son obtenidas las siguientes ecuaciones:

$$\begin{aligned} S'(t) &= -\beta S(t)I(t)/N, && (2.14)\\ I'(t) &= \beta S(t)I(t)/N - \gamma I(t) - \delta I(t), && (2.15)\\ R'(t) &= \gamma I(t), && (2.16)\\ D'(t) &= \delta I(t). && (2.17) \end{aligned}$$

El modelo epidemiológico SIRD es un modelo compartimental que divide a la población en cuatro compartimentos: Susceptibles (S), Infectados (I), Recuperados (R) y Fallecidos (D). Estas son las ecuaciones que describen cómo cambia la población en cada compartimento con el tiempo:
Susceptibles (S): Personas que son susceptibles a la infección.

$$S'(t) = -\beta S(t)I(t)/N. \tag{2.18}$$

- S: Número de individuos susceptibles.
- β: Tasa de contacto, que representa la probabilidad de transmisión de la enfermedad de un infectado a un susceptible.
- I: Número de individuos infectados.
- N: Número de individuos.

Infectados (I): Personas que están actualmente infectadas.

$$I'(t) = \beta S(t)I(t)/N - \gamma I(t) - \delta I(t). \tag{2.19}$$

- β: Tasa de contacto, como se explicó anteriormente.
- γ: Tasa de recuperación, representa la velocidad a la que los infectados se recuperan o mueren.
- δ: Tasa de mortalidad, representa la proporción de infectados que fallecen.

Recuperados (R): Personas que se han recuperado de la enfermedad.

$$R'(t) = \gamma I(t). \tag{2.20}$$

- γ: Tasa de recuperación.

Fallecidos (D): Personas que han fallecido debido a la enfermedad.

$$D'(t) = \delta I(t). \tag{2.21}$$

- δ: Tasa de mortalidad.

Existe otra ecuación que es necesaria para el análisis que será realizado más adelante y estos son los casos acumulados:

$$C'(t) = \beta S(t)I(t)/N. \tag{2.22}$$

Acumulados(C): Casos acumulados, al momento de analizar una base de datos donde las personas que resultan infectadas ingresan a los hospitales, su conteo puede no ser de lo más preciso, ya que dicha base solo menciona cuantas son reportadas como infectadas por día, por ejemplo, en un día puede haber un intervalo entre 1 a 14 infectados. Sin embargo, si queremos hacer un mejor análisis de este modelo, se deberá reconsiderar a los infectados como un conjunto donde se irán añadiendo conforme se vayan reportando como infectados por día:

$$C'(t) = \beta S(t)I(t)/N. \tag{2.23}$$

Este modelo describe cómo las personas se mueven entre los diferentes compartimentos a medida que la epidemia progresa. Las tasas β, γ y δ son parámetros que determinan la velocidad de transmisión, recuperación y mortalidad de la enfermedad, respectivamente. Este modelo puede ajustarse a los datos observados para estimar los valores de β, γ y δ que mejor describen la propagación de la enfermedad en una población específica.
Si los valores de los parámetros β, γ y δ son fijos y no varían con el tiempo, uno puede asumir que estas tasas no cambian a lo largo de la evolución de la epidemia. En ese caso, el modelo SIRD se vuelve determinista y predecible, ya que no hay fluctuaciones en estas tasas. Lo cual podría tener diferentes implicaciones:

- **Predicción Constante**: Con tasas fijas, las predicciones del modelo serán constantes con el tiempo, siempre y cuando las condiciones iniciales y los parámetros no cambien.
- **Forma de las Curvas**: Si las tasas son fijas, la forma de las curvas de los compartimentos S, I, R y D será determinada principalmente por las condiciones iniciales y las tasas fijas. Las curvas no mostrarán cambios de pendiente o aceleración a lo largo del tiempo.
- **Sensibilidad a las Condiciones Iniciales**: Debido a que las tasas son fijas, el modelo será sensible a las condiciones iniciales. Pequeños cambios en las condiciones iniciales pueden tener un impacto significativo en las predicciones del modelo.
- **Limitaciones en la Captura de Cambios**: Si la realidad cambia y las tasas reales de transmisión, recuperación y mortalidad cambian con el tiempo debido a intervenciones, políticas de salud u otros factores, el modelo con parámetros fijos no podrá capturar estos cambios y sus predicciones pueden desviarse de la realidad.

  Si los valores de los parámetros son fijos y no cambian con el tiempo, el modelo SIRD será más simple y predecible, pero puede no capturar los cambios reales en la dinámica de la epidemia a medida que se desarrolla. Para reflejar mejor las condiciones cambiantes, es útil considerar modelos más complejos que permitan variabilidad en las tasas a lo largo del tiempo.
- **Identificación de Tendencias Iniciales**: El modelo SIRD permite analizar las etapas iniciales de una epidemia y predecir cómo se propagará una enfermedad infecciosa dentro de una población. Esto es fundamental para la planificación de recursos médicos y la implementación de intervenciones tempranas.
- **Entendimiento de Dinámicas Fundamentales**: El modelo SIRD desglosa la dinámica de una epidemia en componentes clave: susceptibles, infectados, recuperados y fallecidos. Esto proporciona una comprensión profunda de cómo las tasas de transmisión, recuperación y mortalidad interactúan para influir en el curso de la enfermedad.

- **Simulación de Escenarios y Políticas**: A pesar de sus simplificaciones, el modelo SIRD puede utilizarse para simular diferentes escenarios y políticas de salud pública. Al ajustar los parámetros, se pueden explorar los efectos de medidas como la cuarentena, la vacunación o el aumento de la capacidad hospitalaria.
- **Comunicación de Conceptos Epidemiológicos**: El modelo SIRD es una herramienta útil para comunicar conceptos epidemiológicos a un público amplio. Ayuda a las personas a entender cómo los cambios en el comportamiento individual y las intervenciones gubernamentales pueden afectar la propagación de una enfermedad.
- **Base para Modelos Más Complejos**: Aunque el modelo SIRD es simple, sienta las bases para modelos más complejos que pueden incorporar características adicionales, como demografía, interacción social y evolución de la inmunidad, lo que permite una mayor precisión en la predicción.
- **Alerta Temprana y Preparación**: El modelo SIRD puede proporcionar una alerta temprana sobre la magnitud de una epidemia, permitiendo a las autoridades sanitarias y a la sociedad prepararse para la respuesta y la mitigación.

2.1.5. Número de Reproducción Básico $\mathscr{R}_0$

Una vez abarcadas las definiciones que componen al modelo SIRD, lo siguiente es abarcar uno de los conceptos más fundamentales sobre la compresión e interpretación para un modelo epidemiológico, y es el número de reproducción básico, denotado comúnmente como $\mathscr{R}_0$.
Existen varios métodos para poder determinar este valor, pero en este trabajo, se utilizará la matriz de siguiente generación, antes de abarcar dicho concepto, primero es necesario abarcar ciertas definiciones y realizar un poco de álgebra.

Matrices

Se abarcarán conceptos poco a poco para un mejor entendimiento.
Definición 2.1 *Sea A una matriz de tamaño $n \times n$. La abscisa espectral de $s(A)$ se define como la máxima parte real de los valores propios de A.*
Ejemplo 2.1 Supongamos una matriz A de 2×2 como la siguiente:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

la abscisa espectral se puede obtener después de calcular los valores propios de la matriz A, dichos valores propios se determinan como las soluciones a la expresión $det(A - \lambda I) = 0$, donde I es la matriz identidad.
Para nuestro ejemplo:

$$det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 2 \\ 3 & 4 - \lambda \end{vmatrix},$$

el polinomio característico que se asocia a esta matriz es $P(\lambda) = \lambda^2 - 5\lambda - 2$, donde sus raíces son $\lambda_1 \approx 5.37$ y $\lambda_2 \approx -0.37$, al observar ambas raíces la de mayor parte real es λ_1, por lo que la abscisa espectral $s(A) = 5.37$.
Definición 2.2 *Sea A una matriz de tamaño $n \times n$. El radio espectral $\rho(A)$ se define como el máximo de las normas de los valores propios de A.*
Ejemplo 2.2 Sea A' una matriz de 2×2 de la siguiente manera:

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix},$$

de manera similar al caso de las abscisas, se calculan los valores propios de la matriz A':

$$det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 4 \\ 3 & 1-\lambda \end{vmatrix},$$

el polinomio característico que se asocia a esta matriz es $P(\lambda) = \lambda^2 - 3\lambda - 10$, donde sus raíces son $\lambda_1 = 5$ y $\lambda_2 = -2$, al observar ambas raíces la de mayor norma es λ_1, por lo que el radio espectral $\rho(A) = 5$.

Definición 2.3 *Sea A una matriz de $n \times n$. La matriz A no es negativa, si todas las entradas de A son mayores o iguales a cero, es decir, $a_{i,j} \geq 0$, para todo i, j.*

Ejemplo 2.3 La matriz B de 3×3 no es negativa:

$$\begin{bmatrix} 11 & 5 & 0 \\ 4 & 7 & 1 \\ 0 & 9 & 16 \end{bmatrix},$$

pero la matriz B' es un ejemplo de lo que **no** es una matriz no negativa:

$$\begin{bmatrix} 1 & 55 & 0 \\ 41 & 77 & -1 \\ 7 & 79 & 6 \end{bmatrix}.$$

Definición 2.4 *Sea A una matriz de tamaño $n \times n$. Si la determinante de A es distinto de cero, entonces A es una matriz no singular, es decir, $det(A) \neq 0$. En caso de ser cero, será una matriz singular $det(A) = 0$.*

Ejemplo 2.4 Supongamos una matriz C de 3×3 que se vea de la siguiente forma:

$$\begin{bmatrix} -2 & 4 & 5 \\ 6 & 7 & -3 \\ 3 & 0 & 2 \end{bmatrix},$$

el determinante de C, $det(C) = 217$, por lo que es una matriz no singular. Pero la matriz C' dada por

$$\begin{bmatrix} 3 & 4 & 2 \\ 6 & 8 & 4 \\ 0 & 2 & 5 \end{bmatrix},$$

tiene un determinante igual a cero, por lo que lo hace una matriz singular. Existe otra forma de corroborar si es una matriz singular o no, esto es, verificando si los renglones o columnas son linealmente dependientes.

Definición 2.5 *Sea A una matriz de tamaño $n \times n$. A es una matriz de signo Z, si las entradas fuera de la diagonal principal son menores o iguales a cero, es decir, $a_{i,j} \leq 0$ para todo $i \neq j$.*

Ejemplo 2.5 Veamos la matriz D:

$$\begin{bmatrix} 6 & -8 & -1 \\ -2 & -9 & -3 \\ -4 & -5 & 1 \end{bmatrix},$$

dado que todas las entradas que no están en la diagonal principal son menores a cero, la matriz D es una matriz de signo Z.

Dado que ya están definidas las matrices con signo Z, ahora se pueden definir las M-matrices.

Definición 2.6 *La matriz A es una M-matriz, si tiene un patrón de signo Z y todos los valores propios de A tienen parte real positiva.*

Ejemplo 2.6 Considerando la matriz D':

$$\begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix},$$

es una matriz de signo Z y su polinomio característico es $P(\lambda) = \lambda^2 - 6\lambda + 8$, las raíces del polinomio característico son $\lambda_1 = 4$ y $\lambda_2 = 2$, ambas raíces son positivas por lo que D' es una M-matriz.

Definición 2.7 *Sea A una matriz de tamaño $n \times n$. Sí $A = sI - B$, donde B es una matriz de $n \times n$ no negativa, I la matriz identidad y $s > \rho(B)$, entonces A es una M-matriz no singular.*

Ejemplo 2.7 Sea B una matriz no negativa de 2×2:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

su radio espectral $\rho(B) = 2$. Considerando la matriz D' del ejemplo 2.6 (referenciar):

$$D' = sI - B = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix},$$

donde $s = 4 > \rho(B)$ y $det(D') = 7$. Por lo que D' es una M-matriz no singular. En el caso particular de $s = \rho(B)$, entonces A sería una M-matriz singular. También es importante notar que, cuando la matriz A tiene un patrón de signo Z y además es una M-matriz, entonces A^{-1} es una matriz no negativa.

Ejemplo 2.8 Una vez más, considerando D':

$$\begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix},$$

con $det(D') = 7$, entonces D'^{-1}:

$$\begin{bmatrix} \frac{4}{7} & \frac{1}{7} \\ \frac{1}{7} & \frac{2}{7} \end{bmatrix},$$

se puede observar que todos los valores de D'^{-1} son mayores a cero, por tanto, D'^{-1} es no negativa.

Considerando todo lo visto anteriormente, ahora es posible entender el concepto de la matriz de próxima generación, definida como:

$$K = FV^{-1}, \tag{2.24}$$

donde el número reproductivo básico $\mathscr{R}_0$ es el radio espectral de la matriz FV^{-1}, es decir:

$$\mathscr{R}_0 = \rho(FV^{-1}). \tag{2.25}$$

Con el sistema de ecuaciones diferenciales (2.14)-(2.17), tiene puntos de equilibrio dados por la solución del sistema:

$$\begin{aligned} -\beta S(t)I(t)/N &= 0, && (2.26)\\ \beta S(t)I(t)/N - \gamma I(t) - \delta I(t) &= 0, && (2.27)\\ \gamma I(t) &= 0, && (2.28)\\ \delta I(t) &= 0, && (2.29) \end{aligned}$$

de (2.28) $I(t) = 0$, dicho de otra forma no hay individuos infectados, si no hay infectados, tampoco hay personas que deban recuperarse o fallezcan de dicha enfermedad, también, es correcto decir que la población es constante ($S'(t) + I'(t) + R'(t) + D'(t) = 0$), por tanto, la ecuación (2.13) cuando $I(t) = 0$ resulta en $N = S$, dando un punto de equilibrio para (N, 0, 0, 0). Lo siguiente es calcular la estabilidad del equilibrio, para esto, es necesario calcular los valores propios de la

matriz Jacobiana del sistema:

$$J(S,I,R,D) = \begin{bmatrix} -\frac{\beta I}{N} & -\frac{\beta S}{N} & 0 & 0 \\ \frac{\beta I}{N} & \frac{\beta S}{N} - \gamma - \delta & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & \delta & 0 & 0 \end{bmatrix}, \tag{2.30}$$

y evaluando en el punto de equilibrio, se tiene:

$$J(N,0,0,0) = \begin{bmatrix} 0 & -\beta & 0 & 0 \\ 0 & \beta - \gamma - \delta & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & \delta & 0 & 0 \end{bmatrix}, \tag{2.31}$$

con $\lambda_1 = 0$, $\lambda_2 = 0$, $\lambda_3 = 0$ y $\lambda_4 = \beta - \gamma - \delta$ valores propios, como hay cuatro valores en cero, no se puede asegurar que haya un equilibrio estable, pero para seguir avanzando, se supondrá que es estable. Ahora, para el cálculo de $\mathscr{R}_0$, solo es necesario reacomodar el sistema de ecuaciones (2.14)-(2.17) y calcular las funciones de $\mathcal{F}_i(x)$ (el factor de aparición de las nuevas infecciones en el compartimiento i), $\mathcal{V}_i^+(x)$ (el factor de transferencia de individuos por todos los otros medios que no son nuevas infecciones al compartimiento i) y $\mathcal{V}_i^-(x)$(el factor de transferencia de salida por todos los otros medios que no son nuevas infecciones del compartimiento i):

$$\begin{aligned} I'(t) &= \beta S(t)I(t)/N - \gamma I(t) - \delta I(t), \\ S'(t) &= -\beta S(t)I(t)/N, \\ R'(t) &= \gamma I(t), \\ D'(t) &= \delta I(t). \end{aligned}$$

Por lo que las funciones $\mathcal{F}_i(x)$, $\mathcal{V}_i^+(x)$ y $\mathcal{V}_i^-(x)$ se ven de la siguiente manera:

$$\mathcal{F}(I,S,R,D) = \left(\frac{\beta SI}{N}, 0, 0, 0\right)^T, \tag{2.32}$$

$$\mathcal{V}^+(I,S,R,D) = (0,0,\gamma I, \delta I)^T, \tag{2.33}$$

$$\mathcal{V}^-(I,S,R,D) = \left(\gamma I + \delta I, \frac{\beta SI}{N}, 0, 0\right)^T, \tag{2.34}$$

$$\mathcal{V}(I,S,R,D) = \mathcal{V}^-(I,S,R,D) - \mathcal{V}^+(I,S,R,D) = \left(\gamma I + \delta I, \frac{\beta SI}{N}, -\gamma I, -\delta I\right)^T. \tag{2.35}$$

Como la matriz de próxima generación se centra en el comportamiento infectado, es apropiado restringirse a esa característica, por lo las funciones restringidas son las siguientes:

$$\mathcal{F}(I,S,R,D) = \frac{\beta SI}{N}, \tag{2.36}$$

$$\mathcal{V}(I,S,R,D) = \gamma I + \delta I, \tag{2.37}$$

y para obtener sus respectivas matrices F y V:

$$F = \frac{\partial \mathcal{F}(0,N,0,0)}{\partial I}, \tag{2.38}$$

$$V = \frac{\partial \mathcal{V}(0,N,0,0)}{\partial I}, \tag{2.39}$$

dicho de otra manera, es tomar la derivada parcial de $\mathcal{F}$ y $\mathcal{V}$ en el punto de equilibrio, lo que deja este resultado:

$$F = \beta, \tag{2.40}$$
$$V = \gamma + \delta, \tag{2.41}$$

por lo que la matriz de próxima generación para el modelo epidemiológico SIRD se vería así:

$$K(0, N, 0, 0) = F(0, N, 0, 0)V^{-1}(0, N, 0, 0) = \beta\left(\frac{1}{\gamma + \delta}\right) = \frac{\beta}{\gamma + \delta}, \tag{2.42}$$

y en consecuencia,

$$\mathscr{R}_0 = \frac{\beta}{\gamma + \delta}. \tag{2.43}$$

Ahora bien, a partir de los resultados que arrojará el modelo, se pueden concluir dos situaciones:

- Si $\mathscr{R}_0$ <1, entonces existe el equilibrio libre de enfermedad. De modo que cada solución del sistema se aproxima a este equilibrio, y la enfermedad desaparece de la población.
- Si $\mathscr{R}_0$ >1, entonces hay dos equilibrios: el equilibrio libre de enfermedad y el equilibrio endémico. El equilibrio libre de enfermedad no es tan contundente, en el sentido de que las soluciones del sistema que comienzan muy cerca de él tienden a alejarse. El equilibrio endémico resulta interesante, de modo que las soluciones del sistema se aproximan a él a medida que el tiempo llega a infinito. Por lo tanto, en este caso, la enfermedad se convierte en una epidemia en la población.

Es interesante resaltar que hacer este tipo de predicciones para una enfermedad como esta resulta un poco difícil, debido a las medidas tanto del gobierno por tratar de mantener los contagios al mínimo, como los de la población por respetar sus medidas [14], aunada a la falta de información que se tenía sobre este virus. El número reproductivo básico puede convertirse en una herramienta para enfrentar esta enfermedad de manera más eficiente.

2.2. Filtro Kalman

El filtro de Kalman es uno de los algoritmos de estimación más importantes y comunes [5]. Este filtro produce estimaciones de variables ocultas basadas en mediciones inexactas e inciertas. Además, predice el estado futuro del sistema basándose en estimaciones que se irán actualizando conforme se vayan alimentando las ecuaciones con datos, dicho de otra forma, el filtro podrá predecir el comportamiento del sistema dependiendo de cuanta información sea proporcionada a las ecuaciones correspondientes. Lo primero y más importante es entender como funciona el filtro [8], siendo más un algoritmo que un filtro, se hizo un análisis para verificar las covarianzas entre las columnas, el cual debe ser el primer paso para determinar si los datos a analizar [2] son realmente independientes uno del otro, el resultado arrojó una matriz de covarianza donde indicaba valores cercanos cero en todas las entradas diferentes a la diagonal (ver apéndice A.1), lo que indica que correlación entre las columnas de susceptibles, casos confirmados, recuperados y personas fallecidas [2] es muy baja, estos datos fueron extraídos directamente de una página gubernamental [2] y posteriormente tratados (es decir, eliminar columnas vacías, reemplazar texto por valores, etc.) para un mejor entendimiento. Ahora es plausible preguntarse, ¿Cómo se pueden juntar estos dos valores no-correlacionados? Una herramienta útil, sería usar un estimador lineal con parámetros α y β, de la forma $\alpha x_1 + \beta x_2$. La primera conjetura lógica es que si las dos estimaciones x_1 y x_2 son iguales, entonces $\alpha + \beta = 1$. Dejando al estimador lineal de la siguiente forma:

$$y_\alpha(x_1, x_2) = (1 - \alpha)x_1 + \alpha x_2. \tag{2.44}$$

Ahora, es posible pensar que, α minimiza la varianza de y_α, esto suponiendo que x_1 y x_2 no tengan sesgo, y en consecuencia, haciendo que y_α, tampoco lo tenga. Para poder determinar el MSE (Mean Square Error):

$$\sigma_y^2(\alpha) = (1-\alpha)^2\sigma_1 + \alpha^2\sigma_2. \tag{2.45}$$

TEOREMA 2.1 [[4], Teorema 3.1]: Con $x_1 \sim p_1(\mu_1, \sigma_1^2)$ y $x_2 \sim p_2(\mu_2, \sigma_2^2)$ donde x_1 y x_2 tienen distribuciones p_1 y p_2, con media μ_1 y μ_2 y varianza σ_1^2 y σ_2^2 respectivamente, sean variables aleatorias no relacionadas. Considerando el estimador lineal $y_\alpha(x_1, x_2) = (1-\alpha)x_1 + \alpha x_2$ entonces, la varianza del estimador se minimiza para $\alpha = \dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$.

DEMOSTRACIÓN.

$$\begin{aligned} \frac{d}{d\alpha}\sigma_y^2(\alpha) &= -2(1-\alpha)\sigma_1 + 2\alpha\sigma_2, \\ &= 2\alpha(\sigma_1^2 + \sigma_2^2) - 2\sigma_1^2, \\ &= 0. \end{aligned} \tag{2.46}$$

Por lo tanto:

$$\alpha = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.47}$$

La derivada de segundo orden de $\sigma_y^2(\alpha)$, $(\sigma_1^2 + \sigma_2^2)$ es positiva, mostrando que $\sigma_y^2(\alpha)$ alcanza el mínimo en ese punto.
□
El valor óptimo que puede tomar α se conoce como la ganancia K de Kalman, sustituyendo la ganancia de kalman en un modelo de fusión lineal, se puede obtener el estimador lineal óptimo:

$$y(x_1, x_2) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}x_1 + \frac{\sigma_2^1}{\sigma_1^2 + \sigma_2^2}x_2, \tag{2.48}$$

como un paso previo a la fisión de $n > 2$ estimaciones, es más útil escribirlo de la siguiente manera:

$$y(x_1, x_2) = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}x_1 + \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}x_2. \tag{2.49}$$

Al sustituir el valor óptimo de α en la ecuación (2.45), se obtiene que:

$$\sigma_y^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}. \tag{2.50}$$

Se puede apreciar que las expresiones para y y σ_y^2 son algo complicadas porque ellas contienen sus variables recíprocas, sin embargo, solo basta con hacer nuevas variables ν_1 y ν_2 la precisión respectiva de cada distribución:

$$\begin{aligned} y(x_1, x_2) &= \frac{\nu_1}{\nu_1 + \nu_2}x_1 + \frac{\nu_2}{\nu_1 + \nu_2}x_2, \\ v_y &= \nu_1 + \nu_2. \end{aligned} \tag{2.51}$$

Estos resultados indican que el peso dado a una estimación es proporcional al grado de confianza obtenido en ella y que se tendrá más confianza en la estimación conjunta que en las estimaciones individuales, lo que es intuitivamente razonable. Para utilizar estos resultados, solamente son requeridas las varianzas de las distribuciones.

Deseando usar más estimadores, se supondrán n estimadores y se observará qué sucede.
El filtro de Kalman es un algoritmo de estimación esencial en el campo de la estadística y la ingeniería que permite estimar las variables ocultas de un sistema basándose en mediciones inexactas e inciertas. Este filtro también realiza predicciones sobre el estado futuro del sistema utilizando estimaciones previas.
Basándose en la idea de combinar dos valores estimados, en general no correlacionados, mediante un estimador lineal ponderado. Su funcionamiento se describe en varios pasos:

- **Análisis de Correlaciones**: Se inicia analizando las correlaciones entre las columnas de datos. Si la matriz de covarianza resulta en cero para todas las entradas, indica que las columnas son independientes.
- **Estimador Lineal Ponderado**: La idea es combinar dos estimaciones independientes, en este caso x_1 y x_2, usando una ponderación α de forma que si las dos estimaciones son iguales, α será igual a 1.
- **Minimización de Varianza**: Se busca encontrar el valor de α que minimice la varianza del estimador lineal ponderado. El resultado es que $\alpha = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$, donde σ_1^2 y σ_2^2 son las varianzas de las estimaciones x_1 y x_2 respectivamente.
- **Ganancia de Kalman**: Este valor óptimo de α es conocido como la ganancia de Kalman, que minimiza la varianza del estimador.
- **Estimador Lineal Óptimo**: Con la ganancia de Kalman, se puede obtener el estimador lineal óptimo: $y(x_1, x_2) = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$.
- **Extensión a Más Estimaciones**: El proceso se puede extender para fusionar más de dos estimaciones independientes. Se encuentra que el estimador óptimo pesa cada estimación de acuerdo con la confianza que se tiene en ella.

TEOREMA 2.2 [[4], Teorema 3.2]: Sea $x_i \sim p_i(\mu_i, \sigma_i^2)$ para $(1 \leq i \leq n)$ sea un conjunto de variables aleatorias no correlacionadas por pares. Considerando el estimador lineal $y_{n,\alpha}(x_1, ..., x_n) = \sum_{i=1}^n \alpha_i x_i$ donde $\sum_{i=1}^n \alpha_i = 1$, entonces la varianza del estimador es minimizada por:

$$\alpha_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}. \tag{2.52}$$

DEMOSTRACIÓN
La solución para minimizar la varianza del estimador lineal $y_{n,\alpha}(x_1, ..., x_n) = \sum_{i=1}^n \alpha_i x_i$, sujeta a la restricción $\sum_{i=1}^n \alpha_i = 1$, se obtiene estableciendo las derivadas parciales de la varianza respecto a cada α_i igual a cero y resolviendo para α_i. Primero, se deberá calcular la varianza del estimador:

$$\text{Var}(y_{n,\alpha}) = \text{Var}\left(\sum_{i=1}^{n} \alpha_i x_i\right) \quad = \sum_{i=1}^{n} \alpha_i^2 \text{Var}(x_i) \quad \text{(debido a la independencia)} \; = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2.$$

Ahora, se debe minimizar esta expresión sujeta a la restricción $\sum_{i=1}^n \alpha_i = 1$. Usando el método de los multiplicadores de Lagrange para encontrar los valores α_i que minimizan la varianza bajo esta

restricción:
Se define la función objetivo $L(\alpha_1, \alpha_2, \ldots, \alpha_n, \lambda)$ como:

$$L(\alpha_1, \alpha_2, ..., \alpha_n, \lambda) = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 + \lambda \left(1 - \sum_{i=1}^{n} \alpha_i\right), \tag{2.53}$$

luego, tomando las derivadas parciales de L respecto a cada α_i y λ y siendo igualadas a cero:

$$\frac{\partial L}{\partial \alpha_i} = 2\alpha_i \sigma_i^2 - \lambda = 0 \quad \text{para} \quad 1 \leq i \leq n, \tag{2.54}$$

y

$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{i=1}^{n} \alpha_i = 0. \tag{2.55}$$

Resolviendo la primera ecuación para α_i, se obtiene que:

$$1 = \sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \frac{\lambda}{2\sigma_i^2} = \frac{\lambda}{2} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}. \tag{2.56}$$

Por lo tanto,

$$\lambda = \frac{2}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}. \tag{2.57}$$

Sustituyendo este valor de λ de nuevo en la expresión para α_i:

$$\alpha_i = \frac{\frac{2}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}}{2\sigma_i^2} = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}, \tag{2.58}$$

entonces, la varianza mínima es dada por la siguiente expresión:

$$\sigma_{yn}^2 = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}. \tag{2.59}$$

□
Siendo posible encontrar expresiones más generales a las ecuaciones (2.50):

$$\begin{aligned} y_n(x_1, ..., x_n) &= \sum_{i=1}^{n} \frac{\nu_1}{\nu_1 + ... + \nu_n} x_1, \\ \nu_{yn} &= \sum_{i=1}^{n} \nu_i. \end{aligned} \tag{2.60}$$

Si bien es cierto que se puede utilizar ambos pares de ecuaciones (2.50) o (2.59) dependiendo de cuantos estimadores se traten, se puede ahorrar un poco de tiempo si se van agregando estimadores sobre la marcha. Se demostró que al igual que una secuencia de números puede sumarse manteniendo una suma y añadiendo los números a esta suma de uno en uno a la vez, una secuencia de $n > 2$ estimaciones puede fusionarse manteniendo un cálculo aproximado y fusionando

las estimaciones de la secuencia sin pérdida de calidad en la estimación final. O bien, mostrar que $y_n(x_1, x_2, ..., x_n) = y_2(y_2(y_2(x_1, x_2), x_3), ..., x_n)$. Analizando un poco las ecuaciones (2.59) es posible observar que:

$$\begin{aligned} y_n(x_1, ..., x_n) &= \frac{\nu_{yn-1}}{\nu_{yn-1} + \nu_n} y_{n-1}(x_1, ..., x_{n-1}) + \frac{\nu_n}{\nu_{n-1} + \nu_n} x_n, \\ \nu_{yn} &= \nu_{yn-1} + \nu_n, \end{aligned} \tag{2.61}$$

demostrando $y_n(x_1, ..., x_n) = y_2(y_{n-1}(x_1, ..., x_{n-1}), x_n)$, de seguir usando un método recursivo similar es factible llegar al objetivo inicial. Para poder conectar estas ecuaciones con el filtro Kalman, si $x_1 \sim p_1(\mu_1, \sigma_1^2), x_2 \sim p_2(\mu_2, \sigma_2^2)$, llegando a lo siguiente:

$$K = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{\nu_1}{\nu_1 + \nu_2}, \tag{2.62}$$

$$y(x_1, x_2) = x_1 + K(x_2 - x_1), \tag{2.63}$$

$$\sigma_y^2 = (1 - K)\sigma_1^2. \tag{2.64}$$

Aunque todo el análisis anterior solo fue para estimadores escalares, deseando hacer el mismo análisis, pero ahora con vectores, solo es necesario cambiar las varianzas por matrices de covarianza, dicho de otra forma:

$$\mathbf{y}_{n,A}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) = \sum_{i=1}^{n} A_i \mathbf{x}_i \qquad \text{donde} \qquad \sum_{i=1}^{n} A_i = I. \tag{2.65}$$

Aquí, A es la matriz de los parámetros $(A_i, A_2, ..., A_n)$ y a su vez, donde se asume que todos los vectores $\mathbf{x}_i$ son de la misma longitud.
El teorema (2.2) generaliza esta extensión para cualquier número de estimaciones n. La demostración del teorema utiliza el concepto de la matriz de precisión $(\mathcal{N}_i)$, que es el inverso de la matriz de covarianza. La estimación óptima pondera cada estimación por su matriz de precisión y normaliza por la suma total de las matrices de precisión de todas las estimaciones.
La conexión entre el filtro de Kalman y este análisis se encuentra al considerar x_1 y x_2 como estimaciones de estados y aplicar el concepto de ganancia de Kalman para fusionar estas estimaciones. El filtro de Kalman modela la evolución del estado real del sistema a través de las ecuaciones de transición de estado y de observación.
Existe un resultado similar al teorema (2.1):

TEOREMA 2.3 [4, Teorema 4.1]: Sea $\mathsf{x}_i \sim p_i(\mu_i, \Sigma_i)$ para $(1 \leq i \leq n)$ sea un conjunto de variables aleatorias no correlacionadas por pares. Considerando el estimador lineal $\mathbf{y}_A(\mathsf{x}_1, ..., \mathsf{x}_n) = \sum_{i=1}^{n} A_i \mathbf{x}_1$ donde $\sum_{i=1}^{n} A_i = I$. El valor de MSE$(\mathbf{y}_A)$ es minimizado por:

$$A_i = \left(\sum_{j=1}^{n} \Sigma_j^{-1} \right)^{-1} \Sigma_i^{-1}. \tag{2.66}$$

Por tanto, el estimador óptimo es:

$$\mathbf{y}(\mathbf{x}_1, ..., \mathbf{x}_n) = \left(\sum_{j=1}^{n} \Sigma_j^{-1} \right)^{-1} \sum_{i=1}^{n} \Sigma_i^{-1} \mathbf{x}_i, \tag{2.67}$$

luego, la matriz de covarianza de $\mathbf{y}$ puede ser hallada con un poco de programación:

$$\Sigma_{yy} = \left(\sum_{j=1}^{n} \Sigma_j^{-1} \right)^{-1}. \tag{2.68}$$

□
Es necesario mencionar que la precisión es el inverso de la matriz de covarianza, denotado por $\mathscr{N}$:

$$\begin{aligned} \mathbf{y}(\mathbf{x}_1, ..., \mathbf{x}_n) &= \mathscr{N}_y^{-1} \sum_{i=1}^{n} \mathscr{N}_i \mathbf{x}_i, \\ \mathscr{N}_y &= \sum_{j=1}^{n} \mathscr{N}_j. \end{aligned} \tag{2.69}$$

Estas ecuaciones utilizan la precisión para encontrar el valor óptimo y su varianza, también es una generalización de las ecuaciones (2.59), la demostración de este teorema se hace de manera muy similar al teorema (2.2).
Considerando $\mathbf{x}_i \sim p_1(\mu_1, \Sigma_1), \quad \mathbf{x}_2 \sim p_2(\mu_2, \Sigma_2)$, se puede hacer es estas ecuaciones encontradas se puedan expresar en términos de la ganancia K:

$$K = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1} = (\mathscr{N}_1 + \mathscr{N}_2)^{-1}\mathscr{N}_2, \tag{2.70}$$

$$\mathbf{y}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1 + K(\mathbf{x}_2 + \mathbf{x}_1), \tag{2.71}$$

$$\Sigma_{yy} = (I - K)\Sigma_1 \qquad \text{o} \qquad \mathscr{N}_y = \mathscr{N}_1 + \mathscr{N}_2. \tag{2.72}$$

El filtro de Kalman asume que el estado verdadero del sistema en el paso de tiempo k evoluciona desde el estado en el paso $k-1$ de acuerdo con:

$$x_k = F_k x_{k-1} + B_k u_k + w_k, \tag{2.73}$$

donde:

- F_k es la matriz del modelo de transición entre estados, de dimensión $n \times n$, que se aplica al estado del sistema anterior x_{k-1}, de dimensión $n \times 1$.
- B_k es la matriz de control del modelo, de dimensión $n \times n$, que se aplica al vector de control u_k, de dimensión $n \times 1$.
- w_k es el vector del ruido del proceso, de dimensión $n \times 1$, que se supone que sigue una distribución normal de media cero con matriz de covarianzas Q_k, de dimensión $n \times n$.

El filtro de Kalman también asume que en el paso de tiempo k se hace una observación z_k, el cual es un vector de dimensión $m \times 1$(no siempre se observan todas las componentes de x_k) [9], del estado verdadero del sistema x_k de acuerdo con:

$$z_k = H_k x_k + v_k, \tag{2.74}$$

donde:

- H_k es la matriz del modelo de observación, de dimensión $m \times n$, que se aplica al estado real del sistema x_k.

- v_k es el ruido de la observación, de dimensión $m \times 1$, el cual se supone que sigue una distribución normal de media cero con matriz de covarianzas R_k, de dimensión $m \times m$.

Se supone también que el estado inicial y los vectores de ruido en cada paso son independientes. Por cuestiones de notación y comodidad, denotaremos por $\hat{x}_{k|m}$ a la estimación de la variable de estado del sistema, x, en el paso de tiempo o momento k, teniendo en cuenta $m \leq k$ observaciones. El filtro de Kalman se divide en dos etapas: predicción y actualización. En la primera de las etapas se realiza una predicción del estado del sistema, dotando a dicha predicción de una medida de incertidumbre. De acuerdo con lo visto anteriormente, se conoce un modelo de transición de estados del sistema, expresado a través de la matriz F_k, por lo que, el paso más intuitivo a la hora de dar una predicción del estado de dicho sistema sería aplicar al último vector de estado disponible el modelo de transiciones de estado. Matemáticamente, resultaría la siguiente expresión:

$$\hat{x}_{k|k-1} = F_k x_{k-1|k-1}, \tag{2.75}$$

donde:

- $\hat{x}_{k|k-1}$ es la estimación del estado del sistema en el tiempo k, basada en las observaciones hasta el tiempo $k-1$.
- F_k es la matriz del modelo de transición de estados que se aplica al estado estimado en el tiempo $k-1$.
- $x_{k-1|k-1}$ es la estimación del estado del sistema en el tiempo $k-1$, basada en las observaciones hasta ese momento.

De acuerdo con el verdadero estado del sistema asumido por el filtro de Kalman, a través del vector de control se puede influir en el estado del sistema, haciendo este más predecible, por lo que, a la hora de predecir el estado del sistema, habrá que tomar en cuenta las manipulaciones que se hagan sobre este a través del vector de control, el cual intervendrá en el estado del sistema a través de la matriz B_k. De esta forma, el estado del sistema calculado en la etapa de predicción viene dado por:

$$\hat{x}_{k|k-1} = F_k x_{k-1|k-1} + B_k u_k. \tag{2.76}$$

Una vez calculada la predicción del estado del sistema, se dotará al filtro de una medida de incertidumbre sobre dicho cálculo. Esta medida será la matriz de covarianza, $P_{k|k-1}$, de la diferencia entre el estado real del sistema, x_k, y el estado predicho en la etapa actual, $\hat{x}_{k|k-1}$, es decir:

$$P_{k|k-1} = cov(x_k, \hat{x}_{k|k-1}). \tag{2.77}$$

La etapa de actualización del filtro de Kalman se centra en ajustar la estimación del estado del sistema en función de las nuevas mediciones observadas en el tiempo k. Se utiliza la matriz de observación H_k para relacionar el estado del sistema con las observaciones y se incorpora el ruido de la observación v_k:

$$z_k = H_k x_k + v_k, \tag{2.78}$$

donde:

- z_k es el vector de observación en el tiempo k.
- H_k es la matriz de observación que relaciona el estado del sistema con las observaciones.
- v_k es el ruido de la observación.

Utilizando las mediciones y el modelo de observación, se calcula la innovación, que es la diferencia entre la observación real y la predicción del estado:

$$y_k = z_k - H_k \hat{x}_{k|k-1}, \tag{2.79}$$

luego, se calcula la matriz de covarianza de la innovación S_k, que refleja la incertidumbre de las mediciones:

$$S_k = H_k P_{k|k-1} H_k^T + R_k, \tag{2.80}$$

finalmente, el filtro de Kalman actualiza la estimación del estado utilizando la ganancia de Kalman K_k, que pondera la predicción del estado en función de la confiabilidad de las mediciones:

$$K_k = P_{k|k-1} H_k^T S_k^{-1}. \tag{2.81}$$

La estimación del estado actualizada $\hat{x}_{k|k}$ y la matriz de covarianza actualizada $P_{k|k}$ se calculan como:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k y_k, \tag{2.82}$$
$$P_{k|k} = (I - K_k H_k) P_{k|k-1}. \tag{2.83}$$

2.3. Estimación de Máxima Verosimilitud

Otra pieza necesaria en el desarrollo del modelo epidemiológico usado en este trabajo es la estimación de máxima verosimilitud, esto se debe a que en el código que fue desarrollado, es necesario describir un método para la obtención de los parámetros ajustados β, γ y δ. (ver apéndice A.2)
Considerando que este método abreviado a menudo como MLE, fue propuesto por Fisher [10, 11] (1930), aunque ya de una forma mucho más artificiosa, fue inicialmente atisbado por Bernoulli [12] (1778), cuyo planteamiento fue revisado y modificado por su coetáneo y amigo, el gran matemático Euler [13] (1778). Sin embargo, la resolución de los problemas numéricos planteados por este método en la mayor parte de los casos son de tal magnitud que no ha sido posible su amplia utilización hasta la llegada de los modernos ordenadores.
El método de estimación de la máxima verosimilitud, Fisher (1930) lo atribuye a Gauss, aunque hay precedentes en los trabajos de Lambert (1760) y D. Bernoulli (1778) y en Edgeworth (1908 - 9). No obstante, Fisher fue mucho más lejos que sus predecesores en promocionar su uso como un método universal de estimación y estudiar sus propiedades. Hoy se sabe que muchas de las proposiciones que Fisher demostró sobre las propiedades de los estimadores de máxima verosimilitud no son universalmente ciertas. Aunque, muchos de estos resultados se pueden demostrar rigurosamente bajo ciertas condiciones. Fisher fue, sin embargo, consciente de los defectos de sus demostraciones:

"Por mi parte, gustosamente habría retrasado su publicación hasta que hubiera formulado una demostración completamente rigurosa; pero la cantidad y la variedad de resultados nuevos que este método revela me empujó a publicarlo."

El método de máxima verosimilitud (MLE, por sus siglas en inglés) es una poderosa técnica utilizada en estadísticas e inferencia para estimar los parámetros de un modelo estadístico a partir de

datos observados. La idea fundamental detrás de este método es encontrar los valores de los parámetros que hacen que los datos observados sean los más probables bajo el modelo propuesto. En otras palabras, se busca encontrar los parámetros que maximizan la "verosimilitud"de los datos.
La verosimilitud, denotada como $L(\theta|\text{datos})$, representa la probabilidad de observar los datos a los que se tiene acceso, asumiendo que el modelo subyacente es correcto y que los parámetros del modelo toman valores específicos θ. O bien:

$$L(\theta|\text{datos}) = P(\text{datos}|\theta), \tag{2.84}$$

donde:

- $L(\theta|\text{datos})$ es la función de verosimilitud, que mide cuán verosímiles son los datos bajo ciertos valores de los parámetros θ.
- $P(\text{datos}|\theta)$ representa la probabilidad de observar los datos que tenemos, suponiendo que los parámetros θ son los correctos.

El objetivo del método de máxima verosimilitud es encontrar los valores de θ que maximizan esta función de verosimilitud $L(\theta|\text{datos})$. En otras palabras, se busca que los valores de los parámetros que hacen que los datos observados sean los más probables de acuerdo con el modelo propuesto.
Para encontrar estos valores, se utiliza el cálculo diferencial. Primero, se toma la derivada de $L(\theta|\text{datos})$ con respecto a cada parámetro θ_i. Luego, se iguala cada derivada a cero y se resuelve el sistema resultante de ecuaciones para encontrar los valores de θ que maximizan la verosimilitud. Los valores de θ que se obtienen de esta manera se llaman estimadores de máxima verosimilitud (MLE) y representan las estimaciones óptimas de los parámetros del modelo dados los datos observados. Estos estimadores son ampliamente utilizados en estadísticas y se caracterizan por tener propiedades deseables, como ser insesgados y eficientes en términos de varianza, cuando se cumplen ciertas condiciones.
Aunque existen varios algoritmos para resolver una máxima verosimilitud, en esta caso particular se usará el "Nelder-Mead". El algoritmo de Nelder-Mead opera en un espacio de parámetros multidimensional y se basa en la manipulación de un conjunto de puntos llamado "simplex"que abarca una región del espacio de parámetros.
Supone que se tienen n parámetros en el modelo, es decir, $\theta = (\theta_1, \theta_2, ..., \theta_n)$.
Ahora, los pasos a seguir para elaborar un algoritmo Nelder-Mead es el siguiente:

1. **Inicialización del Simplex**: Comenzando con un conjunto de $n+1$ puntos en el espacio de parámetros. Cada punto representa una estimación inicial de los parámetros θ. Estos puntos forman un simplex en el espacio de parámetros.
$$\mathscr{S} = \{\theta^{(1)},\ \theta^{(2)},\ ...,\ \theta^{(n+1)}\}.$$
2. **Evaluación de la Función de Verosimilitud**: Se calcula el valor de la función de verosimilitud $L(\theta|\text{datos})$ para cada punto del simplex:
$$L^{(i)} = L(\theta^{(i)}|\text{datos}) \qquad \text{para } i = 1,\ 2,\ ...,\ n+1.$$
3. **Ordenación de los Puntos**: Ordenando los puntos del simplex en función de los valores de verosimilitud, de modo que:
$$L^{(1)} \geq L^{(2)} \geq ... \geq L^{(n+1)}.$$
4. **Cálculo del Punto Centroidal**: Se determina el punto centroidal, que es el promedio de los n puntos con los valores de verosimilitud más altos:
$$\theta^c = \frac{1}{n}\sum_{i=1}^{n}\theta^{(i)}.$$

5. **Reflexión**: Encontrando el punto de reflexión θ^r a través del punto centroidal θ^c y el punto del simplex $\theta^{(n+1)}$:
$$\theta^r = \theta^c + \alpha(\theta^c - \theta^{(n+1)}),$$
donde α es un factor de reflexión, generalmente mayor a 1.

6. **Evaluación de la Reflexión**: Se calcula el valor de la verosimilitud en el punto de reflexión:
$$L^r = L(\theta^r|\text{datos}).$$

7. **Comparación de Resultados**:
   - Si $L^{(1)} \leq L^r < L^{(n)}$, reemplazando el punto con θ^r y regresando al paso 3.
   - Sí $L^r < L^{(1)}$, se realza una expansión para explorar más en esa dirección.
   - Si $L^r \geq L^{(n)}$, entonces se hace una contracción para reducir el tamaño del simplex.
   - Si la contracción no mejora la verosimilitud, será necesario hacer una contracción en el origen.

8. **Criterio de Parada**: Repitiendo los pasos 4 a 7 hasta que se cumpla un criterio de parada, como un número máximo de iteraciones o convergencia satisfactoria.

9. **Resultado final**: El vértice con la mayor verosimilitud en el último simplex obtenido se toma como la estimación de MLE de los parámetros del modelo.

2.4. Datos

Es ampliamente reconocido que para verificar un modelo o una hipótesis, se requieren pruebas empíricas que respalden lo que se afirma y no sean simplemente ideas vacías. En esta sección, se abordará la importancia de los datos utilizados en este proyecto.

2.4.1. Aprovechamiento de datos

Para comenzar de manera apropiada, es fundamental tener información acerca del origen de los datos y verificar la confiabilidad de esta fuente. Los datos fueron proporcionados por el CONAHCYT [2], desde donde se descargaron las tablas en bruto de los registros de confirmados y defunciones. Una vez obtenidas estas tablas, se centró el análisis en la población del estado de Puebla, que constaba de un estimado de 6,604,451 (seis millones seiscientos cuatro mil cuatrocientos cincuenta y uno) personas susceptibles a esta enfermedad, o al menos, esa fue la cifra analizada. Posteriormente, los datos fueron procesados; es decir, se fusionaron ambos archivos, se sincronizaron las fechas y se completaron los espacios vacíos en caso de existir alguno. Dichos espacios solo fueron llenados con un cero para evitar, en la medida de lo posible, sesgos en los resultados. Siendo precisos, solo se tuvieron que llenar tres espacios. Este tipo de tratamiento de datos resulta crucial para prevenir problemas futuros durante la programación y evitar errores derivados de datos faltantes o celdas que contengan texto en lugar de números.
Otro punto importante a tomar en cuenta es la posibilidad de que los datos contengan ruido. [15]:

"La presencia de ruido significa que los resultados del muestreo podrían no duplicarse si el proceso se repitiera."

Otro significado que se puede atribuir es que los datos podrían estar sesgados con respecto a la información real en ese momento o podrían no concordar con otras tablas que hayan analizado la misma información en ese instante. De manera más descriptiva, el ruido presente en los datos se debe a que, al contar la cantidad de personas reportadas que ingresaron al hospital y resultaron

positivas a COVID-19, no necesariamente coincide con el número presente en las tablas. Asimismo, al tratarse de un conteo diario, no se conoce la cantidad exacta de hospitales que participaron en este conteo, junto con otros factores que contribuyen a crear sesgos entre los datos reales y los datos reportados. Para gestionar este ruido de manera eficiente, se utilizará el filtro Kalman.
Otorgando una representación más general y visual de los datos, se procederá a crear representaciones gráficas, que serán denotadas como figuras:

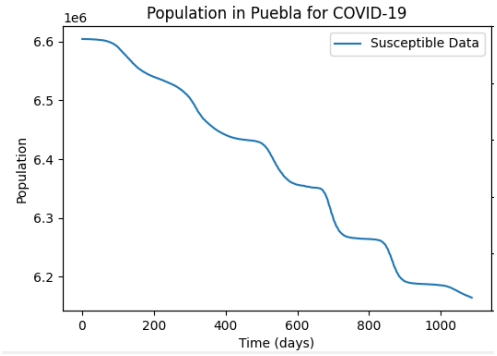

Figura 2.1: Se puede apreciar el cambio de personas susceptibles en los 1024 días del análisis. Los datos disminuyen debido a que son los individuos que se han estado infectando y posteriormente recuperando o falleciendo.

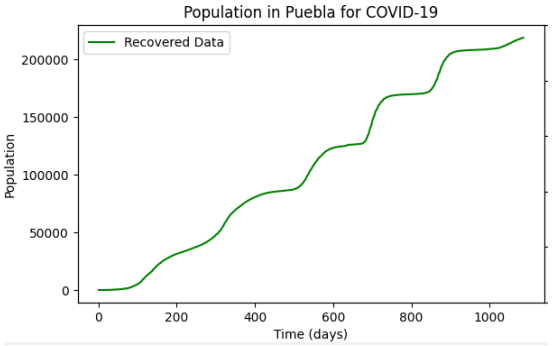

Figura 2.3: Aquellos individuos que se recuperaron de la enfermedad, luego de tenerla por un estimado de dos semanas, son representados aquí. Es importante destacar que son solo estimaciones, esto se debe a que no fueron encontrados los datos que representan a los recuperados.

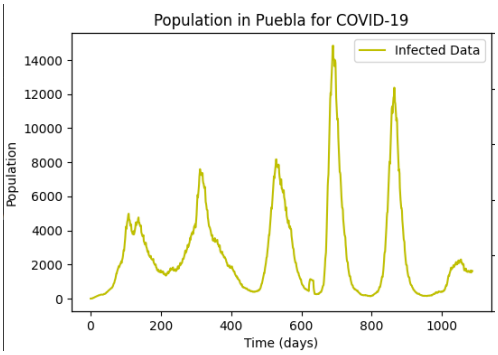

Figura 2.2: Es el conteo diario de individuos que resultaron positivos en hospitales. Resulta interesante destacar cómo los datos pueden observarse en olas, ya que en un momento dado, los casos tienden a disminuir.

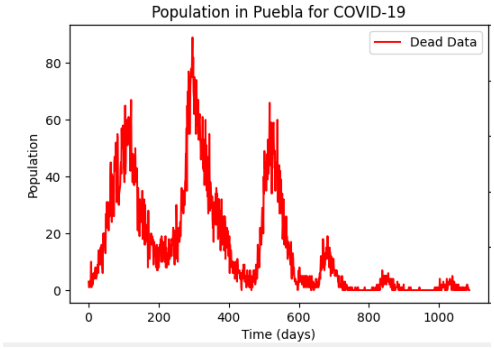

Figura 2.4: Aquellas personas que fallecieron por Covid-19 en Puebla (al menos los reportados en hospitales) son representados aquí, también es importante destacar que aquellos casos, también coinciden con las olas.

En estos tres años se ha obtenido un panorama más general acerca de qué tan bien preparada estaba la población para enfrentar una enfermedad que amenazaba a toda la población. También se ha destacado la importancia de la epidemiología y de los resultados que arroja.
Se puede observar que antes del primer año e incluso a mitad del segundo año de pandemia, la tasa de personas fallecidas era muy alta, alcanzando los 80 fallecimientos en un solo día. Se menciona esto para crear conciencia sobre los eventos ocurridos y cómo se enfrentaron, incluyendo

medidas como el uso de cubrebocas, mantener la "sana distancia", evitar salir a menos que fuera estrictamente necesario, entre otras medidas de seguridad. El primer año fue especialmente crucial, ya que mostró si las medidas realmente funcionaban y si la población las respetaba, además de reflejar el estado "desatado"de la enfermedad, dado que nadie estaba preparado para enfrentar una situación así.

2.5. Ejemplos

Se ha abarcado la teoría necesaria para comprender el funcionamiento del filtro Kalman y tener un mejor entendimiento de como opera un modelo epidemiológico, para ser más precisos, el SIRD, no obstante, con el propósito de otorgar una idea más clara al respecto, se presentarán tres ejemplos.

2.5.1. Ejemplo de un modelo epidemiológico SIRD

Se desea conocer el comportamiento de una enfermedad ficticia que azota una población muy pequeña y ver qué comportamiento va a presentar durante un periodo de 100 días, por lo que se proporcionan los siguientes datos:

- Población total de individuos en la localidad: Un millón.
- Existe un individuo infectado, que fue el que comenzó con la propagación de la enfermedad a la población.
- Capacidad de transmisión del virus: 1.5.
- En caso de que un individuo infectado se llegue a recuperar, este tardará un total de cinco días aproximadamente para hacerlo.
- La enfermedad presenta una tasa de mortalidad del 0.1 %.
- En esta localidad no son considerados los nuevos nacimientos, ni migraciones de población.
- Una persona que se recuperó de la enfermedad adquiere inmunidad a ella y no se vuelve a enfermar.
- Duración del análisis 100 días.

Esta información se puede ver de la siguiente manera. Plantea una población de un millón de personas que enfrentará la propagación de un virus con una capacidad de transmisión $\beta = 1.5$, lo que indica qué tan rápidamente una persona susceptible se infecta al entrar en contacto con una persona enferma. Además, las personas afectadas por esta enfermedad ficticia presentan un coeficiente de recuperación de $\gamma = 0.2$, lo que implica que tardan cinco días en recuperarse después de haberse contagiado, y una tasa de mortalidad $\delta = 0.1$. Posteriormente, informa que no hay migraciones y no se consideran nuevos nacimientos, por lo que la población en todo momento permanece constante, por último, es mencionado que las personas recuperadas adquieren inmunidad a la enfermedad por lo que no es necesario colocarlas de nuevo a los individuos susceptibles. Sin caer en la redundancia o repetición solo se hará una lista de todos los parámetros conocidos:

- $N = 1000000$ (Un millón de individuos).
- $S(0) = 999999$ (Todas las demás personas de la población que aún no se han infectado, pero son susceptibles).
- $I(0) = 1$ (La única persona que introdujo la enfermedad a la población).

- $R(0) = 0$ (Dado que solo hay un usuario que contrajo la enfermedad y no hay migraciones, se asume que no hay individuos recuperados).
- $D(0) = 0$ (De manera similar a los recuperados, tampoco puede haber individuos fallecidos).
- $\beta = 1.5$.
- $\gamma = 0.2$.
- $\delta = 0.1$.
- $t = 100$ (Duración del análisis de la enfermedad).

Con estos parámetros establecidos, se procederá a desarrollar el modelo:

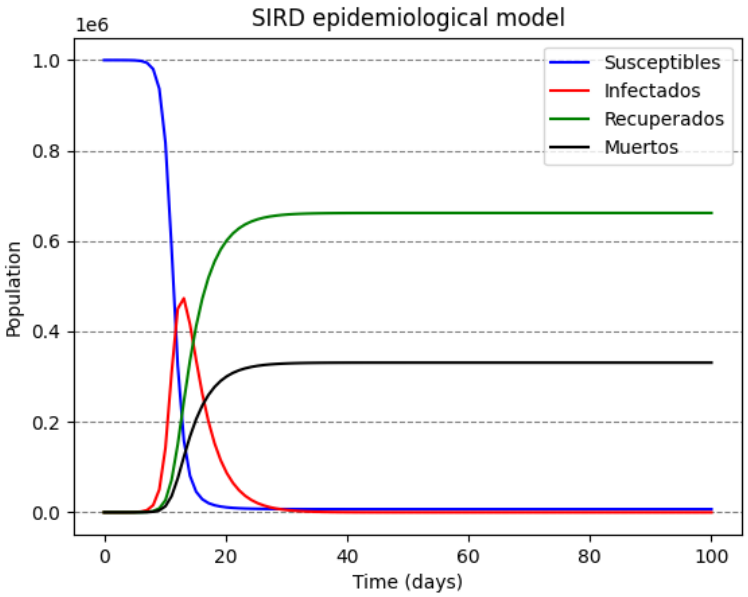

Figura 2.5: Modelo simulado con una escala de 1×10^6, es evidente como los primeros 20 días del análisis son los que presentan mayor actividad y resolución del comportamiento de la enfermedad, después del día 25 todos y cada uno de los individuos de la población contrajeron la enfermedad, algunos se recuperaron un 65 % de la población para ser precisos, mientras que el 35 % restante falleció, a su vez, se observa el pico máximo de individuos contagiados en el día 13, resulta interesante que 100 días de análisis eran innecesarios y que bastaba con a lo sumo, 30 días.

Con toda esta información, es posible calcular el valor de reproducción inicial $\mathscr{R}_0$, el cual se determina a partir de la ecuación (2.43):

$$\mathscr{R}_0 = \frac{\beta}{\gamma + \delta} = \frac{1.5}{0.2 + 0.1} \approx 5$$

Se pueden concluir dos cosas de este ejemplo:

- A partir del día 25 todas las personas susceptibles se infectaron y se recuperaron o fallecieron.
- Que el valor de $\mathscr{R}_0$ sea 5 indica que es muy probable que la enfermedad se convierta en endémica, aunque debido al tipo de análisis realizado, no existen pruebas concluyentes para afirmarlo. Pero lo que si se puede decir con certeza, es que la enfermedad se convirtió en pandemia.

2.5.2. Ejemplos del filtro Kalman

Supongamos que se desea conocer el peso real de un objeto, cualquiera que sea, por ejemplo, una mancuerna oxidada que no indica su peso. Para lograrlo, lo más lógico sería colocar la mancuerna en una báscula, registrar las medidas que arroja y repetir el proceso varias veces. Al final, solo se tendría que calcular el promedio de los valores obtenidos para obtener una estimación más precisa del peso. A pesar de que todo esto suene adecuado, se ha pasado por alto un detalle importante: todos los equipos de medición tienen un índice de error al proporcionar una medida. Para abordar este problema inherente al error del equipo, se empleará el filtro Kalman.
Lo que puede apreciarse en la siguiente figura:

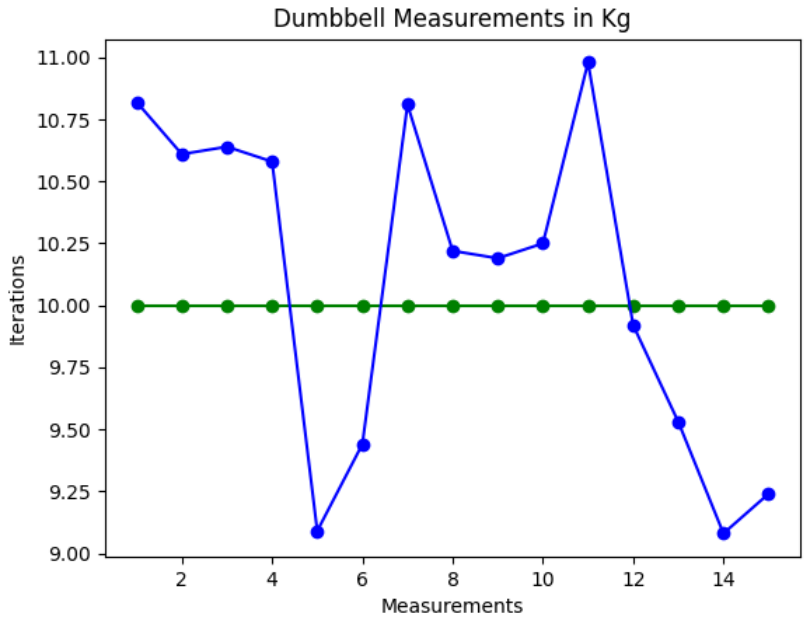

Figura 2.6: Una comparativa entre el peso real de la mancuerna y las medidas que arrojaba la báscula durante 15 veces.

La figura anterior muestra la comparación de tomar el peso 15 veces de una mancuerna (datos de color azul) contra el peso real de la misma (datos en verde), cada dato fue registrado y comparado con su respectivo punto, es decir, en la medida uno, la báscula arrojo que la mancuerna poseía un peso de $10.82Kg$, mientras que su peso real constaba de $10Kg$, el peso real no varía con las mediciones dado que la mancuerna se encontraba en un ambiente controlado, todo lo contrario al peso medido por la báscula, donde las mediciones presentaban diferentes medidas cada vez que se volvía a pesar el objeto, dando un sesgo resultante de incluso un kilo de diferencia al valor real.
Para poder utilizar el filtro, primero hay que poder identificar los parámetros que serán utilizados y tomar en cuenta ciertas consideraciones, como considerar la ganancia K_k constante. Deseando resolver este problema, los parámetros considerados son [8]:

- x: Es el valor verdadero del peso; $10Kg$.
- z_k: Es el valor de la medición en el instante **k**; $10.82kg$ en la primera medición, es decir, z_1.
- $\hat{x}_{k,k}$: Es la estimación de **x** en el tiempo **k** (la estimación se hace luego de tomar z_k mediciones); 10.82, aquí $x_{1,1}$ coincide con z_1 dado que es la primera medición y no contamos con mediciones previas para ajustar ese valor.

- $\hat{x}_{k,k-1}$: Es la estimación previa de **x** que se realizó en el tiempo **k-1** (la estimación que se tomó después de la medición z_{k-1}); en la primera iteración $x_{1,0}$ este valor es cero.
- K_k: La ganancia de Kalman, la piedra angular de todo este proceso; 1, mientras más pasos haya, el valor seguirá bajando de la forma $\frac{1}{k}$, donde k es el número del paso actual.

Por lo que, solo resta aplicar la fórmula característica del filtro Kalman para predecir el sistema:

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k(z_k - \hat{x}_{k,k-1}) = (1 - K)\hat{x}_{k,k-1} + K_k z_k. \tag{2.85}$$

La figura anterior muestra que hay valores que se alejan bastante y otros que se acercan más al valor real de la báscula, este error del aparato de medición es aleatorio, por lo que será denotado como una varianza (σ^2). Si se adopta un enfoque minucioso, es posible buscar el error de la báscula o del dispositivo utilizado para medir el peso. También se puede solicitar esta información al fabricante en caso de que se desee obtener detalles más precisos.
Ahora se presentará una figura que ilustra la aplicación del filtro y cómo se visualizarían los datos después de ser filtrados.

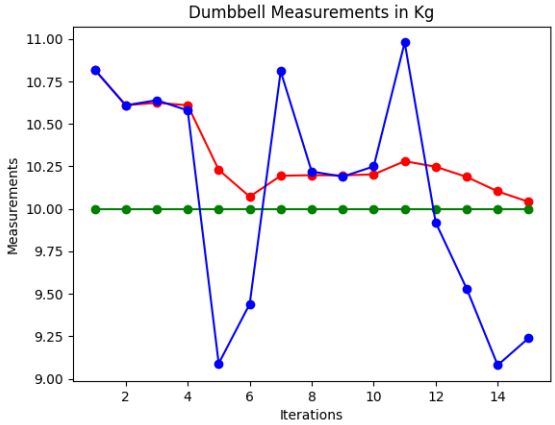

Figura 2.7: Figura que muestra los 3 tipos de datos, las medidas reales, los datos arrojados por la báscula, y los datos filtrados.

La diferencia más significativa entre la figura (2.6) y (2.7) radica en la adición de los datos filtrados (las mediciones de color rojo) y como con cada nuevo dato, la línea de puntos se va suavizando con respecto a las mediciones reales, dicho de otra manera, los datos convergen gradualmente; sin embargo, este fenómeno se atribuye a que la ganancia K_k se va ajustando de manera constante. Esto se debe a la dinámica del sistema, ya que el peso de la mancuerna no experimenta cambios con el tiempo, lo que elimina la necesidad de realizar un análisis de la actualización de la ganancia K_k con la ecuación (2.83).
Además, se asume que el peso real de la mancuerna es de $10kg$. Ahora bien, ¿qué ocurriría si se eliminara el valor real y se agregaran otras 15 predicciones?

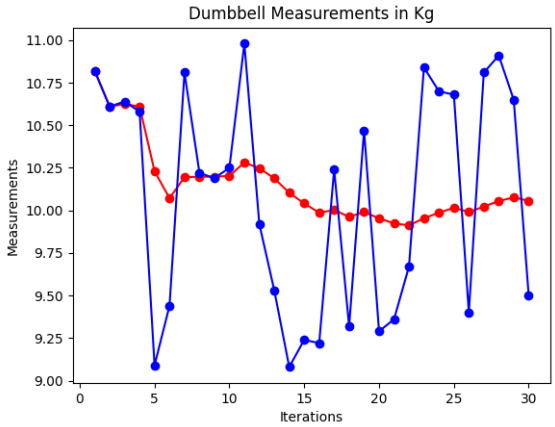

Figura 2.8: Filtro Kalman con 30 mediciones.

En la figura presentada se puede observar las primeras 15 mediciones realizadas en la primera parte de este ejemplo, solo fue necesario realizar otras 15 mediciones para realizar la gráfica y que a medida que se incorporan más mediciones en el filtro, este converge hacia el valor real de la medición. Es una forma interesante de corroborar si estos valores convergen hacia un valor en específico, para la ecuación anterior, de no conocer el valor real, se puede colocar una estimación que puede variar desde un valor cercano a las mediciones observadas o directamente poner cero, en ambos casos se observarán picos en las primeras mediciones, pero eventualmente el valor de la predicción $x_{k,k}$ convergerá al valor real, estas aseveraciones se examinarán mejor en el segundo ejemplo.

Segundo ejemplo

El primer ejemplo ilustra de manera clara el funcionamiento del filtro Kalman, pero se intentará agregar algunos conceptos adicionales.
Si la dinámica del proceso analizado no es constante, en ese caso, es necesario actualizar la ganancia del filtro, y para lograrlo se define la **ecuación de actualización de covarianza**:

$$K_k = \frac{p_{k,k-1}}{p_{k,k-1} + r_k}, \tag{2.86}$$

$$p_{k,k} = (1 - K_k)p_{k,k-1}. \tag{2.87}$$

- $p_{k,k-1}$: Es la incertidumbre estimada que se calculó durante la estimación previa.
- $p_{k,k}$: Es la incertidumbre estimada del estado actual.
- r_k: Es la incertidumbre de medición.

De aquí, que, cuando la incertidumbre de medición es grande, la convergencia sería lenta debido a que la ganancia K_k tendría valores cada vez más bajos. Sin embargo, si la incertidumbre es baja, la ganancia de Kalman será alta y la incertidumbre estimada convergería rápidamente hacia cero. Con lo mencionado, se inicia el ejemplo. Se supone que se analiza pasta de cocina de diversas marcas con el objetivo de determinar si existe un valor estándar para cada uno de los criterios evaluados. Se toman en cuenta cuatro factores: la capacidad de absorción de aceite, la densidad, la textura crujiente y la fragilidad. Utilizando las ecuaciones anteriores, se considera un experimento que está

controlado, es decir, el índice de error de medición humana es de uno ($\sigma = 0.1$), lo que implica una varianza de ($\sigma^2 = 0.01$). En esta situación, la primera estimación en cada caso se realiza de manera distante respecto a los valores medidos. De este modo, el error de inicialización de la estimación es $\sigma = 100$, y la incertidumbre estimada de la varianza del error es $\sigma^2(p_{0,0} = 10000)$. Se observa que esta variación es muy alta. Si se realiza una inicialización con un valor más significativo, se logrará una convergencia más rápida del filtro de Kalman:

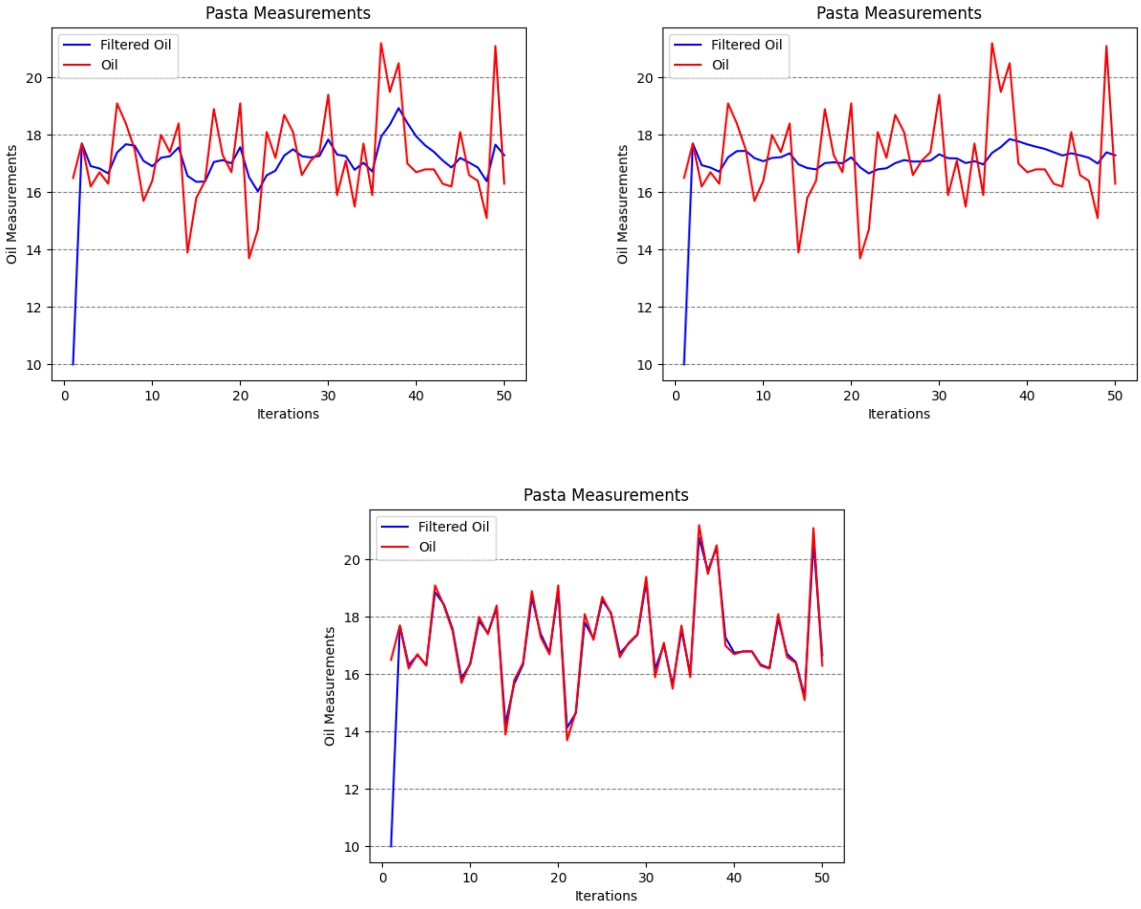

Figura 2.9: Se hicieron distintos casos a manera de comparativa acerca de como el índice de error de medición humano puede afectar la suavidad de la figura, en la primera figura se usó un error de $\sigma = 0.1$ siendo este el caso tomado en cuenta como el ambiente controlado, donde efectivamente se puede observar como los datos si recibieron un ajuste y como cambian las mediciones (rojo) y los datos filtrados (azul), otra observación a tener en cuenta es el primer dato, tomando la primera predicción echa como $x_{0,0} = 10$ se aprecia mejor el salto donde el usuario coloca su predicción y como a partir de ahí y las mediciones el filtro esboza una figura más suave. En la siguiente figura (la de la derecha), es el supuesto de un caso idílico, ya que el error humano es muy bajo ($\sigma = 0.01$), por ende la figura que muestran los datos filtrados es bastante más suave que la primera y se puede observar como converge a un cierto valor real, el cual parece ser 17. Por último, se tomó en cuenta un caso donde el error humano sea grande, muy grande a comparación de los dos casos anteriores, la tercera figura muestra como se vería el filtro si $\sigma = 10$, es evidente que en esta situación los datos filtrados y las medidas casi coinciden, denotando precisamente la importancia de usar un ambiente controlado y de como esto puede afectar los resultados que arrojará este experimento.

Las dos últimas figuras presentadas son únicamente con fines ilustrativos; este ejemplo solo considerará $\sigma = 0.1$.
Dicho lo anterior, se procederá a analizar los demás rasgos:

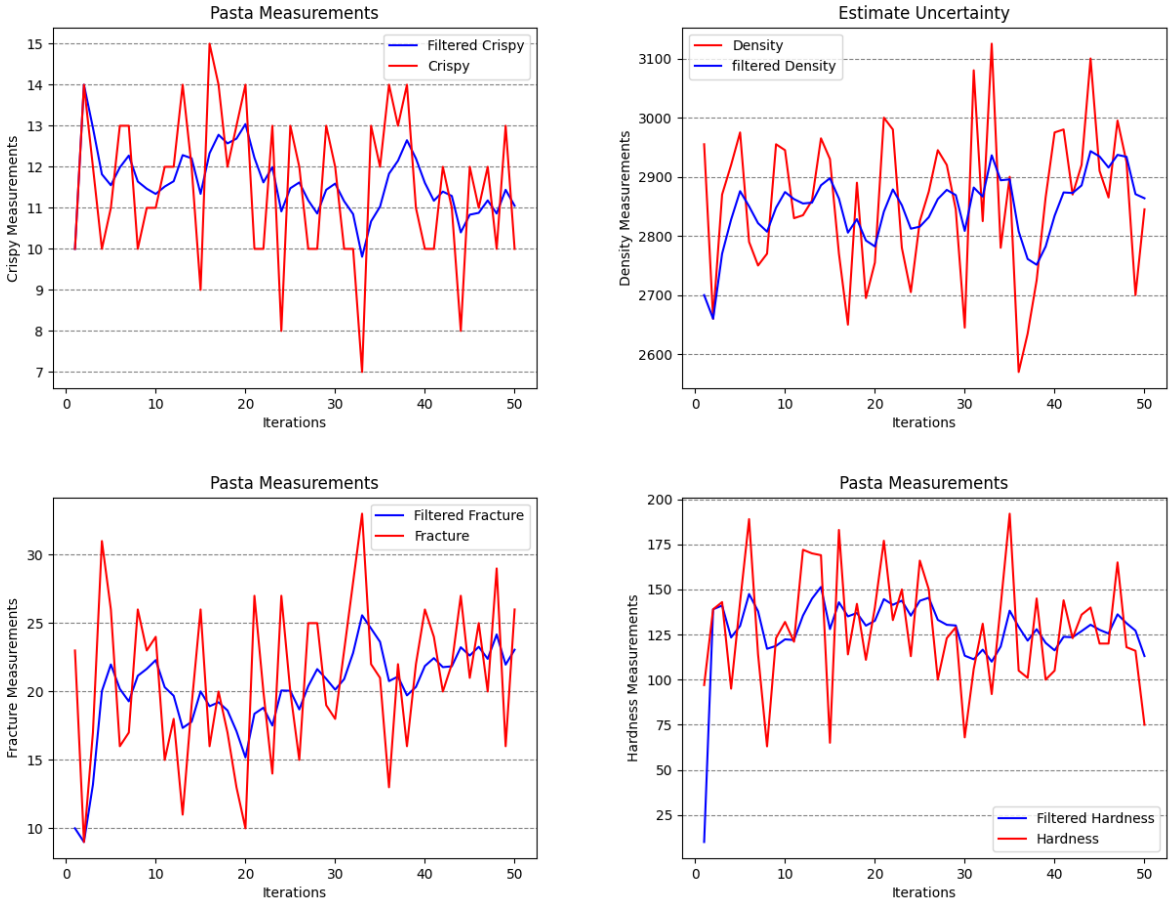

Figura 2.10: La primera figura indica que tan crujiente es la pasta, en primera instancia no se aprecia si esta cualidad converge o no a un valor en específico, se necesitarían contar con más mediciones para poder concluir de manera satisfactoria este argumente, no obstante, si se puede apreciar como el filtro dibuja una figura más suave de los datos. La siguiente figura, (superior derecha) indica que tan gruesa es la pasta, en esta segunda figura se puede apreciar mejor el filtrado de los datos, por los picos más pronunciados presentados en las mediciones. La tercera figura (inferior izquierda), son las fracturas de la pasta, es decir, al someterla a una caída libre desde una altura de dos metros, en cuantos pedazos visibles se dividía la pasta y en consecuencia de la primera y tercera figura, también se hizo un experimento de que tan frágil es entonces la pasta, la cual es la última figura, donde, pese a los picos presentados, si se observa una consistencia con los fatos filtrados y como estos parecen converger a un valor estimado en 125.

Una vez visto los rasgos de la pasta, solo resta examinar cómo cambian tanto la ganancia K_k como el valor de $p_{k,k}$:

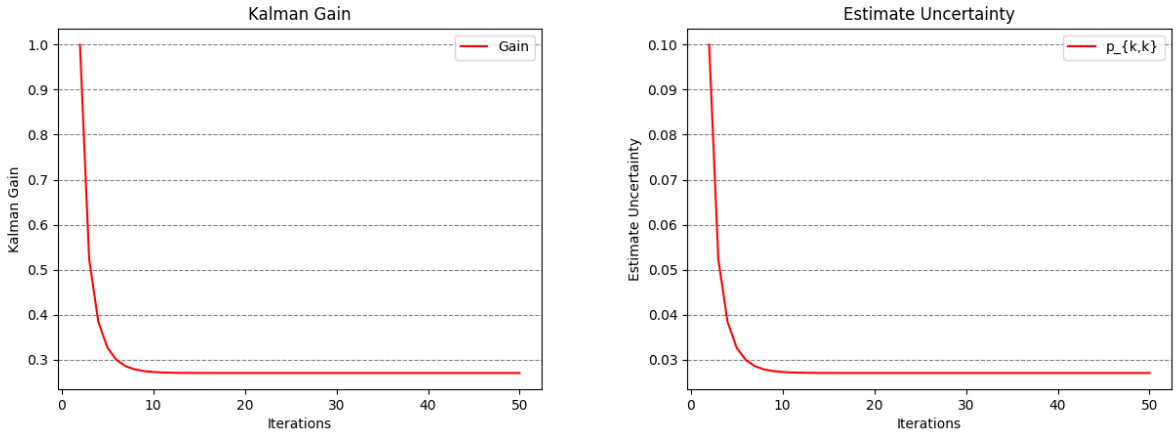

Figura 2.11: La figura de la izquierda muestra como la ganancia K_k va disminuyendo, provocando que el peso de la medición sea cada vez más pequeño, lo cual haría que el filtro se acerque al valor real. Aquí K_k depende de la ecuación (2.87) y de como las actualizaciones de $p_{k,k}$ (figura de la derecha) vayan variando. Ambas figuras comparten una similitud exacta y no es de extrañarse, dado que la ganancia K_k está definida a partir de la incertidumbre estimada, también, es importante destacar que a partir de las primeras 10 mediciones la incertidumbre comienza a tomar un valor fijo $\sigma^2 = 2.70156 \times 10^{-2}$, es decir, $\sigma = 0.16436$, dicho de otra forma, el filtro comienza a darle más importancia a las predicciones que a los valores reales, que es justamente lo mencionado en la figura (2.7).

Capítulo 3

Análisis a la base de datos.

3.1. Modelo Epidemiológico SIRD

En la sección (2.5.1) se abordó un ejemplo de como resolver un modelo epidemiológico SIRD, en esta sección se realizará un proceso similar, además de utilizar una base de datos extraída del CONAHCYT [2] serán tomadas las siguientes consideraciones:

- Nuestra base de datos consta de 5 columnas, que serán:
  1. Susceptibles.
  2. Infectados.
  3. Recuperados.
  4. Fallecidos.
  5. Acumulados.
- Todos los datos de estas columnas son de casos reportados día a día.
- La expresión del número de reproducción básico $\mathscr{R}_0$ utilizado en este caso es:

$$\mathscr{R}_0 = \frac{\beta}{(\gamma + \delta)}. \tag{3.1}$$

- En este caso particular de análisis no se considerarán tasas de:
  1. Nuevos nacimientos.
  2. Migraciones de población.

  Por lo que la población será constante en todo momento del análisis.
- Las personas infectadas que pasan a ser individuos recuperados adquieren inmunidad, por lo que no pueden infectarse nuevamente.
- Las condiciones iniciales respectivas para alimentar el modelo son:
  1. $N = 6604451$, como anteriormente se mencionó, este número permanecerá constante en el proceso de análisis del modelo epidemiológico SIRD.
  2. $S(0) = 6604446$, particularmente estima a $S(0)$ como, $S(0) = N - C(0)$, donde $C(0)$, son los casos acumulados iniciales.

3. $I(0) = 5$, son las primeras personas reportadas positivas el día 13 de marzo del 2020, que fue el día que fue considerado como punto de partida.
4. $R(0) = 0$, dado que la enfermedad tiene un tiempo estimado de 14 días en recuperarse, tenemos que, no puede haber recuperados en el momento inicial $R(0)$.
5. $D(0) = 0$, Se tiene un razonamiento similar con los individuos fallecidos, no se reportan personas fallecidas en nuestro punto de partida.
6. $C(0) = 5$, particularmente es notable que coinciden los infectados iniciales con los casos acumulados, esto es principalmente porque son los valores iniciales al tiempo t_0.
7. Estos datos son exclusivamente para la primera ola, la cual fue acotada hasta el día número 211. Para las demás olas, estos datos iniciales varían un poco, considerando que ya habría personas recuperadas y fallecidas. La importancia de utilizar la primera ola como punto de referencia, se da debido a la primera ola fue la que inicio todo el proceso de la enfermedad y su propagación.
8. Los datos utilizados fueron analizados a través de simulaciones numéricas realizadas con el lenguaje Python 3. Todos los códigos utilizados en este trabajo fueron desarrollados en su totalidad para tener un mejor control de las simulaciones (ver apéndice A.2).

Con todo lo anterior, se da inicio al proceso. En primer lugar, se deben leer los datos, limitándose a los primeros 211 días de la pandemia. A continuación, se asignan las variables necesarias, en este caso:

- N.
- El vector de condiciones iniciales $y_0 = \{S(0), I(0), R(0), D(0), C(0)\}$.
- *cum*. Todo el conjunto de los datos acumulados.

Ahora, se define una función para el modelo epidemiológico, dicha función se ve de la siguiente forma:

$$\begin{aligned} S'(t) &= -\beta SI/N, && (3.2)\\ I'(t) &= \beta SI/N - \gamma I - \delta I, && (3.3)\\ R'(t) &= \gamma I, && (3.4)\\ D'(t) &= \delta I. && (3.5) \end{aligned}$$

Es posible permitir que los parámetros β, γ y δ sean estimados por otra función. Sin embargo, para ello, se define una función de pérdida que compare los datos reales de la variable *cum* con los del modelo de $C'(t)$, la función que define el cambio de los acumulados con respecto al tiempo del modelo (3.2)-(3.5) es:

$$C'(t) = \beta SI/N. \tag{3.6}$$

La función de pérdida se configuraría de la siguiente manera:

$$\sum_{i=0}^{t} (cum - C)^2 . \tag{3.7}$$

Una vez creada la función de pérdida, se pueden determinar los parámetros más óptimos para resolver el modelo. Se utilizará la librería "$curve\_fit$"de Python para comparar valores iniciales con los que el modelo va encontrando y devolver los más relevantes para cada variable. A partir de este proceso, se obtienen los siguientes valores para cada parámetro:

- $\beta \approx 1.108$,

- $\gamma \approx 0.719$,
- $\delta \approx 0.339$,

con estos parámetros se puede determinar el valor de $\mathscr{R}_0$:

$$\mathscr{R}_0 = \frac{\beta}{(\gamma + \delta)} = \frac{1.108}{(0.719 + 0.339)} \approx 1.045. \tag{3.8}$$

Como $\mathscr{R}_0 > 1$, se puede prever que la enfermedad se convierta en una epidemia. No obstante, no se puede garantizar si la enfermedad se vaya a convertir o no en endémica, ya que, según el modelo epidemiológico utilizado, cuando una persona se recupera de la infección, adquiere inmunidad. Por lo tanto, solo se pueden hacer suposiciones al respecto.

Con los parámetros ajustados se puede sacar una figura, ahora sí, de la comparativa entre los datos reales y modelados:

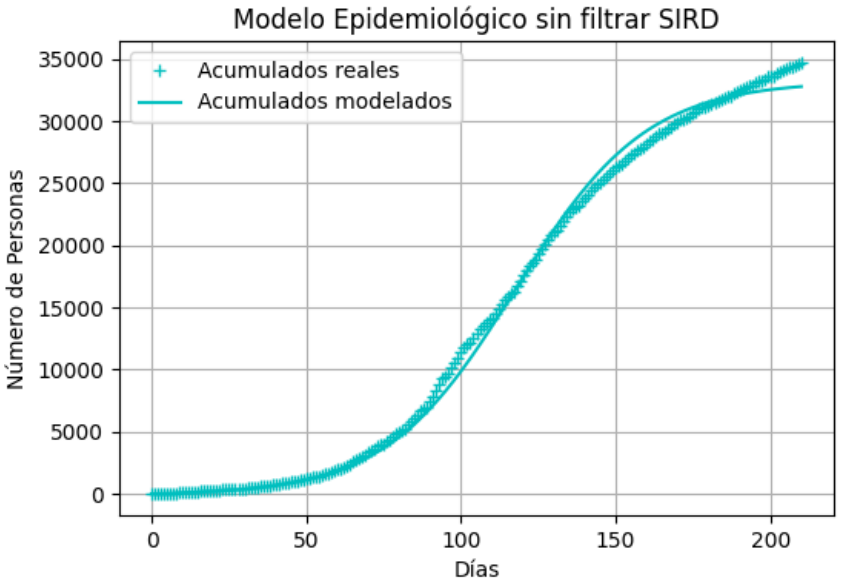

Figura 3.1: Modelo de los datos Acumulados reales vs modelados, es posible apreciar como la proyección del modelo supone una proyección de un máximo de 32783 individuos acumulados infectados, mientras que el máximo de personas acumuladas reportadas fue de 34594 dando como resultado una aproximación.

3.2. Filtro Kalman en la base de datos

Para comprender mejor cómo se implementará el filtro Kalman en la base de datos de interés, es relevante regresar al segundo ejemplo sobre el filtro, del cual se tomará la premisa de que la ganancia K no es constante y varía con las iteraciones. Además, se destaca que, al igual que en el modelo epidemiológico SIRD, se acotaron los datos a los primeros 211; no obstante, esta limitación no es necesaria y se realizó por motivos prácticos. Es posible incluir el conjunto completo de datos, es decir, analizar los 1086 datos de cada columna.
Dicho esto, se puede iniciar la definición de variables. Las matrices de transición de estado (F) y de observación (H) serán la matriz identidad $\mathbf{I}_4$, es decir:

$$F = H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.9}$$

De todas las ecuaciones explicadas anteriormente (2.74)-(2.83), se usarán las siguientes para un desarrollo apropiado del filtro Kalman:

$$\begin{aligned} \hat{x}_{k|k-1} &= F_k x_{k-1|k-1}, && (3.10)\\ P_{k+1|k} &= F_k P_{k|k} F_k^T + Q_k, && (3.11)\\ K_k &= P_{k|k-1} H_k^T S_k^{-1}, && (3.12)\\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k y_k, && (3.13)\\ P_{k|k} &= (\mathbf{I} - K_k H_k) P_{k|k-1}. && (3.14) \end{aligned}$$

- $\hat{x}_{k|k-1}$ es la estimación del estado del sistema en el tiempo k, basada en las observaciones hasta el tiempo $k-1$.
- F_k es la matriz del modelo de transición de estados que se aplica al estado estimado en el tiempo $k-1$.
- $x_{k-1|k-1}$ es la estimación del estado del sistema en el tiempo $k-1$, basada en las observaciones hasta ese momento.
- $P_{k+1|k}$ ecuación de extrapolación de covarianza.
- Q_k es la matriz de ruido del sistema.
- x_k es el estado real del sistema.
- K_k es la ganancia del filtro Kalman.
- H_k es la matriz de observación que relaciona el estado del sistema con las observaciones.
- S_k es la matriz de covarianza de la innovación.
- y_k la innovación o el error entre la medición real y la predicción.
- $\mathbf{I}$ es la matriz identidad.

Particularmente, como es un análisis de 4 dimensiones, el vector de estimación inicial tendrá 4 valores:

$$x_{0|0} = [1.0, 1.0, 1.0, 1.0], \tag{3.15}$$

con un error de estimación inicial de:

$$y_0 = [0.1, 0.1, 0.1, 0.1], \tag{3.16}$$

y una varianza:

$$S_k = \frac{1}{2}\mathbf{I}_4, \tag{3.17}$$

por tanto, la comparativa entre los datos reales y los filtrados entre cada columna del modelo SIRD se vería de la siguiente manera:

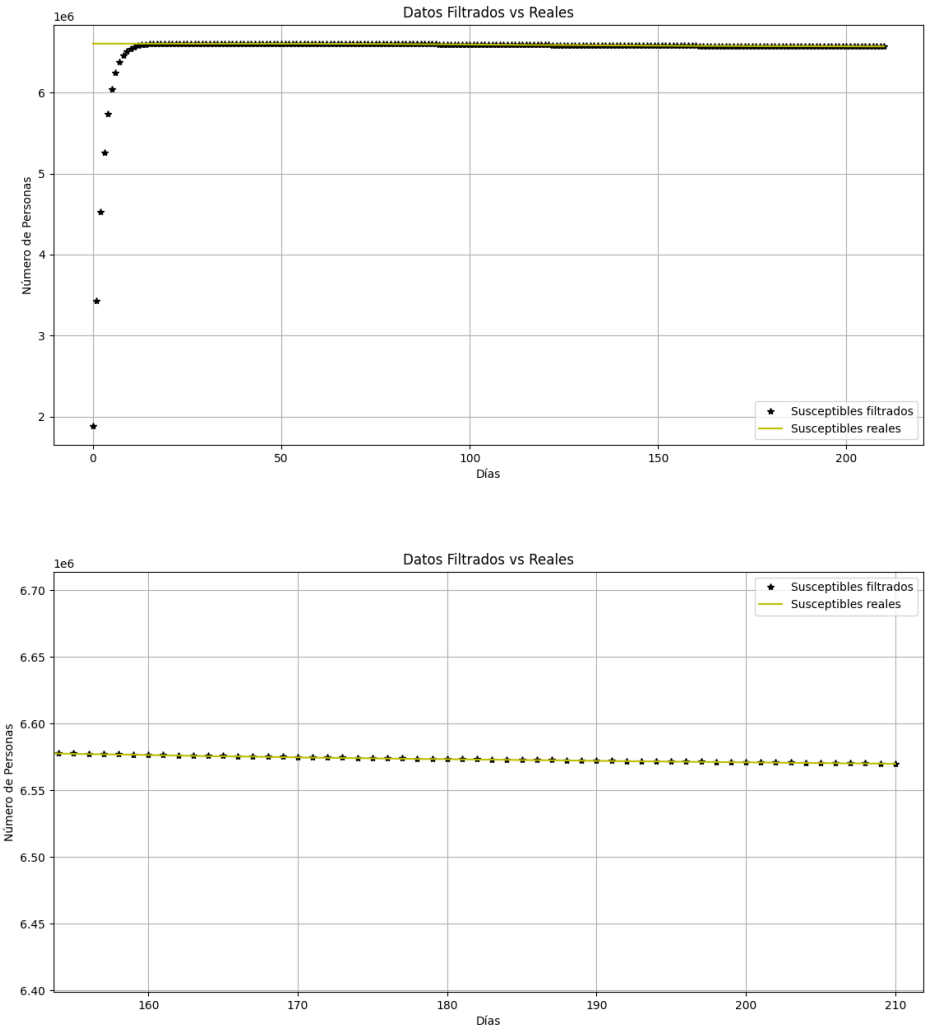

Figura 3.2: La primera figura representa como se ven los datos de los individuos susceptibles reportados (datos en amarillo), a comparación de los obtenidos por el filtro (color negro), los primeros valores arrojados por el filtro distan mucho de los primeros reales en consecuencia de las condiciones iniciales declaradas en las variables (3.15) y (3.16), donde el primer valor es uno y se estima un error de 0.1, también se puede apreciar como los datos no cuentan apenas con ruido, pues están muy juntos unos de otros, para ser más exactos, la diferencia entre un punto y el otro es de 203.86 unidades, tomando en cuanta que se está hablando de escalas de millones, se puede considerar como una buena aproximación. La segunda figura solo representa un acercamiento de la primera figura, además de que sirve para corroborar como el filtro poco a poco converge a los valores reales.

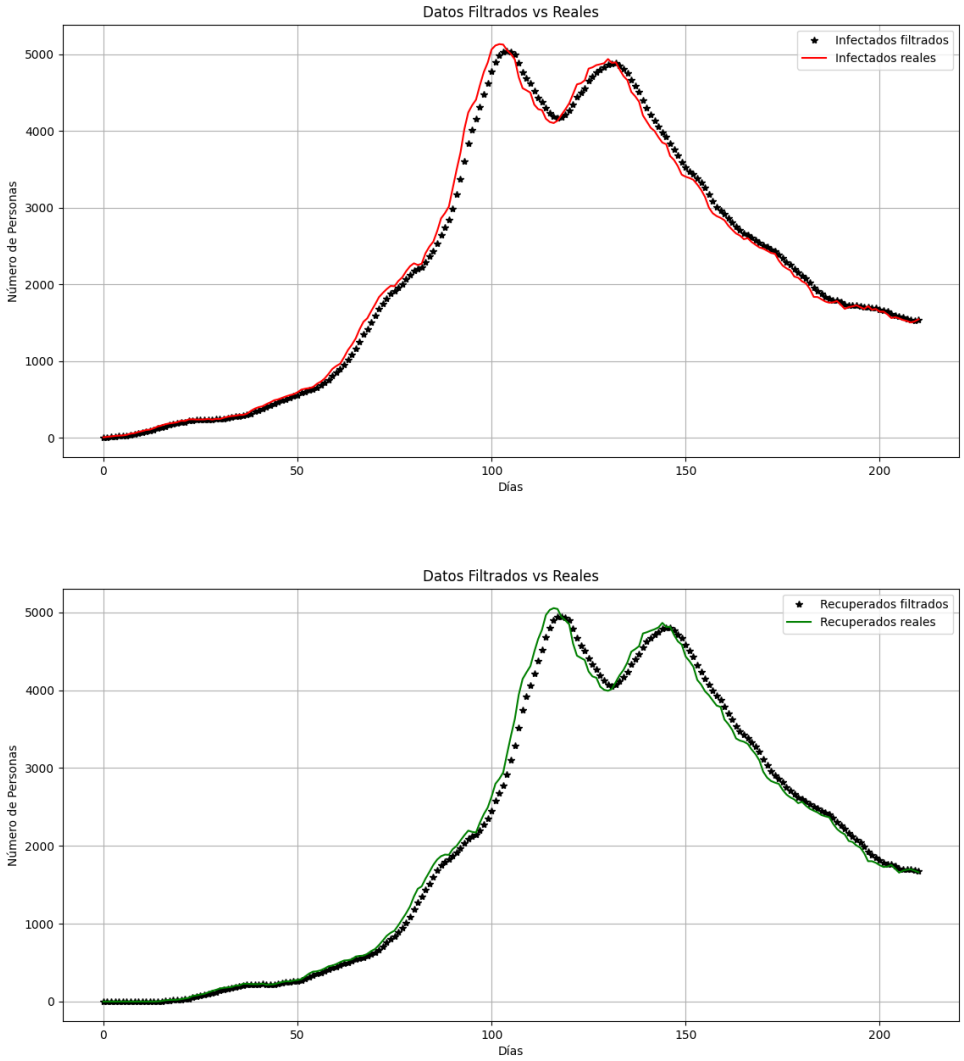

Figura 3.3: Para propósitos de mejor comprensión se colocaron juntas las figuras de los individuos infectados (color rojo), con su respectivo filtro (color negro), con la de los recuperados (verde) y nuevamente con su respectivo filtrado (negro). En el sistema de ecuaciones utilizado para el modelo SIRD (3.2) - (3.5), para ser más específicos la función (3.4) indica la relación entre las personas infectadas y como estas pueden llegar a recuperarse, luego de un tiempo determinado $\frac{1}{\gamma}$, lo que quiere decir en el análisis de este trabajo es que la figura de recuperados está desfasada 15 días con respecto a la de los infectados.

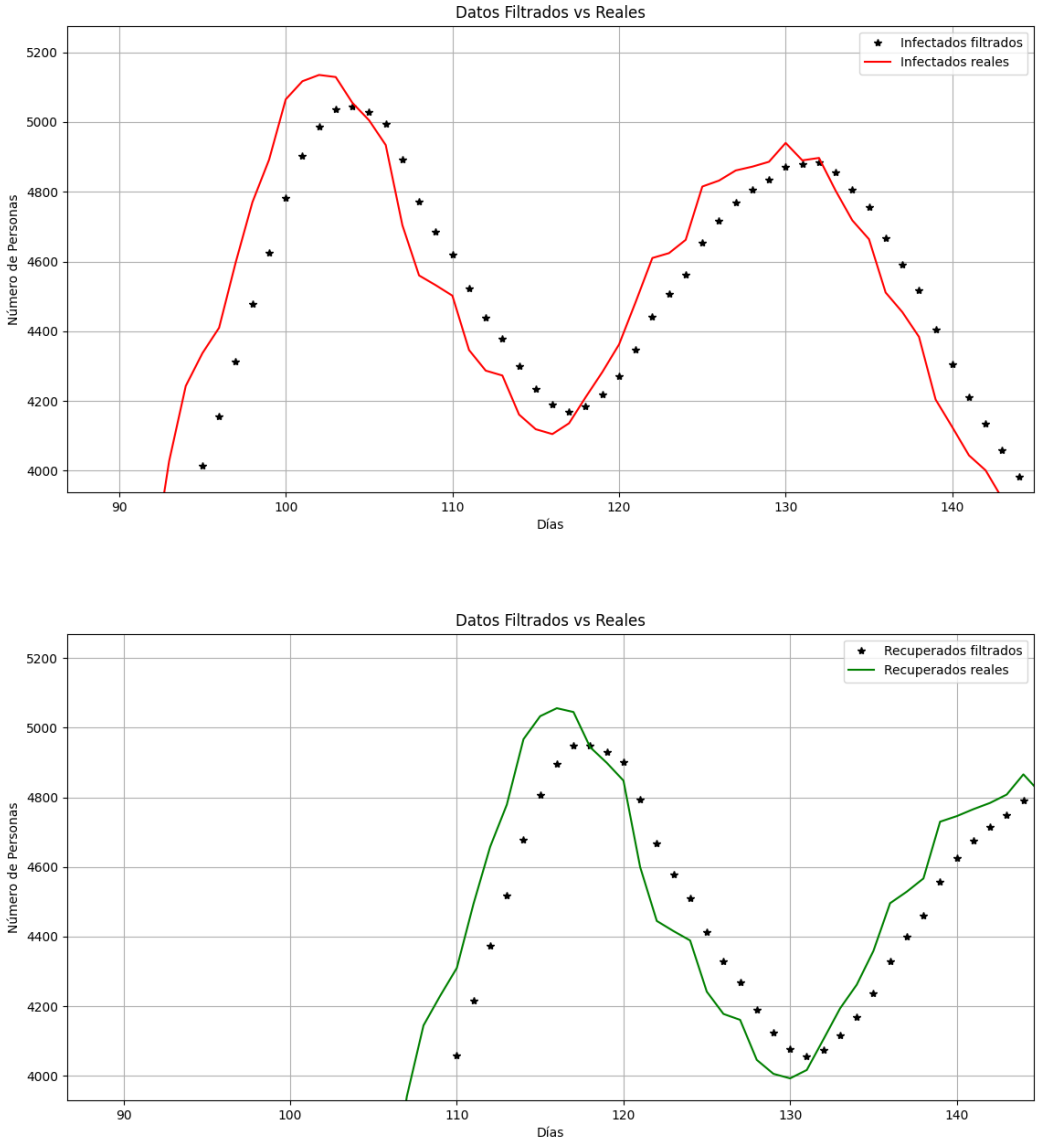

Figura 3.4: Nuevamente se presenta un acercamiento a las figuras en momentos destacables, como son los picos más altos, se destaca como estos no llegan a coincidir incluso con el desfase, la explicación se debe a aquellos individuos que fallecieron y ya no forman parte de los recuperados. En ambas comparativas también se puede ver como los datos cuentan con poco ruido, aun así, el filtro sí presenta una figura más definida y continua que los datos reportados.

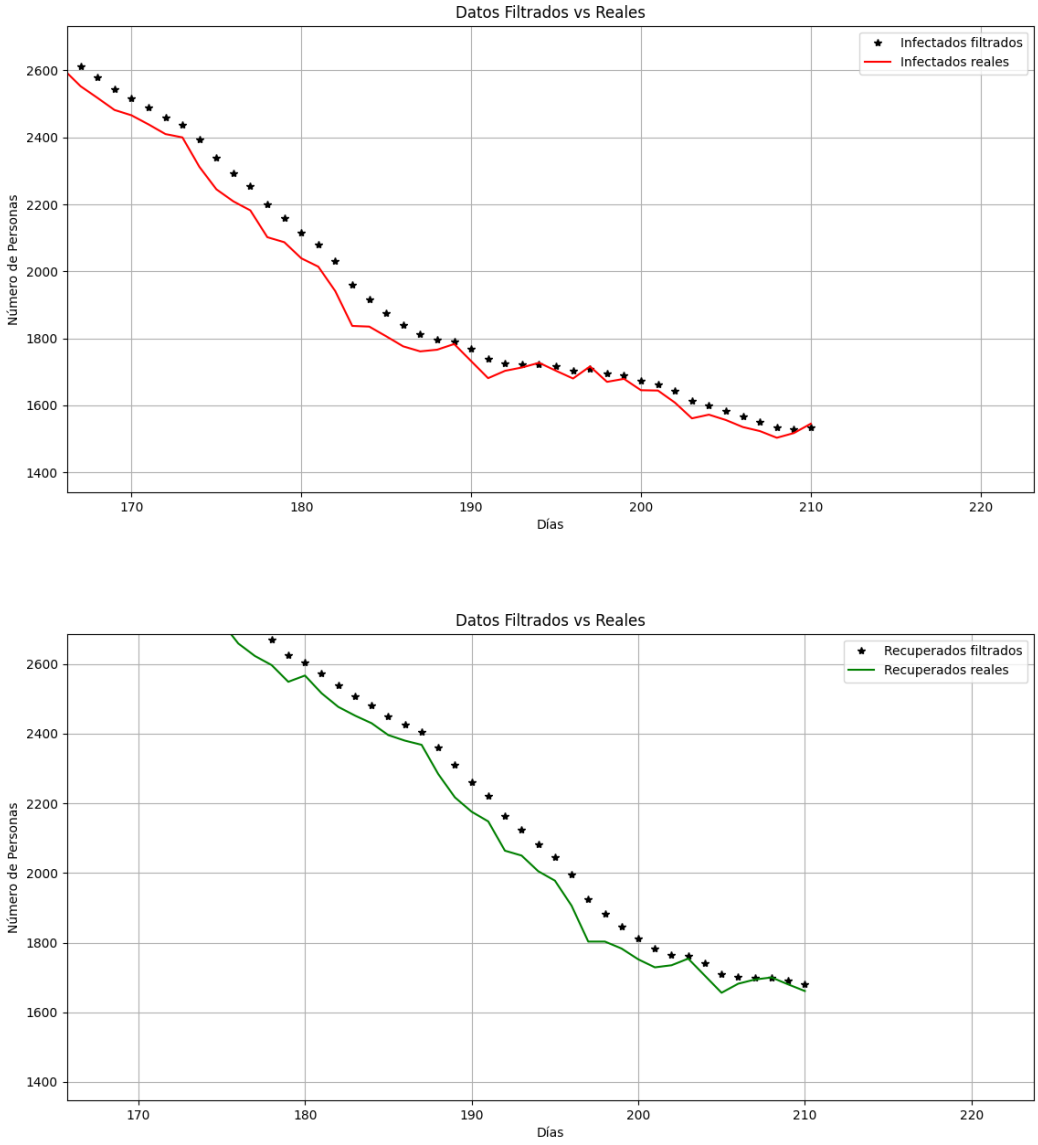

Figura 3.5: Por último, solo resta analizar un poco los finales respectivos de cada figura de individuos infectados y recuperados, donde una vez más los datos filtrados, tanto en infectados como en recuperados, se aprecia una línea punteada más continua y menos ruidosa a la de los datos reportados.

Es importante mencionar las figuras son exclusivamente de los conteos diarios, es decir, representa la cantidad diaria de individuos que se fueron reportando como positivos para COVID-19. Para este estudio no se tomarán en cuenta los individuos recuperados mostrados en la Figura (3.3), debido a que son solo estimaciones a partir del conteo de confirmados diarios, dicho de otra forma, los datos de recuperados presentados, no son datos reales proporcionados por el CONAHCYT [2], es solo una muestra con fines ilustrativos. Por lo que este estudio solo considera a los individuos susceptibles, los casos diarios confirmados (solo para hacer comparativas con el modelo, no debe

confundirse con los individuos infectados, los cuales tampoco son declarados por el CONAHCYT [2]), fallecidos y acumulados.

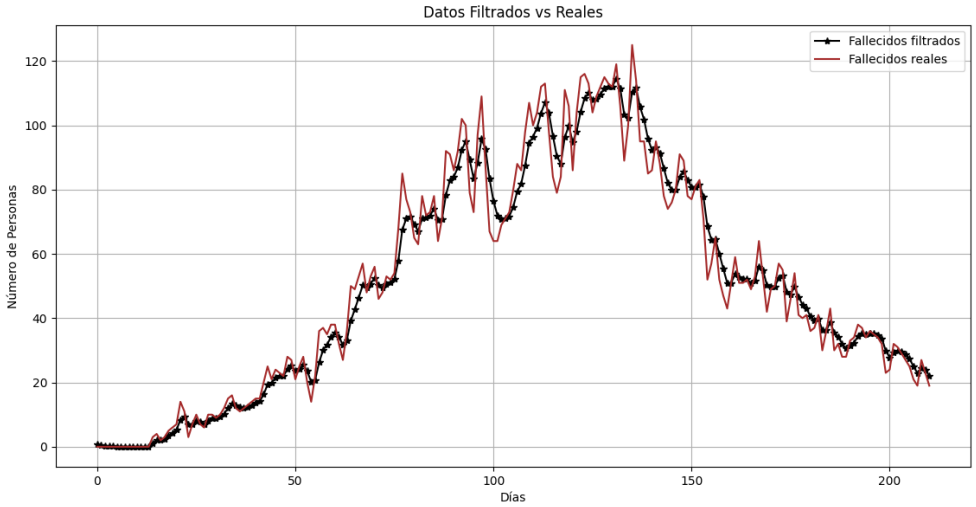

Figura 3.6: Con diferencia la figura con más ruido en este trabajo es la de los individuos fallecidos, debido a esto, es también donde se puede apreciar de mejor manera como se va suavizando la silueta de estos datos con el filtro Kalman (negro), se destaca que estos datos no se ven como los datos presentados en la figura (3.2) y (3.3), donde los datos presentados y los filtrados son prácticamente los mismos, por lo que esta figura es la más importante solo por mostrar el esbozo de una gráfica prácticamente sin ruido, de donde se pueden extraer datos más legibles y con menos sesgo para su manejo.

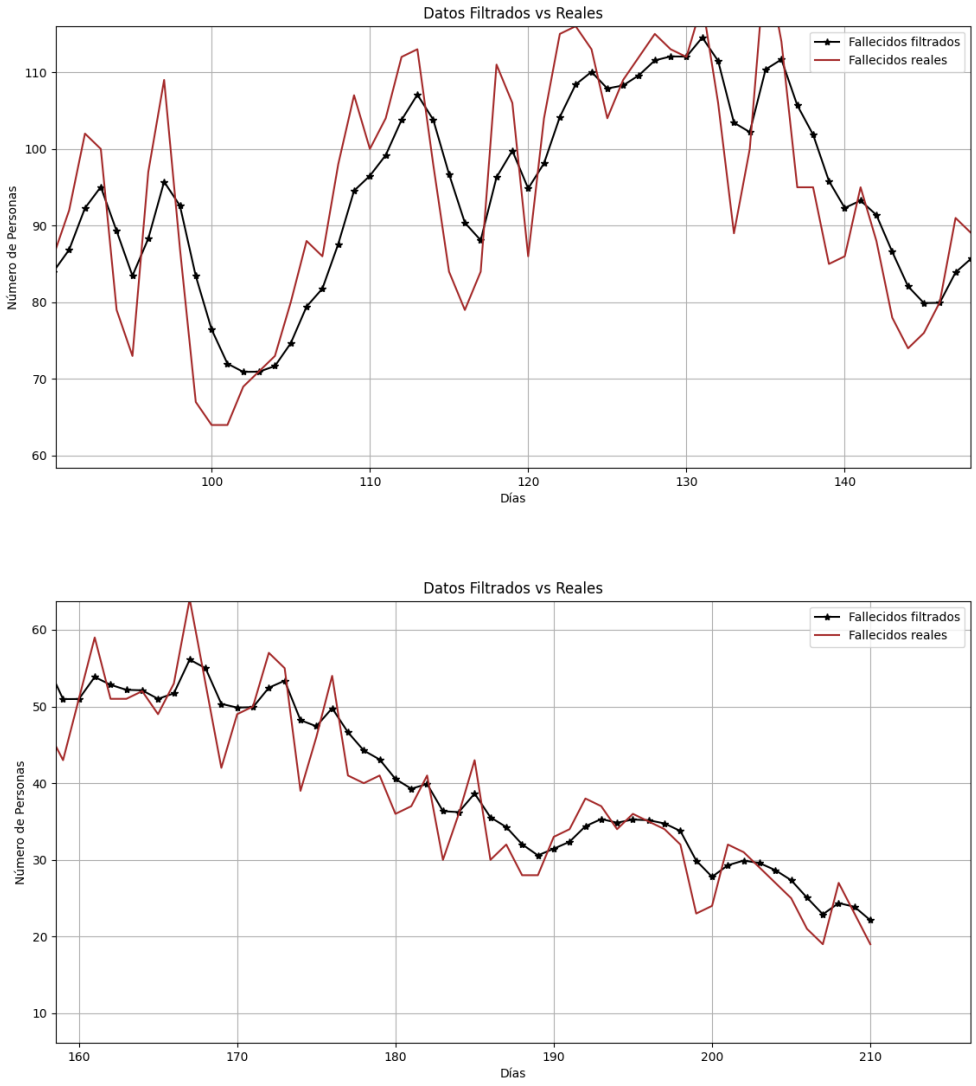

Figura 3.7: Haciendo un acercamiento a los picos de la figura (3.6), se ven unos bordeados mucho más redondos y con valores no dispersos uno del otro, mientras más datos son proporcionados la figura que se va esbozando por el filtro es más precisa a los valores reportados.

En este caso, también se hace la distinción de que es el conteo de fallecidos diarios, mostrando un conteo que se asemeja más a un modelo, se vería de la siguiente manera:

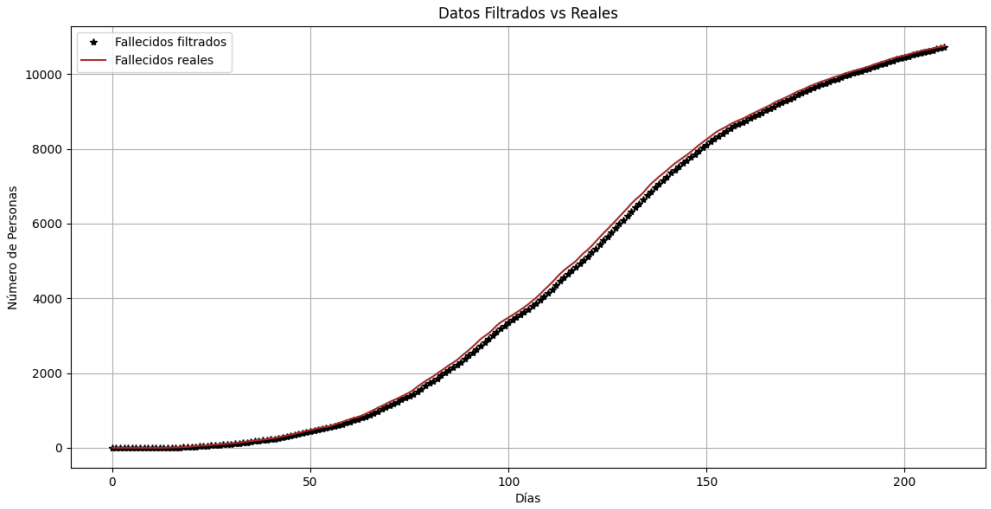

Figura 3.8: Los datos son continuos y prácticamente sin ruido, el filtro es prescindible, por lo que los datos reales (color café) y los datos filtrados (color negro) son similares.

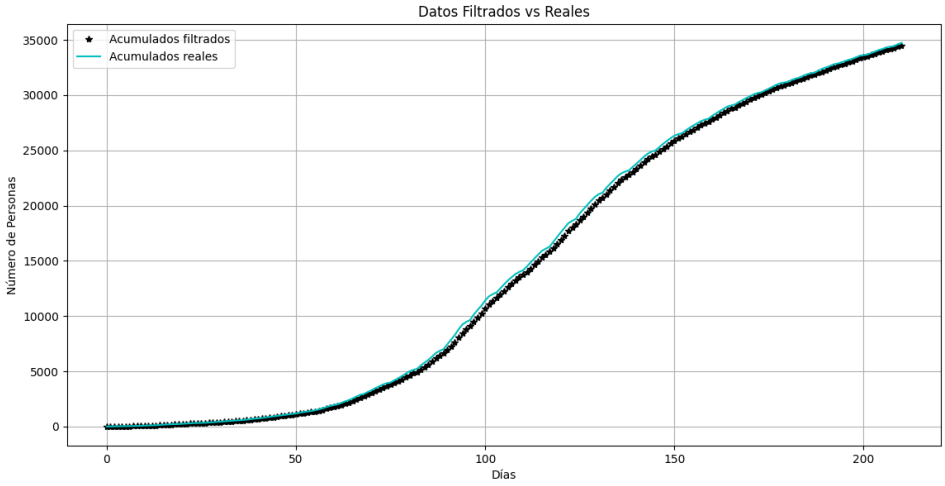

Figura 3.9: En el caso de los datos acumulados, el filtro Kalman difiere por unidades de los valores reales, porque como se puede apreciar en la figura, los datos carecen de una cantidad significativa de ruido para filtrar, dando como resultado una comparativa cuasi idéntica.

Una vez analizados los datos, solo resta comparar los parámetros encontrados cuando los datos no están filtrados contra los que sí. Para esto, la gráfica elegida será la de los individuos acumulados, por el comportamiento de la gráfica, dando este resultado:

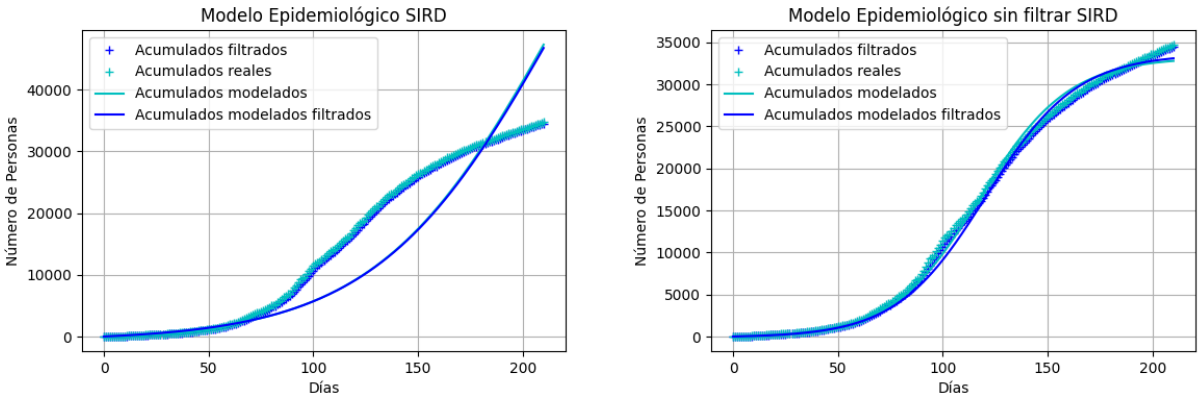

Figura 3.10: La comparación entre el modelo utilizando datos reportados (color cian) y el modelo que emplea datos filtrados (color azul) revela diferencias significativas. En la figura de la izquierda, se observa claramente que la aproximación del modelo, tanto con datos reales como con datos filtrados, no es una buena estimación. En contraste, en la imagen de la derecha, la aproximación del modelo es mucho más cercana a los datos reales, lo que indica una mejor precisión en la predicción. El principal motivo por el cual las figuras presentadas discrepan tanto es por el tamaño de la población, en el caso de la izquierda la población N no fue ajustada, en la derecha, N se consideró como un parámetro más a ajustar.

La discrepancia entre el modelo representado en el lado izquierdo y los datos reales se atribuye a múltiples factores. El tamaño de la población desempeña un preciso papel en la modelación de una epidemia, justo como fue mencionado anteriormente. La elección cuidadosa del tamaño poblacional determina la fidelidad de la simulación y los patrones de propagación de la enfermedad, considerando la duración de la simulación y los posibles parámetros relevantes de la enfermedad. En este caso particular, se ha considerado la población total de Puebla, México, que asciende a $6,604,451$ personas (datos de febrero de 2020). Ambas estimaciones se realizaron bajo la suposición de que el coeficiente de transmisión de la enfermedad, β, alcanzaría un valor de hasta 4, lo que indica una alta tasa de contagio. Sin embargo, se observa un crecimiento insuficiente en los casos acumulados representados en la curva del modelo, al final de la simulación (primera ola), que aún no alcanza su máximo. Para que la figura del lado izquierdo se asemeje más a los datos reales, sería necesario aumentar el límite del parámetro β a valores extremadamente altos, hasta 10, pero incluso así, la aproximación seguiría siendo inadecuada.

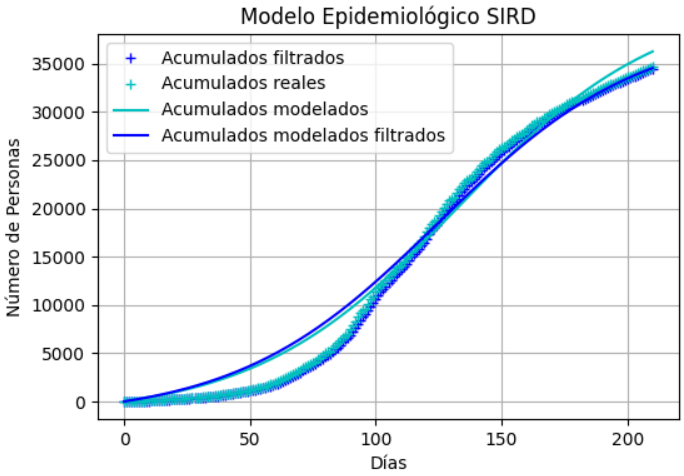

Figura 3.11: El coeficiente de transmisión obtenido aquí es de 7.5 (datos reales) y 8.1 (datos filtrados).

Considerando que los valores de transmisión encontrados en la Figura (3.11) sean correctos, esto representa una enfermedad sin precedentes, donde un solo individuo infectado es capaz de enfermar a 8 sin mayor complicación, y en retrospectiva, dicha enfermedad tardaría un estimado de 5 horas en ser curada por el individuo en caso de que este se pueda recuperar, estos resultados no coinciden en lo absoluto a los observados por el COVID-19, los cuales en sus inicios, un individuo infectado tardaba un mínimo de 14 días en recuperarse. Entonces, para poder obtener resultados más fiables, fue necesario acotar la población N del modelo para que pueda coincidir lo mejor posible con los datos reportados reales, siendo la figura de la derecha de (3.10) el resultado de este nuevo parámetro a considerar.
Con todo lo antes mencionado se concluye lo siguiente respecto a los casos acumulados. El valor encontrado de la población N es $397,158$ individuos para los datos reales y $442,309$ para los datos filtrados, lo que se puede interpretar como la máxima cantidad que individuos a los que 5 personas pueden tener acceso (por trabajo, medios de transporte, lugar donde reside el individuo infeccioso, como las calles aledañas, la colonia, los lugares donde va a comer), lo cual dista mucho del valor de la población original, indicando de manera directa y muy interesante la ley de acción de masas (ya que esta se está cumpliendo en subgrupos de la población general y no en todo el estado, como dicta la teoría). En 1906 William Hamer mientras estudiaba la recurrencia de la aparición del sarampión, plantea que "la tasa a la cual una enfermedad se propaga es proporcional al número de individuos susceptibles por el número de individuos infecciosos", la cual es la definición de esta ley.
Los parámetros hallados en esta comparativa para la Figura de la derecha en (3.10) fueron:

Parámetro	Descripción	Valor	
		D. sin filtrar	D. filtrados
β	Tasa de infección	1.108	1.074
γ	Tasa de Recuperación	1.045	1.030
$\mathscr{R}_0$	Valor de Reproducción inicial	1.0456	1.0451
N	Población del modelo	388863	401557

Empíricamente, se observaba que un individuo que se contagiaba tardaba un mínimo de 14 días en recuperarse, lo que indica un valor γ de mínimo 0.08, sin embargo, el valor encontrado es de 0.715 y 0.706 respectivamente, a lo que se le pueden atribuir dos explicaciones, la primera es considerar

la descripción del libro del coeficiente y aseverar que los individuos que contraían la enfermedad tardaban un tiempo aproximado de 1.5 días en recuperarse y adquirir inmunidad, lo cual no es consistente con los datos recavados en esta tesis; la siguiente explicación sugiere otro punto de vista, donde este coeficiente aún es un inverso de tiempo, es decir, $\frac{1}{\gamma}$ = tiempo en días, pero la interpretación dice que este es lo que tarda un individuo en aislarse y comenzar con su recuperación y evitar en la medida de lo posible generar más contagios, este modelo no considera la posibilidad de que el individuo pueda generar más contagios luego de aislarse.
Se observan solo dos parámetros (coeficiente de infección y tasa de recuperación) pese a que en un modelo SIRD se deben abarcar tres (coeficiente de infección, tasa de recuperación y de mortalidad), esto sucede porque no hay una base de datos de individuos recuperados, dificultando un poco más el proceso, sin embargo, al contar con la base de fallecidos, se puede hacer un ajuste con el modelo SIR para encontrar una interpretación aceptable:

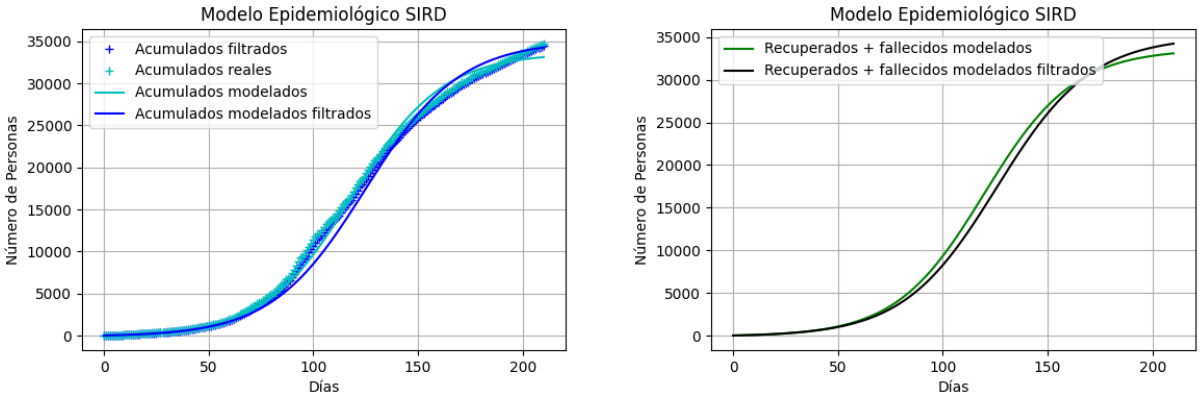

Figura 3.12: De lado izquierdo se encuentra el modelo de individuos acumulados, con la comparación entre los datos reales (color cian y azul, pero con el símbolo de -") y el modelo (las funciones continuas) compartiendo el mismo color entre datos real-modelo para evitar confusiones de interpretación, del lado derecho, una comparativa entre modelos, las predicciones de los datos reales (verde) y de los datos filtrados (negro).

Con estas predicciones para los recuperados más los fallecidos y tomando en consideración los datos sobre individuos fallecidos, es posible realizar un ajuste y armar el modelo SIRD, el cual se vería de la siguiente manera:

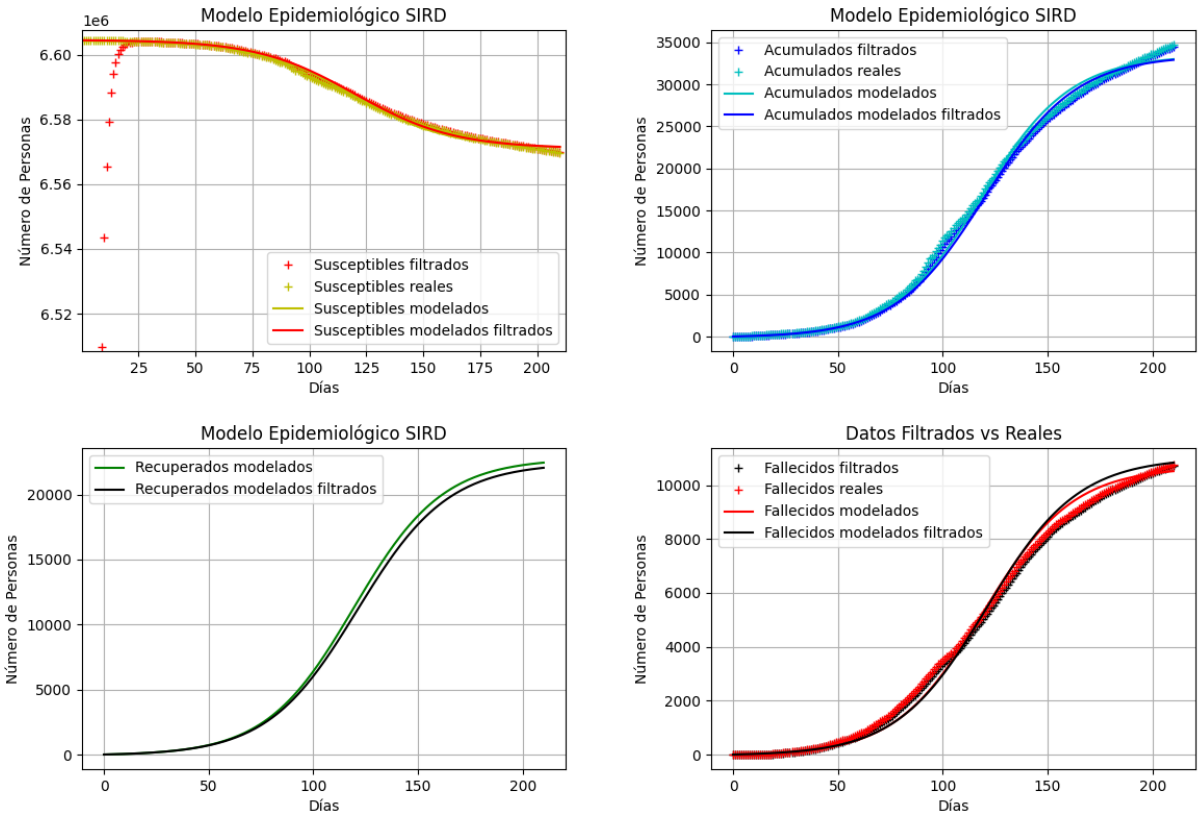

Figura 3.13: La figura de arriba a la izquierda representa la evolución de la población de susceptibles con el número real (6,604,451), donde se observan las comparativas entre los datos reales y su modelo respectivo (amarillo) y los datos filtrados junto a su modelo (rojo); a la derecha de esta figura, se encuentra nuevamente los datos acumulados, la piedra angular de este modelo, las comparativas entre los datos reales (cian) y los datos filtrados (azul); abajo a la izquierda son las predicciones del modelo del comportamiento de los individuos recuperados; finalmente abajo a la derecha se encuentra el ajuste de los fallecidos, al ver el comportamiento del modelo, se puede aseverar que es una buena predicción, ya que los datos no distan mucho de los reportes, los datos reales (rojo) y los filtrados (negro), son similares a los modelados.

Al ver la figura sobre el modelo de los individuos susceptibles, es claro ver como la aproximación de la población ayudó en gran medida para que fuera un modelo viable y así ver el comportamiento de la enfermedad, al observar este conjunto de figuras, es posible preguntarse si no se trata entonces de un modelo SIR en lugar de un SIRD por la presentación de los datos, sin embargo, la presencia del modelo de recuperados y fallecidos en distintas figuras y además representando distintos valores, son prueba suficiente para acreditar esta representación de figuras como un modelo SIRD, con todo lo mencionado anteriormente, para ser llamado un modelo SIRD apropiado, todavía falta una última figura, a partir de todos los datos y parámetros encontrados fue creado un modelo sobre los individuos infectados, el cual es el siguiente:

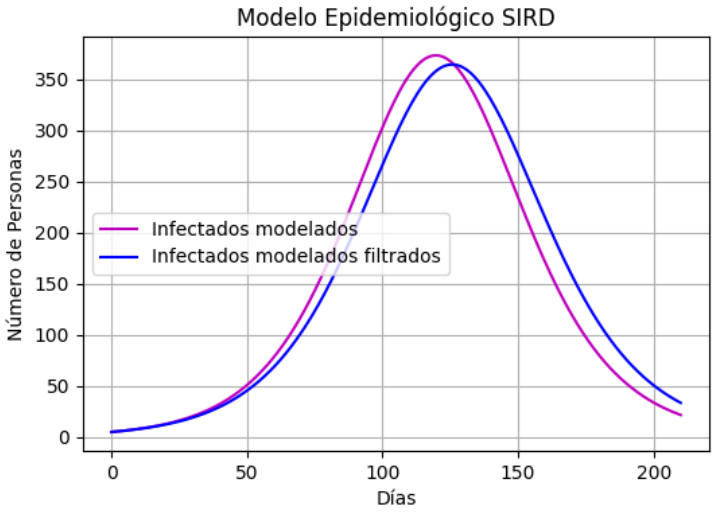

Figura 3.14: Representa la estimación del modelo de como fue el comportamiento de los individuos infecciosos en la primera ola de, a su vez, se puede apreciar el característico comportamiento de los individuos infectados (forma parcial de una campana de Gauss), alcanzando su pico máximo de 373 infectados en el día 120 de la simulación (datos reales, color magenta) con su respectiva comparación con un modelo de datos filtrados (color azul), donde el pico infeccioso fue de 364 individuos en el día 126.

Los parámetros asociados al modelo representado en las figuras (3.13) son los siguientes:

Parámetro	Descripción	Valor	
		D. sin filtrar	D. filtrados
β	Tasa de infección	1.108	1.074
γ	Tasa de Recuperación	0.715	0.706
δ	Tasa de Mortalidad	0.33	0.321
σ_δ	Desviación estándar de la tasa de mortalidad	$0.04 \approx 12\,\%$	$0.01 \approx 5.3\,\%$
$\mathscr{R}_0$	Valor de Reproducción inicial	1.0456	1.0451
N	Población del modelo	388863	401557
σ_N	Desviación estándar de la Población	$8758 \approx 2.2\,\%$	$1712 \approx 0.4\,\%$

Donde los parámetros γ y δ del modelo SIRD son una descomposición del parámetro γ presentado en la tabla (3.2), como esta indicado por la desviación estándar, el parámetro δ es una buena aproximación del modelo que ajusta a los individuos fallecidos, de manera similar, se observa como la estimación mejora en el caso de usar datos filtrados. En el caso de la población, se puede destacar que la aproximación es mejor cuando se usan datos filtrados ya que la desviación es menor en un $1.8\,\%$.

Capítulo 4

Conclusiones

En este trabajo de tesis, se observaron distintos factores que contribuyeron a un mayor entendimiento de las interrogantes planteadas al principio, como la importancia de considerar enfoques diversos, como el filtro de Kalman, para mejorar la calidad de los datos y, por ende, la eficiencia de los modelos en la predicción y gestión de situaciones pandémicas. Considerar a los individuos acumulados en lugar de usar a los infectados, debido a que no se tiene una medición de la prevalencia, es necesario ajustar los modelos a los casos acumulados calculados por medio de la incidencia observada.

La importancia de delimitar la población como un parámetro adicional a estimar resultó crucial para la resolución definitiva del modelo. Gracias a la ley de acción de masas, fue posible otorgarle significado al considerar N como un parámetro. La subpoblación encontrada se interpreta como la mayor cantidad de individuos que los infectados pueden llegar a infectar (susceptibles). Esto confirma directamente la importancia de restringir adecuadamente el tamaño de muestra de la población N al realizar un modelo epidemiológico. Como se demostró previamente, si se considerara la población de todo el estado en lugar de estimarla, el desfase del modelo con los datos reales reportados sería significativo, resultando en un modelo deficiente y erróneo.

Es posible resolver un modelo SIRD usando un SIR como base, ya que, al no tener conocimiento del comportamiento de los individuos recuperados, pero si, de los fallecidos (mostrados como conteo acumulado figura (3.8)) se combinan ambos parámetros (coeficiente de recuperación γ y tasa de mortalidad δ) para convertir el modelo propuesto en uno más sencillo, una vez resuelto el modelo SIR, se usó el parámetro encontrado para poder ajustar la población de fallecidos y estimar el parámetro restante γ, haciendo posible la creación de un modelo SIRD a partir del SIR (figuras (3.13) y (3.14)).

El filtro Kalman mostró su eficiencia con datos que contenían una gran cantidad de ruido, como el conteo de los fallecidos diarios, véase en las figuras (3.6) y (3.7). Al referirse a la palabra ruido", se hace alusión a la diferencia entre los datos reales y los datos estimados, es decir, la figura (3.6) tiene datos con ruido"porque existe una diferencia considerable entre los datos presentados y los modelados. A su vez, demostró ser útil cuando los datos no presentaban cantidades de ruido para tomar en cuenta, como en las figuras (3.3), (3.4) y (3.5), por último, se halló que también es un buen estimador, ya que en unos pocos pasos, es capaz de alcanzar el valor de las mediciones y luego converger al valor real de la medición, el mejor ejemplo de esto es la figura (3.2). La importancia de usar estos pasos anteriores era solo como demostración de la capacidad que el filtro posee para adaptarse a varios escenarios, con grandes o pocas cantidades de ruido en los datos, el filtro no se verá afectado. En donde el paso final y más importante era corroborar si todas las herramientas que posee permiten mejorar la estimación de parámetros en el modelo epidemiológico designado.

Se corrobora que utilizar el filtro si cambia los valores de los parámetros, en algunos casos varían por muy poco, como el caso del valor de reproducción inicial $\mathscr{R}_0$, en otros las comparativas discrepan en cien mil unidades (estimación de la población N), por lo que se puede decir que si tiene

mayor fiabilidad calcular los parámetros de esta manera. Sin embargo, es importante destacar que al proporcionar medidas más exactas gracias al uso del filtro, los resultados mostrados tienen mayor veracidad y precisión. Aquellos datos analizados sin utilizar un filtro pueden provocar mayores sesgos o riesgos al predecir comportamientos en la población o los posibles impactos que esta tendrá.

Futuras líneas de investigación.

Con todas las consideraciones y comentarios hechos anteriormente, existen numerosas posibilidades para poder enriquecer el entendimiento de una pandemia y poder mejorar este trabajo.

- Para el sistema de ecuaciones utilizado del filtro Kalman (3.10) - (3.14), sería interesante ver el comportamiento del filtro al cambiar la ecuación de extrapolación de la covarianza (3.11) por una que muestre la diferencia entre el estado real del sistema con el predicho (2.77).
- Pese a que fue utilizado un modelo SIR para resolver un SIRD, es conveniente corroborar el comportamiento de β en distintos instantes de tiempo, donde se puede hacer semanalmente (para tener un mejor entendimiento de su evolución conforme iba pasando el tiempo de la pandemia) o, hacerlo al principio, a la mitad y al final de la misma, ambos son dos puntos de vista diferentes, pero que otorgarán una vista más amplia sobre la evolución del parámetro infeccioso.
- Retomando el punto dos, ver la evolución del modelo y las estimaciones arrojadas con las respectivas evoluciones del parámetro infeccioso, es decir, se debería ver reflejado el aislamiento de la población, medidas de las autoridades para reducir los contagios y la respuesta de la población ante tales medidas.

Apéndice A

Apéndices

A.1. Apéndice A: Correlación de los datos

```
install.packages("pheatmap")
library(pheatmap)


##############################################################################
#########################   Lectura de archivos   ############################
##############################################################################

setwd("D:\\KALMAN") #Colocar la dirreccion donde se encuentra la base de datos
data <- read.csv("database.csv") #leerla
data <- (subset(data,Day<212 )) #acotarla a la primera ola
columnas <- c(7,3,5) #tomar solo las columnas de interes (casos acumulados,
    reportes diarios y fallecidos diarios)
new_data <- data[columnas] #unir los dos pasos anteriores

##############################################################################
#################   Mapa de calor de correlaciones   #########################
##############################################################################


correlaciones = cor(new_data[,1:3]) #calcular las correlaciones
correlaciones

pheatmap(correlaciones,
         display_numbers = TRUE,
         number_color = "black",
         fontsize_number = 8,     #Graficar
         cluster_cols = FALSE,
         cluster_rows = FALSE)
```

Dependiendo de los datos que se deseen analizar las correlaciones, se pueden obtener cualquiera de estos dos ejemplos:

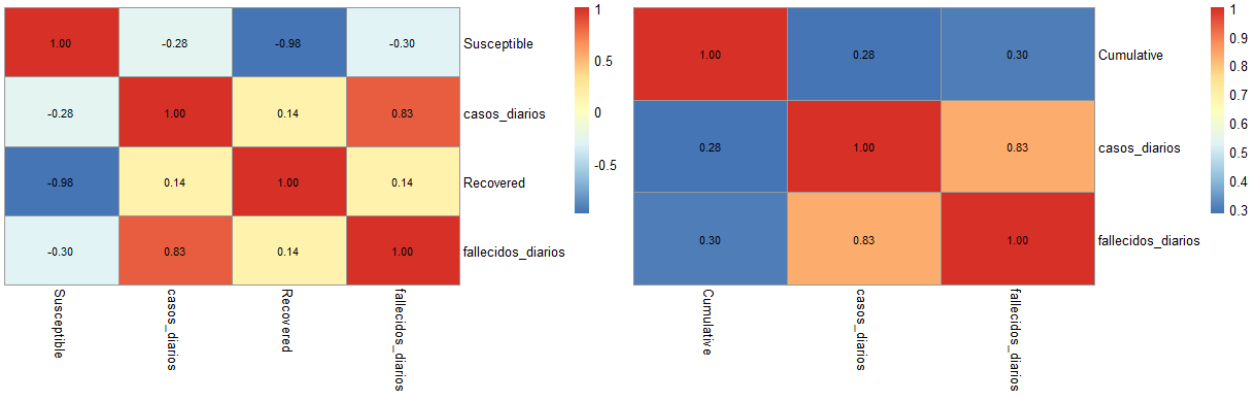

Figura A.1: Esta es una representación más visual de las correlaciones halladas, la figura de la izquierda contempla a los individuos susceptibles, además de sumar la aproximación de los recuperados. La figura de la derecha muestra las correlaciones solo de los datos conocidos al principio del análisis.

A.2. Apéndice B: Código implementado

```
import numpy as np
import pandas as pd
import csv
from scipy.integrate import odeint
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit

# Declarar variables
a = 211
data = pd.read_csv("database.csv")
data = data.iloc[:a, [1, 2, 8, 6]] #Susceptibles, Casos_diarios, Fallecidos,
    Acumulados
data = data.to_numpy()

#Elaborar filtro kalman
def kalman_filter(data, initial_estimate, initial_error_estimate, process_variance,
     measurement_variance):
    estimates = []
    estimate = np.array(initial_estimate)
    error_estimate = np.diag(initial_error_estimate)

    # Definir matrices para el proceso del modelo
    F = np.eye(4)
    #B = np.eye(4)
    H = np.eye(4)
    for measurement in data:
        # Predict
        prediction = np.dot(F, estimate)
        prediction_error = np.dot(np.dot(F, error_estimate), F.T) + process_
            variance

        # Update
        kalman_gain = np.dot(np.dot(prediction_error, H.T), np.linalg.inv(np.dot(np
            .dot(H, prediction_error), H.T) + measurement_variance))
```

```
        estimate = prediction + np.dot(kalman_gain, measurement - np.dot(H,
            prediction))
        error_estimate = np.dot(np.eye(4) - np.dot(kalman_gain, H), prediction_
            error)

        estimates.append(estimate)

    return np.array(estimates)

# Ejemplo de uso
initial_estimate = [1.0, 1.0, 1.0, 1.0]
initial_error_estimate = [0.1, 0.1, 0.1, 0.1]
process_variance = 0.1 * np.eye(4)  # Varianza del proceso (matriz 5x5)
measurement_variance = 0.5 * np.eye(4)

estimates = kalman_filter(data, initial_estimate, initial_error_estimate, process_
    variance, measurement_variance)
#print(estimates)
filter_data = estimates
# Nombres de las columnas
nombres_columnas = ["filtered_sus", "filtered_inf", "filtered_ded", "filtered_cum"]

# Especifica el nombre del archivo CSV donde deseas guardar los datos
archivo_csv = 'filtered_database.csv'

# Abre el archivo CSV y escribe los datos con nombres de columna
with open(archivo_csv, 'w', newline='') as archivo:
    escritor_csv = csv.writer(archivo)

    # Escribe los nombres de las columnas en la primera fila
    escritor_csv.writerow(nombres_columnas)

    # Escribe los datos
    for fila in filter_data:
        escritor_csv.writerow(fila)
filter = pd.read_csv("filtered_database.csv")
filter = filter.to_numpy()

b = 0
c = a - b

f = 10
tiempo = np.linspace(0, c - 1, c) #tiempo en dias de la pandemia
datos_cum = data[b:a, 3]
datos_cumf = filter[b:a, 3]
#N = 6604451 #posible cambio de variable
If0 = filter[1, 1]
Rf0 = 0
Df0 = 0
Cf0 = filter[1, 3]
Sf0 = 6604451 - If0

S0 = data[0, 0]
I0 = data[0, 1]
D0 = data[0, 2]
C0 = data[0, 3]

def modelo_sird(t, beta, N, delta):
    S0, I0, D0, C0 = N-5, 5, 0, 5  # Condiciones iniciales
    y0 = [S0, I0, D0, C0]

    # Sistema de ecuaciones diferenciales
    def sistema(y, t):
        S, I, R, C = y
        dSdt = - beta * S * I / N
```

```
        dIdt = beta * S * I / N - delta * I #- delta * I
        dRdt = delta * I
        #dDdt = delta * I
        dCdt = beta * S * I / N
        return [dSdt, dIdt, dRdt, dCdt]

    # Resolver el sistema de ecuaciones diferenciales
    sol = odeint(sistema, y0, t)

    # Devolver todas las variables del modelo
    return sol[:, 0], sol[:, 1], sol[:, 2], sol[:, 3]


def modelo_sird_fil(t, beta, N, gamma):
    Sf0, If0, Rf0, Cf0 =N - 5.038461538461538, 5.038461538461538, 0,
        5.038461538461538
    yf0 = [Sf0, If0, Rf0, Cf0]


    # Sistema de ecuaciones diferenciales
    def sistema_f(y, t):
        S, I, R, C = y
        dSdt = - beta * S * I / N
        dIdt = beta * S * I / N - gamma * I #- delta * I
        dRdt = gamma * I
        #dDdt = delta * I
        dCdt = beta * S * I / N
        return [dSdt, dIdt, dRdt, dCdt] #dDdt, dCdt]

    # Resolver el sistema de ecuaciones diferenciales
    sol_f = odeint(sistema_f, yf0, t)


    return sol_f[:, 0], sol_f[:, 1], sol_f[:, 2], sol_f[:, 3]


def ajuste_casos_acumulados(t, beta, N, gamma):
    _, _, _, acumulados = modelo_sird(t, beta, N, gamma)
    return acumulados

def ajuste_casos_acumulados_fil(t, beta, N, gamma):
    _, _, _, acumulados_fil = modelo_sird_fil(t, beta, N, gamma)
    return acumulados_fil

# Datos reales (t, casos acumulados)
t_real = np.linspace(0, 210, 211)
datos_reales = datos_cum
datos_reales_filtrados = datos_cumf


lower_bounds = [0, 0, 0]
upper_bounds = [4, 1000000, 4] #[7, 3.3, 1.85][1, 1000000, 0.08, 0.04]
bounds = (lower_bounds, upper_bounds)
parametros_iniciales = [1.1, 600000, 1.0]  # Posibles parametros que sirvan de algo
    :,v[3.2, 3.1, 1.7][.1, 55000, 0.08, 0.04]
parametros_optimos, covariance = curve_fit(ajuste_casos_acumulados, tiempo, datos_
    reales, p0=parametros_iniciales, bounds= bounds)
parametros_optimos_fil, covariance = curve_fit(ajuste_casos_acumulados_fil, tiempo,
    datos_reales_filtrados, p0=parametros_iniciales, bounds= bounds)

print(parametros_optimos) #121,530
print(parametros_optimos_fil)
print(f"r0 = {parametros_optimos[0]/(parametros_optimos[2])}", f"r0_f = {
    parametros_optimos_fil[0]/(parametros_optimos_fil[2])}")
# Graficar resultado2
```

```
t_pred = np.linspace(0, max(t_real), 210)
susceptibles, infectados, recuperados, acumulados = modelo_sird(t_pred, *parametros
    _optimos)
susceptibles_fil, infectados_fil, recuperados_fil, acumulados_fil = modelo_sird_fil
    (t_pred, *parametros_optimos_fil)

susceptibles = susceptibles * 16.6292567
susceptibles_fil = susceptibles_fil * 14.931731
print(infectados)
print(infectados_fil)
fig, ax1 = plt.subplots(figsize=(6, 4))
ax1.plot(tiempo, filter[:, 3], 'b+', label = 'Acumulados filtrados')
ax1.plot(tiempo, data[:, 3], 'c+', label = 'Acumulados reales')
ax1.plot(t_pred, acumulados, 'c-', label = f'Acumulados modelados')
ax1.plot(t_pred, acumulados_fil, 'b-', label = f'Acumulados modelados filtrados')
ax1.legend()

fig, ax1 = plt.subplots(figsize=(6, 4))
ax1.plot(tiempo, filter[:, 0], 'r+', label = 'Susceptibles filtrados')
ax1.plot(tiempo, data[:, 0], 'y+', label = 'Susceptibles reales')
ax1.plot(t_pred, susceptibles, 'y-', label = f'Susceptibles modelados')
ax1.plot(t_pred, susceptibles_fil, 'r-', label = f'Susceptibles modelados filtrados
    ')
plt.show()

fig, ax1 = plt.subplots(figsize=(6, 4))
#ax1.plot(tiempo, filter[:, 1], 'b+-', label = 'Confirmados diarios filtrados')
#ax1.plot(tiempo, data[:, 1], 'm+', label = 'Confirmados diarios reales')
ax1.plot(t_pred, infectados, 'm-', label = f'Infectados modelados')
ax1.plot(t_pred, infectados_fil, 'b-', label = f'Infectados modelados filtrados')
plt.show()

fig, ax1 = plt.subplots(figsize=(6, 4))
#ax1.plot(tiempo, filter[:, 2], 'k+', label = 'Recuperados filtrados')
#ax1.plot(tiempo, data[:, 2], 'g+', label = 'Recuperados reales')
ax1.plot(t_pred, recuperados, 'g-', label = f'Recuperados modelados')
ax1.plot(t_pred, recuperados_fil, 'k-', label = f'Recuperados modelados filtrados')
plt.show()

"""fig, ax1 = plt.subplots(figsize=(6, 4))
ax1.plot(tiempo, filter[:, 2], 'k*', label = 'Fallecidos filtrados')
ax1.plot(tiempo, data[:, 2], 'brown', label = 'Fallecidos reales')
#ax1.plot(t_pred, fallecidos, 'r-', label = f'Fallecidos modelados')
#ax1.plot(t_pred, fallecidos_fil, 'k-', label = f'Fallecidos modelados filtrados')
plt.show()
"""
```

Bibliografía

[1] SALUD COAHUILA, https://www.saludcoahuila.gob.mx/COVID19/que_es.php#:~:text=La%20COVID%E2%80%9119%20es%20la,China)%20en%20diciembre%20de%202019, 20 de Diciembre de 2019.

[2] Conahcyt, https://datos.covid-19.conacyt.mx/#DownZCSV, 22 de Diciembre de 2022.

[3] Maricarmen Hernández, https://www.pressreader.com/mexico/el-sol-de-puebla/20200420/282699049265468, 20 de Abril de 2020.

[4] Yan Pei, Donald S. Fussell, Swarnendu Biswas y Keshav Pingali, https://www.researchgate.net/publication/320345038_An_Elementary_Introduction_to_Kalman_Filtering, 27 de Junio de 2019.

[5] CDC, https://espanol.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html, 26 de Octubre de 2022.

[6] M. Martcheva, An Introduction to Mathematical Epidemiology,Texts in Applied Mathematics 61, DOI 10.1007/978-1-4899-7612-3 , 1 de Enero de 2015.

[7] Dorado, J; Ruíz, J. F., IMPLEMENTACIÓN DE FILTROS DE KALMAN COMO MÉTODO DE AJUSTE A LOS MODELOS DE PRONÓSTICO (GFS) DE TEMPERATURAS MÁXIMAS Y MÍNIMA PARA ALGUNAS CIUDADES DE COLOMBIA, IDEAM. Enero de 2014.

[8] Alex Becker, https://www.kalmanfilter.net/kalmanmulti.html.

[9] Cobo C. Marcos, El filtro de Kalman con aplicaciones en inversiones, 25 de Junio de 2021.

[10] Fisher, R. A. On the mathematical foundations of theoretical statistics. Phil. Trans. , A, 222, 309 - 368., 1 de Febrero de 1922.

[11] Fisher, R. A. Inverse Probability. Proc. Camb. Phil. Soc. , 26, 528 - 535. Octubre de 1930.

[12] Alberto Landro, Mirta L. González. Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad, 1a ed , Ciudad Autónoma de Buenos Aires : Ediciones Cooperativas, 4 - 13, 2016.

[13] Raúl J. Casas Fernández, Ecuaciones Diferenciales Estocásticas aplicadas a las Finanzas, 3, 64 - 70. 2019.

[14] cornell, https://blogs.cornell.edu/info2040/2020/12/17/estimating-the-basic-reproductive-number-for-covid-19-using-the-sird-model/, 17 de Diciembre de 2020.

[15] Javier Vargas Garategúa, https://www.linkedin.com/pulse/statistical-noise-ruido-estad%C3%ADstico-una-definici%C3%B3n-vargas-guarateg%C3%BAa/?originalSubdomain=es, 21 de Mayo de 2020.

[16] Eleanor Lutz y Amy Schoenfeld Walker, https://www.nytimes.com/es/interactive/2022/04/28/espanol/covid-estado-endemico.html#:~:text=La%20pandemia%20del%20coronavirus%20contin%C3%BAa&text=En%20su%20forma%20m%C3%A1s%20b%C3%A1sica,no%20hay%20una%20definici%C3%B3n%20establecida. 28 de Abril de 2022.

[17] Fred Brauer y Carlos Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, segunda edición, 30 de Marzo de 2001.

[18] W. O. Kermack, A. G. McKendrick, A Contribution to the Mathematical Theory of Epidemics, Vol. 115, No. 772. pp. 700-721, 1 de Agosto de 1927.

[19] Féñix Sebástian Rincón Tobo, Javier Antonio Ballesteros y Ángela María González Amarillo, https://hemeroteca.unad.edu.co/index.php/riaa/article/view/2281/3782. 18 de Diciembre de 2018.

[20] Jorge X. Velasco Hernández, Modelos matemáticos en epidemiología: Enfoques y Alcances, 27 de Noviembre de 2007.

[21] Maria Cristina Munuera Raga, Filtro de Kalman y sus aplicaciones, 27 de Junio de 2018.

[22] Alfonso Novales, Filtro de Kalman: teoria y aplicaciones, Diciembre de 2017.

[23] José Gregorio Díaz, Ana María Mejías, Francisco Arteaga , Aplicación de los filtros de Kalman a sistemas de control, junio, 2001.

[24] Atocha Alisea ,Modelos Epidemiológicos y COVID-19 , 3 de noviembre de 2020.

[25] J. Segarra Bofarull , UTILIDAD DE LOS MODELOS EPÍDEMIOLÓGICOS , 2002.

[26] Xia Yingeun y Ma Stefan , Mathematical Understanding of Infectious Disease Dynamics, 2009.